FEBRUARY 2026

NONVIOLENT PEACEFORCE

Digital Harms and Their Impact on Social Cohesion: Examples from Ninewa, Iraq



CONTACT

WAFA ALI
POLICY AND ADVOCACY MANAGER, IRAQ
WAFAALI@NONVIOLENTPEACEFORCE.ORG

REAM DHAHER HAYDER
POLICY AND ADVOCACY OFFICER, IRAQ
RHAYDER@NONVIOLENTPEACEFORCE.ORG

Context

Digital spaces shape how communities access information, interpret events, and relate to one another. In fragile and conflict-affected environments, online narratives often circulate quickly and before they can be verified, influencing perceptions long before facts are confirmed. Research demonstrates that harmful digital content, including misinformation, hate speech, and dangerous speech, can escalate tension and contribute to violence when left unaddressed. As a result, the digital environment becomes not only a space for communication but a critical arena where trust, cohesion, and safety are constantly formed and reshaped.

In Iraq, particularly in IS-affected contexts such as Ninewa, communities continue to grapple with the long-term impacts of conflict, displacement, and unresolved grievances amid fragile governance structures. In this environment, social media has become a primary source of information and a key space where fear, frustration, and collective memory are expressed. While digital platforms can facilitate communication, they have also enabled the rapid spread of misinformation, hate speech, and polarising narratives. These dynamics have reinforced harmful stereotypes, fuelled tensions between communities, and weakened trust in authorities, further weakening social cohesion in areas still recovering from the legacy of the so-called Islamic State (IS) violence.

Challenges

The examples below highlight how misinformation and harmful narratives circulating online can undermine social cohesion and strain relationships in areas where Nonviolent Peaceforce (NP) engages with conflict-affected communities in Ninewa.

1. Misinformation and Hate Speech Undermining Social Cohesion

In Ninewa, misinformation, hate speech, and polarising digital narratives continue to shape perceptions and behaviour in ways that place strain on already fragile social cohesion. In such contexts, online content often circulates faster than verified information, amplifying fear and uncertainty across communities. These dynamics affect trust between communities and authorities and influence decisions related to safety, movement, and return.

Community-level examples collected through Community Peace Teams (CPTs) in Ninewa, including Tel Afar, Baaj, and Sinjar, stress on how misinformation and false online narratives have immediate and tangible consequences for community safety, trust, and social relations.

Several CPT members highlighted how rumours during periods of violence continue to shape collective memory and fear. One participant from Tel Afar recalled how, during IS control, unverified reports claiming mass killings and enslavement spread rapidly, prompting family separation and long-lasting distress. In one case, false information about a woman being killed circulated online, leaving her family in prolonged fear until contact was restored months later.

Together, these patterns demonstrate how digital narratives in Ninewa can both reflect and exacerbate existing tensions. By reactivating past trauma, shaping perceptions before facts are established, and influencing everyday interactions, misinformation and hate speech complicate efforts to rebuild trust and strengthen social cohesion in communities still recovering from conflict.

2. Exploitation of grievances through divisive narratives

Actors promoting violent divisive and extremist narratives continue to exploit digital platforms to manipulate frustration, trauma, and unresolved grievances, particularly among youth. Across Ninewa, online content promoting extremist and violent narratives spread fear and deepen existing divisions within already fragile communities.

In Sinjar, one high-follower Facebook page amplified IS-supportive messages and alarming claims about insecurity, triggering anxiety among community members and reinforcing mistrust between communities still coping with past violence. The event explains how quickly online extremist messaging can destabilise communities, erode resilience, and strain social cohesion.

3. Limited presence of positive digital narratives

Across many districts in Ninewa, online spaces are dominated by fear-based or divisive narratives, with very little content reflecting positive or peaceful community relationships. In areas recovering from IS, including Sinjar, Baaj and Tel Afar, the shortage of positive digital narratives leaves few counterweights to hate speech, misinformation or rumours. This gap particularly affects youth, whose primary information sources often amplify anger over empathy. NP staff and CPTs have observed that community Facebook pages tend to prioritise crisis-related and divisive content over constructive stories, allowing harmful narratives to travel quickly while messages that support social cohesion remain limited; research also backs these observations, illustrating human tendency to engage in negative news more than positive ones.

CPT members consistently noted that negative and fear-based narratives spread significantly faster than positive ones. Several participants emphasised that online spaces rarely highlight peaceful coexistence or recovery efforts, leaving communities affected by divisive messaging. In response to rumours circulating within diaspora networks about tensions in Tel Afar, one CPT member documented local peace initiatives, including art exhibitions, and shared these footage of it online. These posts were later amplified by diaspora contacts, helping to counter misinformation and present an alternative narrative focused on recovery and coexistence. CPT members stressed that youth-led digital engagement presents an opportunity to promote positive narratives and challenge misinformation before it escalates.

4. Limitations of platform moderation for local languages

Content written in local languages such as Kurmanji Kurdish or in dialect-specific Arabic is not always recognisable by the platforms. As a result, posts involving online extortion, hate speech or threats often remain online for long periods despite being reported. This invisibility in moderation systems leaves communities without effective digital protection.

When harmful content remains online, it exposes families and individuals to increased risks – particularly youth and women-, creates fear, stigma and reputational damage for individuals and families. It discourages community members from seeking help and fuels mistrust within and between groups. This dynamic can reinforce secrecy around sensitive incidents, increase community tensions and deepen existing social divides.

5. Gendered impacts of online and offline harassment

Several CPT members, particularly women, reported facing online and offline harassment linked to their civic engagement. Female CPT members described being mocked for their voluntary work or questioned for participating in public activities traditionally viewed as male spaces. Male CPT members similarly reported ridicule for unpaid community initiatives, reflecting broader societal norms that discourage civic action and participation. These dynamics silence voices that are vital for bridging divides.

Interrupting digital harms in Ninewa: Interventions and response options

Community-based responses to online violence

Across Ninewa, NP works closely with CPTs to understand how misinformation, hate speech and online extortion shape daily life and can translate into protection concerns. CPTs analyse how harmful online narratives influence movement, reinforce stereotypes, deepen mistrust and contribute to rising tensions.

NP's interventions to address this combine capacity strengthening, community engagement and coordination with relevant authorities. In Sinjar, CPTs monitored Facebook pages circulating harmful content and reported them through a national partner of NP which has direct coordination with META, Tech4Peace (T4P) NGO, resulting in the taking down of twenty-one pages that had been fuelling anger, fear and division. In another instance, CPTs identified a high-follower page spreading IS-supportive messages and reported it through verified mechanisms, leading to its permanent removal. These actions helped reduce fear, prevent further circulation of harmful content and limit attempts to target youth with manipulative messaging.

Moderation as violence interruption

NP also engages with local Facebook administrators who influence how information circulates within communities. After participating in NP sessions on rumour control and verification, several administrators in Baaj and Sinjar began sharing only confirmed updates, which reduced misinformation and helped prevent tension during sensitive periods. More broadly, community-led moderation works best when it is supported by an enabling environment: clear coordination arrangements for responding to sudden rumours surges, trusted channels that can issue timely, verified information, confidential and accessible referral options for individuals facing harassment, and sustained digital and media literacy initiatives. In parallel, rights-based legal and policy protection, designed with safeguards to prevent misuse, can strengthen public trust and make reporting and response mechanisms more credible.

Supporting the safety of women and girls online

Supporting safe reporting pathways for women and girls affected by digital harm is another core area of NP's work. In South Mosul, a young woman facing online blackmail was accompanied by the Women Peace Team (WPT) to relevant authorities, enabling her to seek support without fear or stigma. In Sinjar, NP supported the Community Police to promote a hotline for reporting electronic blackmail, which quickly became a trusted tool. In Baaj, NP facilitated a meeting between twenty-seven women and police and national security actors to discuss online violence, resulting in the establishment of a dedicated reporting hotline in a context where women seldom engage directly with security institutions.

NP also brings communities, authorities and service providers together to address misinformation, clarify rumours and strengthen communication. These dialogues provide CPTs with safe opportunities to raise digital-related concerns and help duty bearers respond more effectively during sensitive periods.

At national and international levels, NP elevates community perspectives so that protection and governance reflect the lived realities of conflict-affected communities. In practice, this include documenting and synthesizing community feedback on digital harms (harassment, rumor dynamics, misinformation). Convening dialogue between community actors and relevant authorities and service providers. And supporting locally led awareness and referral pathways so people know where to seek support when facing online threats. These efforts help communities to better recognize and manage digital-related risks, strengthens local protection mechanisms, and support safer, more connected relationships in Ninewa's post-IS context.

Recommendations

1. Responding to Misinformation and Hate Speech at Community Level

- Integrate digital narrative and trend observations into existing Early Warning and Early Response mechanisms to identify emerging risks early and enable timely, non-violent, community-led protective actions.
- Encourage trauma-sensitive and inclusive communication practices among local stakeholders to avoid unintentionally reinforcing fear, marginalisation, or division.
- Recentre civilians and communities to inform policies on addressing digital risks,
- Inclusive adoption of multi-stakeholders protocols that improve access to timely verified information, protect safe civic dialogue, and strengthen trusted coordination during periods of heightened tension.

2. Reducing the influence of harmful divisive narratives online

- Strengthen digital literacy for youth, women and community members to help them identify manipulative content and engage more safely online.
- Support community reporting pathways with NGOs and CPTs to flag harmful pages and share concerns through safe, trusted channels that protect survivors' rights.
- Adopt targeted protection measures for women targeted by online violence.
- Formalise survivor-centred protocols (consent, confidentiality, do-no-harm) through collaborations between civil society and community police and judicial actors.
- Provide guidance for safe evidence capture and storage.

3. Promote positive, community-led digital narratives

- Support CPTs and community leaders to create and share stories that highlight shared experiences and constructive community interaction.
- Amplify positive local experiences and initiatives to counter divisive or fear-based content.
- Enhance civil society capacity in digital storytelling and conflict-sensitive communication to help generate safe, constructive narratives.
- Encourage administrators of local digital platforms to integrate more solution-focused and community-strength content alongside crisis updates.

4. Engage with tech platforms to improve local-language recognition

- Social media platforms (Meta, TikTok and others) should invest in moderation that recognises Kurmanji Kurdish and Iraqi Arabic dialects, including: dialect glossaries built with local linguists to enhance platform capacity to recognise harmful content expressed in dialects, slang or coded language.
- Regulators and civil society can facilitate data-sharing and oversight that is rights-respecting, to prevent misuse while improving protection.