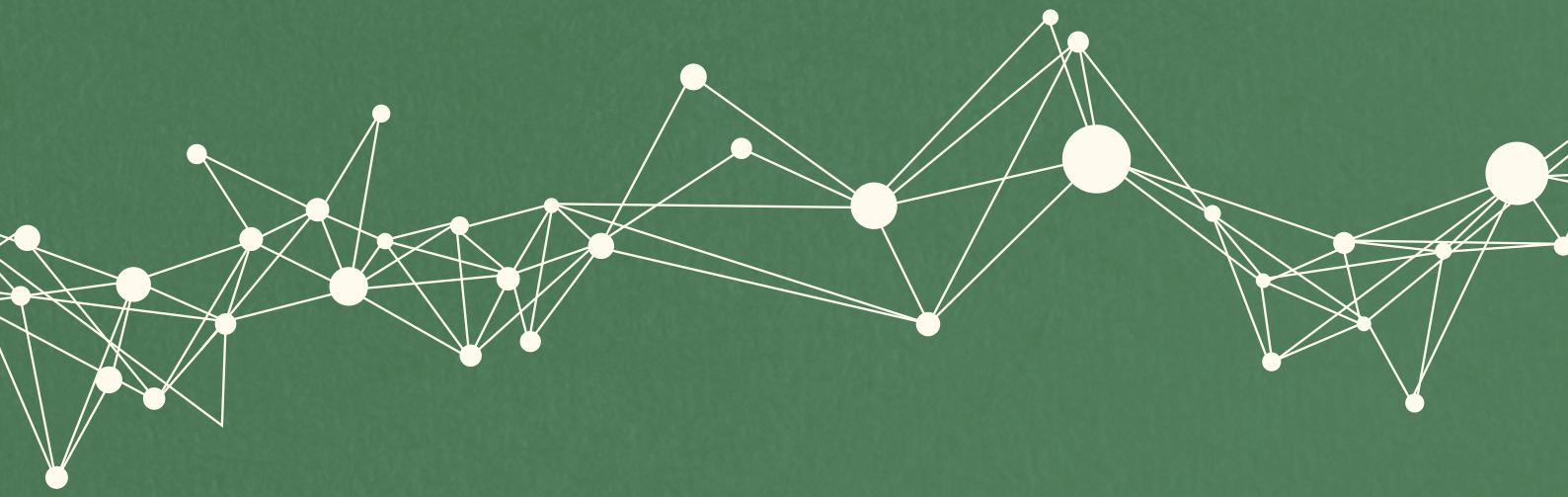


Algorithms of Violence Against Women:

A Human Rights Report on Digital Violence Targeting Women, Feminists, and Women Human Rights Defenders

A Human Rights Analysis of a Monitoring Sample of 250 Documented Cases

Report writing
Ruaa Khalaf
Editing
Tabark Majeed
Design
Ayat Tawfiq
Translation
Batool K. Hamdi

Dedication

To the soul of Yanar Mohammed,

Since your assassination, it has become difficult to treat hate speech and incitement against women as mere passing words on a screen.

Your departure was one of the reasons that pushed us to look more closely, to monitor more, and to ask: how do words create a climate that permits threats, defamation, and violence? And how do incitement campaigns become a part of the daily lives of women and women human rights defenders?

We dedicate this report to your soul, and to every woman who paid a price because she raised her voice, defended a right, or refused to stay silent.

Because words can leave a mark that reaches beyond the screen.

And because confronting hate begins with seeing it, naming it, and documenting it.

Executive Summary:

This report provides a human-rights analysis of a monitoring sample comprising 250 cases of digital hate speech documented and analysed to understand patterns of targeting directed against women, women activists, women human rights defenders, and civil society organisations in Iraq. The report does not treat this discourse as a series of isolated verbal abuses; rather, the monitoring reveals a recurring digital pattern of stigmatisation, defamation, incitement, and the social and political delegitimisation of the targeted groups.

The report adopts a human-rights approach based on the premise that hate speech, when linked to identity, gender, or human-rights and civil-society work, can become a tool for producing discrimination, silencing victims, and undermining their participation in the public sphere. The United Nations Strategy and Plan of Action on Hate Speech defines hate speech as any kind of communication in speech, writing, or behaviour that uses pejorative or discriminatory language with reference to a person or group on the basis of who they are, including gender or another identity factor. The same document notes that hate speech is amplified in the digital age through technology, often targets women and the most vulnerable groups, and intersects with disinformation and misinformation.

At the same time, the report maintains the necessary distinction between protected freedom of expression, including political and social criticism, and speech involving incitement, threats, discrimination, or collective stigmatisation. The International Covenant on Civil and Political Rights protects freedom of expression, while requiring the prohibition of advocacy of national, racial, or religious hatred that constitutes incitement to discrimination, hostility, or violence. The Rabat Plan of Action provides a threshold test for assessing the severity of speech based on context, the speaker's position or status, intent, content and form, extent of dissemination, and likelihood of harm.

The monitoring findings show that Facebook was the principal digital space in which the speech under study spread, accounting for 188 of the 250 cases, or 75.2% of the sample. Instagram accounted for 12.4%, followed by TikTok at 4.4%, while X and live streaming each accounted for 4%. This concentration shows that, within the monitored sample, Facebook goes beyond the role of a technical intermediary and constitutes the main environment in which the discourse is produced, circulated, expanded, and transformed into social pressure on women, activists, and civil society organisations.

As for the accounts producing or disseminating the speech, public pages ranked first at 46.4%, followed by anonymous accounts at 25.2%, personal pages at 24.4%, and media accounts at 4%. The data also show that 83.2% of these accounts were unverified, compared with only 16.8% that were verified. This does not mean that the discourse operates only behind a veil of anonymity; a significant part of it is produced in semi-public spaces through public and personal pages, giving it greater capacity for circulation, influence, and reproduction.

As for the accounts producing or disseminating the speech, public pages ranked first at 46.4%, followed by anonymous accounts at 25.2%, personal pages at 24.4%, and media accounts at 4%. The data also show that 83.2% of these accounts were unverified, compared with only 16.8% that were verified. This does not mean that the discourse operates only behind a veil of anonymity; a significant part of it is produced in semi-public spaces through public and personal pages, giving it greater capacity for circulation, influence, and reproduction.

This conclusion is consistent with international standards on the protection of human rights defenders. The UN Declaration on Human Rights Defenders recognises the right of individuals and groups to promote and protect human rights and affirms the State's responsibility to protect them from violence, threats, retaliation, discrimination, and pressure arising from the legitimate exercise of that role. The Committee on the Elimination of Discrimination against Women further recognises that gender-based violence can occur in both public and private spheres, including digital environments, and that in many cases it targets women human rights defenders, activists, journalists, and politicians.

Regarding the nature of the discourse, the monitoring showed a clear overlap among several forms of verbal, symbolic, and digital abuse: 147 cases were classified as hate speech, 115 as misinformation, 94 as damage to professional reputation, and 92 as defamation. The remaining cases included 74 instances of incitement to violence, 73 involving conspiracy theories, 73 containing moral accusations, 21 instances of incitement to killing, and 21 direct insults, as well as cases involving security-related accusations, sexualised speech, or verbal harassment.

These patterns do not operate independently. Misinformation supplies material for defamation; defamation opens the way to social stigmatisation; stigmatisation fuels incitement; and incitement can raise the level of risk from symbolic harm to a security threat. The report therefore treats digital hate speech as a complex structure rather than a single offensive word or passing comment. It combines the production of a hostile narrative about women and organisations, the attribution of negative moral, religious, or political characteristics to them, and the subsequent generalisation of those characteristics to an entire category.

In terms of tone, 208 cases, or 83.2% of the sample, relied on a clear emotional charge designed to provoke fear, anger, or hatred. Of these, 126 cases used an emotional or angry tone, 72 used a sarcastic or mocking tone, and 31 appeared outwardly rational while containing implicit abuse. This finding is particularly significant because inciting discourse does not always require explicit and direct wording. It may take the form of sarcasm, 'moral advice,' 'social warning,' or 'defence of values,' while functionally isolating the targets, undermining their reputations, and raising the cost of their public participation.

The data reveal a strong gendered and moral structure in the monitored discourse: misogynistic discourse was recorded in 103 cases[CL1] , religious accusations in 105 cases, moral accusations in 104 cases, and the linking of behaviour to shame or dishonour in 41 cases. Clear generalisation appeared in 199 cases, including 126 cases generalising accusations to feminists, women human rights defenders, and women activists; 83 cases generalising them to civil society organisations; and 31 cases generalising them to women in general. This means that the discourse does not merely attack an individual or organisation; it reproduces collective stigma aimed at undermining the legitimacy of feminist, human-rights, and civil-society work.

Regarding severity, 125 cases - half of the sample - were classified as high risk, 63 as medium risk, and 21 as low risk, while 41 cases remained unclassified in this field. Eighty-six cases contained direct threats, compared with 164 cases involving indirect threats. These ranged from incitement and legal or political threats to physical and cyber threats, hacking, blackmail, accusations of disloyalty, and threats to damage reputation. This indicates that risk assessment cannot be limited to the presence of an explicit threat; it must also account for the cumulative effect of discourse that places targeted women in a position of suspicion, hostility, and exclusion.

When assessing potential impact, the data show that 197 cases, or 78.8% of the sample, could lead to social harm in the form of stigma or reputational damage, while 53 cases, or 21.2%, could result in an actual security risk. In the Iraqi social context, social harm cannot be treated as secondary. Moral and religious defamation against women and activists may lead to isolation, threats, withdrawal from public work, self-censorship, or exposure to family, community, and security risks.

The data also reveal a clear shortfall in digital response. Of the sample, 139 cases (55.6%) were reported, while 81 were not reported and 30 remained unresolved. Yet content was removed in only 53 cases, while it remained online in 197 cases. This finding calls for more serious scrutiny of platform reporting and review mechanisms, particularly for Arabic-language content, local dialects, and gendered and political contexts.

The Committee similarly stresses the need for social media companies to maintain systems that are responsive to local and regional contexts and capable of effectively addressing technology-facilitated gender-based violence.

One of the report's most prominent findings is that all 250 monitored cases involved repeated attacks on the same target. Indicators of repetition included the use of the same phrases in 165 cases, multiple accounts posting the same or similar content in 53 cases, and synchronised timing in 32 cases. These indicators alone are not sufficient to prove centralised management of all attacks, but they reveal recurring and coordinated or semi-coordinated patterns. This requires monitoring to move beyond the individual post toward analysis of networks, repeated phrases, timing, multiple accounts, and recurring targets.

The report concludes that digital hate speech against women, activists, and civil society organisations in Iraq amounts to a digital practice with direct human-rights consequences and cannot be reduced to a mere reflection of temporary social tension. It affects the rights to dignity, non-discrimination, personal security, freedom of expression, freedom of association, participation in public affairs, and the right to defend human rights. It also constricts civic space and reproduces structures of gendered control through rapidly spreading, low-cost digital tools that are difficult to hold accountable.

Accordingly, the report recommends a multi-level approach: legal and institutional protection for women human rights defenders, activists, and organisations; investigation and response mechanisms that reflect the seriousness of digital threats; stronger documentation and monitoring capacities within civil society organisations; requirements for digital platforms to improve Arabic-language reporting and review tools in ways suited to the Iraqi context; and a clear gender perspective in addressing hate speech, incitement, and defamation.

The UN Guiding Principles on Business and Human Rights affirm that the responsibility to respect human rights applies to companies, while States have a duty to protect against abuses committed by third parties, including companies, through prevention, investigation, and effective remedy. The UN B-Tech Project further clarifies that these principles provide a framework for identifying, preventing, mitigating, and addressing human-rights harms linked to digital technologies.



Introduction:

This report examines digital hate speech as a social and political practice with direct human-rights consequences, rather than as a passing linguistic deviation or an excess of verbal violence on social media platforms. When such speech targets women, activists, women human rights defenders, and civil society organisations, it goes beyond the boundaries of 'harsh opinion' or 'public debate' and becomes a pattern of stigmatisation, incitement, and delegitimisation that threatens the rights to dignity, equality, non-discrimination, personal safety, freedom of expression, and participation in public affairs. The importance of this report derives from an Iraqi context that has witnessed, in recent years, a rise in hostile discourse against women, women human rights defenders, and civil society work, alongside a decline in legal and institutional guarantees protecting women's rights. The amendment to the Personal Status Law entered into force in February 2025 after a controversial political and judicial process. Human-rights organisations considered that it undermines equality before the law and weakens protections for women and girls, particularly in matters of marriage, divorce, custody, and inheritance, while opening the way to parallel religiously based legal regimes in family matters. In the same context, the assassination of feminist activist and human rights defender Yanar Mohammed in March 2026 represented an extremely serious sign of the movement of violence against women defenders from threats and stigmatisation into physical elimination. Amnesty International reported that Yanar Mohammed was shot inside her home in northern Baghdad on 2 March 2026 and that the Prime Minister ordered an investigation the same day. Amnesty International and Human Rights Watch also called for an effective investigation and accountability and linked the killing to a broader climate of impunity and the targeting of human rights defenders in Iraq.

Yet the monitored digital discourse reflects a recurring transformation of references to CEDAW from an international human-rights framework into a tool of stigma and accusation, to the point that the Convention is used in some discourse as a political and moral insult against feminists, organisations, and women's rights defenders. This transformation does not stop at the Convention itself; it extends to targeting women's right to demand equality, protection, and justice, and recasts human-rights work as suspicious, foreign, or contrary to social values.

This escalation cannot be separated from the legal and social environment in which women operate in Iraq. Iraq is a State party to the Convention on the Elimination of All Forms of Discrimination against Women (CEDAW) and underwent periodic review before the CEDAW Committee in February 2026, during which issues were raised concerning violence against women, impunity for so-called 'honour' crimes, and the continued existence of legal provisions allowing mitigating circumstances in some such crimes.

This climate deepened as attacks on the vocabulary of women's rights moved from social controversy into institutional procedures and discourse. In August 2023, the Communications and Media Commission issued a directive asking media outlets to replace the term 'homosexuality' with the expression 'sexual deviance,' and it also decided to prohibit the use of terms such as 'gender,' 'gender identity' and 'homosexuality' in its communications. Amnesty International regarded this as a step that could fuel discrimination and attacks and argued that demonising the word 'gender' reflects disregard for gender-based violence in a context of widespread impunity.

In February 2024, the Federal Supreme Court issued a ruling in Case No. 326/Federal/2023 concerning the invalidity of provisions in instructions governing the formation and functions of departments within the General Secretariat of the Council of Ministers. These included provisions related to the 'Gender Section,' coordination with gender sections and units in government institutions, and integration of the concept of gender into government plans and policies. This was followed by a trend in some official institutions to replace labels associated with 'women's empowerment' with terms such as 'Iraqi Women's Affairs.' In March 2024, the Ministry of Electricity announced that its Iraqi Women's Empowerment Department would be renamed the Iraqi Women's Affairs Department, and other ministries and institutions followed.

In this context, targeting terms such as 'gender,' 'women's empowerment,' and 'equality' goes beyond linguistic or terminological disagreement and becomes part of a wider environment that narrows the symbolic and legal space available to feminist and human-rights work. When these terms disappear from official discourse or become objects of suspicion, they become more easily transformed in the digital sphere into tools of stigmatisation and accusations of betrayal against women defenders and organisations. This is precisely what the monitoring data reveal: hateful discourse deploys these terms as markers of moral, religious, and political accusation rather than as concepts open to discussion or disagreement, repeatedly linking rights claims with immorality, foreign allegiance, threats to the family, or violations of social values.

The monitoring material indicates that the absence of a clear and coherent official discourse confronting this distortion, combined with weak institutional responses to campaigns of incitement and defamation, has widened the space for impunity. The problem lies not only in the existence of hateful or misleading content, but in the fact that it operates within a complete protection gap: there is still no effective national system for protecting human rights defenders, and no enforceable law against domestic violence in federal Iraq despite years of delay. Existing provisions in the Penal Code continue to create an environment in which penalties may be reduced in some forms of violence against women and girls, an issue raised by the CEDAW Committee in its review of Iraq.



Although some institutions have announced communication channels or mechanisms for dealing with digital violence, the monitoring data do not show a response proportionate to the level of risk. Threats go beyond insult and extend to defamation, moral and religious accusations, incitement, accusations of disloyalty, legal and political threats, and physical or cyber threats. In many cases, these acts are not treated as early-warning indicators of broader violence; instead, they are allowed to accumulate in the digital sphere until they become part of a hostile social climate toward women, particularly those who enter public life, defend rights, or work within civil institutions.

This discourse is not confined to activists or women engaged in public affairs. The monitored sample shows that targeting also extended to women who are not necessarily involved in human-rights, social, or political work, demonstrating the breadth of the space in which digital hatred against women operates. The issue therefore goes beyond the targeting of feminist activism alone: simply by being present at work, within an institution, or in the digital sphere, a woman may become a potential target of defamation, moral surveillance, and symbolic punishment.

The circulation of leaked clips, or clips attributed to individuals holding administrative or political authority and containing allegations of abuse of power relationships in the workplace, has also widened hostile discourse against women working in the public and private sectors. Regardless of the details of each case and its legal validity, the more dangerous pattern is the shift of public discussion away from accountability for authority and abuse of power and toward stigmatising, mocking, and defaming the women concerned. In this way, a woman moves from being a potential victim of an unequal power relationship to becoming the object of social accusation, while discussion of institutional responsibility, protection of women workers, the duty to investigate, and prevention of retaliation and defamation recedes.

The monitoring database was analysed as a set of indicators of a recurring discursive structure rather than as isolated incidents. The following sections detail the main patterns revealed by the database in relation to platforms, targeted groups, and levels of risk.

This report does not seek to restrict freedom of expression or shield public actors from legitimate criticism. Distinguishing criticism from hateful discourse is essential to any human-rights approach.

Freedom of expression, however, does not protect incitement to violence, discrimination, or hostility, nor does it justify campaigns of collective stigmatisation targeting women because of their gender or their human-rights or civil-society work. The report later applies two relevant standards: the Rabat Plan of Action threshold test and the definition set out in the UN Strategy and Plan of Action on Hate Speech.

The report therefore reads digital hate speech against women in Iraq as part of a broader crisis: declining legal guarantees, weak protection mechanisms, absence of effective accountability, restrictions on human-rights and feminist vocabulary, and growing social and political discourse that uses religion, morality, and shame to strip women, women defenders, and organisations of legitimacy. This crisis does not threaten women alone; it threatens civic space as a whole by making public participation conditional on accepting defamation, danger, and silence.

Report Objectives and Questions:

This section sets out the report's objectives, central questions, and the analytical hypothesis on which its examination of the monitoring database is based. The report does not merely count offensive expressions as isolated occurrences; it seeks to understand the wider structure within which the discourse operates: who produces it, where it spreads, whom it targets, what language it uses, and what human-rights, social, and security effects it may have.

The report's approach is guided by international standards distinguishing protected freedom of expression from speech that may develop into incitement to discrimination, hostility, or violence, as described in the Executive Summary above.

First: Report Objectives

The report aims to monitor and analyse digital hate speech directed against women, activists, women human rights defenders, and civil society organisations, and to determine whether this discourse appears as scattered individual cases or as a recurring pattern containing indicators of coordinated or semi-coordinated campaigns.

It also aims to identify the digital platforms most frequently used to disseminate such discourse, particularly in light of the clear concentration of monitored cases on Facebook, so as to understand the platform's role in expanding stigmatisation, defamation, and incitement.

The report seeks to analyse the groups most frequently targeted and determine whether targeting is directed at women in general or concentrates more heavily on women when they perform public roles, such as human-rights activism, human-rights defence, civil-society work, or participation in public debates related to equality and legal protection.

The report aims to monitor and analyse digital hate speech directed against women, activists, women human rights defenders, and civil society organisations, and to determine whether this discourse appears as scattered individual cases or as a recurring pattern containing indicators of coordinated or semi-coordinated campaigns.

It also aims to identify the digital platforms most frequently used to disseminate such discourse, particularly in light of the clear concentration of monitored cases on Facebook, so as to understand the platform's role in expanding stigmatisation, defamation, and incitement.

The report seeks to analyse the groups most frequently targeted and determine whether targeting is directed at women in general or concentrates more heavily on women when they perform public roles, such as human-rights activism, human-rights defence, civil-society work, or participation in public debates related to equality and legal protection.



The report aims to assess the response of digital platforms and relevant institutions, particularly in light of the gap between the number of reports submitted and the amount of content removed, and what this gap reveals about shortcomings in digital protection mechanisms, especially in dealing with the Iraqi and Arabic contexts and coded gendered discourse.

It seeks to highlight the human-rights and social impact of this discourse as a practice that may silence women, push activists toward self-censorship, weaken civil-society work, and expose targeted women to social exclusion or security risks, rather than as merely limited moral or emotional harm.

Finally, the report aims to provide practical recommendations to governmental, judicial, and legislative bodies; digital platforms; media outlets; civil society organisations; and international actors in order to reduce digital hate speech, improve protection mechanisms, and ensure accountability without undermining freedom of expression or legitimate criticism.

Second: Report Questions

The report seeks to answer a set of central questions:

- Which digital platforms contain the greatest concentration of hate speech against women, activists, women human rights defenders, and civil society organisations in the monitored sample?
- What types of accounts or pages most frequently produce or circulate this discourse? Does it originate mainly from anonymous accounts, or does part of it operate through public pages, personal pages, and accounts with a visible public presence?
- Who are the groups most frequently targeted? Is the discourse concentrated against women in general, or against women engaged in human-rights and civil-society work or participating in the public sphere?
- Which patterns are most prevalent in the monitored discourse: hate speech, defamation, misinformation, incitement, accusations of disloyalty, threats, moral stigmatisation, religious accusations, or conspiracy theories?
- How are moral, religious, and gendered forms of language used to produce stigma against women and organisations? How do human-rights concepts such as 'CEDAW,' 'gender,' 'equality,' and 'women's empowerment' become tools of defamation and accusation in the digital sphere?
- Does the discourse target only specific individuals, or does it generalise accusations to entire groups such as feminists, women human rights defenders, civil society organisations, and women in general?
- What is the level of risk in the monitored cases, and what kinds of threats do they contain? Are direct threats predominant, or indirect threats that open the way to incitement, stigma, and social persecution?

- What potential impact does this discourse have on the targeted women? Is it limited to social harm and reputational damage, or does it extend to security, professional, psychological, and institutional risks?
- To what extent do the data show indicators of repetition or coordination, such as use of the same expressions, multiple accounts publishing similar content, synchronised attacks, or repeated targeting of the same entity?
- How do digital platforms and relevant institutions respond to reports or complaints? Do the data reveal a gap between the scale of risk and the scale of the response?
- What is the relationship between the rise in digital hate speech and the broader legal and institutional context, including amendments to the Personal Status Law, the absence of legislation protecting human rights defenders, and weak accountability for violence against women?
- How can a human-rights response be developed that does not use hate speech as a pretext to restrict freedom of expression while also ensuring that women, women defenders, and organisations are not left exposed to defamation, incitement, and threats?

Third: Analytical Hypothesis of the Report

The report proceeds from the hypothesis that digital hate speech against women, activists, and civil society organisations in Iraq operates not merely as abusive speech but as a mechanism of social and political pressure aimed at regulating women's participation in the public sphere. This reading assumes that moral and religious stigmatisation, defamation, misinformation, and incitement do not appear in the sample as separate elements but as interconnected links within a single discursive structure whose purpose is to weaken trust in women who defend rights, disrupt the work of civil society organisations, and raise the cost of demanding equality and legal protection.

The report also assumes that weak official and digital responses, the absence of effective accountability, and the deterioration of certain legal and institutional guarantees do not necessarily produce hate speech themselves, but create an environment in which it can spread, recur, and become a tool of silencing and symbolic and social punishment.

Methodology and Data Limitations

This report is based on a monitoring database containing 250 documented records of digital hate speech targeting women, activists, women human rights defenders, and civil society organisations in Iraq. These records were treated as analytical material revealing recurring patterns in discourse, rather than merely as scattered examples of digital abuse or isolated hostile comments.

The monitoring process collected and classified cases according to a set of variables designed to illuminate the digital, social, and human-rights structure of the discourse. These variables included: the platform on which the content was published; type of account or page; level of engagement; reason for targeting; classification of the discourse; targeted group; nature of the discourse; level of risk; type of threat; potential impact; reporting status; deletion status; and indicators of repetition or similarity between attacks.

The analysis was not limited to identifying the presence of abusive or hateful speech; it also examined the context in which such discourse operates. Attention was therefore paid to the nature of the language used and whether it involved defamation, misinformation, incitement, moral stigmatisation, religious accusations, accusations of disloyalty, or direct or indirect threats. The analysis also considered whether the discourse targeted a particular individual or generalised accusations to a broader group such as feminists, women human rights defenders, civil society organisations, or women in general.

The report treated each record as an independent monitoring unit while recognising that some cases may contain more than one discursive pattern at the same time. A single post may combine hate speech with defamation, misinformation with incitement, or a moral accusation with an indirect threat. The report therefore does not regard the classifications as closed boxes, but as overlapping indicators that help explain levels of risk and potential impact.

The human-rights classification of the cases was based on distinguishing legitimate criticism, including harsh political and social criticism, from speech involving collective stigmatisation, incitement, threats, defamation, or discrimination on the basis of gender or human-rights or civil-society work. The report therefore does not assume that every negative statement or expression opposing women's issues automatically constitutes hate speech. It focuses on cases that use degrading, hostile, or discriminatory language, or associate women and organisations with negative moral, religious, or political attributes in ways that affect dignity, safety, and public participation.

The analysis also included indicators of repetition, an important element in the database. Cases were tracked for repeated attacks on the same target, use of identical phrases, publication of similar content across multiple accounts, and synchronised timing. The report does not treat these indicators as conclusive proof of centralised management of all attacks, because such a conclusion would require additional technical and investigative tools. It treats them instead as indicators of recurring, coordinated, or semi-coordinated campaigns warranting further monitoring and analysis.

To assess severity, the report considered the nature of the threat in each case and whether it was direct or indirect, physical, cyber, political, legal, or social. It also analysed the potential impact of each case, whether in the form of social harm, stigma, reputational damage, or an actual security risk. The report uses a contextual reading because some expressions that appear general or indirect can become more serious in a conservative or polarised social environment, especially when they concern women's reputations or accuse them on moral or religious grounds.

The report also monitored the response dimension through two main fields: whether the content was reported, and whether it was removed after reporting or monitoring. This helps clarify the gap between the scale of harmful discourse on the one hand and the level of digital or institutional response on the other. The report does not use these data to reach a final judgment on the performance of every platform in each individual case, but to identify a general pattern of weak response compared with the level of risk and recurrence of attacks.

It should be noted that, like any field or digital monitoring database, the dataset contains some fields requiring editorial review before final publication. Some entries contain preliminary or incomplete formulations such as 'Option 5',,'' or truncated phrases such as 'accusations..' and 'information..'. These issues do not invalidate the database for analytical purposes, but they require a final cleaning stage before general tables are approved or charts are published. This stage should include standardising terminology, completing incomplete phrases where the original material allows, removing non-informative entries, and reviewing fields that could affect statistical classification.

In handling the data, the report observes the principle of avoiding additional risk to victims or targets. Even in the final published version, names, links, images, or identifying details that could lead to renewed targeting of women or organisations should be avoided. Examples should be presented briefly and with identifying data removed, including only what is necessary for human-rights analysis and avoiding the recirculation or amplification of abusive speech.

This sample does not claim to represent all digital hate speech against women and organisations in Iraq or to measure the overall scale of the phenomenon nationwide. It is a specific monitoring sample that reveals significant patterns within the digital space that was observed. The findings should therefore not be read as a comprehensive national statistic, but as analytical evidence of recurring trends in targeting, language, affected groups, weak response, and indicators of repetition and coordination.

The strength of this methodology lies in its combination of quantitative and qualitative analysis. On the one hand, the figures identify the most prominent platforms, groups, and patterns; on the other, qualitative analysis explains the function and impact of the discourse and how moral, religious, and gendered stigmatisation is used to weaken the participation of women, women defenders, and organisations in the public sphere. The report therefore moves beyond the question 'How many cases?' to a broader set of questions: How does this discourse operate? Whom does it target? What is its impact? And why does it become a human-rights risk rather than merely a digital dispute?



- Presentation of Data and Monitoring Findings:

Question in the monitoring form: On which platform was the content published?

Category/Response	Number of Cases	Percentage of Sample
Facebook	188	752%
Instagram	31	124%
TikTok	11	44%
X	10	40%
Live Stream	10	40%

Analysis: The data show a clear concentration of the discourse on Facebook, making it the principal space for the production and circulation of hate speech within the sample. This does not mean that other platforms are absent, but it identifies where digital responses and reporting and review mechanisms should be concentrated.

- Types of Accounts Producing or Disseminating the Discourse:

Question in the monitoring form: What type of account or page published the content?

Category/Response	Number of Cases	Percentage of Sample
Public Page	116	464%
Anonymous Account	63	252%
Personal Page	61	244%
Media Account	10	40%

Analysis: The prominence of public pages as the most frequent source indicates that the discourse does not circulate only through anonymity or anonymous accounts; it also appears in public spaces with the capacity to influence audiences and expand dissemination.

- Account Verification Status

Question in the monitoring form: Is the account verified?

Category/Response	Number of Cases	Percentage of Sample
No	208	832%
Yes	42	168%

Analysis: The predominance of unverified accounts shows that weak digital identification does not prevent the production of high-impact discourse. At the same time, the presence of verified accounts in the sample shows that part of the discourse also circulates through accounts with an openly declared presence.

- Media Attached to the Content

Question in the monitoring form: Was media attached to the post or comment?

Category/Response	Number of Cases	Percentage of Sample
Image	185	740%
Video	65	260%

Analysis: The predominance of images over video in the sample suggests that the discourse often relies on material that is easy to copy and circulate. Images, particularly when combined with inciting or mocking text, help visually reinforce stigma and spread it quickly.

- Question in the monitoring form: What was the reason for targeting?

This is a multiple-response field; a total of 293 indications were extracted from 250 monitoring records.

Category/Response	Number of Cases	Percentage of Sample
Response to an Event	74	296%
Publication of Specific	62	248%
Advocacy Campaign	31	124%
During a Discussion	21	84%
Feminist Activity	20	80%
Media Participation	11	44%
Posting about the Death	11	44%
This is a post against the	11	44%
Incitement against Law	11	44%
Against the 1 May	11	44%
Inciting Society against	10	40%
Reputational Damage	10	40%
Mocking an Activist	10	40%

Analysis: The reasons show that the discourse often escalates following a particular event, specific content, or advocacy activity. The presence of multiple reasons in some cases indicates that targeting is not linked to a single cause but to an accumulated context of positions, activities, and circulating content.

- Question in the monitoring form: How was the discourse classified?

This is a multiple-response field; a total of 741 indications were extracted from 250 monitoring records.

Category/Response	Number of Cases	Percentage of Sample
Hate Speech	147	588%
Misinformation	115	460%
Damage to Professional	94	376%
Defamation	92	368%
Incitement to Violence	74	296%
Conspiracy Theory	73	292%
Moral Accusations	73	292%
Incitement to Killing	21	84%
Direct Insult	21	84%
Security-related	11	44%
Incitement against the	10	40%
Sexualised	10	40%

Analysis: Hate speech does not appear separately from misinformation, defamation, and damage to reputation. A single case may combine several patterns, indicating a complex discursive structure that moves from stigmatisation to incitement and from reputational harm to social or security risk.

- Question in the monitoring form: What group was targeted?

Category/Response	Number of Cases	Percentage of Sample
Women	124	496%
Civil Society	84	336%
Women in General	32	128%
Educational	10	40%

Analysis: The concentration of targeting on women activists, women human rights defenders, and civil society organisations shows that the discourse does not attack women only as individuals; it intensifies when their presence is connected to human-rights or civil-society work or demands for equality and protection.

- Question in the monitoring form: Is the target an organisation, an activist, a public figure, or another type of actor?

Category/Response	Number of Cases	Percentage of Sample
Unclassified/Blank	126	504%
Local Activist	52	208%
Local Organisation	42	168%
Public Figure	20	80%
Invalid Value/Needs	10	40%

Analysis: The high number of unclassified entries in this variable reflects the optional nature of the field or incomplete data entry, but it does not diminish the importance of classified cases showing local activists, local organisations, and public figures among the targets.

Type of Generalisation and Collective Stigmatisation

- Question in the monitoring form: If generalisation is present, whom does it target

Category/Response	Number of Cases	Percentage of Sample
Yes	199	796%
Unclassified/Blank	30	120%
Maybe	11	44%
No	10	40%

Analysis: Generalisation primarily targets feminists, women human rights defenders, and activists, followed by civil society organisations.

Level of Risk of the Discourse

- Question in the monitoring form: What is the level of risk of the discourse?

Category/Response	Number of Cases	Percentage of Sample
High	125	500%
Medium	63	252%
Unclassified/Blank	41	164%
Low	21	84%

Analysis: The fact that approximately half of the sample falls within the high-risk category calls for treating the discourse as an early-warning indicator of broader risks. Risk should not be limited to explicit threats, because repeated stigmatisation may accumulate social and security consequences.

Nature of the Threat

- Question in the monitoring form: Is the threat direct or indirect?

Category/Response	Number of Cases	Percentage of Sample
Indirect Threat	164	656%
Direct Threat	86	344%

Analysis: Indirect threats predominate in the sample. This type is not necessarily less serious because it leaves the door open for audiences to reproduce stigma, incitement, or social persecution.

- Question in the monitoring form: What type of threat is involved?

Category/Response	Number of Cases	Percentage of Sample
Incitement	81	324%
Legal/Political Threat	52	208%
Physical Threat	22	88%
Cyber	20	80%
Accusation of Disloyalty	11	44%
Mockery and Disrespect	11	44%
Invalid Value/Needs	11	44%
Community/Social	11	44%
Threat to Damage	11	44%
No Clear Threat	10	40%
Other	10	40%

Analysis: Incitement and legal or political threats are the most prevalent categories, alongside physical and cyber threats. Invalid or empty values show the need for final cleaning of this field before detailed tables are published.

- Question in the monitoring form: What is the potential impact of the case?

Category/Response	Number of Cases	Percentage of Sample
Social	197	788%
Actual Security Risk	53	212%

Analysis: Social stigma and reputational harm predominate. These effects cannot be minimised in the Iraqi context, because social harm can turn into family or professional pressure, security risk, or withdrawal from the public sphere.

- Question in the monitoring form: Does the discourse rely on provoking fear, anger, or hatred?

Category/Response	Number of Cases	Percentage of Sample
Yes	208	832%
No	32	128%
Maybe	10	40%

Analysis: Most cases rely on clear emotional mobilisation, which gives the discourse greater capacity for dissemination and escalation. Emotionally charged speech does not merely convey a position; it generates a hostile response toward the targets.

Reporting the Content

- Question in the monitoring form: Was the content reported?

Category/Response	Number of Cases	Percentage of Sample
Yes	139	556%
No	81	324%
Maybe	30	120%

Analysis: The reporting rate indicates attempts to address harmful content through platform mechanisms. This indicator, however, can only be fully understood when compared with deletion data, because reporting alone does not mean that protection has been achieved.

- Question in the monitoring form: If the content was reported, was it removed?

Category/Response	Number of Cases	Percentage of Sample
No	197	788%
Yes	53	212%

Analysis: The gap in removals shows that the digital response is not proportionate to the scale of harmful discourse. The fact that most content remained online despite reports reveals shortcomings in review procedures or in understanding the local and gendered context of the discourse.

- Question in the monitoring form: Was repetition of the attack observed?

Category/Response	Number of Cases	Percentage of Sample
Yes	250	1.000%

Analysis: The fact that all records indicate repeated attacks makes this one of the report's strongest findings. Repetition transforms abuse from a passing incident into a continuing pattern of pressure on the target.

- Question in the monitoring form: What are the indicators of repetition?

Category/Response	Number of Cases	Percentage of Sample
Use of the Same Phrases	165	660%
Multiple Accounts with	53	212%
Synchronised Timing	32	128%



The indicators in this section converge to produce a single picture: recurring discourse concentrated on Facebook, primarily targeting women activists, women human rights defenders, and civil society organisations, and combining hate, defamation, misinformation, and incitement within a single highly emotionally charged structure. The next section moves from presenting these figures to reading them through a human-rights lens: which rights are affected, how the discourse operates as a recurring structure of pressure, and what it reveals about the position of women in Iraq's public sphere.

Human-Rights Analysis of the Findings

The monitoring findings show that digital hate speech against women, activists, women human rights defenders, and civil society organisations in Iraq operates as part of a broader environment that restricts women's presence in the public sphere, rather than as linguistic behaviour detached from its context. The sample of 250 cases points to a recurring discursive structure in which hatred overlaps with defamation, misinformation, incitement, and moral, religious, and gendered stigmatisation. This overlap is what gives the discourse its human-rights seriousness: its effects go beyond attacking a particular opinion and affect the dignity of targeted groups and their rights to participation, protection, and non-discrimination.

From a human-rights perspective, the problem does not lie in the existence of harsh criticism or social and political disagreement over issues concerning women and organisations. Criticism, however severe, remains part of the freedom of expression protected under the International Covenant on Civil and Political Rights. The monitored data, however, reveal a clear transition from criticism to stigmatisation, from disagreement to incitement, and from public accountability to the delegitimisation of women, women defenders, and organisations. This transition marks the decisive boundary between legitimate expression and discourse that produces discriminatory or threatening effects, in accordance with the standard established by the Rabat Plan of Action.

In the Iraqi context, content published on Facebook does not remain confined to a phone screen. Digital reputation quickly travels into the family, workplace, social circles, and the political or civil environment in which women and organisations operate. When the target is a woman or a woman human rights defender, the impact of a post can extend beyond emotional harm to family pressure, social persecution, or withdrawal from the public sphere.

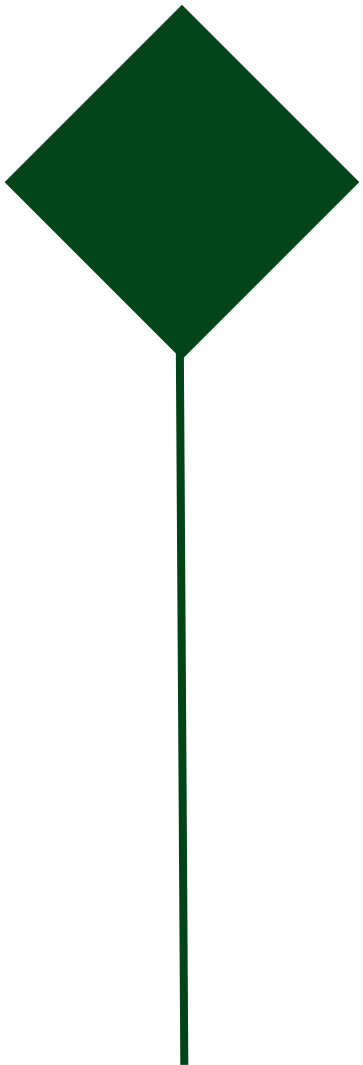
Data on the publishing accounts show that the discourse does not rely on anonymity alone. Although unverified accounts are strongly represented (83.2%), public pages are the most common type of account producing or disseminating the discourse, at 46.4%. This gives the phenomenon two faces: one part of the discourse benefits from anonymity to evade accountability, while another operates openly and gains social, political, or media legitimacy. This openness can sometimes be more dangerous than anonymity because it makes hatred appear as acceptable public discourse rather than as marginal, isolated behaviour. From a human-rights perspective, the problem therefore goes beyond the technical dimension and concerns a digital environment that allows women, women defenders, and organisations to become recurring targets. This expands platform responsibility beyond post-hoc deletion to understanding the local context, identifying repetition, and addressing gendered, religious, and moral stigmatisation when it is used to silence women or expose them to danger. The UN Guiding Principles on Business and Human Rights reinforce this responsibility by defining companies' duty to respect human rights and to identify and address risks linked to their activities.

Targeted Groups and the Meaning of Targeting: Targeting in the sample is concentrated on women activists and women human rights defenders (49.6%), followed by civil society organisations (33.6%), and women in general (12.8%). This ordering is more than a statistical detail; it reveals the logic of the discourse itself. Women are not targeted only because of their gender, but because of their public presence and their capacity to speak, organise, object, and defend the rights of women and affected groups. This carries a central feminist significance: the problem is not women's existence, but their presence as actors, claimants, organisers, defenders, and critics. The monitored discourse thus functions as a mechanism of social control rather than as a spontaneous reaction, conveying a deterrent message that public participation has a cost and that claiming rights may be met with defamation, accusations of disloyalty, and moral accusation.

The same logic is repeated when a woman becomes an activist, a defender, an employee of an organisation, or a participant in public debate: her exposure to defamation, accusations of disloyalty, and moral accusation increases. Hate speech thereby becomes a tool of social control that raises the cost of participation, makes public visibility risky, and pushes women to think twice before speaking, appearing, or organising. Targeting civil society organisations adds another dimension: the discourse does not merely attack a particular institution, but attempts to undermine trust in civil-society work as a whole by linking it to foreign allegiance, immorality, threats to the family, or violations of religion and values. An attack on an organisation thus becomes an attack on the very idea of defending rights. This type of targeting affects the rights to public participation, freedom of association, freedom of expression, and the right to defend human rights, in contradiction with the UN Declaration on Human Rights Defenders, which recognises the right of individuals and groups to promote and defend rights and places on the State the duty to provide a safe environment and protect them from threats, retaliation, or pressure because of this role.

Patterns of Hate Speech, Defamation, and Incitement: Hate speech in the sample does not appear as a single, pure category. It coexists with misinformation, defamation, damage to professional reputation, incitement to violence, conspiracy theories, and moral accusations. This combination is itself a source of risk: discourse may begin with misleading information or an unverified allegation, then turn into defamation, collective stigmatisation, and direct or indirect incitement. This sequence is not incidental.

The sample also shows a strong pattern of damage to professional reputation, which is particularly significant. Targeting women in the public sphere does not operate only through their bodies or their alleged 'morality,' but also by attacking their competence, integrity, and the legitimacy of their work. Gendered hatred therefore intersects with the weakening of professional and civil roles: the targeted woman is not only described as 'contrary to values,' but is also represented as unqualified, suspicious, paid, or part of a foreign agenda.



This sequence is not incidental .Misinformation provides the raw material for attack; defamation moves the attack into the field of reputation; moral or religious accusation gives the discourse a more dangerous social charge; and incitement turns audiences from recipients into potential participants in harm. Each post therefore cannot be treated as an isolated unit because harm is often produced by the accumulation of discourse rather than by a single phrase. The discourse does not merely describe targets in negative terms; it constructs a full narrative around them in which women and organisations are said to threaten the family, religion, values, sovereignty, security, or 'social order.' Within this narrative, symbolic abuse becomes a precondition for accepting exclusion, incitement, or punishment. The discourse therefore cannot be analysed as a set of scattered insults; it is an ongoing production of suspicion toward feminist and civil-society work.

the UN Strategy and Plan of Action on Hate Speech, such discourse should not be treated as isolated insults; it attacks or uses pejorative or discriminatory language against a person or group on the basis of identity, affiliation, gender, or other factors. At the same time, responses must remain consistent with the protection of freedom of expression and must not become a pretext for suppressing opinion. This distinction guided the report's monitoring and analysis.

Religious discourse in the sample does not appear as theological or jurisprudential debate; it often functions as a sorting mechanism between those considered 'within the values' and those placed 'outside them.' When women activists and organisations are placed outside the circle of accepted values, they become easier to target and audiences become more willing to tolerate verbal or symbolic violence against them. The discourse thus operates as a process of withdrawing social protection.

The Gendered, Moral, and Religious Structure of the Discourse: One of the most serious findings is that the discourse does not stop at attacking the positions of women or organisations; it seeks to strip them of moral and social legitimacy. Terms associated with religion, shame, dishonour, honour, family, gender, CEDAW, empowerment, and equality are often not used for discussion or disagreement, but as markers of accusation that turn human-rights concepts into 'suspicions' and transform demands for protection and equality into material for accusations of disloyalty and defamation. This structure is distinctly gendered: moral stigmatisation does not affect women and men in the same way. In a conservative society, accusing a woman on moral or religious grounds may open the door to social violence, family pressure, professional punishment, and forced withdrawal from public life. Moral defamation against women is therefore not merely verbal abuse; it is a tool of subordination.

From the perspective of the Convention on the Elimination of All Forms of Discrimination against Women (CEDAW), violence against women is not limited to physical violence but includes forms of gender-based violence that impede women's equal enjoyment of rights. General Recommendation No. 35 of the CEDAW Committee identifies such violence as a form of discrimination and links it to unequal power relations and harmful stereotypes.

The findings on generalisation are especially significant in this context. The widespread generalisation directed against feminists, women human rights defenders, civil society organisations, and women in general shows that the discourse goes beyond targeting a particular individual or organisation. Generalisation is the point at which speech shifts from individual abuse to hate speech with collective impact: when 'feminists,' 'organisations,' or 'women' are collectively described as traitorous, immoral, or a threat to the family, every individual within that group becomes a potential target even if she was not involved in the original incident. This is precisely what makes the discourse dangerous to civic space as a whole: it produces a general climate of suspicion and exclusion that erases differences between individuals. The target ceases to be a person with a name, context, and experience and becomes a symbol of a group that is being demonised; simply being a woman activist, an organisation worker, or a defender of a legal right may be enough to place her under the same accusation.

This pattern damages civic space from within: organisations become 'all suspicious,' feminists become 'all against the family,' women defenders become 'all part of an agenda,' and women working in public life become 'all objects of suspicion.' The discourse therefore does not need to silence every woman individually; it is enough to create a general atmosphere in which women anticipate harm in advance. From a human-rights perspective, generalisation is what gives the discourse its discriminatory character because it does not criticise a specific act; it attaches a negative trait to an entire group, making the harm a matter of non-discrimination, dignity, and public participation rather than personal reputation alone.

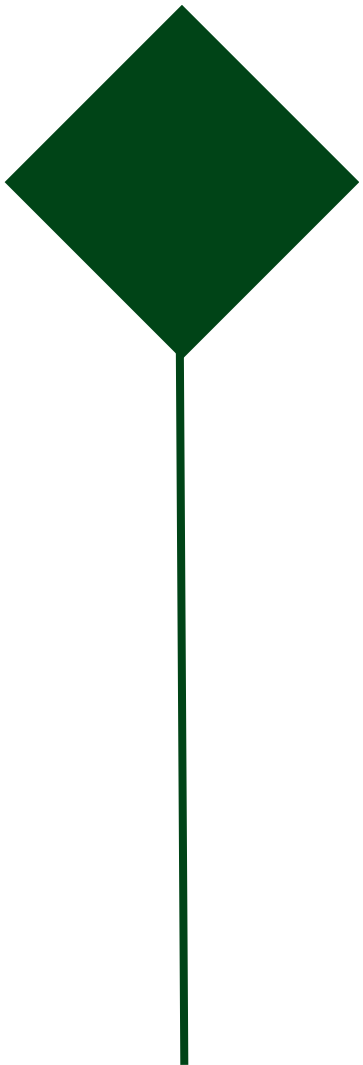
Level of Risk and Nature of the Threat: The sample shows that the level of risk is neither low nor marginal. A large proportion of cases - close to half the sample - fall within the high-risk category, alongside a clear predominance of indirect threats (65.6%, compared with 34.4% direct threats). This means that the discourse operates in a grey zone between words and harm, a particularly dangerous zone in cases involving women because threats do not always present themselves in direct form. Direct threats are clear: an explicit call for violence, a physical threat, blackmail, or pursuit. Indirect threats work differently, through insinuation, social incitement, accusations of disloyalty, references to 'deterrence,' or opening the way for hostile comments. In a charged context, insinuation alone may be enough to trigger a wave of harm.

This reading requires contextual rather than purely literal precision. Assessing severity solely through the wording of a phrase is insufficient; the phrase must be weighed against its context, audience, repetition, and the status of the publisher, in line with the Rabat Plan of Action standard mentioned earlier. The same phrase may be less serious in a limited context and more serious when it is published by a public page, repeated, or invokes religion, clan, or a hostile audience. The danger lies not only in what has already happened but in what may happen: hate speech against women may not always end in direct assault, but it creates an environment in which assault becomes possible, justified, or socially intelligible. That in itself constitutes a serious human-rights impact.

Human-Rights, Social, and Security Impact:

The broadest impact in the sample is social harm: stigmatisation, reputational damage, isolation, and psychological pressure. This impact should not be minimised because reputation in the Iraqi context is not a purely personal matter; it is a social, professional, and security resource. When a woman is targeted by moral or religious defamation, the damage may reach her family, work, relationships, safety, and ability to remain active in public life. These campaigns therefore produce widespread self-censorship: they do not need to silence every woman directly. It is enough for other women to see what happens to activists and women defenders and then withdraw from speaking, appearing, or participating. Hate speech thus functions as an indirect tool for emptying the public sphere of women or keeping them within it under conditions of fear, caution, and silence.

Reporting, Removal, and Platform Response: The gap between reporting and removal reveals a deep problem in response. More than half of the cases (55.6%) were reported, yet the removal rate remained limited at only 21.2%.



Security impacts are also present, although numerically less frequent than social harm, and they cannot be separated from it. Social stigma may be a precursor to security risk; moral defamation may turn into a family or community threat; and digital incitement may become real-world pursuit. The distinction between 'social harm' and 'security risk' should therefore remain analytical rather than rigid, because in practice the two pathways feed into one another. Taken as a whole, this discourse affects a wide set of rights: dignity, non-discrimination, personal safety, privacy, freedom of expression, freedom of association, the right to work, and the right to public participation. From a feminist perspective, it is a form of gender-based digital violence because it targets women on the basis of stereotypes and imposed social roles and affects them disproportionately.

Platform responses should therefore move beyond deleting individual posts and instead read the pattern of the campaign itself: Are phrases repeated? Is the same entity being targeted? Are multiple accounts participating? Does the discourse use locally coded language? Does it involve gendered stigmatisation or indirect threats? Without this kind of analysis, the response remains superficial and the campaign returns in another form. The UN Guiding Principles on Business and Human Rights establish the responsibility of companies to respect human rights, including by identifying, preventing, mitigating, and addressing adverse impacts connected with their operations or services. This places a particular responsibility on digital platforms when their technical infrastructure becomes an environment for reproducing harm against women, women defenders, and organisations.

This gap does not reflect only technical weakness in platform mechanisms; it also reveals shortcomings in understanding the local context of gendered discourse. Much of the discourse does not use direct words sufficient to trigger automated review tools, yet it carries clear risk through insinuation, sarcasm, moral and religious stigmatisation, accusations of disloyalty, accusations of foreign allegiance, and social incitement. Review systems that rely on explicit vocabulary therefore often fail to capture the actual danger. The gap creates a double effect: targets feel that reporting does not protect them, while publishers realise that the likelihood of accountability is limited. Digital impunity thus becomes part of the problem itself, encouraging repetition and increasing the boldness of pages and accounts.

Indicators of Repetition and Coordinated or Semi-Coordinated Campaigns: The repeated targeting of the same entity in every monitored case (250 cases, the entire sample) is one of the report's most sensitive and significant findings. This repetition appears through identical phrases, multiple accounts posting the same content, or synchronised timing. These indicators alone are not enough to prove that a single centralised structure directs all the attacks, a conclusion that would require additional technical and investigative evidence not available here. They do, however, reveal a pattern that goes beyond passing individual behaviour: the target does not face one post but a recurring wave of pressure that changes the description of the case from 'abuse' to 'targeting' and from a 'comment' to a 'campaign.' Even without a clear operations room, the practical result is almost the same: amplification of the message, surrounding the target, pushing audiences to participate, and producing the impression that the attack represents a social 'consensus' rather than the view of a small group. The difference between abuse and repeated targeting is fundamental in human-rights analysis. Abuse may be individual and limited; repeated targeting has a cumulative effect that exhausts the victim, weakens her ability to respond, disrupts organisational work, and pushes targets toward withdrawal, silence, or reduced visibility. Indicators of repetition should therefore be central to risk assessment. A single post may not appear sufficient to establish a threat, but repetition dozens of times, copying across multiple accounts, or synchronisation with defamation campaigns increases the level of risk and changes the nature of the violation.

From a feminist perspective, the discourse as a whole reveals an attempt to return women to a position of silence through modern digital tools grounded in traditional structures: honour, shame, obedience, reputation, and moral accusation. The phenomenon should therefore not be confined to the label of 'cyberbullying.' It is a form of gender-based digital violence that intersects with political and social misinformation, weak legal protection, and shortcomings in the responses of platforms and institutions. These findings do not justify general censorship of expression. What they require is a precise human-rights response that protects freedom of criticism while ensuring that women, women defenders, and organisations are not left exposed to incitement, defamation, and collective stigmatisation. The objective is not to silence public debate, but to prevent it from being transformed into a tool for silencing women.



Analysis of Responsibilities:

- Primary responsibility lies with the State because it bears the obligation to protect rights, prevent discrimination, investigate threats, and ensure accountability. Institutions cannot treat hate speech merely as disagreement on social media when the discourse contains incitement, threats, or organised defamation. In such cases, the duty to protect is engaged, particularly where the targets are women human rights defenders or women working in public life.
- Security and judicial authorities are responsible for treating complaints seriously without turning the fight against hate speech into a tool for restricting criticism. What is required is a clear standard: protect legitimate criticism while pursuing threats, incitement, organised defamation, blackmail, and hacking. Failure to maintain this distinction may produce two harmful outcomes: impunity for perpetrators on the one hand, or restrictions on freedom of expression on the other.
- The Council of Representatives is responsible for closing legislative gaps, particularly with regard to the protection of women and men human rights defenders, combating domestic violence, and regulating digital
- legislation should protect women and organisations from incitement and threats, not give authorities an overly broad tool for punishing opinion.
- Digital platforms, foremost among them Meta given the centrality of Facebook in the sample, are responsible for improving reporting and review tools and developing contextual understanding of Arabic, local dialects, and gendered stigmatisation. Post-hoc deletion is not enough. Platforms need to monitor recurring campaigns, connect reports concerning the same target, and treat indirect threats and moral defamation as genuine human-rights risks

- Media outlets and public pages are responsible for not recirculating hateful discourse or giving it professional cover. Republishing unverified accusations, turning targeted women into objects of ridicule, or packaging defamation as 'public opinion' are all practices that expand the harm.
- Civil society organisations, for their part, carry a dual responsibility: documenting violations, supporting targets, developing digital-security protocols, and avoiding the republication of abuse during advocacy. Good advocacy does not re-expose the victim to danger; it moves the issue from the level of scandal to the level of rights, protection, and accountability.

Conclusions:

The report shows that digital hate speech against women, activists, and civil society organisations in Iraq operates as a recurring structure of pressure rather than as a series of separate individual incidents. It draws on moral, religious, and gendered stigmatisation and uses defamation, misinformation, and incitement to raise the cost of women's participation in public life. The data show that targeting intensifies when women become actors: activists, defenders, organisation workers, or participants in public debate. This makes the discourse a tool for regulating the public sphere in ways that restrict women's participation and weaken civil society. The findings also show that indirect threats are no less serious than direct threats in the Iraqi context. Insinuation, moral defamation, accusations of disloyalty, and generalisation can produce broad social and security effects, particularly when repeated and disseminated through public pages or multiple accounts. The gap between reporting and removal indicates that platforms remain unable to identify many forms of gender-based digital violence, especially when they appear in coded or sarcastic language or are framed through religion and morality. This leaves targets facing recurring campaigns without adequate protection. Digital hate speech cannot be separated from the wider context: declining legal guarantees, weak accountability, the absence of a law protecting women and men human rights defenders, and growing suspicion toward human-rights and feminist vocabulary. These factors do not create the discourse by themselves, but they provide an environment in which it can spread. The most important conclusion is that the phenomenon is not merely 'cyberbullying' It is a form of gender-based digital violence that intersects with political and social misinformation and targets women and organisations as actors in the public sphere.

Recommendations to the Government:

- Adopt a national policy to address technology-facilitated gender-based violence based on protection and accountability rather than general surveillance or restrictions on freedom of expression.
- Issue an official position rejecting the use of concepts such as CEDAW, gender, women's empowerment, and equality as tools of stigmatisation or accusations of disloyalty against women and organisations.
- Establish a publicised and safe mechanism for receiving reports from women, women human rights defenders, and organisations concerning digital threats, while guaranteeing confidentiality and preventing additional risks to those who report.

To the Iraqi Council of Representatives:

- Enact a law protecting women and men human rights defenders that includes protection from digital threats, defamation, incitement, and retaliation.
- Adopt a law combating domestic violence that provides effective protection for women and girls within the family and the community.
- Review digital legislation to ensure that it is not used against criticism, journalism, or civil-society activism, while clearly criminalising threats, incitement, blackmail, and violations of privacy.



To the Judiciary and Security Authorities:

- Issue a code of conduct preventing media outlets and public pages from reproducing gendered or moral defamation against women.
- Review any directives or practices that demonise human-rights and feminist terminology or use such terminology to restrict public debate.
- Ensure that content regulation targets incitement, threats, and discrimination rather than freedom of expression or legitimate criticism.

To the High Commission for Human Rights:

- Launch periodic monitoring of digital hate speech against women, women human rights defenders, and organisations.
- Publish public reports on patterns of gender-based digital violence and its impact on public participation and the protection of women defenders.
- Coordinate with feminist and human-rights organisations to prepare a national guide for monitoring and documenting hate speech against women.



To the Communications and Media Commission:

- Treat digital threats against women and women human rights defenders as risk indicators rather than as personal disputes or passing posts.
- Train police and judicial units to understand gender-based digital violence, including indirect threats, moral defamation, blackmail, and coordinated campaigns.
- Ensure that victims are not blamed or pushed into silence or tribal and social settlements in cases involving threats and defamation.

To Digital Platforms, Particularly Meta:

- Develop review tools that take into account the Iraqi context, local dialects, and moral, religious, and gendered stigmatisation.
- Establish an expedited reporting pathway for cases concerning women human rights defenders and civil society organisations.
- Link repeated reports targeting the same entity rather than treating every post as a separate case.
- Treat indirect threats and coordinated campaigns as human-rights risks even when they do not contain explicit violent language.

To Media Outlets and Public Pages:

- Refrain from republishing content that contains moral, sexual, or religious defamation against women.
- Adopt editorial rules that distinguish accountability of authorities or institutions from stigmatising or defaming women.
- Do not turn targeted women into material for ridicule, sensationalism, or digital engagement.

To Local and International Civil Society Organisations:

- Develop rapid-response protocols for cases in which women staff members, partners, or activists are subjected to digital hate campaigns.
- Train teams in safe documentation, preservation of evidence, risk assessment, and protection of personal data.
- Avoid republishing abusive expressions or images of victims during advocacy; instead, present the human-rights pattern and violation without recirculating the harmful material.
- Support protection programmes for women human rights defenders in Iraq, particularly in response to digital violence.
- Fund independent monitoring mechanisms for gendered hate speech and coordinated campaigns.
- Support local organisations in digital security, documentation, legal advocacy, and psychosocial support for targeted women.



Sources:

1. ohchr.documents
2.](The Declaration on Human Rights Defenders
3. 1.([ohchr.org](General recommendation No. 35 (2017) on gender-based violence
4. ohchr/publications

Algorithms of Violence Against Women:

A Human Rights Report on Digital Violence Targeting Women, Feminists, and Women Human Rights Defenders

November - 2026