Funded by
the European Union

Mapping Digital Hate: A Landscape Review of Online Hate Speech in Jordan

Zain Hakim & Ihsan Al Khleifat

Content warning: This report contains mentions and examples of hateful content which some readers may find distressing.

Implemented by

Saba Hamlet
for gender equality

ISD Institute for Strategic Dialogue

IDARE
for sustainable development
أنا أتجرأ للتنمية المستدامة

This research series was part of The Strengthening Prevention and Response Mechanisms to Intersectional Online Hate Speech in Jordan. A project funded by the European Union and implemented by ISD Jordan in partnership with I-DARE Sustainable Development and Saba Hamlet for Gender Equality. It aims to address the challenge of online hate speech in Jordan through a comprehensive, evidence-based programme of research, policy engagement and .community action

Implemented by

Saba Hamlet
for gender equality

ISD | Institute for Strategic Dialogue

I DARE
for sustainable development
أنا أتجـــــرأ للتنميـــــة المستدامـــــة

Amman | Berlin | London | Paris | Toronto | Washington DC

This document has been produced with the financial assistance of the European Union. The contents of this document are the sole responsibility of The Institute for Strategic Dialogue and can under no circumstances be regarded as reflecting the position of the European Union.

www.isdglobal.org

Contents

Executive summary

Over two years of research across X and Instagram, ISD Jordan has built an evidence-based picture of online hate speech in Jordan, documenting who is targeted, how harmful content spreads, what drives it and what needs to change. This report brings together an overview of key trends and dynamics of online hate speech in Jordan, summarising the findings of a research series, conducted between September 2024 and April 2026 as part of the EU-funded project "Strengthening Prevention and Response Mechanisms to Intersectional Online Hate Speech in Jordan," implemented by ISD Jordan in partnership with I-DARE Sustainable Development and Saba Hamlet for Gender Equality.

The project was designed to address a set of clearly identified gaps: the absence of a nationally agreed definition of hate speech, policy frameworks, a lack of evidence-based research on hate speech in the Jordanian context, limited Arabic-language resources for prevention and response and institutional awareness of the nature and scale of the problem. Two years on, this research series represents a direct and substantive response to each of those gaps by providing an evidence-based national framework developed through a multi-stakeholder process and generating recommendations grounded in data.

The research had four core objectives: to identify and map key trends and narratives related to online hate speech across social media platforms; to examine spikes linked to political, social or cultural events; to assess which groups are targeted and how; and to generate actionable, data-driven recommendations for policymakers, platforms, civil society and media actors. Across each of these objectives, the research produced concrete findings.

As with other countries in the Middle East and North Africa (MENA) region, Jordan is heavily affected by events in its neighboring states, including conflicts, as well as political and economic instability. Regional trigger events -such as the ongoing Gaza conflict and the events in Sweida, Syria generated spikes of sectarian and nationalist content online. These conflicts profoundly shape Jordanian society across ideological, economic, political and social dimensions. The cumulative effect is that hate speech becomes not merely reactive but also embedded in how Jordanians process regional instability—reflecting anxieties about resources, sovereignty and survival that transcend the immediate conflict.

Across the dataset, ISD identified three persistent typologies of harmful content on Jordan's social media platforms: gender-based abuse targeting women in public life, religious hate speech directed disproportionately at Shia communities and other minorities and exclusionary nationalist discourse. These are not peripheral phenomena. They are structural features of Jordan's online environment, present across platforms, persistent across time and concentrated against specific communities.

These findings point to something important: Jordan's online environment cannot be understood, or addressed, as a purely domestic problem. ISD's research consistently documented how pan-Arab narratives, sectarian framings originating elsewhere in the Arab world, and content from regional influencers circulate freely on Jordanian social media platforms, where they are adapted and amplified within local contexts. The country's information ecosystem is inseparable from broader regional flows. Any response that treats online hate speech in Jordan as a domestic issue alone will be incomplete.

ISD's research also identified two structural conditions that make harmful content harder to detect, limit and address. The first is the Arabic-language moderation gap. Major platforms have systematically underinvested in Arabic-language tools, human review and accessible redress mechanisms. ISD found direct evidence that this gap is being actively exploited: Arabizi—Arabic written in Latin characters and numerals—is used as a deliberate and widespread workaround to evade automated detection. Users posting hateful content have adapted their behaviour to the known limitations of platform moderation systems, and those systems have not kept pace. The second structural condition is algorithmic amplification. Content that provokes strong emotional responses such as identity-polarising, divisive content targeted at specific communities, receives disproportionate reach on platforms optimised for engagement. ISD's monitoring found this dynamic operating across all three typologies, with harmful narratives gaining visibility.

Key findings

- **Regional conflicts result in upticks in hate speech in Jordan. Crises in Gaza and Syria triggered immediate spikes in sectarian and nationalist hate speech in Jordan, but the impact extended beyond one-off reactions.** ISD research noted that these conflicts shape Jordanian society across ideological, economic, political and social fault lines, embedding hate speech into how citizens process regional instability. The rhetoric reflected deeper anxieties about resource scarcity, sovereignty and national survival that persist beyond individual triggering events.

- ISD researchers identified three forms of harmful online content present in the Jordanian digital ecosystem, which were primarily focused on gender, religion and ethnicity:

    - **Misogyny in Jordan operates as both an attitude and a patriarchal system of power, manifesting online as a spectrum of harmful content.** Researchers found 15 percent of X comments and five percent of Instagram comments targeting women included threats and incitement to violence, sustained harassment and sexualized bullying, and pervasive derogatory and dehumanizing language.

    - **Sectarianism is a key element of the online Jordanian ecosystem but ultimately forms a relatively small share of overall digital harms.** Researchers noted that 17 percent of data collected from X disproportionately targeted Shia communities, alongside other religious minorities (such as the Druze). The posts combined dehumanising rhetoric and an occasional explicit incitement to violence.

    - **Nationalism-related discourse coursing through the Jordanian digital ecosystem was derogatory and exclusionary.** Researchers found 10 percent of content on X targeted Jordanians mainly across Jordanian-Palestinian and tribal lines. This content intensified during political or economic strain and undermined social cohesion by reinforcing identity-based hierarchies.

- **Hate speech in Jordan operates as part of a wider ecosystem of digital harm. ISD's research consistently found hate speech intersecting with disinformation, and coordinated harassment campaigns.** These forces amplify one another. Disinformation creates the conditions in which hate speech spreads more easily, while coordinated inauthentic behaviour extends its reach.

- **Arabic-language users are exploiting loopholes in moderation to spread harmful content.** Researchers found accounts using Arabizi: an informal script using numbers and Latin characters to write Arabic terms. Arabizi aids users to avoid detection, thus encouraging the spread of harmful content, which faces less scrutiny than equivalent content in English, and platforms have not been sufficiently concerned with closing that gap.

- **The legal framework governing digital platforms in Jordan lacks definitional precision and appears to be overused for criminalizing expression.** ISD's assessment found that key provisions of the 2023 Cybercrime Law rely on broad, undefined terms resulting in interpretive ambiguity and inconsistent enforcement. The 3,170 prosecutions recorded in 2024 reflect a law being applied beyond its intended scope.

Methodology

This research was conducted over a two-year period and employed a mixed-methods approach combining quantitative social media analysis with qualitative content assessment.

The study examined online discourse across several major social media platforms, including X (formerly Twitter), Facebook, YouTube and Instagram. Data collection and analysis methods were adapted to the technical affordances and data access limitations of each platform. X served as the primary source for quantitative analysis due to its greater accessibility for large-scale data collection and geolocation capabilities. Complementary qualitative analysis was conducted across the other platforms to capture broader patterns of discourse within Jordan's digital ecosystem.

Publicly available social media data was collected using social listening tools, primarily Brandwatch. The research team developed structured search queries using predefined keywords and hashtags in both Arabic and English, tailored specifically to the Jordanian context. These keyword sets were developed by the ISD research team for each individual paper within the project. Queries were designed to capture discussions related to hate speech, polarisation and other narratives relevant to Jordanian online spaces. The resulting datasets were compiled and visualised through a publicly accessible analytical dashboard that enabled ongoing monitoring of trends and patterns in the data.

Quantitative analysis focused primarily on X, where the platform's data structure and geolocation functionality allowed for more precise filtering of content originating from Jordan. These features enabled the research team to assess content volume, engagement levels and trends in online conversations. Metrics such as post frequency, repost activity and interaction rates were used to identify peaks in discussion and examine how specific narratives evolved over time. This quantitative component helped identify key moments of heightened online activity, which were then examined more closely through qualitative analysis.

Qualitative analysis was conducted across all platforms and played a central role in interpreting the narratives and discourse emerging from the data. This process involved a manual review of sampled posts to assess tone, context and framing. Attention was given to identifying recurring themes, discourse patterns and forms of harmful or polarising language. Instagram content was examined primarily through qualitative methods due to the platform's limited accessibility for large-scale data extraction. In these cases, researchers conducted targeted searches using the predefined keyword lists and analysed posts, comments and visual content to understand how narratives were presented and circulated.

Several methodological challenges affected data collection. One key limitation was the linguistic diversity of Arabic-language social media discourse. Online discussions often include multiple dialects, informal language variations and hybrid writing systems such as Arabizi. Users also frequently incorporate emojis, abbreviations or mixed Arabic-English expressions within their posts. These factors complicate automated data collection and keyword-based searches as relevant content may appear in many different forms. While the research team attempted to account for these variations through extensive keyword development, it was not always possible to capture every relevant linguistic variation.

Another challenge arose from the fact that most platforms do not provide reliable geolocation metadata that would allow researchers to easily restrict datasets to a specific country. As a result, some data collection processes required additional manual filtering and contextual interpretation to determine whether content was relevant to Jordanian discourse. Despite these limitations, the combination of platform-specific methods, multilingual keyword queries and qualitative validation allowed the research team to capture a broad and representative snapshot of online discussions relevant to the study.

This research was further supported through a collaboration with CASM Technology, which acted as the project's technological partner. CASM contributed to the classification and structuring of the dataset using advanced machine learning techniques, helping to identify key themes, narratives and patterns across large volumes of social media data. They also supported the development of the project's digital analytical dashboard, which enabled the visualisation of findings across all research papers, including the most prevalent narratives, recurring keywords and available demographic insights, thereby enhancing the accessibility and analytical value of the research outputs.

All research activities adhered to ethical data collection standards. The research relied exclusively on publicly accessible data and did not involve the collection of private user information. Personal identifiers were not included in the analysis. The research followed established principles for responsible social media research, including anonymisation and respect for user privacy.

Findings generated through this methodology formed the basis of a series of research papers produced throughout the project. Each publication underwent an internal review and quality assurance process before being released through the project's public website.

Glossary

Harmful online content: Text, images, videos, audio or interactive material disseminated via digital platforms that is likely to inflict physical, emotional, psychological, social or economic harm, incite hostility or discrimination, facilitate illegal or dangerous behaviour, or undermine the safety, dignity or rights of individuals or communities. Hate speech: ISD classifies hate speech as any activity or content that seeks to dehumanise, demonise, harass, threaten or incite violence against an individual or community on the basis of their race, ethnicity, sexual orientation, gender, sex, disability, migrant status or religion.

Harassment: Harassment targeting a specific individual or group with the intent to threaten, provoke or cause distress

Disinformation: The deliberate sharing of deceptive information designed to mislead others, manipulate public opinion and influence public policies. It can play a significant role in fuelling hostility toward specific groups. Misinformation: False information shared without harmful intent. Similar to disinformation, it can fuel hostility towards specific groups.

Moral shaming: When individuals/communities shame a target based on social norms, cultural expectations or communal codes of conduct, rather than formal religious doctrine. It is tied to social ethics: i.e. what the community views as respectable, appropriate or honourable. Moral shaming is related to urf or aib.

Sexual harassment: Unwelcome sexual advances, requests for sexual favours and other verbal or nonverbal conduct of a sexual nature.

Dehumanising speech: Verbal or written expressions that deny the inherent dignity, humanity or moral worth of individuals or groups by portraying them as less than uman (such as comparisons to animals, filth or demons). Threatening language and incitement to violence: The incitement of violence based on gender, gender identity or sexuality including sexual violence, physical assault and murder.

Derogatory speech: Speech that is insulting, derogatory, uses offensive stereotyping and potentially leads to emotional distress but does not reach the threshold for legal action.

Discriminatory speech: Speech that discriminates against individuals based on personal characteristics that can lead to marginalisation and exclusion.
Arabizi: Also referred to as Latinised Arabic, which is a hybrid form of Arabic using Latin script and numerals that is often used to bypass content moderation filters.

Jordan's Online Ecosystem

The nature of many of the digital harms found in Jordan are shaped by a distinct combination of local socio-political dynamics and broader structural factors. These contextual factors are influenced by gender, identity and regional crises that are a mix of internal and external issues in Jordan.

Deep-rooted patriarchal norms and traditional gender roles contribute to the normalisation of misogynistic behaviour online. In 2025, ISD Jordan research found that 15 percent of all posts and comments collected on X over a seven-month period contained misogynistic content. Women in public life—such as politicians, journalists and activists—faced disproportionate levels of targeted abuse. This abuse included dehumanising language, threats and content designed explicitly to silence their participation in public life. During the September 2024 parliamentary elections, in which almost 25 percent of the candidates were women, female candidates faced intense online harassment questioning their competence, emotional stability and legitimacy in leadership roles. According to a survey by the Salamat programme for digital safety, almost 55 percent of female journalists in Jordan said they had experienced digital violence at some point in their careers.

Identity is a second significant driver of community tensions in Jordan, which hosts one of the largest refugee populations in the world relative to its size. Religious-based hate speech represents represents a further persistent category of harmful content. This speech particularly targets Shia communities and other minority faith groups and periodically intensifies during regional developments.

Regional geopolitics function as a recurring and powerful trigger. The ongoing conflict in Gaza has generated sustained spikes in hate speech across multiple categories, reinforcing both nationalist discourse and religious-based hostility. Pan-Arab narratives circulating across the wider region feed directly into Jordan's online environment, making the country's information ecosystem inseparable from broader regional flows of content and disinformation. Rabbit holes formed through algorithmic curation reinforce these effects, limiting users' exposure to diverse perspectives and accelerating the spread of polarising content within insular communities.

Platform dynamics and the Arabic-language moderation

A structural challenge underlying these dynamics is a consistent underinvestment by major social media platforms in Arabic-language content moderation. Although Arabic is spoken by approximately 380 million people, research consistently documents that content moderation resources, tools and accountability mechanisms lag significantly behind those available in English-language markets. The Middle East Institute has also found that platforms do not publish robust data on how their moderation practices are enforced in MENA. They also found that content moderation systems enable functional discrimination: platform policies, appeals processes and community standards are frequently unavailable in accessible Arabic, preventing users from understanding or advocating for their rights.

The Arabic language itself presents particular technical challenges. Modern Standard Arabic, the dialect primarily used to train automated classifiers, differs substantially from the colloquial dialects used by the wider population. The Tech Global Institute has documented that this means that Arabic classifiers are significantly under-resourced, with publicly available analyses and Oversight Board comments repeatedly identifying substantial gaps in Arabic moderation accuracy. This challenge is actively exploited: ISD's research identified how the use of Arabizi can be used as a workaround to evade automated detection when posting hateful content. Platforms have also moved toward greater reliance on automated tools at the expense of human review. This decision disproportionately harms Arabic-language users given the known limitations of current automated systems in this context.

Legal context of the Jordan digital sphere

Jordan addresses online hate speech through a layered framework of existing legislation rather than a dedicated law. The principal instruments are the No. 17 of 2023, Article 150 of the Penal Code No. 16 of 1960 and the Press and Publications Law. Each contains provisions applicable to online speech, reflecting a broader effort to adapt legal frameworks to the realities of a rapidly evolving digital environment.

The 2023 Cybercrime Law is the most significant of these instruments. The law introduced protections against cyber-extortion, phishing, unauthorised access and the creation of counterfeit accounts. The law also introduced a degree of platform accountability, requiring social media companies with Jordanian user bases to establish local offices and respond to judicial requests. The law has been deployed in a context where digital harms—particularly against women, who are the primary victims of cyber-extortion according to Jordan's own Cybercrime Unit—are growing.

The law's principal hate speech provision, Article 17, makes it an offence to disseminate content that promotes sectarian strife, insults religious beliefs or hate speech. Related provisions across Articles 14, 15 and 16 address character assassination, incitement to immorality, and undermining national unity. The Press and Publications Law makes electronic publishers liable for both their own content and user-generated comments posted on their platforms.

However, these advances are undermined by the law's central weakness: the absence of clear, legally precise definitions for the speech-related offences it criminalises. Articles 15, 16 and 17 rely on broad and undefined terms. None provides a legally precise definition of hate speech. Terms such as "sedition", "sectarian strife", "undermining national unity" and "contempt for religions" remain undefined in the text of the law. Although this is not unique to Jordan it nevertheless creates interpretive ambiguity. The relevant international benchmark, Article 20 of the ICCPR, sets a specific threshold: speech must constitute "advocacy of national, racial or religious hatred that constitutes incitement to discrimination, hostility or violence". Aligning domestic definitions more closely with international standards would help distinguish targeted, harmful speech from protected expression.

Jordan's National Centre for Human Rights recorded 3,170 prosecutions under the 2023 Cybercrime Law in 2024, the majority for online defamation; a figure that reflects the law's broad usage. These numbers point to the importance of clearer definitions. Tensions also sit within Jordan's own constitutional framework. Article 15 of the Constitution guarantees freedom of expression, while Article 7 protects rights and public freedoms from infringement. Developing clearer definitions of hate speech would serve both commitments: strengthening protections for individuals targeted by harmful speech while providing greater legal clarity for all.

Trigger events and amplification dynamics: how offline events translate into online hate

A growing body of research demonstrates that offline political, social or violent events frequently act as "trigger events" that produce measurable spikes in online hate speech. A 2025 study by George Washington University revealed that real world events are often followed by surges in several types of online hate speech on both fringe and mainstream social platforms. These dynamics occur when real-world tensions activate identity-based grievances that are subsequently amplified within digital communication networks. Studies of social media discourse show that major offline incidents—including protests, elections, terrorist attacks or geopolitical conflicts are often followed by rapid increases in hateful rhetoric online.

Large-scale analyses of social media activity further support these findings. Studies examining millions of posts across multiple platforms show that real-world events can generate sudden spikes in various forms of hate speech, including racism, misogyny, xenophobia and religious hostility. These increases are not always limited to groups directly associated with the triggering event. For example, ISD's research on religious hate speech found that the conflict in Sweida, Syria involving the Druze community triggered hostility towards both Druze individuals and other religious groups including Shia and Alawite communities.

Research conducted by Jordanian journalists found that social media platforms in Jordan contribute directly to the spread and escalation of hate speech. The study concludes that social media has a statistically significant impact on the spread of hate speech. In turn, the spread of hate speech incites discord within society.

Trigger events tend to generate large volumes of emotional discourse online, creating fertile conditions for the spread of hateful narratives. For instance, in 2016, hate speech towards the Jordanian Christian journalist Nahed Hattar resulted in his death. Public hate speech and death threats against Hattar on Jordanian social media platforms included calls for his execution. Hattar shared a cartoon on his Facebook page that Sunnis in Jordan perceived to be mocking God. Later, he was shot dead in front of a courthouse in Amman, in an unprecedented assassination that sent shock waves throughout the kingdom.

Research also shows that social media platforms allow hate speech to spread rapidly and influence collective values, identity and intergroup relations. A study on university students in 2023 shows that social media contributes significantly to the spread of hate speech and the exacerbation of conflicts. The study found that the prevalence of hate speech on social media platforms leads to increased conflict, societal extremism, a decline in religious and political values, and the incitement of discord within society, along with the spread of division, rivalry and sectarianism. Statistically, the study found a significant relationship between social media usage and the spread of hate speech.

In the Jordanian context, these dynamics highlight how offline developments (including political controversies, regional conflicts or domestic social tensions) can quickly translate into waves of online hostility. When such trigger events occur, social media platforms often become spaces where grievances are rapidly amplified, narratives are contested and hostile rhetoric spreads across networks. Understanding these amplification dynamics is essential to identify periods of heightened risk and thus develop strategies to monitor and mitigate the spread of hate speech online.

Jordan does not operate within an isolated digital ecosystem: political crises, sectarian narratives and identity conflicts from the broader Arabic-speaking information sphere circulate across national boundaries. As a result, online hate speech in Jordan is shaped not only by domestic developments but also by regional dynamics. Narratives originating in other parts of the Middle East frequently enter Jordanian online conversations where they are adapted, reframed and amplified within local context.

One key driver of this dynamic is conflict spillover from neighbouring crises, particularly those related to Palestine and Syria. Political and humanitarian developments in Palestine have long played a central role in shaping public discourse in Jordan: a significant portion of the population has Palestinian heritage, and the Palestinian cause occupies an important place in political and social identity. Periods of heightened violence or political escalation in the Israeli–Hamas conflict often generate intense emotional reactions across Jordanian social media. Much of this discourse centres on political solidarity or humanitarian concerns. However, such moments can also create conditions in which hostile rhetoric and identity-based narratives proliferate online. The Syrian conflict has also had a profound influence on the regional information environment in which Jordanian social media operates. The war in Syria produced a large-scale refugee crisis. It also led to the spread of highly polarised regional narratives along sectarian, political and ideological lines. These narratives continue to circulate widely across Arabic-language media and social platforms. In some cases, they contribute to the spread of religious or sectarian hate speech online. This is particularly evident when regional developments reignite debates around identity, loyalty or geopolitical alignments.

Pan-Arab identity narratives further shape how these regional conflicts are interpreted and discussed online. Within the Arabic-speaking information sphere, political events are frequently framed through broader narratives of collective identity, historical grievances and regional solidarity. While such narratives can foster shared political consciousness across borders, they may also contribute to the circulation of polarizing or sectarian rhetoric. Religious identity (particularly distinctions between Sunni, Shia and other religious communities) can become a focal point in online discourse during periods of regional crisis. As a result, sectarian language and religious hostility sometimes emerge in online conversations that are only indirectly related to events within Jordan itself.

Another factor shaping the spread of hate speech in Jordan is the influence of regional media and influencer ecosystems. Social media platforms have enabled journalists, political commentators, activists, and online personalities across the Middle East to reach audiences that extend far beyond their national borders. Content produced by high-profile regional commentators, religious figures or political influencers can quickly circulate within Jordanian digital spaces, where it may shape public discourse or reinforce existing narratives. In some cases, such content introduces polarising frames or identity-based rhetoric that can contribute to the escalation of online hostility. These cross-border dynamics illustrate how online discourse in Jordan is embedded within a wider regional information environment. Hate speech circulating within Jordanian social media spaces often reflects the interaction between local grievances and transnational narratives originating elsewhere in the Arabic-speaking world. Consequently, efforts to understand and address online hate speech in Jordan must account for broader regional networks through which narratives, identities and grievances circulate online as well as domestic political and social dynamics.

Typologies of online hate speech in Jordan

The ISD team identified three main types of harmful online content in the Jordanian digital sphere:

- Gender-based abuse and harassment directed at women and other non-gender conforming minorities. Content of this nature may be legal or illegal and result in, or be likely to result in, physical, sexual, psychological or economic harm.
- Religious-based hate and abuse.
- Nationalism-based othering and exclusionary discourse; speech that asserts superiority of one group over another or delegitimises beliefs.

Harmful content targeting women in public

Misogyny can be viewed as both "an attitude and as a system of power". At the level of attitudes, it refers to hostility, prejudice or resentment directed toward women. As a system, it operates to reinforce patriarchal hierarchies and sustain gender inequality within social, political and cultural structures. Research shows that misogyny exists along a continuum, ranging from normalised forms of sexism to more extreme expressions, including explicit advocacy of violence.

A growing body of research shows that misogyny can constitute a "form of extremism in its own right, often referred to as male supremacism". Within this framework, violence against women is rationalised through ideological narratives that depict women as threats to male authority, objects to be controlled, or adversaries whose perceived transgressions justify punishment. As Jordan is patriarchal society, a significant deviation from traditional gender-roles is felt as a threat to male supremacy and authority. In fact, within the last two years, ISD researchers in Jordan have found a disproportionately high level of misogynistic and harmful content targeting women—especially those in the public eye. This also extends to individuals who do not conform to dominant gender norms. This frequently manifested as gender-based insults, moral condemnation, sexualized harassment and attempts to discredit professional competence or personal integrity.

An analysis of posts and comments from X, Facebook and Instagram found deep-seated patriarchal values and traditional gender roles in Jordan have contributed to the normalisation of harmful behaviour online. Most harmful content aimed at women was derogatory in nature. However, ISD also identified dehumanising language, sexualised harassment, and threatening language and incitement to violence.

Fifteen percent of comments on X and five percent on Instagram were identified as harmful content targeting women. This content spread across three major themes:

Threatening language and incitement to violence towards women who do not adhere to traditional gender roles or behave in a manner perceived as immoral or indecent

This frequently framed Jordanian women's autonomy as a threat to social order, family honour or religious values. Studies on digital gender violence show that such threats are often justified through moralistic narratives. These portray potential punitive action against women as socially or religiously legitimate rather than criminal.

A report by the Jordanian National Commission for Women highlights that online threats frequently mirror offline honour-based discourse. This includes calls for "punishment" against women who are perceived to transgress norms, reflecting deeply embedded patriarchal control mechanisms. Such rhetoric contributes to a climate of intimidation that can deter women from participating in public debate, reinforcing gender exclusion in civic life.

Harassment, trolling and bullying language

Beyond explicit threats, sustained harassment campaigns constitute a significant portion of harmful content. These often involve coordinated ridicule, repeated hostile comments and attempts to provoke emotional distress. These behaviours are consistent with definitions of cyberbullying documented in ISD's research.

A study by the United Nations Development Programme found that women in Arab societies experience higher levels of online harassment than men. This is particularly true for women with visible public profiles. Attacks frequently target their appearance, marital status or perceived morality rather than providing substantive arguments.

In Jordan, ISD and other digital rights organisations has shown that trolling often escalates during politically sensitive periods or when women challenge dominant narratives. This finding suggests harassment functions as a mechanism of social policing. The persistence and volume of such attacks can create what scholars describe as a "hostile communication environment", discouraging participation in public life and normalising abuse for women online.

ISD's report on harmful content targeting public figures and influencers found that nine percent of comments from Jordanian Instagram users contained sexual harassment exclusively targeting women. Comments frequently relied on misogynistic insults, sexual slurs, references to prostitution, revealing clothing or repeated references to specific body parts. In several cases, users implied influencers or public figures were involved in pornographic or sexually exploitative activities/encouraged them to take part.

Derogatory and dehumanising language

Derogatory and dehumanising language was the most common form of content. Posts and comments of this nature used language to silence, shame and discredit women. These included:

Gendered and misogynistic slurs, insults and derogatory terms (e.g. "spinster", "whore") or claims that women are "lacking intellect".

These expressions draw on longstanding cultural stereotypes that define women's place primarily through their marital status, sexual propriety, and perceived emotional or intellectual inferiority. Research indicates that such labels function as tools of reputational control, intended to undermine credibility and social standing rather than merely express disagreement.

The term aanis ("spinster"), for example, carries strong stigma in many Arab societies; it frames unmarried women as socially incomplete or burdensome. ISD's research on media discourse in Jordan also notes that accusations of sexual impropriety are disproportionately deployed against women to delegitimise their public presence, reflecting gendered double standards in moral judgment.

Dehumanising language that compares women to animals or vermin (including dogs, lizards, cows or cockroaches) and inanimate objects (such as brooms and shoes).

Dehumanisation intensifies hostility by symbolically stripping women of dignity and personhood. This makes abuse appear more acceptable.

ISD's research shows the use of animalistic metaphors silences women by making their online presence appear insignificant or irrelevant.

Other research efforts show that animalistic metaphors are commonly used to portray women as irrational, impure or controllable. This reinforces hierarchies that place men above women in social value.

Such language normalises domination and diminishes empathy toward the target. Scholars have linked these patterns to broader misogynistic narratives that treat women as property or sources of shame. This is particularly true in contexts where honour is tied to female behaviour.

Dismissive and silencing language such as claims that women do not belong in certain industries or public life, e.g. "women shouldn't have a say in war".

ISD's analysis of Jordanian users on X found patterns of gender-based intimidation meant to suppress women's participation in professional and civic discourse. Comments and posts included gendered insults that attacked women on the basis of their morality, honour or virtue. Others questioned women's intellectual capabilities, often justifying this through interpretations of religious texts which reinforce patriarchal narratives designed to limit women's roles in public life. These tactics contributed to a hostile environment that undermines gender equality, restricting women's engagement and their right to participate in societal and political spheres.

This discourse reflects exclusionary beliefs about gender roles. It asserts that women's participation in politics, security or public decision-making is illegitimate. UN research in Jordan on gender attitudes in Jordan indicates that such views remain prevalent, especially regarding leadership in traditionally male-dominated fields.

Online expressions of these beliefs often justify exclusion by invoking biological determinism, religious interpretations or appeals to tradition. Research shows that repeated exposure to such narratives contributes to symbolic marginalisation, reinforcing the perception that women's voices are less authoritative or relevant in matters of national importance. Consequently, dismissive language functions not only as individual prejudice but as a mechanism that reproduces structural gender inequality in digital public spheres.

Religious hate speech

ISD's research into religion-related hate speech on Jordanian social media focused primarily on X and YouTube as we were able to perform large-scale monitoring of publicly accessible discourse on them. The analysis found that religious hate speech constituted a relatively limited proportion of overall online content but was often characterised by a high degree of intensity and hostility.

The findings further show that a substantial share of the material collected consisted of derogatory and dehumanising language directed at Shia and other Islamic sects as well as religious minorities (accounting for 35 percent of all collected comments on X). Approximately 17 percent of the collected posts and comments across the dataset were classified as hate speech.

Sixty-three percent of posts and comments on X from the dataset contained harmful content directed at the Shia community. Other religious minorities were also targeted: the Druze community accounted for 12 percent of harmful content, while Alawites, Baha'is and Sufis collectively comprised seven percent, and Christians five percent. These figures indicate a disproportionate concentration of sectarian hostility toward Shia users relative to other minority groups. This suggests that anti-Shia sentiment remains the most salient form of online religious intolerance.

A subset of the identified content included explicit hate speech. These were characterised by direct threats and incitement to violence such as terms including "kill", "slaughter", "burn" and "exterminate". Such rhetoric framed non-Sunni religious communities as deviant, heretical or politically disloyal to Jordan's Sunni majority, imagining sectarian identity as fundamentally incompatible with national belonging. These narratives thus reinforced exclusionary conceptions of citizenship rooted in religious homogeneity.

Most harmful posts and comments used demeaning, dehumanising and overtly insulting language rather than explicit calls to violence. This included derogatory descriptors such as "impure" and "filthy" and slurs likening minority groups to "dogs", "pigs" or other animals.

Users also frequently challenged the legitimacy of minority religious beliefs and practices, portraying them as aberrant or illegitimate. These patterns reflect an environment where sectarian prejudice is normalised and minority faith identities are systematically marginalised.

Such discourse often intensifies during periods of regional conflict and political crisis, reflecting the transnational nature of sectarian narratives. Escalations in neighbouring countries (particularly conflicts framed along Sunni-Shia lines) tend to trigger surges in hostile online rhetoric as users project regional tensions onto local minority communities. In these moments, digital platforms can serve as arenas for symbolic conflict where external political developments are refracted through domestic identity politics. Consequently, online sectarian hostility should be understood not only as interpersonal prejudice but also as a manifestation of wider regional dynamics that shape perceptions of loyalty, belonging and threat.

Exclusionary and nationalistic speech

ISD analysts found that approximately 10 percent of all posts and comments collected on X relating to Jordanian nationalism contained narratives that were demeaning, derogatory or exclusionary in nature. These exchanges frequently involved nationalist insults between Jordanians (including Jordanians of Palestinian descent) and Palestinians. ISD also identified hostile rhetoric between Jordanians from different tribal, regional and socio-economic backgrounds. While not always rising to the level of explicit hate speech, such content reflected deeply entrenched identity tensions embedded within the public discourse.

These derogatory and exclusionary narratives often generated retaliatory responses, creating cyclical patterns of hostility that intensified during periods of political uncertainty or economic strain. In this sense, online nationalist discourse appeared reactive and context-dependent, evolving in tone and frequency alongside broader developments in Jordan's domestic and regional environment. Economic grievances, unemployment and debates over political representation are often catalysts for identity-based blame and mutual recrimination.

ISD researchers noticed that this pattern of exclusion, othering and discrimination was used to delegitimise, shame, belittle and silence individuals, reinforcing social boundaries and discouraging participation in public discourse. In one report focusing on hate speech and harassment towards influencers and public figures in Jordan, ISD found that if an influencer's post was perceived as shameful or controversial, some users questioned their identity. This included asserting that the individual was "baljik" going as far as stating "they would give up their own Jordanian citizenship in protest of this individual".

The narratives themselves commonly invoked themes of exclusion, disgust and contested authenticity. For example, some Palestinian users referred to Jordanians as "uncivilized nomads", drawing on classed and urban–rural stereotypes. Conversely, some Jordanians characterised Palestinians as 'inauthentic' calling them "Baljak", a derogatory term used to question their national loyalty or implying opportunistic affiliation.

In other instances, users portrayed certain tribes or family lineages as inherently more intelligent, loyal or prestigious than others. These claims often relied on essentialist assumptions about lineage and heritage, reinforcing hierarchical social distinctions rooted in origin and ancestry rather than individual merit.

In some cases, nationalist rhetoric adopted explicitly racialised undertones. One example is the use of the term ghorani in reference to Jordanians from the Jordan Valley (al-Ghor). Although the term is geographically descriptive in origin, it was at times used in a derogatory manner to denote darker-skinned Jordanians, carrying implicit assumptions about class, race and social status. In these contexts, the label functioned as a racialised insult, reinforcing exclusionary attitudes and perpetuating colour-based discrimination within nationalist discourse.

These discursive patterns reinforced rigid in-group/out-group distinctions and perpetuated symbolic hierarchies tied to origin, heritage and perceived political entitlement. Nationalist narratives were not limited to overtly political debates but also surfaced in cultural and social spaces. Football, one of the most popular sports in Jordan, emerged as a key arena for the expression of identity-based tensions. Rivalries between Al-Wehdat SC and Al-Faisaly have long reflected broader social and demographic divisions particularly those linked to questions of origin and belonging. Online discussions surrounding matches frequently included chants, slogans and coded phrases that reinforced nationalist or tribal antagonisms.

For example, supporters have circulated slogans such as "Jordan is for Jordanians" or chants emphasizing tribal pride and lineage, implicitly excluding other identity groups. These chants often originate on the streets in Jordan before migrating to social media platforms, where they become vehicles for broader political and social grievances.

Collectively, these findings suggest that online nationalism in Jordan is not solely a matter of patriotic expression but frequently intersects with unresolved questions of belonging, representation and social hierarchy. Although much of the discourse remains rhetorical rather than overtly violent, its cumulative effect contributes to an exclusionary communication environment in which identity is politicised and social cohesion is periodically strained.

Hate speech as part of a wider digital harm ecosystem

As mentioned earlier, hate speech should be understood as part of a broader ecosystem of digital harms rather than as an isolated phenomenon. In online environments, it often intersects with other harmful activities, including the spread of disinformation, misinformation, coordinated harassment campaigns. Research on digital harms has shown that social media platforms facilitate the coexistence of multiple forms of harmful content, noting that online environments generate numerous challenges such as disinformation, online hate, cyberbullying, discrimination, biases and other facets of harm, which frequently overlap within the same digital space.

These different forms of harmful activity can reinforce one another. Disinformation can fuel hostility toward specific groups, while harassment campaigns and coordinated amplification can increase the visibility and impact of hateful narratives. Together, these dynamics contribute to the spread of divisive discourse, target individuals or communities, and erode trust within the broader information environment. As a result, hate speech both shapes and is shaped by wider patterns of digital manipulation and abuse, which increases the risk of social polarisation and can contribute to offline tensions or harm.

The interaction between disinformation, misinformation and hate speech

Disinformation and hate speech frequently interact within digital information environments, reinforcing one another and amplifying the spread of harmful narratives. Disinformation can play a significant role in fuelling hostility toward specific groups.

A prominent example of disinformation in Jordan in Jordan was during several episodes of unrest in 2022, which included protests and labour disputes. Social media users shared old or unrelated videos and falsely claimed they depicted current events inside Jordan as evidence of violence or instability. Some clips were taken from other countries or earlier incidents. Authorities and media outlets later debunked these posts and warned the public against spreading unverified information.

More recently, following the outbreak of the Gaza war in October 2023, Jordan experienced a surge of misinformation on social media. Much of this content was related to protests, border tensions and government policies toward Israel. Some posts falsely claimed that there were escalating tensions inside Jordan or misrepresented protest activity. Fact-checking groups (such as Akeed) documented a sharp rise in such content, linking the increase directly to the regional conflict and its strong resonance among Jordanians.

When misleading narratives portray particular communities as threats to national identity, security or social stability, they can create conditions in which hate speech becomes more prevalent and socially acceptable. Research on the relationship between disinformation and online harms notes that these phenomena often coexist within the same digital ecosystems: misinformation, harassment, and hostile discourse can mutually reinforce one another.

Algorithmic amplification on social media platforms in the Middle East and their effect on Jordan

Prior research conducted across social media platforms suggests that divisive and emotionally charged content (including hate speech) tends to be disproportionately amplified by social media algorithms because it generates higher levels of engagement such as comments, shares and reactions. Platforms optimised for attention and interaction often prioritise content that provokes strong emotional responses, incentivising the circulation of polarising narratives. This dynamic has been documented globally, including in Middle East, where content that reinforces identity polarisation, religious antagonism or political grievances has been found to be algorithmically favoured.

Researchers studying online discourse in the region have noted that social media algorithms frequently elevate posts that frame political debates through narratives of cultural authenticity, national belonging or moral

transgression. Such dynamics can intensify existing societal tensions by rewarding content that frames social or political disagreements in antagonistic or exclusionary terms rather than encouraging deliberative discussion. Unfortunately, there is insufficient data on algorithmic amplification and echo chambers on Jordanian social media platforms.

A Jordan-specific example that ISD research found is that this dynamic can be observed in social media debates surrounding Jordanian–Palestinian identity politics, which periodically surface during political crises, protests or discussions about electoral representation. During these moments, viral posts on social media platforms such as X, Facebook and Instagram framed political grievances through a lens of national authenticity, loyalty and belonging. In some cases, users circulated content questioning whether Palestinians in Jordan are "truly Jordanian" while other posts accuse East Bank Jordanians of exclusionary nationalism. These narratives often rely on stereotypes or derogatory labels. They can gain significant visibility because they provoke strong emotional reactions and engagement online. Researchers studying digital discourse in Jordan note that such identity-based narratives can reinforce existing social cleavages and intensify polarisation when amplified through social media dynamics.

ISD's research also highlights this effect in the context of misogynistic and gendered hate speech. Even after content is reported or flagged by users, moderation systems may fail to remove it quickly or effectively reduce its reach. This can allow such narratives to persist within online networks. In these cases, algorithmic recommendation systems can continue promoting content that generates engagement. This may unintentionally amplify misogynistic discourse and facilitate its spread across multiple digital communities.

Interplay between regional events and harmful content on Jordanian social media platforms

Conflict and political crises in the region have intensifying effects on online hostility. Research on Arabic content reporting shows that major events (particularly those framed as sectarian or related to identity) fuel spikes in hateful narratives directed toward communities within Jordan's digital sphere and beyond.

ISD's research on religious hate speech shows how hate speech towards the Druze community intensified among Jordanian users of X after the events in Sweida, Syria in July 2025. This aligns with broader studies finding that violent political or extremist events correlate with increased volume and severity of hate speech online.

Conclusion and recommendations

Conclusion

This report has presented a comprehensive, evidence-based analysis of online hate speech in Jordan across a two-year period, drawing on large-scale quantitative monitoring of social media platforms and qualitative analysis of discourse patterns across X, Facebook, Instagram and YouTube. Its findings confirm that online hate speech in Jordan is a pervasive, multidimensional and structurally enabled phenomenon—one that demands coordinated responses from government, platforms, civil society and the research community. The three typologies documented across this series—gender-based abuse, religious hate speech and exclusionary nationalist discourse—are expressions of deeper social tensions that existed before social media, now amplified and accelerated by the platforms through which public life flows. Women in Jordan who enter public debate, whether as politicians, journalists, or influencers, do so knowing they are likely to face hate designed to silence them. Religious minorities face hostility that intensifies in direct response to events happening in the region. Questions of national identity and belonging generate cycles of mutual recrimination that create divisions rather than work through them.

These harms do not happen in a vacuum. They are shaped by companies that have designed and built platforms with systems optimised for engagement without sufficiently investing in the Arabic-language moderation infrastructure needed to limit the damage caused. They are shaped by a legal environment that broadly criminalises speech but often lacks clear definitions and consistent application, which can leave both victims and the wider public uncertain and vulnerable.

The recommendations that follow are grounded in two years of evidence. They are also grounded in a straightforward observation: the people most harmed by online hate speech in Jordan: women, religious minorities, those caught in identity-based conflicts, are not well served by the current combination of under-resourced platforms, imprecise law and fragmented institutional responses.

Recommendations

The recommendations set out below are grounded in the findings of this research and informed by the national action plan developed through the project's multi-stakeholder process. They are addressed to the range of actors best positioned to act on them.

For the Jordanian government and policymakers

Establish a legally precise definition of hate speech aligned with the context: Develop a legally precise definition and a consistent legal threshold against which harmful content aligned with the International Covenant on Civil and Political Rights Article 20 standard can be assessed, and that protects targeted individuals without criminalising legitimate free expression. [11]

Develop trigger event monitoring and a response protocol. This research documented clearly how regional crises produce sustained, measurable spikes in hate speech in Jordan's online spaces. A formal protocol with early-warning monitoring of regional developments would enable faster, coordinated responses—by the electronic crime unit or other similar entities—during periods of elevated risk.

Ensure gender-based online harm is addressed as a distinct legal priority. Findings show that women face threatening behaviour and derogatory speech on social media platforms that disproportionately target their professional public life, such as during elections.

[1] ISD's team, through a collaboration with various stakeholders in Jordan—including Civil Society Organisations (CSOs), law enforcement representatives, UN agency representatives, and local Ministries—worked on an inter-institutional secretariate to develop a national definition for online hate speech and an action plan to counter it in Jordanian digital spaces. Through this effort, the secretariate meetings developed the following definition for digital hate speech in Jordan: any expression, content, or behavior in the digital space—whether explicit or implicit—that promotes, incites, or justifies hatred, discrimination, or hostility against an individual or group based on identity, gender, religion, origin, language, geographic affiliation, opinion, or health status, and that leads to exclusion, incitement to violence, or harm to social peace.

Ensure gender-based online harm is addressed as a distinct legal priority. Findings show that women face threatening behaviour and derogatory speech on social media platforms that disproportionately target their professional public life, such as during elections. This points to the need for specific provisions addressing gender-based digital violence, including incitement, threats and discrimination based on protected characteristics (i.e. sex or gender), rather than relying solely on general cybercrime provisions.

Mandate Arabic-language moderation transparency: Platforms serving tens of millions of Arabic speakers publish no data on how moderation truly performs in this language.

Require algorithmic impact assessments for Arabic-language markets. Platforms optimised for engagement structurally reward divisive, identity-polarising content. This dynamic is well documented globally but unexamined for Arabic-language users specifically.

For social media platforms

Platforms should publish regular, disaggregated transparency reports on content moderation in the MENA region, including data on enforcement rates and appeal outcomes by language, in line with the accountability standards they apply in other markets.

Close the "Arabizi" detection gap. ISD researchers consistently observed that harmful content may be written in Arabizi specifically to evade detection. This is not a technical edge case, but a widespread and deliberate workaround that platforms should address.

For civil society organisations

Civil society organisations should continue to play a central role in monitoring online hate speech, documenting harmful narratives and providing evidence-based advocacy for policy reform. Building on the monitoring infrastructure developed during this project, organisations should sustain and expand their capacity to track harmful content in close to real time, with particular attention to periods of heightened risk associated with political events, elections or regional crises.

Invest in long-term digital safety infrastructure for women in public life: Dedicated helplines and rapid-response legal support for women facing online abuse, embedding digital safety training into journalist associations and parliamentary onboarding, and ensuring that civil society organisations working on this issue have the resources to operate year-round rather than within project cycles.

Design digital literacy programmes around how harmful content actually spreads: Programmes should move beyond individual fact-checking skills to build genuine understanding of the features of social media - how algorithms work, how echo chambers and filter bubbles form and reinforce harmful narratives, and how regional trigger events reliably produce spikes in online hostility.

For media actors and journalists

Media organisations should develop and embed clear editorial standards for the responsible reporting of hate speech and online abuse, including guidance on when and how to cover incidents without inadvertently amplifying harmful narratives or further harming those targeted.

Journalists—particularly those reporting on Jordanian digital spaces—should have access to ongoing training on digital safety, platform reporting tools and the legal protections available to them.

Investigative and public interest journalism plays an important role in holding platforms and institutions accountable for their responses to online hate speech. Media actors should be supported in this function through strengthened legal protections for press freedom and through collaboration with civil society organisations conducting research and monitoring amplification and echo chambers on Jordanian social media platforms.

Implemented by