PUBLIC SPHERE SAFETY KIT

2026 EDITION

Practical Tools for Safer Civic Life

Practical tools for safer engagement in Lebanon's civic and political life during periods of heightened risk.

For Women Human Rights Defenders · Women in public and civic life · LGBTQI+ defenders · Allies · Community moderators

HuMENA For Human Rights and Civic Engagement
HuMENA pour les Droits de l'Homme et l'Engagement Civique
هيومينا لحقوق الإنسان والمشاركة المدنية

INNOVATION FOR CHANGE MIDDLE EAST & NORTH AFRICA

CIVICUS

Digital Democracy Initiative

humena.org info@humena.org

NAVIGATION

Table of Contents

ORGANISATION

About HuMENA

HuMENA for Human Rights and Civic Engagement is an independent, non-profit organization that works to protect and expand civic space in the Middle East and North Africa. Based in Brussels with a regional office in Beirut, it supports human rights defenders and civil society actors, especially those in exile and the diaspora, through advocacy, research, and training.

Its work spans the core civic freedoms of expression, peaceful assembly, and association, both offline and online. HuMENA helps individuals and movements organize safely and sustain their activism under repression.

The organization works closely with women, refugees, LGBTQI+ people, and other marginalized communities, and builds gender equity and social justice into its programs. Documentation, advocacy, and artivism are central to its approach.

FOREWORD

This Kit is shaped by the conditions Lebanese women, women human rights defenders, LGBTQI+ activists, and civic actors have been living with for the last several years — and by the specific risks of online violence intensifying around moments of public visibility.

It comes out of a project HuMENA has been carrying through 2026 with support from the Innovation for Change MENA Hub, under the Digital Democracy Initiative. The Kit's content was developed in dialogue with WHRDs, civic actors, and LGBTQI+ practitioners through validation workshops held in Lebanon in mid-2026, and reflects the patterns they named, the gaps they flagged, and the protections they identified as most needed in the current environment.

Lebanon's political, legal, and digital environment evolves continually, and HuMENA reviews and revises this Kit as conditions change. For inquiries and partnerships, write to lebanon@humena.org.

> If you are using this Kit during an incident, act on the most immediately applicable guidance first and reach out for support afterward.

ACKNOWLEDGEMENTS

The Public Sphere Safety Kit benefited from the time, insight, and lived experience of Women Human Rights Defenders, LGBTQI+ activists, women civic actors, and allied practitioners who took part in HuMENA's validation workshops in Lebanon. By their own request, individual participants are not named here. The patterns they identified, the gaps they corrected, and the protections they prioritised shaped every section of this document.

The Kit was developed and edited by Christy Iskandar (lead author and researcher) with the HuMENA team, under the supervision of Mostafa Fouad (Executive Director). Production was made possible with support from the Innovation for Change – MENA Hub and CIVICUS under the Digital Democracy Initiative.

KEY TERMS

Glossary

Short definitions of acronyms and Lebanese-context terms used throughout this Kit.

CLDH
Lebanese Center for Human Rights — non-partisan rights documentation and legal services.

Doxxing
The deliberate exposure of a person's private or identifying information online, typically to enable harassment, intimidation, or offline harm.

ISF
Internal Security Forces — the Lebanese national police. Includes the Cyber Security Crime Unit. See Section 9.2 Important Safety Considerations Before Reaching Out for safety considerations before approach.

LGBTQI+
Lesbian, gay, bisexual, transgender, queer, intersex, and other sexual and gender minorities.

NCII
Non-Consensual Intimate Images. The sharing or threatened sharing of sexually explicit images or videos without consent. See Section 3.6 Non-Consensual Intimate Images and Deepfake Sexual Content.

NSSF
National Social Security Fund — the Lebanese social-security institution.

Outing
Disclosure of a person's sexual orientation, gender identity, or other sensitive identity-related information without their consent.

PSEA
Prevention of Sexual Exploitation, Abuse, and Harassment — a set of standards applied to staff and partners of humanitarian and civic organizations.

StopNCII.org
Free hash-matching service operated by the Revenge Porn Helpline that pre-emptively blocks known intimate images on participating major platforms. stopncii.org

TFGBV
Technology-Facilitated Gender-Based Violence. Any act carried out, supported, or amplified through digital technologies that causes physical, sexual, psychological, social, political, or economic harm on the basis of gender.

WHRD
Woman Human Rights Defender — a woman who advocates for human rights, or a human rights defender who works on women's and gender rights.

Article 534
Article 534 of the Lebanese Penal Code, used historically to criminalise same-sex relations and to detain LGBTQI+ individuals. Cited throughout this Kit as a risk factor when LGBTQI+ users consider approaching state institutions.

Article 582
Article 582 of the Lebanese Penal Code (criminal defamation). Used to prosecute individuals who publicly criticise officials or political figures, including counter-statements made in response to TFGBV. See Section 5.5 Decide on the Next Response Path.

1

SECTION 1

About This Kit

Lebanon's election periods are marked by heightened risks and digital violence in the public sphere, particularly for women in public life, Women Human Rights Defenders (WHRDs), and LGBTQI+ defenders. These attacks are often used to punish public visibility, deter civic and political participation, enforce self-censorship and withdrawal, and undermine representation in the public sphere.

In response, this safety kit has been developed as a practical, action-oriented resource that provides tools to support safer engagement in civic and public life during politically tense periods, and elections. It is designed as a concise and adaptable reference, offering accessible guidance that can be applied by users based on their needs when faced with Technology Facilitated Gender Based Violence (TFGBV), doxxing, outing, impersonation, and coordinated harassment.

The kit was developed through a practice-based and participatory process grounded in lived experiences of Technology Facilitated Gender-Based Violence (TFGBV) in Lebanon. To refine its content, two closed sessions are planned with Women Human Rights Defenders (WHRDs), women civic actors, and LGBTQI+ defenders. The first session is designed to review and inform the kit's outcomes and guidance; the second will follow with the same stakeholders to confirm the relevance, accuracy, and safety of the content. The kit will be revised after these sessions and a final version released in the next reporting period.

1.1 Purpose of the Kit

The Public Sphere Safety Kit aims to enhance practical preparedness and effective first response to TFGBV and promote safer participation in the public sphere during high-risk political periods. It seeks to strengthen safer online practices and improve responsiveness to digital threats to mitigate the impact of online violence on civic engagement.

It is intended to strengthen preparedness and enable safer, more informed responses to online violence during politically sensitive periods. The kit can be used as a rapid reference guide and preventive tool to identify risks when experiencing or anticipating TFGBV.

1.2 Who This Kit Is For

This kit is intended for: Women Human Rights Defenders (WHRDs), women in public and civic life including candidates and civic actors, and LGBTQI+ defenders and activists. It also targets allies and support networks of LGBTQI+ communities, as well as civic actors, community organizers, and those managing digital spaces including moderators, administrators of community pages, and digital community managers navigating election-related risks. More broadly, the kit is relevant to civic groups and individuals engaged in Lebanon's electoral process who may face similar forms of digital attacks and require practical tools to support safer participation.

PRIMARY AUDIENCES

- Women Human Rights Defenders (WHRDs)
- Women in public and civic life, including candidates and civic actors
- LGBTQI+ defenders and activists
- Allies and support networks of LGBTQI+ communities

ALSO FOR

- Civic actors and community organizers
- Moderators and administrators of community pages
- Digital community managers navigating election-related risks
- Civic groups and individuals engaged in Lebanon's electoral process

1.3 What This Kit Covers, and What It Does Not Cover

This kit provides practical, accessible tools to support basic preparedness, early identification risks, and immediate first response actions in the face of online harassment during periods of elections. It focuses on user-led, low-resource strategies that can be applied by individuals, civic actors, and those managing digital spaces to enhance safety and reduce exposure to harm.

It does not, however, constitute or replace specialized legal, psychosocial, or digital security support, nor does it offer comprehensive or case-specific solutions. The kit does provide referral guidance, but it is intended as a complementary resource that can inform safer practices, while recognizing that certain situations may require professional support or intervention.

2

SECTION 2

Why Political and Election-Linked Periods Increase Digital Risk

This Kit is published in 2026, in the aftermath of the 2023–2024 war and against the backdrop of Lebanon's ongoing economic crisis, shifting political alliances, and an environment in which civic space has both narrowed and re-organised. Online violence in this period intensifies not only around scheduled elections but around every moment of public visibility — protests, parliamentary testimony, public statements on the reconstruction process, advocacy on humanitarian access, and any act of organising that becomes legible to opposing actors. The election-period emphasis throughout this Kit reflects the most acutely studied pattern, but readers should treat the same tools as applicable year-round, with intensity tracking the visibility of the target rather than the calendar.

Periods of heightened political activity, particularly elections, are often accompanied by increased digital campaigning and public visibility. As candidates, activists, and citizens engage more actively online, they become more exposed to hate speech, harassment, and coordinated online attacks, especially when political ideologies and cultural values clash. These periods are characterized by heightened polarization where online spaces become arenas for contestation, amplifying tensions and increasing the likelihood of abuse.

Moreover, the anonymity and wide reach of social media exacerbate the issue, as it enables perpetrators to launch large-scale, coordinated attacks with little to no accountability. Recent shifts in platform moderation policies, especially for Meta and X, have weakened safeguards against hate speech, contributing to a more permissive online environment that proliferates gendered harassment and misogyny under the guise of freedom of expression. In fact, within 3 months of changes to Twitter's policies, attacks on women journalists tripled compared to the same period in previous years. These conditions embolden perpetrators and reduce the likelihood of accountability[1].

Structural gaps in protection mechanisms further deepen the problem and erode accountability. Between 2020 and 2023, **80% of digital violence cases in Lebanon targeted women**, and an estimated 30-40% of online abuse escalates into physical threats or other forms of offline harm[2]. Despite the severity of such cases, legal protections remain limited, with insufficient mechanisms to effectively address online abuse or hold perpetrators accountable. This is largely due to an outdated legal framework that does not adequately respond to cybercrimes, including gaps in criminalizing sextortion, prosecuting online abuse, or ensuring meaningful protection for victims.[3] Therefore, those targeted, particularly women political activists, are left increasingly exposed to escalating harm, with limited legal recourse, or justice.

Lebanon's cultural and social norms further shape exposure to digital risk during political periods. Entrenched patriarchal attitudes continue to frame politics and leadership as predominantly male fields, often casting women as less legitimate actors, while issues related to sexual orientation ad gender identity remain highly sensitive in public discourse. In such a context, individuals who challenge these norms, particularly women and LGBTQI+ persons, are more likely to be targeted through gendered, moral, and reputational attacks that seek to discredit their public participation.

As a result, periods of political competition and tensions create an enabling environment where digital violence can thrive and may push women and other marginalized groups to withdraw from public engagement or be discouraged from entering political life altogether. This not only reinforces gender disparities in political leadership but also contributes to the gradual shrinking of civic space and the exclusion of diverse voices from the public sphere[4].

2.1 Digital Violence as a Participation-Silencing Tactic

Digital violence is used during politically sensitive periods as a deliberate tactic to undermine and discredit opponents, suppress participation, and shape public discourse. Attacks such as defamation, smear campaigns, threats, and harassment aim to punish visibility and delegitimize individuals, particularly those expressing dissenting views or challenging established political actors.

The 2022 Lebanese parliamentary elections witnessed a surge in systemic digital violence directed often at women participating in public life. In most cases, political parties and networks deployed such organized attacks and coordinated campaigns, often through "electronic armies" of managed accounts, to discredit women candidates, and influence public perceptions. These practices are closely tied to broader political competition, where digital attacks are used to strategically weaken opponents and shape narratives.[1] Women candidates, particularly those running against established parties and promoting differing opinions, are disproportionately targeted as a means of undermining their credibility. The reach of these digital attacks challenges political position, delegitimizes women's presence in the political sphere, and reinforces culturally embedded misogynistic values. They amplify gendered and sexualized content that entrenches discriminatory attitudes and normalizes exclusion.[2]

> For instance, candidate Dima Abou Daya, who was running for the Shia seat on the "Zahle for Sovereignty" list in 2022, faced a sustained and coordinated campaign of online abuse after challenging dominant political actors in her region.[3] She reported receiving severe threats from anonymous and fake accounts, which posed direct threats to her safety and forced her to temporarily leave her hometown. Various online platforms and websites also amplified smear narratives targeting her personally and politically. As attacks escalated and spread to offline harm, her parents issued a public statement disavowing her candidacy and disowning her.[4]

Dima Abou Daya's case clearly illustrates how digital violence can be coordinated, sustained, and deeply destructive. It impacts personal safety online and offline, affects social standing, and disrupts political engagement. In fact, dominant political parties, often built on patriarchal ideologies, hold significant influence over digital spaces and public discourse. This influence is mobilized to coordinate and amplify targeted harassment, through organized networks and online supporters, these actors can sustain campaigns of intimidation and reputational attacks that go beyond political contestation to directly target individuals. In doing so, digital violence punishes visibility and pushes individuals out of public debate.

2.2 Why Women, WHRDs, and LGBTQI+ Defenders Face Heightened Exposure

The shifting legal and social conditions facing LGBTQI+ communities in the MENA region, and the specific transformation of Lebanon between 2020 and 2023, are documented in HuMENA's own published research [5]. The political pressure on women in public life and on LGBTQI+ defenders described in the rest of this section sits within that broader pattern.

Exposure to digital violence during politically sensitive periods disproportionately affects Women, WHRDs, and LGBTQI+ defenders due to the intersection of gender, social norms, and political visibility. Their participation in public life often challenges established norms and power structures, making them more likely to be targeted through deliberate and identity-based attacks.

Women are particularly vulnerable to these forms of abuse due to the deeply rooted gender norms framing them illegitimate political actors. As a result, they are more likely to be targeted with gendered, personal, and reputational attacks that seek to undermine their credibility, and deter their participation in public life.[1] Online abuse directed at women is often sexualized and invasive, focusing on their appearance, behavior, and perceived morality rather than their political positions or qualifications. In many cases, women candidates are subjected to commentary on their looks, clothing, or personal lives, shifting attention away from merit and reinforcing harmful stereotypes that limit their role in public life.

Lebanon's cultural and social norms shape how women are perceived and treated in public life. Engrained patriarchal values portray women as unfit or less capable of holding authority. During intense political periods, such narratives are amplified and mobilized to weaken women's participation, discredit and undermine their legitimacy, particularly those challenging established political structures. This patriarchal culture normalizes gendered abuse in digital spaces and delegitimizes women's capacity to participate in public life, further limiting equitable political representation.

WHRDs face additional layers of risk due to the nature of their work, particularly when addressing and advocating for sensitive issues such as accountability, governance, and gender equality. Their activism often places them in direct opposition to powerful actors, increasing the likelihood of coordinated attacks, surveillance, and intimidation. This pattern is reflected across the MENA region. A study conducted by Fe-Male on online violence against WHRDs in the MENA region (53 out of 115 participants were Lebanese activists) found that **78.3% of respondents reported experiencing violence online**, most commonly in the form of sexist, racist, or homophobic messages.[2] Additionally, 97.4% have experienced at least one form of gender-based discrimination, including commentary on appearance or lifestyle, and threats of a sexual nature.[3] These patterns highlight the extent to which digital violence against WHRDs is both widespread and deeply gendered, reinforcing efforts to silence their engagement in public life.

Within these cultural norms, sexual orientation and gender identity remain highly sensitive in Lebanon, particularly in political contexts, where visibility can carry significant social, personal, and legal repercussions. While consensual same-sex conduct is not explicitly criminalized in the Lebanese penal code, article 534 has historically been used to punish consensual same-sex relations, despite several rulings between 2007 and 2018 affirming that consensual same-sex relations are not illegal. In July 2023, members of the parliament submitted a draft law to repeal article 534.[4] However, signatories to the proposal were subsequently subjected to online harassment campaigns from political and religious actors. This prompted one parliamentarian to withdraw his signature. As a result, most individuals who may identify as LGBTQI+ do not run openly, and discussions around their identities often emerge through speculation rather than self-identification. Nevertheless, digital violence does intersect with perceived or actual sexual orientation during election periods, or periods of heightened political tensions. Candidates and LGBTQI+ activists have at times been subjected to rumors, insinuations, or accusations about their sexuality to discredit them. This form of speculative outing is used strategically to shift attention away from political platforms and instead frames candidates as socially deviant or morally unfit within Lebanon's conservative setting. As a result, outing is used as a tool of digital violence during elections to police who are culturally considered acceptable in public life, reinforcing homophobic rhetoric and exclusion from political space.

In a political context where even women's participation is restricted, LGBTQI+ individuals face greater barriers to entry. The social stigma, potential legal risks, and the likelihood of targeted digital attacks makes openly running for office impossible, often forcing individuals to conceal aspects of their identities or refrain from engaging in political life altogether. This further restricts civic space and limits inclusive political participation.

SECTION 3

Types of Attacks Covered by This Kit

This section gives users a common language for the harms they may be facing. Accurate naming matters because it shapes the response path, urgency, and support needs.

The main forms of online attacks covered in the kit include TFGBV, doxxing, outing, impersonation, and coordinated harassment. This section aims to define each type of violence to allow users to identify the threat, determine the level of risk, the urgency of response, and the type of support that may be needed. In practice, it is important to note that digital threats are often interconnected and may occur simultaneously.

3.1 Technology-Facilitated Gender-Based Violence

TFGBV refers to any act carried out, supported, or amplified through the use of information and communication technologies and digital media that results in, or is likely to result in, physical, sexual, psychological, social, political, or economic harm, as well as violations of rights and freedoms.[6] While originally defined in relation to violence against women, the term TFGBV is used more broadly to include harm affecting individuals based on their gender identity or sexual orientation.[7]

In online spaces, TFGBV manifests through a range of interconnected behaviors, including gendered abuse, sexualized threats, humiliation, and intimidation which may often occur simultaneously.[8] These frequently take the form of degrading language targeting a person's gender or identity, unwanted sexual comments or threats intended to instill fear, or the sharing and manipulation of content to shame and discredit individuals publicly. Such abuse is frequently repeated, coordinated, or amplified across platforms, allowing it to spread quickly and reach wider audiences. Perpetrators may use different tactics, such as harassment, stalking, or the non-consensual sharing of sexual content, to increase pressure and maintain control. As a result, these forms of violence can escalate over time, intensify visibility and impact, and create sustained environments of fear that discourage individuals from participating in public and civic life.

TFGBV IN PRACTICE

- The use of degrading or gendered and sexist language
- Unsolicited sexual messages or threats, targeted comments on appearance or morality, and/or
- Attempts to provoke fear, shame or silence.

3.2 Doxxing

Doxxing refers to the deliberate exposure of a person's private or identifying information online without their consent.[6] This may include details such as phone numbers, home addresses, workplace information, family connections, or other sensitive data. While some of this information may already exist online, its targeted disclosure in a hostile context can significantly increase risk by making individuals easier to locate, contact, or harass. Doxxing can escalate quickly and spread into offline threat and harm. Certain types of disclosures, such as home addresses, real-time location, personal contact information, are especially dangerous, as they heighten vulnerability and undermine an individual's ability to maintain safety and privacy online and offline. This risk is particularly heightened during periods of political tensions, such as elections, as it may intensify targeting and amplify the severity of attacks.

DOXXING IN PRACTICE

- Sharing of phone numbers, email addresses, home, work or real-time locations without consent and meant to mobilize people to cause harm,
- Sharing of sensitive information of family members,
- Receiving increased numbers of unknown calls and messages that include threats or harassment, and/or
- Threats implying knowledge of private and personal information.

3.3 Outing

Outing refers to the disclosure of a person's sexual orientation, gender identity, or other sensitive aspects of their identity without their consent.[6] In practice, this may involve sharing private information, images, or personal details in ways that expose individuals to others before they are ready or willing to do so themselves. In Lebanon, where laws and social norms may be used to target LGBTQI+ individuals, outing can create immediate and serious risks, including social stigma, family rejection, loss of employment, and potential legal or physical harm. These risks may be further heightened during periods of increased public visibility, where targeted exposure can be used as a tool of intimidation to silence and exclude individuals in the public sphere.

OUTING IN PRACTICE

- The spread of online rumors, insinuations, or direct claims about someone's sexual identity or orientation,
- Pressure to confirm or deny one's identity,
- The circulation of private conversations, images, or details related to someone's sexual identity or orientation, and/or
- Blackmailing, and threats to reveal someone's identity unless they comply or remain silent.

3.4 Impersonation

Impersonation refers to the creation of false identities, fake accounts, manipulated profiles, or false attribution to pose as a targeted person without their consent. In practice, impersonation is used to spread false information in the name of victims, damaging their reputation, provoking harassment, or exposing them to further abuse or security risks. This may take the form of fraudulent accounts created to publish harmful or misleading content, fabricated statements or images falsely attributed to the target, or deceptive communications intended to mislead others, compromise trust, or solicit information under false pretenses. In periods of heightened political tension, impersonation may be used to undermine credibility, distort political views or affiliations, or incite backlash. This can erode personal safety, damage professional and political standing, and ultimately deter participation in public and civic life.

IMPERSONATION IN PRACTICE

- Fake accounts using someone's name, image, or identity without authorization,
- Releasing statements or content in someone's name,
- Edited screenshots or quotes falsely attributed to the target,
- Reports from friends about suspicious communication and messages received from accounts in the target's name, and/or
- Messages sent to others pretending to be someone else to gather information.

3.5 Coordinated Harassment

Coordinated harassment is targeted abuse carried out by multiple individuals, groups, or networks acting in a deliberate and organized manner to intimidate, silence, discredit, or overwhelm a person. Unlike isolated incidents of abuse, coordinated harassment involves repeated or simultaneous attacks that may be amplified across accounts, platforms, or communication channels, increasing both the scale and intensity of harm. This could be executed in the form of mass targeting through abusive messages, trolling campaigns, or the deliberate spread of harmful content intended to provoke fear. These attacks may intersect with disinformation, impersonation, threats, or surveillance, and can rapidly escalate through amplification due to organized networks or automated accounts. In the case of coordinated harassment, harm is generated collectively rather than through a single perpetrator and often exceeds the capacity of routine responses. This often changes the severity of responses needed, requiring rapid risk assessment, pattern documentation, collective protection measures, and escalation to institutional, legal or security support if possible. As a result, coordinated harassment can create a sustained environment of intimidation that forces individuals to self-censor, withdraw or disengage from public life, affecting both personal safety and inclusive civic and political life.

COORDINATED HARASSMENT IN PRACTICE

- Sudden increase in hostile comments or messages from different accounts or people,
- Similar or identical content posted across multiple accounts,
- Organized attacks and campaigns occurring simultaneously across platforms and accounts,
- The use of accounts that appear to be automated, anonymous, or newly created sharing hostile content and messages, and/or
- Multiple individuals or group of people mobilizing others to attack, report, or discredit a person simultaneously.

3.6 Non-Consensual Intimate Images and Deepfake Sexual Content

NCII refers to the sharing, distribution, or threatening distribution of intimate, sexual, or nude images of a person without their consent.

The non-consensual sharing of intimate images, including real images and AI-generated or manipulated sexual content, is one of the most severe forms of digital violence faced by women in public life. It is designed to cause psychological harm, reputational damage, and social pressure to withdraw from public engagement. In the MENA region, this tactic is increasingly documented against women politicians, candidates, and human rights defenders, including through AI-generated deepfake content that fabricates intimate images of real individuals without any real image ever existing. The threat of exposure is often as damaging as the exposure itself.

Women and LGBTQI+ activists and candidates can be targeted with fabricated NCII, and its fabricated nature does not reduce its harm. In fact, deepfakes spread rapidly, are difficult to definitively debunk without forensic analysis, and cause the same psychological, reputational, and social damage as real images.

SHARING OF NCII AND DEEPFAKES IN PRACTICE

- Real images or videos originally created consensually but shared without consent.
- Images or videos obtained through hacking, device compromise, or coercion.
- AI-generated or manipulated content, including deepfakes, that depicts a real person in sexual or intimate scenarios that never occurred.
- Manipulated content combining real photos of you with fabricated sexual content.
- Threats to share such content as a form of blackmail or coercion, even when no image has yet been distributed.

⚠ IMPORTANT

IMPORTANT: Before pursuing a formal legal complaint to state institutions (i.e. ISF), check Pursuing criminal complaint for NCII: section, and refer to specialized organizations before reaching out to authorities [refer to 9.4 Resource List].

LEBANON-SPECIFIC CONSIDERATIONS

Pursuing criminal complaints involving NCII in Lebanon carries the same risks discussed in Section 9.2 Important Safety Considerations Before Reaching Out, below: state institutions, including the ISF, may not be safe or accessible, particularly for LGBTQI+ individuals or for women whose images can themselves be used against them under prevailing social norms. Consult Legal Agenda, CLDH, or ABAAD before considering any formal complaint, and never approach state institutions with intimate evidence in hand without legal counsel present.

IMMEDIATE RESPONSE PRIORITIES

- **Preserve all evidence** — including the abusive content itself, even when it is distressing — before submitting platform reports, as content is often removed quickly after a report and may be needed for legal action or pattern documentation.
- For deepfake or AI-generated content, **document the discrepancy with verifiable real-world evidence** (genuine photographs from the same period, location records, witness statements) to support takedown requests and any public clarification.
- **Report through each platform's dedicated NCII removal pathway.** Major platforms have expedited review processes for this content category, though response times vary.
- Use StopNCII.org (stopncii.org), a free hash-matching service operated by the Revenge Porn Helpline, to pre-emptively block known intimate images from being shared on Meta, TikTok, X, Reddit, OnlyFans, Bumble, and other major platforms. This works even if the content has not yet been shared publicly.

3.7 Overlapping and Escalating Attacks

These forms of online abuse do not always occur as isolated or neatly separate incidents. In practice, harm often overlaps, reinforces one another, and evolves over time, with one form of abuse creating conditions for another to emerge or intensify. What may begin as targeted harassment, for example, may escalate into coordinated attacks, impersonation, and even doxxing, increasing exposure and vulnerability. Similarly, tactics may be deployed simultaneously, with perpetrators using multiple methods in combination to intensify pressure and maximize harm.

During elections and heightened political tensions, these patterns may escalate rapidly. The layering of harms can increase both the severity and unpredictability of threats, particularly when digital abuse creates offline safety risks, defamation, and barriers to continued participation. For this reason, incidents should be understood and assessed as potentially interconnected rather than treated as single events. Recognizing incidents as layered can support more effective risk assessment, documentation, and response.

OVERLAPPING AND ESCALATING ATTACKS IN PRACTICE

- Harassment escalating into doxxing, and offline threats,
- Impersonation combined with coordinated attacks to amplify harm, and/or
- Coordinated attacks leaking personal information, leading to doxxing and offline threats on the target and their families.

3.8 Targeting of Women Journalists

Women journalists in Lebanon — particularly those covering politics, corruption, gender, sectarian dynamics, or LGBTQI+ rights — are among the most consistently targeted civic actors in the country's online sphere. Maharat, SKeyes, and the Samir Kassir Foundation have documented sustained patterns of harassment, doxxing, and impersonation directed at women journalists across print, broadcast, and independent online outlets. The patterns differ from those facing candidates and WHRDs in several ways:

- Attacks frequently come from networks aligned with the subjects of their reporting — political parties, security services, business interests — rather than from anonymous publics.
- The professional consequence is more direct: editors, advertisers, and outlets receive pressure designed to limit the journalist's assignments or force a retraction.
- Family members and sources of the journalist are often targeted in parallel, exploiting the journalist's professional network.
- Press-freedom organizations (Maharat, SKeyes, Reporters Without Borders) should be the first referral alongside the women's-rights and digital-security organisations covered elsewhere in this Kit.

3.9 Diaspora and Cross-Border Targeting

Lebanese WHRDs, candidates, and LGBTQI+ activists frequently face TFGBV with a cross-border dimension. Three patterns are common: Lebanese women in Lebanon attacked by networks operating from outside the country (the Gulf, North America, or Europe); Lebanese women in the diaspora attacked by networks operating from inside Lebanon; and coordinated campaigns that move across both jurisdictions, exploiting the practical limits of national law enforcement.

Implications for response:

- Platform reporting becomes more important than legal complaint, because platforms operate across borders while courts do not.
- Documentation should record the apparent geographic origin of accounts where this is visible (account creation language, time-zone patterns, country indicators), as this informs which organisations can usefully respond.
- For diaspora users, local women's-rights and LGBTQI+ organisations in their country of residence may be a faster route to support than Lebanese organisations operating across distance.

SECTION 4

Rapid Risk Identification and Severity Triage

Digital attacks can escalate quickly, particularly during periods of heightened political tensions when online violence is politically driven. Early identification of risk is critical to prevent harm, reduce exposure, and determine the appropriate course of action depending on the severity of the case.

4.1 Questions to Ask Immediately

When facing a potential attack, the following questions can help determine the level of risk and urgency of response required. These questions are not exhaustive but are intended to guide immediate situational awareness:

What kind of attacks are happening?

This helps determine the nature of the threat and the likely direction of escalation.

- Is this harassment, TFGBV, doxxing, impersonation, outing, or coordinated harassment? [refer to 3. Types of Attacks Covered by This Kit]
- Are different types of attacks happening simultaneously?
- Is this a single incident or has this been repetitive?

Where is the attack taking place?

This helps assess exposure and spread. Public and cross-platform attacks tend to escalate faster and reach wider audiences, increasing safety risks.

- Is it occurring on public platforms, and across multiple accounts?
- Is it through private messages and involves blackmailing?
- Is it occurring in closed groups?

Has any personal or sensitive information been publicly shared?

This helps assess immediate safety risk, particularly the potential of offline harm, and direct contact.

- Have phone numbers, addresses, workplace, family details, and/or private content been shared or referenced?
- Have repetitive spam calls or messages been received from unknown numbers?

Is the attack coordinated or repeated?

This helps identify organized targeting and harassment. Coordinated attacks are more difficult to control, escalate faster, and often require stronger response measures.

- Are multiple accounts posting similar messages?
- Is the same content appearing across different platforms?

Are there threats, either direct or implied?

This helps determine severity. The presence of threats, especially credible or repeated ones, signals escalation.

- Do messages include language suggesting physical harm, and/or death threats?
- Do messages refer to sexual violence?
- Do messages hint at reputational damage?
- Are messages intended for intimidation?

Is the situation escalating?

This helps track progression, and ensure response addresses proper risk levels.

- Has the frequency and intensity of the attack been increasing?
- Did the tone become more violent?
- Are threats and blackmail increasing?

Does the attack involve others around you?

This helps assess the widening impact. When attacks extend beyond the individual, risks increase significantly, and response needs to shift.

- Are family members, colleagues, or associates being mentioned?
- Are they being contacted?
- Are they being targeted in messages and content?

Is there any indication of offline risk?

This helps identify immediate danger. Any link between online and offline activities significantly raises the level of urgency.

- Has anyone attempted to make direct contact, or locate the target?
- Has there been any reference of the target's movements, or suggest physical proximity?

Can the person or group behind the attacks be identified?

This helps assess capacity and intent. Attacks linked to organized actors or networks may be more sustained, strategic, and harder to contain.

- Are there anonymous, newly created accounts designed specifically to conduct these attacks?
- Are there automated accounts?
- Are there accounts linked to known individuals, groups, or political actors?

The answers to these questions should be considered together rather than in isolation. Even if a situation seems to have low impact, the presence of multiple risk indicators, such as visibility, coordination, or personal data exposure, may signal a higher level of risk and require a more urgent response.

4.2 When Attacks Originate from Intimate Partners or Family

Technology facilitated attacks against women candidates, WHRDs, LGBTQI+ defenders, and civic actors can sometimes, especially during politically active periods, originate from current or former intimate partners, or from networks acting on their behalf.

Attacks originating from intimate partners require different responses and assessment. The associated risks with Intimate Partner Violence (IPV) include:

- **Deeper knowledge of routines and private life.** A current or former partner has direct knowledge of your home address, daily schedule, workplace, frequently visited locations, family members, and personal relationships. This significantly increases the risk of offline harm.
- **Access to private media and communications.** IPV aggressors may have or have previously obtained access to private images, personal communications, or sensitive information that can be used for blackmail, non-consensual sharing of intimate images, or targeted exposure. This access may predate the attack by months or years.
- **Mixed and compounding motives.** IPV-origin attacks are often driven by a combination of personal grievance, desire for control, and political or reputational motivation. This combination can make attacks more sustained, personalized in their content, and harder to contain.
- **Access through shared or linked accounts.** If devices, accounts, cloud storage, or family-sharing systems were ever shared with or accessible to a partner, they may still have access and they may actively exploit it. Refer to section 5.3 Reduce Immediate Exposure for immediate account security steps.

MITIGATION MEASURES

Standard response recommendations in this Kit require careful adaptation when the attack originates from an intimate partner or domestic violence context. The following are appropriate steps to follow:

- **Be selective about who you tell, especially family members.** Before disclosing to family members, assess whether they are likely to be supportive, whether they have existing relationships with the perpetrator, or whether they might apply pressure to reconcile. Disclose only to family members you are certain are safe, and able to support you without involving the perpetrator.
- **Contact KAFA or ABAAD first**, both organizations can offer support and advise on next steps. Refer to section 9.4 Resource List.
- **Assume the perpetrator already knows your routines in detail.** Simple changes to the routine may not be sufficient. Consider: avoid locations the perpetrator knows you frequent; do not announce your whereabouts publicly and even to people you trust until you have assessed who in your circle may be in contact with the aggressor.
- **Do not approach formal state institutions without first consulting KAFA or ABAAD** to assess whether formal reporting is safe. Institutional responses often do not adequately protect the targeted person, and may take a lot of time to respond.
- **Limit who in your network is aware about the situation.** In IPV contexts, information shared within a community can reach the perpetrator through shared social circles, mutual contacts, or well-meaning people who do not understand the risk. Your support network in this context may not include family and friends, but rather organizations that can support and minimize the risk.

4.3 Low, Medium, and High-Risk Situations

This section builds on the risk analysis tool, and explains what each level means in practice, and the type of response it requires. Risk levels should be treated as guidance for action, not fixed categories. Situations can shift quickly, and reassessment should be continuous as conditions evolve.

• Low Risk — Limited exposure, low severity, no immediate threat.

Low-risk situations typically involve isolated incidents with limited visibility and no indication of escalation or harm beyond online spaces. These may include single instances of harassment, low-engagement content, or interactions confined to private messages or small groups. There is no exposure of sensitive personal information, no coordinated targeting, and no direct or implied threat.

In practice: *the situation can be monitored and should not be ignored. Low-level incidents can escalate if left unaddressed.*

RECOMMENDED MEASURES TO TAKE

- **Stop all communication (if any) with the aggressor** and avoid engaging with abusive content. Block them on all platforms that can be used to potentially make contact (e.g.: Gmail, Instagram, Facebook, X, TikTok, Telegram, WhatsApp). *This helps victims reclaim their safety and prevent abuse.*
- **Document everything to preserve evidence** and this includes anything that indicates violence, abuse or exploitation: screenshots, emails, messages, photos received. *This validates the victims' experiences and helps if they choose to pursue legal pathways and report the abuse.*
- **Use platform tools to restrict or report incidents.** On Facebook, Instagram, Telegram, WhatsApp, and TikTok any user or content can be reported to platforms [refer to 8. Platform Reporting Guidance]. *This helps limit engagement and possible contact with the aggressor.*
- **Review and adjust basic account safety settings** by checking who can see your posts and personal information, who can contact or message you, tagging and comments permissions, and enable two-factor authentication if not already active. *This ensures sensitive information is protected, reduces the risk of account compromise, impersonation, doxxing, or unauthorized access.*
- **Consider seeking help** because isolation often exacerbates digital violence. It is important to break the cycle and talk to someone trusted: a friend, family member, or a professional.[9] *This establishes a support system in which the victim does not feel alone and subsequently reduces the psychological impact of the abuse.*

> **Note for LGBTQI+ defenders:** even low-risk situations involving rumors, insinuations, or identity-based content should be assessed with specialized organizations before deciding on a response pathway given the legal risk specific to this context [Refer to 9.4 Resource List].

• Medium Risk — Increased visibility, persistence, and early signs of escalation.

Medium-risk situations involve sustained or expanding attacks that may include repeated harassment, emerging coordinated attacks, impersonation, or partial exposure of personal information. At this stage, content is gaining visibility, it is spreading across platforms and attracting engagement. Threats may not yet be severe or immediate, but the situation shows signs of progression and escalation.

In practice: *the situation requires active response, monitoring, and containment without waiting as it can increase exposure and reduce control over the situation.*

RECOMMENDED MEASURES TO TAKE: (IN ADDITION TO THE MEASURES MENTIONED IN THE LOW-RISK SECTION)

- **Systematically document patterns of abuse** (not just single incidents) keeping record of what is happening over time. Have screenshots of messages and posts, usernames of all accounts involved, URLs, dates and times. *This helps establish patterns of targeting, monitors risks of escalation, and ensures the preservation of evidence in case victims decide to seek legal aid.*
- **Monitor for patterns** of similar messaging, timing, and accounts acting together. *This helps identify whether the situation is shifting towards coordinated harassment.*
- **Avoid actions that may unintentionally amplify harm** and encourage your network to do the same. These actions include resharing hostile content, engaging directly with abusive posts or accounts, or drawing additional attention to harmful profiles, or hashtags. *This reduces the risk of increasing visibility, extending the reach of coordinated campaigns, or intensifying the impact.*
- **Limit public exposure** by reviewing posts, visibility and tagging options to reduce the amount of personal and identifying information online. Review recent posts, adjust audience settings, and restrict who can tag you or comment. *This minimizes the information available online that can potentially be used for further targeting.*
- **Inform trusted individuals and networks** for support to reduce isolation and allow others to help monitor and respond to the situation. *This creates a support system as the situation escalates and allows others to monitor and document evidence without further exposing the victim to the abusive content.*
- **Inform audience or network of any fake accounts** impersonating victims to delegitimize them and limit reputational damage. *This limits the effectiveness of impersonation, and the spread of misinformation.*

• High Risk — Serious threats, exposure, coordination, or likelihood of harm.

High-risk situations involve significant indicators of harm, including doxxing, threats, viral coordinated harassment campaigns, outing threats, or sharing of NCII and deepfakes. This level of intensity includes rapid escalation, wide dissemination, and targeting that extends beyond the individual to family, friends or colleagues and has offline implications.

In practice: *response needs to be immediate, online and physical safety are a priority.*

RECOMMENDED MEASURES TO TAKE: (IN addition TO THE MEASURES MENTIONED PREVIOUSLY)

- **Prioritize personal safety and reduce exposure:** avoid real-time posting (especially posts that include location), turn off location sharing on all platforms, and avoid engaging with aggressors. *This helps reduce immediate vulnerability and limits opportunities for further targeting.*
- **Report to relevant authorities** (where safe and appropriate) or organizations that can support. Evaluate whether formal reporting is a viable safe option to seek legal and security support to ensure protection against threats and offline harm. *This contributes to seeking justice and accountability while ensuring the physical safety of victims.*
- **Consider changing the phone number** in cases of severe exposure and in case victims are receiving persistent harassment, repeated calls, messages or threats. Changing the phone number may help regain control and reduce immediate harm. Before doing so, victims should inform trusted contacts and secure access to essential accounts linked to the old number. *This helps interrupt ongoing targeting and prevents further direct contact.*

If you are unsure about the level of risk, it is safer to treat the situation as **high-risk** and move to protective action.

4.4 Signs of Immediate Danger

Certain indicators suggest that a situation may escalate quickly or has already moved beyond digital harm into immediate safety risk. When one or more of the following signs are present, the situation should be treated as high-risk and urgent, requiring immediate protective action and support.

- **Direct threats of physical harm or violence** including explicit threats of assault, killing, or sexual violence against victims or others, whether sent privately or publicly.
  → Stop engagement, preserve evidence, and inform a trusted person or network immediately.
- **Harassment and threats extended to family members, colleagues, and friends** including attempts to contact, or expose people in the victim's personal life or professional network.
  → Inform those affected and targeted and coordinate a shared response.
- **Doxxing and sharing of personal data** such as home address, phone number, workplace, or real-time location, allowing others to locate or contact victims.
  → Alert close contacts and take steps to ensure physical safety whether by contacting authorities or temporarily relocating.
- **Stalking, offline tracking, surveillance and approach** as a result of doxxing.
  → Avoid predictable routines, ensure the victim is never alone if possible, and inform a trusted person of the situation.
- **Threats or sharing of NCII and deepfakes.**
  → Do not delete the content yourself before document it [refer to 12.4 Documentation Checklist], do not engage with the perpetrator or respond to threats, contact trusted person and organizations immediately [KAFA/ABAAD – refer to 9.4 Resource List], use platform-specific removal tools [refer to 8.4 Reporting NCII and Deepfake Content].

Rapid Risk Scoring Matrix

Score each indicator based on your current situation. Add up your total and refer to the score interpretation below. Score 0 for low, 1 for medium, and 2 for high, sum up your score and use the score interpretation table to identify your risk level.

RISK INDICATOR	LOW (0)	MEDIUM (1)	HIGH (2)
Type of attack	Single incident, low-level harassment	Repeated harassment, impersonation or overlapping attacks	Multiple attack types occurring together, and seem to be coordinated
Visibility / Spread	Private or limited audience	Public or growing visibility	Viral, cross-platform, and rapidly spreading
Doxxing and exposure of personal information	No personal data shared	Partial or indirect exposure	Sensitive data shared: phone number, email addresses, home address, and workplace locations
Coordination	No signs of coordination	Indication of repeated content and emerging patterns	Clear coordinated attack across multiple accounts, using same messaging
Threats	No threats	Implied or indirect threats	Direct threats of violence, death, and sexual harm
Escalation	Stable, not increasing	Increasing frequency and intensity	Rapid escalation or sudden surge
Impact on others	Only the victim is targeted	Mentions of others	Family, colleagues and contacts targeted
Offline risk	No offline implications	Indirect references	Clear signs of tracking, contact, and proximity
Source of attack	Isolated or unknown individual	Multiple and suspicious accounts	Organized actors, networks or potentially identifiable groups
Does the attack involve a current, former partner, or their network?	No	No	Yes
NCII and deepfakes were shared	No	No	Yes
Total Score		**____ / 20**	

SCORE RANGE	RISK LEVEL	WHAT IT MEANS IN PRACTICE
0 – 6	LOW RISK	Limited exposure, no immediate threat. Monitor and apply basic protection measures.
7 – 12	MEDIUM RISK	Situation is escalating. Respond and contain to prevent further escalation.
13 – 20	HIGH RISK	Serious risk of digital and physical harm. Immediate protective action and support required.

In relation to threats, doxxing, offline risk, NCII and IPV always treat the situation as **high risk regardless of total score.** If unsure how to score the situation, treat it as high risk and move to protective action.

5

SECTION 5

First 24 to 72 Hours Response Flow

During elections and periods of political tensions, online attacks can escalate rapidly, and become highly visible as they aim to disrupt public engagement, pressure individuals into silence or withdrawal from public life. The first 24 to 72 hours following an incident are often critical in limiting escalation, preserving evidence, reducing exposure, and protecting online and offline safety. This section provides immediate response measures that can help users stabilize the situation, make informed decisions, and identify appropriate next steps while continuing to navigate public and political participation as safely as possible.

⚠ IMPORTANT

IMPORTANT: The guidance in this section applies across all users of this Kit. However, the appropriate application of specific steps, particularly those involving public response, formal reporting, trusted support, and escalation, varies depending on your context. Women candidates, WHRDs, LGBTQI+ defenders each face different risk landscapes that shape which pathways are safe and effective. Where a recommendation requires adaptation for a specific audience, this is noted directly.

5.1 Stabilize the Situation

Online attacks during political campaigns are often designed to provoke fear, urgency, confusion, and reputational harm to discredit candidates, activists and journalists. Immediate reactions made under pressure may unintentionally increase exposure and amplify harmful content. The first step is therefore, to slow escalation, and assess the situation clearly to regain control before deciding on next steps.

- **Pause** before responding publicly or privately to the attack. Avoid reacting to the content, engaging directly with aggressors, or attempting to defend yourself immediately while distressed.
- **Conduct an immediate risk assessment** using the Rapid Risk Identification and Severity Triage tool to determine the level of risk of the incident and identify the type of response required.
- **Identify the main form of attack** and whether multiple tactics are being employed simultaneously and determine if the attack is isolated or coordinated.
- **Do not delete accounts, messages, or evidence** immediately unless there is an urgent safety concern. Keep the content to help preserve important information and evidence that can later support reporting, legal action, organizational support, or security assessment.
- **Temporarily step away from monitoring abusive content** and ask a trusted person to assist in reviewing messages, comments or posts and collecting evidence. This helps reduce psychological strain and limits repeated exposure to harmful content.
- **If there are signs of offline risks**, and threats referencing location, movement, family members, workplace, or attempts to establish direct contact, then prioritize physical safety:
  - Avoid sharing real-time location posts, stories or content.
  - Temporarily vary routines and movement patterns where possible. Avoid predictable schedules, and repeated routines. Do not publicly announce attendance to meetings, events, protests, or campaign activities.
  - Do not attend in person meetings alone and inform a trusted person about the risks. If appropriate inform authorities.

5.2 Preserve Evidence

Digitally violent content can be deleted quickly or moved across multiple platforms and closed groups. Coordinated campaigns often rely on temporary accounts, mass reposting, anonymous profiles, or disappearing content. Therefore, preserving evidence early is very essential for documenting patterns of abuse, assessing escalation, supporting reporting processes, and seeking organizational, legal, psychosocial, or security support where available or necessary.

When documenting incidents, both the abusive content and the broader context surrounding it should be captured. Individual incidents often appear minor in isolation, but patterns of repetition, coordination, and visibility become clearer over time.

> Refer to section 12.4 Documentation Checklist on what to document.

5.3 Reduce Immediate Exposure

Political and civic visibility increases vulnerability during coordinated online attacks. However, reducing exposure does not necessarily mean withdrawing from participation, but rather limiting opportunities for further targeting, and protecting accounts, while the situation is assessed and stabilized. The below are mandatory measures to take as soon as online violence, harassment, and targeting are identified.

1. Secure Accounts and Communication Channels

- **Change passwords immediately to all important accounts:** create new, long and unique passwords for every account, and avoid reusing passwords across platforms:
  - Email accounts
  - Social media accounts (Facebook, Instagram, WhatsApp, X, TikTok)
  - Banking
- **Enable two-factor authentication (2FA)** on all major accounts and communication platforms to make unauthorized access significantly difficult in case passwords were compromised.
- **Secure email accounts**, especially email addresses connected to password recovery or account verification: ensure it has a unique, high-security password; update recovery options.
- **Review active sessions and remotely log out of browsers and devices:** use the "security settings" or "active sessions" section on platforms like Google, Facebook, and Apple to sign out of all devices and browser sessions.

2. Lock Down Devices and Privacy Settings

- **Review privacy and visibility settings** across all platforms and applications. Restrict:
  - Who can contact you,
  - Who can send direct messages,
  - Who can tag or mention the account,
  - Who can view stories and posts,
  - Who can access personal information.
- **Disable real-time location sharing** and review location permissions on devices and applications. Turn off GPS sharing where not necessary and review which applications have access to location, camera, microphone, contacts, and files.
- **Disable geotagging** on camera applications and social media platforms to avoid revealing locations information.
- **Update operating systems, applications, and security software** to the latest available versions to ensure protection against security vulnerabilities.
- **Review shared accounts, linked devices, and connected services.** Remove unnecessary access to shared cloud storage, streaming accounts (if google account is linked), shared calendars, device synchronization, or family-sharing systems.
- **Avoid linking applications** through social media login tools such as "Login with Facebook", to limit spread of compromise between platforms.
- **Use secure and end-to-end encrypted communication platforms** (e.g. Signal).
- **Remove Images' EXIF data:** EXIF (Exchangeable Image File Format) is metadata automatically embedded in photos and videos by your device when you take them. It can contain: GPS coordinates of where the photo was taken, date and time of capture, device make and model, camera settings.

How to remove it:

- **On iPhone:** iOS 13 and above allows you to strip location data when sharing. Go to Photos ⇒ select image ⇒ share ⇒ tap the location tag at the top ⇒ select "Don't include" before sending. For full EXIF removal, use an app such Metapho or Photo investigator.
- **On Android:** Go to Gallery ⇒ select image = > details ⇒ Edit (pencil icon) ⇒ delete location information. Or use an app such as Photo Exif Editor or Exif eraser.
- **On Desktop/laptop:**
  - **Windows:** right-click the file ⇒ properties ⇒ details tab ⇒ "Remove Properties and Personal Information" ⇒ select all ⇒ ok.
  - **MAC:** use preview or a free tool such as Image Option which strips metadata automatically.

5.4 Identify Trusted Support

⚠ IMPORTANT

This section uses the phrase "trusted contacts" deliberately. The Kit recognizes that for many WHRDs and especially for LGBTQI+ activists in Lebanon, the family is itself part of the risk — outing to family members, honor-based violence, and family-enforced withdrawal from public life are real harms that follow many digital attacks. If your family is part of the risk, replace every reference to "family" below with "chosen trusted contacts". Prioritize LGBTQI+ organizations (Helem, MOSAIC) and women's rights groups (ABAAD, KAFA) over family disclosure where outing or honor-based response is a possibility. Refer to 9.4 Resource List.

Identifying trusted support early inevitably reduces the impact of online violence on isolation and psychological wellbeing. Support systems can help with documentation, reduce emotional pressure, strengthen decision-making, and improve safety planning. Support does not strictly have to be formal, even one trusted person can help stabilize the situation and reduce pressure. It is important to identify support as early as possible rather than waiting for the situation to escalate.

A TRUSTED SUPPORT NETWORK CAN BE COMPOSED OF

- Friends and family members who can help monitor messages, document abuse, and provide emotional support.
- Colleagues, campaign teams, and trusted organizational staff who can assist with communication, reporting, and coordination of response plans.
- Digital safety and cybersecurity organizations that can support account security and risk assessment.
- Legal aid and human rights organizations when legal threats, doxxing, and defamation are involved.
- Mental health and psychosocial support providers can help if attacks are causing distress, anxiety, fear, or exhaustion.
- LGBTQI+ organizations and women's rights groups that understand the specific risks linked to outing, gendered abuse, and targeted harassment.
- Trusted journalists, moderators, and platform contacts where public clarification or urgent reporting may be necessary.

A TRUSTED SUPPORT NETWORK CAN ASSIST WITH

- Helping document and organizing evidence.
- Monitoring abusive content so the targeted person is not continuously exposed to harmful material.
- Assisting with reporting abusive accounts or impersonation.
- Contributing to risk assessment and escalation.
- Supporting decisions around public responses, clarification statements, to ensure engagement is not emotional and have an objective third-party opinion.
- Help coordinate safety measures and accompany victims in public settings in case of offline risks.
- Accompanying the targeted person when seeking legal aid, institutional, or organizational support.

CAREFUL CONSIDERATIONS WHEN ESTABLISHING A SUPPORT NETWORK

- Consider whether the person or organization can be trusted to maintain confidentiality.
- Avoid sharing sensitive information widely or publicly.
- Use secure and end-to-end encrypted communication channels to coordinate response.
- Be cautious when forwarding screenshots or evidence that contain personal information.
- Prioritize people and organizations who understand the political, gendered, and LGBTQI+ dimensions of online attacks in Lebanon.

> Online violence during politically sensitive periods is not something candidates or activists are expected to manage entirely alone. Coordinated support can help reduce harm, improve safety, and strengthen the ability to respond effectively and cool-headedly. It will also help maintain safe political participation without increasing exposure or isolation.

5.5 Decide on the Next Response Path

The next step after immediate stabilization is deciding how to respond moving forward. Not every situation requires the same response strategy. Politically sensitive environments, response decisions should therefore prioritize safety, sustainability, and informed decision-making. This section outlines possible response pathways that may be used individually or in combination depending on the situation.

CONTINUE MONITORING WITHOUT PUBLIC ENGAGEMENT

This response path may be appropriate in lower-risk situations where the attack has limited visibility, no direct threats, the content is not spreading quickly, and there are no signs of escalation. Response involves:

- Continuing documentation of evidence and monitoring of the situation, with the assistance of support networks.
- Blocking and restricting abusive accounts.
- Platform reporting harmful content where relevant.
- Avoiding direct engagement with aggressors.
- Reassessing risk regularly for signs of escalation.

ESCALATE TO ORGANIZATIONAL, LEGAL, OR SECURITY SUPPORT

Escalation can be useful in medium to high-risk situations, particularly in cases of direct threats, doxxing, outing, and sustained coordinated harassment targeting family members and colleagues, and attempts to access accounts.

- Inform employers, organizations, campaign teams, and family members.
- Seek digital security support.
- Contact legal aid, or human rights organizations.
- Reporting threats to state authorities such as the ISF Cyber Security Crime Unit carries real risk in Lebanon and should not be done without prior consultation with an independent civil society organization (see Section 9.4 Resource List).
- Seek emergency safety planning support.

⚠ IMPORTANT

IMPORTANT — for LGBTQI+ users: do not contact the ISF before consulting Helem, MOSAIC, or Legal Agenda. Article 534 of the Lebanese Penal Code has been used to detain LGBTQI+ individuals on the basis of digital evidence shared in good faith during such complaints. See Section 9.2 Important Safety Considerations Before Reaching Out for the full warning on state institutions. Reporting to authorities carries real risk in Lebanon. Consult an independent civil society organization (see 9.4 Resource List) before approaching ISF.

PUBLIC CLARIFICATION AND RESPONSE

⚠ IMPORTANT

Before publishing any clarification that names an individual, organization, or political party, consult legal counsel — Legal Agenda or CLDH are appropriate first contacts. Lebanon's defamation provisions (Article 582 of the Penal Code and Decree-Law 2007/22 on audiovisual media) have been routinely used to prosecute targets who respond publicly to their attackers. A counter-statement intended to defend reputation can itself trigger criminal defamation charges, particularly during politically sensitive periods. Plan the response and its legal exposure together, not separately to avoid counter-defamation risk.

In some situations, individuals may decide that a limited public response is necessary especially in cases involving: impersonation; misinformation; reputational attacks; false allegations; coordinated disinformation campaigns.

Whether a public clarification is appropriate, safe and effective depends significantly on who you are and what context you are operating in. Before deciding to issue a public response, use the following rubric:

- **If you are a woman candidate or civic actor with party or organizational support:** A public clarification is most likely to be effective in your context. You have institutional backing, an established public platform, and electoral legitimacy that gives a response credibility. Party officials, campaign managers, or coalition partners can amplify your clarification and provide institutional weight (refer to questions to consider below).
- **If you are a Women Human Rights Defender (WHRD) without party or institutional backing:** Exercise significant caution before issuing a public clarification. Without institutional support, a public response can draw additional attention to the attack, invite further escalation, and position you as an individual target without collective protection. If a public response is genuinely necessary issue it through organizational or collective channel rather than a personal account, keep it brief and factual, and activate your support networks and seek additional support [refer to 9.4 Resource List].
- **If you are an LGBTQI+ defender or activist:** A public clarification carries specific and serious risks in the Lebanese context that must be assessed before any response is issued. If the attack involves outing, rumors about your sexual orientation or identity, a public response can draw attention to the claim, invite further targeting, and in some cases compound legal risk. Do not issue any public statement before contacting Helem, Proud or MOSAIC [refer to 9.4 Resource List].

Any public response should be carefully considered before posting:

- Avoid emotional or reactive responses made under pressure; step away from social media and harmful content before responding publicly.
- Avoid engaging directly with aggressors.
- Avoid sharing or reposting harmful content.
- Focus on correcting false information or communicating essential facts.

> **Questions to Consider:** Will responding increase visibility or prolong the attack? Will responding lead to a counter-defamation suit? Is clarification necessary for safety, reputation, or public accountability? Can the response remain factual, brief, and non-engaging? Is there organizational or trusted support available to help draft or review the response? *Not every attack requires a public response. In some cases, silence, containment, and documentation may be safer and more effective.*

REASSESS CONTINUOUSLY

Response paths are not fixed. Situations can escalate or stabilize quickly, especially during politically sensitive periods.

- Are attacks still spreading?
- Are threats becoming more serious and frequent?
- Are offline risks emerging, and have personal information been shared?
- Is additional support or escalation needed?

SECTION 6

Safer Participation Options

Periods of political tensions, elections, and heightened public visibility can increase the likelihood and intensity of online attacks. Safer participation, however, does not necessarily require withdrawing from public life. In some situations, temporary adjustments to visibility, communication habits, and platform engagement can reduce exposure while allowing continued participation in campaigning, advocacy, and civic engagement. Safety planning should, therefore, aim not only to reduce harm, but also to support continued participation where possible.

6.1 Adjusting Visibility Without Full Withdrawal

This section outlines practical ways to reduce exposure to online violence while supporting continued engagement in public life. They focus on temporary adjustments to visibility, communication practices, and platform use that mya help reduce targeting, escalation, or exposure to harm without fully withdrawing.

Visibility Adjustments

- Avoid posting real-time updates, livestreams, or identifiable locations during events, meetings, or campaign activities. Share content after leaving a location.
- Temporarily pause, archive, or limit access to posts containing sensitive and identifying information, including workplace details, phone numbers, personal routines, family information, or frequently visited locations.
- Review older posts, photos, tagged content, and platform biographies that may unintentionally reveal identifying details or patterns that could be used for doxxing, surveillance or targeted harassment.
- Consider using page-based, campaign-based, or organizational communication channels and accounts during periods of escalation rather than relying on personal accounts.

Communication Practices

- Prioritize scheduled or pre-prepared content during periods of sustained harassment to reduce the pressure to remain constantly online or feel the need to respond to attacks.
- Limit direct engagement with abusive content, hostile hashtags, quote-post cycles, or coordinated harassment campaigns that are intended to provoke visibility, outrage, or escalation.
- Encourage supporters, colleagues, and trusted networks to avoid amplifying abusive content, including through reposting screenshots, hostile hashtags or coordinated attacks.
- Establish trusted moderation support (using Trusted Networks) so abusive comments, direct messages, or reporting processes do not need to be managed alone.

Platform Use

- Temporarily disable comments, limit replies, and activate moderation tools on platforms where attacks are concentrated.
- Restrict who can tag, mention, stitch, duet, repost, or otherwise amplify content where platform features are being weaponized to increase exposure or harassment.

6.2 Individual and Collective Participation Options

Public-facing roles, statements, and communications do not always need to be attributed to a single individual. During periods of heightened targeting, redistributing visibility across a collective or organizational structure can reduce the concentration of risk on one person while preserving the impact of the message or campaign. In fact, collective voice and solidarity are not only protective strategies, but can also help counter the narrative that an individual has been silenced, isolated, or discredited.

Collective and Organizational Voice

- Where possible, issue statements, press releases, campaign communications, or public positions through an organizational account, coalition page, or collective profile rather than through individual personal accounts. *This shifts the public-facing target from a named individual to a broader entity, making coordinated harassment or doxxing harder to concentrate on one person.*
- Rotate public-facing roles among trusted colleagues or team members during periods of escalation, so that no single person absorbs all direct exposure from a campaign. *This includes relevant media appearances, public events, social media responses, and any activity that generates audience engagement.*
- Use shared authorship or unattributed collective statements when individual attribution is not necessary for the message's legitimacy or impact. *Consider agreeing in advance with your organization or coalition on when collective attribution is appropriate, so the decision does not need to be made under pressure.*
- Designate a communications focal point or spokesperson from within a team or organization who has the capacity, preparedness, and consent to manage public exposure during high-risk periods, rather than defaulting to the most visible individual.

Page-Based and Campaign-Based Communications

- Establish or activate page-based communication channels, such as campaign pages, organizational profiles, or advocacy accounts, that are not tied to a personal identity, phone numbers, or email addresses ahead of periods of likely escalation rather than only in response to attacks.
- Maintain clear internal agreements about who has administrative access to shared pages or accounts, how content is approved and posted, and what happens to access if a team members is targeted or temporarily steps back.
- Review what identification information is accessible through the organization's public footprint to protect personal information and contact details of founding members, officers on "About" sections, or media coverage.

Network Structures

- Build relationships with peer organizations, advocacy networks, or coalitions before a crisis, so that requests for collective amplification, counter-messaging, or public solidarity statements can be mobilized quickly when needed.
- Pre-agree on protocols for mutual support with trusted allies. For instance, agreeing that if one member of a network is targeted, others will amplify their core message, report abusive content, or issue public statements of support without waiting to be asked.

6.3 Protecting Immediate Circles

Safer participation planning cannot focus solely on the targeted individual. Online attacks, particularly doxxing, outing, impersonation, and coordinated harassment, frequently extend to family members, romantic partners, close colleagues, and other people within a target's immediate network. Protecting those around the targeted individual is a corner stone of safety planning.

Inform Family and Close Contacts

- Inform trusted family members, partners, and close friends about the nature of the risks faced once digital violence starts. This would allow them to be prepared in case attacks extend to them, and support the targeted individual.
- Inform trusted colleagues of active or anticipated threats where this is appropriate and safe, so they can avoid inadvertently sharing information, can recognize impersonation attempts, and can provide profession support without being caught off-guard.
- Establish clear boundaries with professional contacts, and campaign teams about what information related to the target's work, schedule, or location can be shared externally, and review what the organization or employer publishes about the target in staff directories, event listings, or press materials.

Reduce the Digital Footprint of Immediate Circles

- Encourage family members and close contacts to review the privacy settings on their own accounts, particularly where tagged photos, location data, relationship information, or workplace details may be visible and could be used to locate or pressure the targeted individual through them.
- Ask close contacts to remove or limit public content that identifies the target's shared locations, routines, living arrangements, or family structure. Especially when content involves tagged photos at home, school locations for children, or visible residential addresses.
- Information about children, including their names, schools, appearance, or daily routines, should be treated as sensitive and kept out of public-facing content, since children are particularly vulnerable to indirect targeting.

Emotional and Psychological Considerations

- Immediate circles may experience their own fear, distress, or secondary trauma as a result of attacks, even if they are not directly targeted. Having open, honest and regular conversations is essential to ease these concerns.
- Where possible, identify in advance who within the target's immediate circle can provide practical support (managing accounts, handling communications), who can provide emotional support, and who may need protection from the full extent of what is happening.
- Set boundaries with close contacts about how much they engage with abusive content directed at the victim, including asking them not to read through harassment threats, screenshots of attacks, or abusive direct messages on the target's behalf, as this can transfer harm to them.

6.4 What Others Must Do

Safer participation in the public sphere cannot rest solely on the targeted person's shoulders. Platforms, political parties, employers, and donors all carry responsibilities that the previous sections do not place on the target. The Kit names them here to balance the burden and to give targets, allies, and advocates a clear set of demands to make of the institutions around them.

WHAT PLATFORMS MUST DO

- **Transparent moderation:** published community standards in Arabic, consistently enforced regardless of political affiliation, with appeal pathways that respond within stated timeframes.
- **Named regional contacts:** identified human moderators for Lebanon and the MENA region who can be reached for high-risk incidents without going through generic reporting forms.
- Faster pathways for NCII, doxxing, and coordinated harassment, including pre-emptive hash matching.
- Public reporting on takedown decisions affecting Lebanese civic actors, with anonymized case data shared with civil society.

WHAT POLITICAL PARTIES MUST DO

- Codes of conduct that explicitly prohibit affiliated members from participating in or amplifying TFGBV, with named sanctions for breach.
- Internal complaint mechanisms that women in the party and on its lists can use safely.
- Public statements condemning specific attacks on women candidates and LGBTQI+ actors when they occur — silence is itself a position.
- A standing commitment that the party will not benefit from or campaign on the basis of TFGBV directed at opponents.

WHAT EMPLOYERS AND ORGANIZATIONS MUST DO

- Paid leave for staff affected by sustained digital attacks, treated as an occupational risk and not as a personal matter.
- Legal and psychosocial support covered through the employer, not left to the target to fund.
- Adjusted public-facing duties during periods of high risk, without penalty to performance evaluation.
- Clear safeguarding protocols for staff whose workplaces them at elevated risk of public targeting.

WHAT DONORS AND FUNDERS MUST DO

- Recognize digital safety as a legitimate, fundable line in grants rather than an unfunded organizational overhead.
- Avoid timelines that force grantees into high-visibility public moments during politically sensitive periods without support for the safety risks that visibility produces.
- Treat reported incidents as a sign of strong monitoring, not as a sign of programmatic weakness.

7

SECTION 7

Artificial Intelligence (AI) and Digital Safety

AI tools are increasingly present in everyday digital life, from chatbots and image generators to content moderation systems and search algorithms. For women candidates, LGBTQI+ candidates, WHRDs and civic actors, AI presents both practical risks and specific vulnerabilities that are worth understanding. This section covers how AI can amplify digital attacks, how to protect personal information when Using AI tools, and practical steps to reduce exposure.

7.1 How AI Can Amplify Digital Violence

AI tools are increasingly being used to scale, automate, and intensify digital attacks. Understanding how they are deployed is the first step to recognizing them.

Deepfakes and Synthetic Media

AI image and video generation tools can be used to create realistic but fabricated content. This includes fake videos, manipulated images, or synthetic voice recordings, that falsely attribute statements, appearances, or actions to a targeted person. In political contexts, deepfakes are used to fabricate compromising or sexualized content, spread disinformation about a candidates' positions, or produce materials intended to humiliate or discredit. The content does not need to be convincing to cause harm, even obviously manipulated material can spread rapidly and damage reputation before it is corrected. For NCII please refer to 3.6 Non-Consensual Intimate Images and Deepfake Sexual Content.

What to look for

- Images or videos of a targeted person in a situation they have never been in, intimate or otherwise.
- Audio recordings attributed to a targeted person that they never made.
- Manipulated screenshots of statements they never wrote.
- Synthetic content combining the targeted person's face and voice with fabricated scenarios.

> **Mitigation measures:** Establish a verified presence across key platforms so the audience has a trusted reference point to distinguish authentic content from fabricated material. If synthetic content appears, issue a clear and brief factual correction through your own verified channels. Do not share the fabricated content even to debunk it. Seek forensic analysis to confirm synthetic origin and strengthen any platform report or legal complaint. Report to platforms using their specific synthetic or manipulated media reporting pathways if they exist [refer to 8.4 Reporting NCII and Deepfake Content].

AI-Powered Impersonation and Text Generation

Large language models can generate large volumes of convincing text quickly, enabling the mass production of fake statements, fabricated quotes, or coordinated messaging that appears to come from real people. This lowers the cost and effort of coordinated harassment campaigns, allowing small groups or individuals to simulate the appearance of widespread organized opposition.

> **Mitigation measures:** Set up alerts for your name and variations of it to monitor fake accounts or fabricated statements using your name. Report impersonation accounts to platforms immediately and ensure all documentation is preserved.

Automated Harassment at a Large Scale

AI tools can be used to generate and distribute abusive content at a volume and speed that exceeds human capacity to manage. Some tactics include flooding comment sections, generating variations of the same attack to evade content filters, or producing personalized harassment at a large scale.

> **Mitigation measures:** Activate comment filters, keyword blocking, and moderation tools on platforms where automated harassment is concentrated. Temporarily restrict who can comment, reply, or interact with the content.

Facial Recognition and Surveillance

AI-powered facial recognition tools can be used to identify individuals from photos or videos posted online and cross-reference them across platforms. For LGBTQI+ individuals, activists, and those who have not publicly disclosed their identities, this poses a specific risk of involuntary identification, exposure, and outing.

> **Mitigation measures:** Be cautious about tagging others in photos and allowing others to tag you. Disable facial recognition features where platforms offer this option in privacy settings.

7.2 Protecting Personal Information when Using AI Tools

AI tools, including chatbots, writing assistants, image generators, and translation tools, process and in some cases store the information provided to them. During politically sensitive periods, or targeted attacks, the information shared with AI tools warrants the same care as any other digital communication.

What AI tools can retain and how it may be used

- Conversations with AI chatbots may be stored, reviewed by developers, or used to train future models depending on the platform's data policies.
- Images, documents, or personal details shared with an AI tool may be retained beyond the session.
- Some AI platforms operate across jurisdictions with varying data protection standards.

Steps to protect your information

- Read the privacy policy and data retention terms of any AI tool before using it for sensitive work. Look specifically for whether conversations are stored, for how long, and whether they are used for training.
- Do not share identifying personal information: full name, location, phone number, workplace, or specific details.
- Do not use AI tools to process sensitive documents, legal materials, evidence files, or confidential communications.
- Use AI tools through private or incognito browsing where possible to limit session data retention.
- Where available, opt out of data sharing for model training (most major AI platforms offer this in account settings).
- Be cautious about suing AI tools on devices that may already be compromised as input data may be accessible beyond the AI platform itself.
- If you use AI tools for campaign or advocacy work, establish a team protocol about what categories of information can and cannot be processed through AI platforms.

> ⚠ IMPORTANT
>
> **Specific considerations for AI-Generated Attacks:** do not share details about your identity, relationships, legal situation, or safety concerns with AI tools. Even private disclosures to a digital tool carry risk if that data is accessible to third parties.

8

SECTION 8

Platform Reporting Guidance

Digital platforms provide reporting mechanisms that can, in some cases, reduce the visibility of harmful content, remove material that violates platform policies, or contribute to a broader documentation trail. However, platform reporting is not a guaranteed solution, and its effectiveness varies significantly depending on the platform, the nature of the violation, and the speed of response. This section outlines when reporting may be worth pursuing, the importance of evidence documentation before reporting, and what reporting can realistically achieve.

8.1 When Reporting May Be Useful

Platform reporting is most likely to produce a meaningful result when the content or behavior being reported falls clearly within a platform's stated policies, and when the report is submitted with sufficient documentation. The following situations are where reporting tends to help:

Content removal

- Non-consensual sharing of intimate images (NCII), where most major platforms have dedicated and expedited removal pathways [refer to 8.4 Reporting NCII and Deepfake Content].
- Impersonation accounts that use names, images, and identity to deceive harass, or damage the person's reputation.
- Direct, unambiguous and explicit threats of violence or targeted harassment directed at a named individual.
- Doxxing posts that publish personal information with clear and explicit intent to facilitate harm.
- Coordinated inauthentic behavior, such as mass reporting campaigns, bot-driven pile-on, or accounts created solely to harass, where patterns of abuse can be documented.

Reducing Visibility

- Reporting content can sometimes trigger automatic demotion or reduced algorithmic amplification while a review is pending, even if remove is not immediate.
- **X/Twitter:** Can place a repeatedly reported account in "read-only mode", limiting posting, reposting, and liking. This is a consequence of a finding after review, not an automatic protective measure during review, and it applies platform-wide, not specifically towards targeted individuals.
- **Instagram:** No restrictions apply during review of a report. It has a limit feature that lets users restrict interactions from accounts that do not follow them, or newly followed them, but this needs to be manually activated by users.
- **TikTok, YouTube, Facebook:** have no mechanism of this kind.

Repeated reports from multiple users about the same content or account can increase the likelihood of review and action by platforms.

Building a Documentation Trail

Even when a report does not result in removal, a submitted report creates a record that an incident was flagged. This helps in documenting patterns of abuse if individuals want to pursue legal proceedings or seek organizational support. Reporting is less likely to be effective when the content falls into gray areas of platform policies. This is exacerbated when platforms have weak and inconsistently enforced community standards, or when harm is generated across many accounts simultaneously during coordinated harassment campaigns.

8.2 What to Document Before Reporting

Sections 12.4 of this toolkit cover what to document in full. Before submitting any platform report, please revisit these guidelines [12.4 Documentation Checklist].

It is important to report a case after documenting evidence, and not before. Platforms and account admins tend to remove content quickly once a report is processed, including content that has not been documented yet. Screenshots, URLs, timestamps, and account details should be saved before a report is filed, as access to the original material could be lost once it is taken down or the account is deactivated.

8.3 What Reporting Can and Cannot Do

It is important to understand the limits of platform reporting processes to identify when to use it.

REPORTING CAN	REPORTING CANNOT
Initiate a review process that may result in content removal, account suspension, or reduced amplification.	Guarantee removal, suspension, or any specific outcome since platforms have inconsistent enforcement and significant backlogs.
Create a timestamp record of an incident within the platform's own system.	Specify what action, if any was taken, after reviewing a report. Most platforms only confirm a report was "reviewed" without specifying the outcome.
Trigger faster responses, in some cases, where platforms have dedicated pathways for high-risk categories such as NCII, credible and explicit threats, or child safety.	Stop harassment in real time. Review processes can take hours or days, during which content remains visible and can potentially be amplified.
Contribute to a patter of evidence if abuse is ongoing and multiple reports are submitted over time.	Address cross-platform abuse. Reporting on one platform has no effect on the same actors or content appearing elsewhere.
Support legal or organization processes where a record of reported incidents strengthens a case.	Substitute for parallel protective measures such as adjusting privacy settings, blocking, documenting evidence, or seeking legal support when necessary.
	Reliably address coordinated or politically motivated abuse, particularly when reporting tools are themselves weaponized through mass false reporting against activists and candidates.

Counter-reporting

In some coordinated harassment campaigns, the aggressors can use platform reporting mechanisms against their targets by submitting false reports to get accounts suspended or content removed. If an account gets suspended or content is removed as a result of false reporting, most platforms have an appeals process. Document the suspension notice, the content that was removed, and any context that demonstrates the report was made in bad faith. Legal support organizations working on digital rights may also be able to assist with appeals.

8.4 Reporting NCII and Deepfake Content

1. StopNCII.org: Pre-blocking tool

StopNCII.org is a free tool that allows individuals to create a digital fingerprint, called a hash, of an intimate image or video without uploading the image itself. That hash is shared with participating platforms, which use it to automatically detect and prevent the image form being uploaded or shared across their networks. StopNCII currently works with Meta platforms (Facebook, Instagram, Threads), TikTok, Snapchat, Reddit (list of partners). For more details on how it works and steps to follow visit: https://stopncii.org/how-it-works/ (available in Arabic).

2. Platform-Specific Emergency Removal Channels

Most major platforms have dedicated removal pathways for NCII that operate faster than standard content reporting. Use these specific pathways rather than general harassment reporting tools:

- **Meta (Facebook/Instagram):** Meta has a dedicated NCII reporting tool that allows individuals to report intimate images of themselves for expedited removal. Go to the image or content you want to report, press on Options (usually in the top or bottom right of the content), press Report, find the section "Reporting Specific harms" and press "Fill in the Form".
- **TikTok:** TikTok has a specific reporting category for NCII under its community guidelines. Reports under this category are prioritized for review. Report content directly through the platform using "Report" function ⇒ "nudity and sexual activities" ⇒ "non-consensual intimate imagery".
- **X/Twitter:** X has a specific reporting pathway for intimate media shared without consent. Select "Report Post" ⇒ "It displays a sensitive image" ⇒ "An unauthorized photo or video" ⇒ "It includes unauthorized, intimate content". *Important: following changes to X's moderation policies, enforcement consistency and responsiveness have declined. Document all reports submitted and use StopNCII.org.*
- **YouTube:** YouTube's privacy complaint tool covers intimate images shared without consent.
- **Any Google service (including Youtube):** use the Ask Google to Remove Explicit Content Form (link).

8.4 Telegram-Specific Guidance

Telegram occupies a particular place in the Lebanese digital landscape. Closed and semi-public channels are where leaked photographs, addresses, and smear material frequently circulate first, before crossing to more policed platforms. The Kit treats Telegram separately because both the threat surface and the reporting options differ from those of the major platforms covered in Section 8.4 Telegram-Specific Guidance.

Monitoring closed channels

- If the target or an ally is a member of the channel where harmful content appears, screenshots can be captured directly. Document the channel name, member count, and channel admins where visible.
- For channels the target cannot access, trusted allies who are members can monitor on the target's behalf and report back with summaries rather than forwarding harmful content (which could otherwise expose the target further).

Telegram reporting in practice

- Telegram's reporting interface accepts complaints for spam, violence, child abuse, illegal pornography, and copyright. Other categories of harmful content — including doxxing, defamation, and coordinated harassment — do not have dedicated pathways and are inconsistently actioned. Expect partial response at best.
- Channels (one-to-many broadcast spaces) are subject to Telegram's terms; supergroups and direct messages have weaker enforcement. Where a channel's name infringes a person's rights or impersonates them, that is the most likely successful report category (impersonation).
- The most reliable mitigations on Telegram are external: documentation for legal use, coordinated counter-messaging across other platforms, and platform-pressure work through digital-rights organizations (SMEX, Access Now – Refer to 9.4). Direct reporting alone rarely resolves Telegram-hosted harassment.

9

SECTION 9

Support Pathways and Referrals

Lebanon's political landscape shapes the safety of every support pathway listed in this section. During elections periods and heightened political tensions, formal institutions may be aligned with the very actors responsible for attacks. Some civil society organizations operate under pressure, with limited resources, and in some cases, under direct threat. Sectarian political networks also have both the reach and motivation to monitor, intercept, or instrumentalize formal complaints. This section reflects those realities, and outlines when additional support may be needed, what kind of support exists, and how to access it safely within the Lebanese context.

This is not a case-management mechanism but rather is a starting point for knowing where to turn and what to consider before you do.

9.1 When to Seek Additional Support

The guidance in this kit covers actions that can be taken independently or with a trusted network. Some situations require more. Consider seeking additional support when:

1. The situation involves legal dimensions

- Credible threats of physical and sexual violence exists and the targeted person is considering a formal complaint.
- Personal data, intimate images, or identifying information has been shared and the targeted person wants to understand options for removal or legal redress.
- Accounts or devices have been compromised in ways that may constitute a criminal offence.
- Targeting appears to involve organized political actors, party networks or actors.

2. The situation is evolving into offline harm

- Online threats are becoming increasingly specific, personalized, or time sensitive.
- There are signs of offline surveillance, tracking, and threats to physical safety.

3. The psychological impact requires professional support

- Sustained harassment is affecting the targeted person's ability to function, sleep, work, or continue their public role.
- Experience of acute stress, fear, or exhaustion as a result of prolonged exposure to abuse.
- Targeted person's circle is also experiencing emotional distress.
- Forensic analysis is needed to identify the source, origin, or technical pattern of an attack particularly in cases involving suspected account compromise, or device infiltration.

4. The technical situation exceeds capacities

- Devices, accounts, or communications have been compromised and individuals require technical support.
- Managing a high volume of coordinated attacks across multiple platforms.
- Need for a security assessment that goes beyond the steps covered in this kit.

9.2 Important Safety Considerations Before Reaching Out

Formal state channels carry specific risks that may increase risks of harm. Lebanese state institutions, including the ISF, the judiciary, and related bodies, are not equally safe or accessible for all. During periods of political sensitivity, these institutions operate within a sectarian political environment in which complaints can be redirected, suppressed or weaponized depending on the political affiliations of those involved. Reporting to formal authorities may be appropriate in some cases and actively dangerous in others depending on who is behind the attack and whether the targeted person's identity and public role makes them a target of the very structures they are reporting to.

⚠ IMPORTANT

Before approaching any formal state body, consult a trusted legal counsel or civil society organization with political independence to assess whether doing so is safe within the context of the attack.

In Lebanon this involves real risks that must be weighed carefully.

- Formal complains require disclosure of the existence of intimate images to police and judiciary, which may involve additional exposure and loss of control over the narrative.
- In cases where the perpetrator is a former partner (IPV), family pressure, mediation, or institutional responses may prioritize reconciliation over protection (contact ABAAD or KAFA – refer 9.4 Resource List).
- For LGBTQI+ individuals, any formal process involving intimate images carries the additional risk of exposure under Article 534 of the Lebanese Penal Code (contact Helem, Proud, or MOSAIC before formal reporting – refer to 9.4 Resource List).
- Legal processes in Lebanon can be slow and the outcome uncertain, particularly where digital protections are scarce, and procedures are inconsistently applied.

LGBTQI+ users face additional and specific legal risks. Article 534 of the Lebanese Penal Code criminalizes same-sex relations and has been used to detain, prosecute, and coerce LGBTQI+ individuals. For queer users, formal reporting to the ISF or other state bodies carries the risk of having their identities used against them, regardless of the nature of the original attack. It is better not to contact formal state institutions without first consulting LGBTQI+ CSOs (i.e. Helem or Proud), which have legal teams experienced in navigating this risk.

Outreach creates a communication trail. In fact, contacting any organization creates a record and can be trailed back. End-to-end encrypted communication channels are essential to ensure protection of data, and targeted individuals should avoid using their regular email or social media accounts to seek help. [Refer to section 5.3 Reduce Immediate Exposure for guidance on secure communications]. Before contacting any support organization, ensure that the evidence is preserved, documented properly and securely. [Refer to 12.4 Documentation Checklist].

9.3 Types of Support That May Be Relevant

— LEGAL SUPPORT

Relevant when targeted people are considering a formal complaint, need to understand their rights under Lebanese Law, are facing defamation or impersonation with professional repercussions, and need documentation for an appeal or escalation. Look for organizations with experience in digital rights, freedom of expression, GBV, or electoral law that operated independently of sectarian political structures.

DIGITAL SECURITY SUPPORT

Relevant when accounts or devices have been compromised, security audit is needed, coordinated attacks are being managed across multiple platforms, and guidance on secure communications beyond this kit is required.

PSYCHOSOCIAL AND WELLBEING SUPPORT

Relevant when sustained harassment or public targeting is affecting victims' mental health, personal relationships, or ability to continue working. Look for trauma-informed providers with experience working with human rights defenders, survivors of GBV, or LGBTQI+ individuals.

COMMUNITY AND PEER SUPPORT

Relevant when practical help is needed to manage abuse, or connection with others who have faced similar situation can provide moral support. Networks of women in politics, LGBTQI+ coalitions, and WHRDs networks can provide peer support, collective advocacy, or solidarity.

EMERGENCY SUPPORT

Relevant when the targeted person or someone in their immediate circle is in danger. Use trusted support networks.

9.4 Resource List

Contact details below were collected at the time of Kit production in 2026. Phone numbers, helpline hours, and email addresses change frequently in Lebanon's operating environment. Before relying on any contact in this list for an urgent incident, verify the current contact information through the organization's website or social media. If you find a contact that no longer works, report it to info@humena.org so it can be corrected in the next edition.

EMERGENCY SUPPORT AND DIGITAL RIGHTS

EMERGENCY · 24/7

Front Line Defenders (FLD)

+353 1 210 0489 (emergency line, 24/7 — Dublin)

International organization dedicated to the protection and security of Human Rights Defenders (HRDs) at risk worldwide. Provides rapid and practical support including emergency grants, temporary relocation, digital security capacity building, and psychological support. Operates a 24-hour emergency line in Arabic, English, and French. Secure contact form available · email: info@frontlinedefenders.org · website

LEGAL SUPPORT

LEGAL

Legal Agenda

+961 1 383 606

Independent legal research and strategic litigation organization. Works on freedom of expression, GBV, LGBTQI+ rights, and political accountability. Appropriate for all civic actors, WHRDs, LGBTQI+ defenders who want to seek legal aid. email: info@legal-agenda.com · website

LEGAL

Lebanese Center for Human Rights (CLDH)

+961 1 24 00 23

Non-partisan human rights organization providing legal services and documentation of abuses. Works on arbitrary detention, torture, and broader human rights violations. Appropriate for cases involving serious rights violations, treats, surveillance, or intimidation by state or non-state actors. CLDH can also provide psychological support. email: info@cldh-lebanon.org · website

GENDER-BASED VIOLENCE AND WOMEN'S RIGHTS SUPPORT

GBV / WOMEN'S RIGHTS

ABAAD – Resource Center for Gender Equality

+961 81 788 178 (helpline)

Feminist organization working on gender equality, GBV, and women's political participation across Lebanon. Operates safe spaces and a 24/7 safe line. Has documented experience working in crisis and emergency settings. Appropriate for women candidates, activists, and civic actors facing gendered targeting. email: abaad@abaadmena.org · website

WOMEN'S POLITICAL PARTICIPATION

The Lebanese Women Democratic Gathering (RDFL)

+961 71 500 808 (helpline)

Feminist secular organization that works on women's political participation, gender equality, and elimination of violence against women. RDFL has five offices across Lebanon (Beirut, Tripoli, Baalbek, Keserwan, and Saida) and has geographic reach outside Beirut. They also have a 24/7 helpline. Appropriate for women candidates and civic actors, has strong referrals of support for women in politics. website

GENDER-BASED VIOLENCE AND WOMEN'S RIGHTS SUPPORT (CONT.)

WOMEN'S RIGHTS

Women Alive

+961 78 842 090

Women and girls' rights organization working on GBV through advocacy, trainings, and awareness. Can provide GBV guidance, and psychosocial support (specifically in Tripoli). email: info@womenalive.net

GBV / IPV SUPPORT

KAFA (Enough Violence and Exploitation)

+961 3 018 019 (helpline 24/7)

Feminist organization that works closely with victims of GBV and IPV in Lebanon. Provides integrated social, legal, and psychological support for survivors through a network of counsellors, lawyers, and social workers. Operates a 24/7 helpline for immediate assistance. email: kafa@kafa.org.lb · website

LGBTQI+ SUPPORT

LGBTQI+ SUPPORT

Helem

+961 81 478 450 (helpline)

An LGBTQI+ rights organization that provides emergency response, legal assistance, case management, mental health support, and community services for LGBTQI+ individuals in Lebanon. Operates a 24/7 emergency hotline (available on Signal). email: info@helem.net · website

LGBTQI+ SUPPORT

MOSAIC

+961 76 945 445 (helpline)

An LGBTQI+ rights organization working to end homo-hate and trans-hate in Lebanon through community empowerment, collective advocacy, and dignified services for LGBTI and queer persons. email: mosaic@mosaic-mena.org · website

LGBTQI+ SUPPORT

Proud Lebanon

+961 76 608 205

An LGBTQI+ rights organization that provides legal assistance, case management, mental health support, and community services to LGBTQI+ and marginalized communities in Lebanon. email: Support@ProudLebanon.org · website

SEXUAL HEALTH & LGBTQI+

MARSA

+961 1 380 515

Sexual and reproductive health centre in Beirut providing confidential medical, social and psychological services to people of all backgrounds, with a focus on LGBQTI+ communities. email: visutus@marsa.me · website

DIGITAL SECURITY AND RIGHTS

DIGITAL RIGHTS

SMEX

+961 81 633 133 (helpdesk)

Digital rights organization working on freedom of expression, privacy, and online safety. SMEX can help with organizational-level digital rights support. SMEX operates a helpdesk that provides direct cybersecurity support to activists, journalists, human rights defenders, and marginalized communities facing online threats and digital attacks. SMEX also offers forensic analysis support. email: helpdesk@smex.org · website

DIGITAL SECURITY · 24/7

Access Now — Digital Security Helpline

+1 888 414 0100 (24/7)

Free, 24/7, Arabic-capable rapid incident response for human rights defenders globally, with Lebanese context experience. Recommended first contact for any account compromise, device infiltration, or coordinated cross-platform attack. Can provide technical support, legal aid, and ways to navigate digital attacks safely. email: help@accessnow.org / legal@accessnow.org · website

PSYCHOSOCIAL SUPPORT

MENTAL HEALTH · 24/7 LIFELINE

Embrace

1564 (lifeline)

Mental health organization that provides affordable therapy, psychiatric care, psychosocial support, and community-based mental health services. They operate a 24/7 national hotline that provides anonymous emotional support, crisis intervention, and referral to mental health services. email: info@embracelebanon.org · website

10

SECTION 10

Guidance for Allies, Friends, and Colleagues

When a woman candidate, LGBTQI+ activist, or civic actor is targeted with digital violence, the response of the people around them – colleagues, campaign teams, party members, friends, and organizational networks – can significantly shape the outcome of these attacks. A well-informed, coordinated ally network can reduce harm, absorb pressure, and help the targeted person maintain their public role. An uninformed or reactive circle can amplify attacks, strip the targeted person of agency, and create new risks.

During politically sensitive periods, campaigns and advocacy organizations are often operating under pressure, with high visibility. Decisions about how to respond to an attack carry political and personal consequences. This section is designed to help the people surrounding a targeted individual act in ways that are genuinely supportive, grounded in the targeted person's best interests, and informed by an understanding of how digital violence works.

10.1 What Supportive Allies Should Do

1. Start with the targeted person, not the attack

The most important step an ally can do is to check in with the person being targeted before taking action. Ask what they need, what they have already done, and what kind of support they want. Do not assume. The instinct to act quickly by posting in solidarity, reporting content, or contacting the media may feel helpful but can override the targeted person's own strategy, and expose information they have not chosen to share. It can also increase visibility and intensify harm in ways that should be avoided.

- Ask directly: "What do you need from me right now?" and "Is there anything you would like me not to do?"
- Follow their lead on whether the situation should be addressed publicly, kept internal, or handled quietly.
- Respect their decision if they choose not to respond publicly, not to report, or not to involve external organizations, even if you disagree with that choice.

2. Provide practical, task-based support

Managing a coordinated attack is exhausting and often relentless especially while maintaining public life engagement. Allies can take on specific tasks that reduce the burden on the targeted person without requiring them to remain constantly online or emotionally engaged with the abuse. These tasks include:

- Helping document evidence of the abuse across platforms, and accounts.
- Monitoring specific platforms, hashtags, or accounts for new developments and report back with summaries rather than screenshots of abuse.
- Handling reporting processes with the targeted person's consent to prevent them from further engaging with harmful content.
- Managing moderation on shared pages, campaign accounts, or organizational channels: filtering comments, blocking accounts, and handling incoming messages.
- Helping draft communications, statements, or responses if these are needed, but the targeted person should decide what is sent and when.

3. Offer visible solidarity

Public expressions of solidarity can help counter the narrative that the targeted person has been discredited, or silence, and would reduce isolation. During elections, and campaigning, collective responses from parties, coalitions, or civil society networks carry political weight that can reduce the public reputational harm of online attacks. However, public solidarity should only come with the targeted person's consent.

- Publish public statements of support, reflecting what the targeted person want those statements to say.
- Highlight the targeted person's work, role, and continued participation rather than the details of the attack since this can amplify the abusive content or draw additional attention to it.

> **For party officials, campaign managers, organizational leaders:** it is essential to offer visible support with a clear and timely institutional response that centers and reinforces the targeted person's continued participation.

4. Allies should also take care of themselves

Supporting someone through sustained digital violence can be emotionally demanding. Allies who are absorbing abusive content, managing communications, and emotionally supporting the targeted person may experience their own trauma and burnout. Allies will be a more effective source of support if they maintain their own boundaries, share the load with others in the network, and seek psychological support when needed.

10.2 What Allies Should Avoid

1. Avoid resharing, screenshotting, or amplifying abusive content

One of the most common and damaging mistakes allies make is resharing abusive content to draw attention to the situation or express outrage. This directly increases the reach and visibility of the attack.

2. Do not act without the targeted person's knowledge or consent

Acting on behalf of someone who is being targeted can undermine their autonomy, disrupt their own strategy, and in some cases worsen the situation.

- Avoid contacting journalists, legal organizations, or external bodies on the targeted person's behalf without their explicit knowledge and agreement.
- Avoid issuing public statements, posts in solidarity, or make the situation visible in new spaces without checking with targeted person first.
- Avoid making decisions about how to frame, characterize, or respond to an attack on behalf of the targeted person whether to describe the nature of the attack, identify the perpetrators, or speculate about motives without their input or consent.

3. Avoid minimalizing, rationalizing or reframing the attack

Responses intended to reassure can unintentionally devalue the targeted person's feelings, and reactions to the abuse, exacerbating the psychological harm.

- Avoid phrases such as "just ignore it", "don't give them the attention they want", "it will blow over" or "this is just what happens in politics". These responses, however well-intentioned, minimize the real harm of sustained digital violence and place the burden of managing it back on the targeted individual.
- Avoid suggesting that the targeted person's own behavior, public visibility, statements, or that adjusting their conduct would reduce harm. This is a form of victim-blaming that is common in response to TFGBV and particularly harmful in political contexts where women and LGBTQI+ people are already pressured to limit their public engagement and presence.

4. Do not out, expose, or disclose information without consent

In attacks targeting LGBTQI+ individuals, or in situations involving the non-consensual sharing of intimate images, allies may have access to sensitive information about the targeted person's identity, relationships, or personal life. This information must be handled with absolute discretion.

- Do not disclose, confirm, or discuss the targeted person's sexual orientation, gender identity, relationship status, or any other personal information without their explicit consent, even in the context of defending them.
- Do not share information about the nature of an attack involving intimate images, outings, or identity-based abuse beyond those who already know and need to know.
- Information regarding sexual orientation and identity, as well as disclosure of intimate images can expose targeted people to legal risks (law 534 used to criminalize homosexuality), and social harm (honor killing of women).

5. Avoid treating the situation as a political opportunity

During election periods, attacks on women candidates can quickly become politically useful to the people around them for generating donations, reinforcing campaign narratives, or positioning allies favorably with certain audiences. Acting on that usefulness without the targeted person's knowledge and consent is harmful, regardless of intent.

- Do not frame the attack primarily as an attack on the party, the campaign, or the cause. The targeted person's safety and experience come before political narratives.
- Do not assume that because someone consented to one public statement or one use of their story, that consent extends to everything that follows. How survivors want their experience to be discussed or not discussed can change. Keep checking and asking rather than assuming.

11

SECTION 11

Guidance for Page Admins, Moderators, and Community Platforms

Community pages, group administrators, and online moderators play an often underestimated role during digital attacks. They are not passive hosts of content but rather are active shapers of what their communities see, amplify, normalize, or reject. During politically heightened periods, when coordinated harassment campaigns, disinformation, and exposure-based attacks can scale rapidly, the decisions made by page admins and moderators carry real consequences for the people being targeted.

This section is dedicated to those who manage community spaces. These spaces include campaign pages, advocacy group accounts, civil society platforms, constituency forums, party communication channels, or any other online space with a moderation function. It sets out why their role matters, what minimum standards of non-amplification look like, and how to moderate more safety during high-tension periods.

11.1 Why Page Admins and Moderators Matter

Moderation is not a neutral act. Every decision a page admin or moderator makes shapes what behaviour is treated as acceptable within that space and what harm is allowed to spread. Therefore, decision such as leaving or taking down a comment, removing or keeping a post, setting community rules or leaving them vague should be taken carefully and with considerations to their implications.

During elections and periods of heightened political activity, community pages and online spaces become sites where attacks are launched, amplified, or contested. Abusive comments, doxxing, rumours, and exposure-based content do not only originate on large public platforms, but are also circulated through local community pages, party group chats, and networks. In many cases, this secondary circulation causes as much harm as the original attack because it reaches audiences who know the targeted person either personally or professionally. It also lends the content an appearance of credibility, and strips the targeted person of control over their own narrative within their own community.

Moderators who allow this content to circulate, even passively, become part of the mechanism of the attack. Conversely, moderators who act quickly, consistently, and transparently can significantly limit the reach of harmful content and signal clearly to their community that this kind of behaviour cannot be tolerated.

This is particularly significant for spaces connected to political campaigns, parties, or civic organizations, where community's trust in leadership, and the safety of the people those organizations are meant to support depends in part on how the online space is managed.

11.2 Minimum Non-Amplification Practices

The following practices are a baseline standard for any page admin or moderator whose platform touches on political, civic, or advocacy content. It is also particularly for those managing spaces where women candidates, LGBTQI+ activists, or human rights defenders are members, subjects of discussion, or public figures associated with the page.

On abusive comments and content

- Remove comments that contain threats, slurs, doxxing information, sexualized language, or identity-based abuse directed at any individual.
    - Do not leave these comments visible while deciding whether they really violate the page's standards: if they are clearly harmful, act.
- Do not engage with abusive comments publicly in ways that increase their visibility, such as quoting them in a response, pinning a rebuttal that repeats the original content, or sharing them as examples of what is being condemned.
- Apply moderation standards consistently across political affiliations, identities, and subjects. Inconsistent moderation, where abuse directed at women and LGBTQI+ people is tolerated while other content is removed, implies that some members of the community are less protected than others.

On rumors, leaks, and exposure-based material

- Do not share, repost, or allow the circulation of content that exposes personal information about any individual, including their home address, workplace, family details, sexual orientation or gender identity.
    - This also applies to content framed as news, gossip, political commentary, or public interest.
- Treat unverified rumors about a targeted person's private life as harmful content, not as open questions for community discussion.
    - The act of hosting that discussion, even without taking a position, amplified the harm and lends it a platform.
- If the content shared on the page turns out to contain doxxing information, or outing material, remove it immediately and do not archive or reshare it for any purpose, including as evidence of what was posted.

On reporting and transparency

- Establish a clear, accessible process for community members to report harmful content to admins and moderators and respond to reports within a reasonable timeframe (could be through a link present in the bio, or through DMs to specific monitoring accounts).
- When content is removed, consider if a brief, non-specific notice to the community is appropriate to state that the content was removed for violating community standards, without repeating or describing the removed material.
- Keep internal records of moderation decisions, particularly during high-tension periods, so that patterns of abuse can be identified and, where necessary, documented.

On admin access and account security

- Review who has administrative access to the page or platform, and remove access from anyone who is no longer active, whose account may have been compromised, or whose role within the organization has changed.
    - Keep account access and posting centralized through one person or a small communications team.
- Use strong, unique passwords and two-factor authentication on all admin accounts.
    - During elections periods, the risk of page compromise or infiltration increases, particularly for pages associated with campaigns or politically active organizations.
- Agree internally on a protocol in case an admin account is compromised, specifying who has authority to remove access and how quickly that can be done.

11.3 Safer Moderation in High-Tension Periods

Elections, political crises, and periods of heightened political controversy create specific moderation challenges. The volume of content increases, the stakes of individual decisions are higher, and coordinated actors may deliberately target community spaces as part of a broader harassment or disinformation campaign. The following guidance is tailored to those conditions.

Anticipate and prepare before tensions peak

- Establish moderation standards before an attack occurs: review and update community guidelines so they clearly address harassment, doxxing, identity-based abuse, and the sharing of personal information.
- Pin the community guidelines visibly so that members are aware of the standards and cannot claim ignorance when those standards are enforced.
- Identify in advance which members of the team have moderation responsibility, what their authority is, and how decisions will be made quickly if harmful content appears outside of working hours or during a fast-paced escalation.
- Consider temporarily increasing moderation capacity before anticipated periods of high tension.

Recognize the signature of coordinated attacks

- Recognizable patterns of coordinated digital attacks:
  - Sudden surge of new or inactive accounts engaging with the page
  - Repeated use of the same phrases or hashtags across multiple comments
  - Mass reporting of the page's content
  - Rapid spread of specific piece of content (a screenshot, a rumor, a video) across multiple spaces simultaneously.
- If these patterns are identified, do not engage with individual comments as though they are independent.
  - Document the harassment.
  - Temporarily restrict who can comment, post, engage with the page while the situation is being assessed.
- Alert the targeted person and your broader team that a coordinated campaign appears to be spreading, so they can take their own protective steps.

Manage exposure without spreading harmful content

- When removing or responding to harmful content: avoid actions that create new exposure for the targeted person. This includes:
  - Writing public posts that describe the nature of the attack in detail
  - Tagging the targeted person in your moderation response
  - Sharing screenshots of what was posted to showcase that it has been removed.
- When communicating internally about an incident whether to a moderation team, organization leadership, or legal counsel, do so through a secure, end-to-end encrypted channel and without forwarding or sharing the harmful content itself beyond those who need to see it.
- Be aware that moderation actions can themselves become the subject of attacks.
  - Removing content may prompt accusations of censorship, political bias, or suppression, especially during periods of heightened political activities.
  - Have clear, publicly visible community standards that predate the incident.

Support targeted individuals without taking over

- When a person associated with the community is being targeted through or around a specific platform, admins should check in with them before taking any public action.
- Document what is happening on the platform carefully (refer to section 12.4 Documentation Checklist), and ensure this documentation is shared with the targeted person so they keep a record of the abuse.

Post-High-Tensions Situations

- Conduct a brief internal review of how moderation functioned.
    - What decisions were made quickly and well?
    - What took too long?
    - What gaps in the community standards or processes became visible?
- Update guidelines and internal protocols based on lessons learnt, so that admins are better prepared for the next high-tension period.
- Check in with any members of the moderation team who absorbed significant exposure to harmful content during the period and ensure they receive the support needed.

12

SECTION 12

Quick Checklists

The checklists allow users to have the kit's guidance translated into fast, actionable reference tools that cna be revisited quickly under pressure, during an active incident, or in preparation for a high-risk period such as an election or advocacy campaign.

12.1 Pre-Exposure Preparedness Checklist

- ☐ Enable two-factor authentication (2FA) on all active accounts, prioritizing email, social media, and any accounts linked to the campaign or organization.
- ☐ Review which accounts are still active and close or deactivate any you no longer use.
- ☐ Use a strong, unique password for every account.
- ☐ Review which third-party apps and services have access to your accounts and revoke access for accounts you no longer use or do not recognize.
- ☐ Create a separate email address for public-facing communications, and campaign sign-ups to protect your primary email address.
- ☐ Review the privacy settings on all active social media accounts.
- ☐ Remove or restrict access to your home address, phone number, and daily location from all public-facing profiles, bios, and platform directories.
- ☐ Identify what personal information is publicly accessible and where it appears (Google search your name, check if your phone number and email addresses are found online).
- ☐ Use a secure, end-to-end encrypted messaging app for sensitive communication with your team and trusted contacts.
- ☐ Establish or activate a campaign or organizational page for public-facing communication to avoid using personal accounts during high-visibility period.
- ☐ Identify at least two people in your immediate network who are aware of your situation, understand the risks you might face, and have agreed to provide support if an incident occurs.
- ☐ Remove EXIF metadata from photos before posting them publicly, particularly images taken at events, meetings, campaign activities, homes, or any location you want to keep private.

12.2 Risk Triage Checklist

Use this checklist when an incident begins or when you need to reassess a developing situation. A single indicator may not determine the level, but a combination of several signals, even in a seemingly low-level situation, may point to high risk than it first appears. Reassess using this checklist regularly as the situation develops since risk levels are not fixed.

1. RAPID DIAGNOSTIC QUESTIONS

Nature of the attack

- ☐ Is this harassment, TFGBV, doxxing, impersonation, outing, coordinated harassment, or a combination of some or all types?
- ☐ Are different attacks happening simultaneously?
- ☐ Is this a single incident or has it been repeated over time?

Location and Spread

- ☐ Is the attack occurring on public platforms?
- ☐ Is it spreading across multiple platforms or attacks?
- ☐ Is it occurring through private messages, closed groups, or direct contact?
- ☐ Are blackmail and the threat of sharing private content present?

Doxxing

- ☐ Has a phone number, home address, workplace, or family details been shared or referenced publicly?
- ☐ Has private content (images, messages, personal data) been exposed?
- ☐ Have repeated spam calls or messages been received from unknown numbers?

Coordination

- ☐ Are multiple accounts posting similar messages or content simultaneously?
- ☐ Is the same content appearing across different platforms?
- ☐ Are there signs of automated accounts or newly created accounts acting together?
- ☐ Are accounts linked to known individuals, political actors, or organized groups?

Threats

- ☐ Are there messages containing language suggesting physical harm or death threats?
- ☐ Are there references of sexual violence?
- ☐ Are there threats of reputational damage, blackmail, or outing?
- ☐ Are messages intended to intimidate or force withdrawal from public life?

Escalation

- ☐ Has the frequency or intensity of attacks been increasing?
- ☐ Has the tone become more violent or threatening over time?
- ☐ Are threats or blackmail attempts increasing in frequency?

Impact on others

- ☐ Are family members, colleagues, or associates being mentioned in attacks?
- ☐ Are they being directly contacted or targeted?
- ☐ Have people in your immediate circle begun receiving harassment?

Offline risk

- ☐ Has anyone attempted to make direct contact or locate you physically?
- ☐ Are there references to your movements, schedule, or physical proximity?
- ☐ Have you noticed signs of offline surveillance, tracking, or approach?

2. DETERMINE THE RISK LEVEL (if uncertain treat the situation as high-risk)

LOW RISK

Limited exposure, low severity, no immediate threat

- ☐ Single or isolated incident with limited visibility
- ☐ No exposure of sensitive personal information
- ☐ No coordinated targeting across multiple accounts or platforms
- ☐ No direct or implied threats
- ☐ No signs of escalation
- ☐ No involvement of family members or colleagues

Immediate Actions

- ☐ Stop all communication with the aggressor and block on all platforms where contact could be made
- ☐ Document everything
- ☐ Use platform tools to restrict or report abusive content and accounts
- ☐ Review and tighten basic account privacy settings
- ☐ Enable two-factor authentication

MEDIUM RISK

Increased visibility, persistence, and early signs of escalation

- ☐ Repeated or sustained harassment
- ☐ Content gaining visibility or spreading across platforms
- ☐ Emerging signs of coordination across multiple accounts
- ☐ Partial exposure of personal information
- ☐ Threats present but not yet immediate or severe
- ☐ Colleagues or associates beginning to be affected

Immediate Actions

- ☐ Systematically document evidence
- ☐ Monitor for coordinated patterns: similar messaging, timing, and accounts activity
- ☐ Limit public exposure: review posts, adjust audience settings, restrict tagging and comments
- ☐ Avoid actions that amplify harm: do not reshare hostile content, engage with abusive posts, or draw attention to harmful hashtags and profiles
- ☐ Notify audience and networks about fake accounts in case of impersonation

HIGH RISK

Serious threats, personal data exposure, coordination, and likelihood of physical harm

- ☐ Doxxing: home address, phone number, workplace, and/or location shared publicly
- ☐ Direct threats of physical harm, sexual violence, or death
- ☐ Viral or rapidly escalating coordinated harassment campaigns
- ☐ Outing threats or non-consensual sharing of intimate images
- ☐ Attack extending to family members, colleagues, or close contacts
- ☐ Blackmail or attempts to coerce withdrawal from public life

Immediate Action [refer to 12.3 Incident Response Checklist]

3. RISK LEVEL SUMMARY

Compare your current situation across each indicator to estimate where it sits on the risk scale.

INDICATOR	LOW RISK	MEDIUM RISK	HIGH RISK
◆ **Visibility**	Limited	Growing	Wide or viral
◆ **Threats**	None	Implied or emerging	Direct and credible
◆ **Personal Data**	Not exposed	Partially exposed	Publicly shared
◆ **Coordination**	No	Early signs	Confirmed
◆ **Others Affected**	No	Beginning	Yes – actively targeted
◆ **Offline Risks**	No	No	Present or imminent

LOWER → HIGHER SEVERITY

Read indicators together, not in isolation. A single signal may sit at low risk while the overall situation is higher — if several indicators point to medium or high, treat the situation at the higher level.

12.3 Incident Response Checklist

1. Pause and Assess

- ☐ Step back from the screen. Do not respond publicly or privately to the attack while distressed. Do not engage with aggressors.
- ☐ Identify the main form of attack. Is more than one tactic being used simultaneously.
- ☐ Preserve messages, content and evidence. Do not delete them and do not delete or deactivate accounts.
- ☐ Document all abusive content and store it safely and chronologically.
- ☐ Ask a trusted person to take over monitoring abusive content.
- ☐ Stop posting real-time location, stories, or posts.
- ☐ Vary your routines and movement patterns where possible in case of offline risks.
- ☐ Take a trusted person with you to in-person meetings and events.

2. Secure Accounts

- ☐ Change passwords immediately on all key accounts: email, social media, banking, and accounts linked to the campaign.
- ☐ Review active sessions and remotely log out of all browsers and devices.
- ☐ Disable real-time location sharing on devices and applications. Turn off GPS sharing.
- ☐ Disable geotagging on camera and social media applications.
- ☐ Review shared accounts. Remove unnecessary access to shared cloud storage, calendars, and device synchronization.
- ☐ Update operating systems, applications, and security software to the latest available versions.

3. Activate Trusted Support Networks

- ☐ Inform at least one trusted person of the situation (a friend, a colleague, or a family member).
- ☐ Ask your trusted network for specific help: monitoring content, collecting and documenting evidence, or handling communications.
- ☐ Inform your campaign team, organizational colleagues, or employer.
- ☐ Contact a digital security organization if you suspect account compromise, device infiltration, or technically complicated attacks.
- ☐ Contact a legal aid or human rights organization when serious threats of outing, doxxing, defamation or outing are identified.
- ☐ Contact a psychosocial support provider if attack is causing distress, fear or exhausting. Refer to section 9.4 Resource List for relevant organizations.

4. Decide on a Response Path (based on risk level)

- ☐ Monitor without public engagement (low-risk attacks):
  - ☐ Continue documenting and monitoring with support.
  - ☐ Block and restrict abusive accounts.
  - ☐ Report content to platforms where relevant.
  - ☐ Reassess regularly for signs of escalation.
- ☐ Escalate to organizational, legal or security support (medium to high risk):
  - ☐ Inform employers, organizations, campaign teams, and family members.
  - ☐ Seek digital security support.
  - ☐ Contact legal aid or human rights organizations.
  - ☐ Report to authorities where safe and appropriate.
- ☐ Issue a limited public clarification (coordinated misinformation, false allegations, impersonation):
  - ☐ Do not engage directly with aggressors or reshare harmful content.
- ☐ Do not respond emotionally or under pressure.
- ☐ Focus on correcting false information with factual, brief, non-engaging language.
- ☐ Reassess the situation regularly.

12.4 Documentation Checklist

Document before reporting as content may be removed once a report is submitted. Work through each category below and tick off what has been captured.

Screenshots

- ☐ Abusive posts, comments, replies, and stories
- ☐ Private or group messages
- ☐ Images, videos, voice notes, or livestreams containing harmful content
- ☐ Username, date and time, platform and full content must be visible in each screenshot
- ☐ View-once or disappearing content photographed using a second device
- ☐ Threats of physical harm or violence
- ☐ Targeting of family members, friends, or colleagues
- ☐ Photographs taken without your consent

URLs and Links

- ☐ Direct link to every abusive post, profile, account, channel, or group
- ☐ Each platform where the content appears is documented separately

Account Details

- ☐ Usernames, display names, and account handles of all accounts involved
- ☐ Profile pictures
- ☐ Follow counts
- ☐ Account creation dates
- ☐ Any visible political party affiliation or identifying information

Incident Log

- ☐ Date and time the attack began and on which platform
- ☐ Date and time threats were identified
- ☐ Date and time personal information was shared
- ☐ Date and time harassment spread to additional platforms or audiences
- ☐ Any other escalation points noted as they occur

Calls and Direct Contact

- ☐ Unknown numbers that have spam called or texted
- ☐ Record frequency of calls and messages
- ☐ Voice notes recordings

Storage and Handling of Evidence

- ☐ All evidence is stored in a secure, private folder
- ☐ Back up on external drive or encrypted storage
- ☐ Organized chronologically
- ☐ Preserved even if reporting is not yet decided

12.5 Moderator and Page Admin Checklist

For use by anyone managing a community page, campaign account, or civic platform, particularly during elections and politically sensitive periods.

- ☐ Document abusive content
- ☐ Remove comments containing abusive content
- ☐ Do not quote, report, or publicly engage with aggressors
- ☐ Do not reshare, repost, or allow circulation of harmful content
- ☐ Apply moderation standards consistently across all identities, affiliations and subjects
- ☐ Treat unverified rumors about a person's private life as harmful content, not discussion topics
- ☐ Ensure community members have a clear, accessible way to report harmful content to admins
- ☐ Review and update who has administrative access
- ☐ Keep posting responsibility centralized to one person or small trusted team
- ☐ Use strong and unique passwords and enable two-factor authentication on all admin accounts

13

SECTION 13

Final Notes and Key Resources

13.1 Core Takeaways

Online attacks during politically sensitive periods and elections are designed to restrict public participation, and they disproportionately target women and LGBTQI+ persons. This is because their presence in political life challenges entrenched social norms, disrupts established political power dynamics, and advances changes that conservative actors and dominant political parties actively resist. Digital violence is used to punish their visibility and engagement in public life. This kit was developed in response to that reality. It provides practical tools for safer participation so that those facing targeted online violence can remain engaged in political and civic life without being forced to fully withdraw.

Once online violence is identified, targeted people and their support network should pause before acting, assess the situation, identify what type of attack is taking place, its reach, its level of risk and coordination, and whether it is escalating. Acting under pressure without this assessment could inadvertently amplify harm. Individuals should immediately document evidence of digital attacks even if they are not yet considering reporting the incident or seeking formal complaint pathways. Content can disappear quickly the moment a platform review is triggered, or an account is deactivated, therefore, preserving screenshots, URLs, account details and threatening messages is crucial. Documentation can also support legal processes, organizational support, and monitoring the escalation of the attack over time.

Risk levels of online attacks are not fixed, and situations evolve rapidly overtime especially during politically active situations. It is, thus, important to reassess the situation regularly, update and adapt response strategies as the situation changes. Depending on the risk level of the attack, continued engagement in political life can amplify the impact on physical and psychological wellbeing, prompting targeted individuals to reduce exposure. However, limiting exposure is not the same as withdrawing from public life. This kit pushes for temporary adjustments to visibility, platform use, and communication habits that prioritize safety while sustaining participation. Safety planning should always aim to keep people engaged where possible, on their own terms and at a pace they control. Additionally, digital coordinated campaigns should not be handled alone. Trusted contacts, allies, moderation support, peer networks, family, and friends can assist in managing the attack, collecting evidence and support targeted people's mental health state. The targeted individual's circle shape whether an attack is contained or amplified, as they can either unintentionally contribute to the harm, or support the targeted person. Therefore, identifying and preparing a support network before an incident occurs is a crucial protective strategy.

It is important to note, that platform reporting and formal state channels are tools that carry real limitations and, in the context of Lebanon, severe risks. Platform reporting does not guarantee removal of harmful content, does not protect against physical threats, and cannot address coordinated cross-platform abuse. They can, however, help limit harm and visibility and should be used alongside other safety measures. State institutions, including the ISF's Cyber Security Unit, can make formal complaints dangerous rather than protective depending on who is behind the attack, and the identity of the person attacked. This is particularly problematic for LGBTQI+ defenders. Reporting should, then, be considered carefully, and pursued only where it is safe and appropriate to do so.

> The tools in this kit will not eliminate these risks, but when used consistently, carefully and collectively they can reduce harm, support continued engagement and ensure that digital violence does not hinder public participation.

13.2 Key Resource Links

GUIDES TO ONLINE PROTECTION

- SMEX – How can women protect themselves online?
- UNHCR – Respond to online harm and reach out for help
- UNFPA – Reporting Tip sheet on Digital Violence: A practical reference guide for journalists and media

GUIDES TO PLATFORM REPORTING

- TikTok
- Facebook
- Instagram
- X/Twitter
- WhatsApp

USEFUL CONTACTS

Legal Agenda	**+961 1 383 606**	**Helem**	**+961 81 478 450**
CLDH	**+961 1 24 00 23**	**MOSAIC**	**+961 76 945 445**
ABAAD	**+961 81 788 178**	**Proud**	**+961 76 608 205**
RDFL	**+961 71 500 808**	**SMEX**	**+961 81 633 133**
Access Now	**+1 (888) 414 0100**	**Embrace**	**1564**

CITATIONS

References

References are numbered in the order they first appear in the text. Click any [N] marker in the body to jump to the corresponding entry below.

[1] Skeyes Media (2023, March). Representative of UN Women: Online Bullying discourages women in Lebanon from engaging in politics and running in elections. Retrieved from: skeyesmedia.org

[2] Fe-male (2021, June). Online Violence Against Women Human Rights Defenders in the MENA: Experiences and Perceptions. Retrieved from: fe-male.org/archives/13653

[3] Daraj Media (2024, November). Biased Algorithms: How Digital Platforms Reinforce Abuse Against Female Politicians in Lebanon. Retrieved from: daraj.media

[4] Human Rights Watch (2023, September). Lebanon: Attack on Freedoms Targets LGBTI People. Repressive legislation; unlawful crackdown. Retrieved from: hrw.org

[5] HuMENA for Human Rights and Civic Engagement (2024). A Report on the Legal and Social Situation of the LGBTQI+ Community in the Middle East and North Africa Region and its Radical Transformation Between 2020 and 2023 — Lebanon as a Model. Retrieved from: humena.org

[6] Amnesty International. Online Violence. Retrieved from: amnesty.org

[7] UN Women (2024, September). Technology Facilitated Gender-Based Violence: Developing a shared research agenda. Retrieved from: unwomen.org

[8] UNFPA (2024, December). An infographic guide to Technology-Facilitated Gender-Based Violence. Retrieved from: unfpa.org

[9] New Research Illuminates Escalating Online Violence on Musk's Twitter. International Center for Journalists (2023, July). Retrieved from: icfj.org

BIBLIOGRAPHY

Further Reading

Additional sources informed the research behind this Kit and are listed here for readers who want to explore further. These materials are not cited inline in the body.

- SMEX (2024, March). 80% of Women in Lebanon Face Digital Violence. Retrieved from: smex.org
- UN Women (2025, March). Shaping Perceptions: Media influence on women's political participation during and beyond elections in Lebanon. Retrieved from: lebanon.unwomen.org
- Maharat (2022, August). Media and Gender Monitoring of the 2022 Elections VAWP. Retrieved from: maharatfoundation.org
- Arab News (2022, April). Lebanese female candidates stand up to Hezbollah, are disowned by families. Retrieved from: arabnews.jp
- UNFPA. Technology-facilitated Gender Based Violence: A Growing Threat. Retrieved from: unfpa.org/TFGBV
- Human Rights Council (2020). Combating Violence Against Women Journalists. Retrieved from: docs.un.org/en/A/HRC/44/52

Stay safe.
Stay engaged.

The tools in this kit will not eliminate these risks, but when used consistently, carefully, and collectively they can reduce harm, support continued engagement, and ensure that digital violence does not hinder public participation.

General: info@humena.org
Advocacy: advocacy@humena.org
Lebanon: lebanon@humena.org
Web: humena.org

HuMENA For Human Rights and Civic Engagement
HuMENA pour les Droits de l'Homme et l'Engagement Civique
هيومينا لحقوق الإنسان والمشاركة المدنية

INNOVATION FOR CHANGE
MIDDLE EAST & NORTH AFRICA

CIVICUS

Digital Democracy Initiative