1500 K St .NW
Washington, DC 20005
Tel : + 1 203 745 4247 | www.iranhrdc.org

**Public Comment Submission**

**Oversight Board Cases 2026-041-IG-UA and 2026-042-IG-UA**

**Submitted by Iran Human Rights Documentation Center**

**July 14, 2026**

## I. Introduction and Standing

The Iran Human Rights Documentation Center (IHRDC) is a non-profit organization dedicated to documenting human rights abuses in Iran. Founded in 2005 by Iranian lawyers and scholars, IHRDC has published dozens of in-depth reports on human rights violations in Iran. It holds ECOSOC consultative status with the UN Economic and Social Council, advocates for human rights in Iran before the Human Rights Council, and engages systematically with UN mechanisms to advance accountability and protection for victims of abuse.

Iranian state-sponsored influence operations should not be understood merely as efforts to spread inaccurate information. They form part of a broader architecture of domestic and transnational repression combining propaganda, coordinated inauthentic behavior, surveillance, intimidation, internet filtering, communications shutdowns, arbitrary detention, and physical violence. The cases before the Board are not ordinary content moderation disputes. They raise structural questions about how Meta's platforms are weaponized as instruments of that repression — questions the IHRDC has documented across the 2019 Aban uprising, the 2022 Mahsa Amini protests, and the protest cycle beginning in late December 2025.

## II. How Platforms Should Detect and Address State-Sponsored Influence Operations

Meta's content-centric framework — evaluating posts individually against community standards — is structurally inadequate against covert state propaganda. Both posts at issue were cleared despite accounts bearing username coding associated with a known pro-Iranian hacking group, despite account creation less than two months before posting, and despite content glorifying security forces during mass civilian killings. Each signal, viewed in isolation, may be insufficient to warrant removal; assessed together, they constitute a pattern that should have triggered heightened review.

Platforms must therefore move from content-based to activity-based detection, assessing behavioral and network-level indicators in combination rather than in isolation: coordinated account creation and activation; shared technical infrastructure, including IP pools and device fingerprints; unnatural synchronization of posting, commenting, or mass reporting; false presentation as independent journalists or civil society actors; sudden amplification bursts following content deployment; and username conventions associated with known hacking or threat

1500 K St .NW
Washington, DC 20005
Tel : + 1 203 745 4247 | www.iranhrdc.org

groups. Meta's own enforcement record — including its removal of 513 Iran-linked pages, groups, and accounts in March 2019 — confirms both the scale of these operations and the necessity of this network-based approach.

The IHRDC's corresponding recommendations — covering cross-signal behavioral clustering, crisis-context escalation, cross-platform intelligence sharing, government-linked actor designation, emergency shutdown protocols, and public transparency reporting — are set out in Section VI below.

## III. How Iranian State-Sponsored Operations Typically Function

Iran's influence operations are institutionally organized, drawing on the Islamic Revolutionary Guard Corps' cyber command, the Ministry of Intelligence and Security, and contracted civilian networks known as the "Cyber Army." Their persistence across successive Meta removal actions confirms that these are permanent institutional functions rather than isolated incidents, deployed through a consistent two-layer architecture.

The first layer cultivates fictional personas masquerading as independent journalists, women's rights advocates, or ordinary citizens. Operators populate these accounts with AI-generated photographs and elaborate backstories, engaging normally with target communities for months before introducing political messaging. Iranian networks have impersonated local residents across numerous countries and constructed fictitious news outlets with custom branding to sanitize state narratives as independent reporting.

The second layer amplifies. Once a narrative is launched — such as content normalizing security forces — a coordinated set of automated or state-directed accounts immediately likes, shares, and comments to manipulate platform algorithms into prioritizing that content in user feeds, while simultaneously suppressing opposing voices through mass reporting, impersonation, and coordinated complaints. Influence operations also insert unrelated content into protest-related hashtags to degrade their utility for organizing and documentation.

Accounts within these networks routinely seek private contact with activists or relatives of dissidents to collect personal information, deliver malicious links, or determine physical location. For dissidents abroad, this activity forms part of a broader pattern of transnational repression — digital surveillance that precedes threats to relatives inside Iran, reputational attack, and physical harm.

This pattern is directly visible in the cases before the Board. The motorcycle convoy post, published at the height of the crackdown and tagged with #Protest and #Unrest, was not incidental political expression. It was a timed operational deployment calibrated to maximize its deterrent effect on the protest community at the moment of greatest vulnerability.

1500 K St .NW
Washington, DC 20005
Tel : + 1 203 745 4247 | www.iranhrdc.org

## IV. Impact on Iranian Civilians During the Recent Protests

The events that began in late December 2025, culminating in the nights of January 8 and 9, 2026, have been described as the deadliest government crackdown in Iran's modern history. Credible reports indicate that security forces carried out widespread and unlawful lethal attacks against unarmed protesters, resulting in mass civilian casualties alongside other grave violations, including arbitrary arrests, enforced disappearances, torture, and ill-treatment. The precise number of those killed, injured, or detained remains contested and independent verification is ongoing; numerous reports nonetheless indicate that the death toll reached into the tens of thousands.

Efforts to establish the full scope of these violations have been significantly impeded by the Iranian authorities' sweeping internet shutdown and communications restrictions, which prevented civilians from documenting abuses, obstructed communication with journalists and human rights defenders, and substantially delayed independent verification of victims and incidents. In January 2026, the government imposed a near-total internet shutdown lasting approximately 88 days — reportedly the longest deliberate civilian blackout on record — imposed at the outset of the crackdown to conceal the truth of the killings of January 8 and 9.

This shutdown substantially magnified the impact of state-sponsored online propaganda by producing an extreme information asymmetry. While civilians faced violent suppression on the streets, state-linked channels amplified images of orderly security convoys and narratives attributing the violence to protestors acting as foreign agents. Coordinated pro-government accounts buried organic citizen journalism beneath regime-favorable content, while selective connectivity gave state-linked actors a structural communications advantage: government accounts continued publishing while independent witnesses could not respond. Meta's review framework contains no adequate mechanism for adjusting enforcement thresholds to reflect the severity of a surrounding information blackout. It should.

In this context, leaving unverified pro-regime content active functions as coercive propaganda, signaling to dissidents that the state retains pervasive control and inducing fear and self-censorship. Such content may also operate as a surveillance instrument: user interactions generate behavioral data that can be used to identify individuals for targeting. The IHRDC has documented dozens of cases in which Instagram activity, including private messages obtained through device seizure, was cited in interrogations and used to justify detention.

Women dissidents are targeted through sexualized defamation, and threats that exploit social stigma and heighten the risk of offline harm. Ethnic, religious, and LGBTQ activists are portrayed as separatists or enemies of religion — labels that carry severe consequences in a legal system where broadly framed national security offenses may result in the death penalty.

1500 K St .NW
Washington, DC 20005
Tel : + 1 203 745 4247 | www.iranhrdc.org

## V. Coded Threats, Incitement, and Social Media Tactics

Rather than relying on explicit language that would trigger automated moderation, state-linked operations employ implicit, culturally specific coercion. Terms such as "rioter," "foreign agent," or "enemy of God" signal detention or lethal punishment within the Iranian context. During January 2026, senior officials publicly threatened charges of *moharebeh* — waging war against God, an offense that may carry the death penalty — even as security forces were conducting mass killings. Captions such as "Just a bit of patience, Iran," paired with military imagery, or footage of security convoys set to music pledging total sacrifice to the regime, function as coded threats of imminent state violence while simultaneously framing brutality as heroic national defense. Meta should not require the explicit phrase "kill this person" before acting; threat assessment must account for the speaker's apparent affiliation, the documented history of state violence, and whether coordinated accounts are converging on the same target.

Indirect threats within this ecosystem include publishing a dissident's home address or family photographs, stating that a person will "meet the fate of traitors," naming relatives who remain in Iran, or combining legal terminology with personal information to imply execution. These tactics extend to diaspora dissidents through coordinated harassment and state-linked cyber operations designed to silence criticism voiced on social media platforms.

Incitement frequently emerges from an ecosystem of coordinated posts rather than any single statement: one account brands an individual a foreign agent, a second publishes identifying information, a third demands punishment, and state media then amplifies the allegation. Each element may individually evade a narrowly drawn content rule, yet the campaign as a whole creates a foreseeable risk of arrest or violence. Meta must permit escalation grounded in cumulative and networked harm, enabling reviewers to weigh related posts, account affiliations, reporting patterns, and offline consequences together.

## VI. Policy Recommendations

1. Adopt an activity-centric CIB detection framework requiring cross-signal behavioral clustering — account age, technical fingerprints, username patterns, amplification networks, and content timing assessed jointly — with mandatory human review triggered by multiple converging signals.

2. Establish a Crisis Context Protocol escalating flagged content to regional expert teams during active state-level crises, triggered by credible reporting from recognized human rights organizations and UN bodies.

3. Implement cross-platform threat intelligence sharing to address influence operations that migrate across ecosystems, with appropriate protections for vulnerable sources.

4. Expand Government-Linked Actor designation, applying elevated scrutiny as a baseline standard regardless of whether individual participants are shown to be direct government employees.

5. Implement emergency shutdown protocols that expand Persian-language review capacity during blackouts and prevent automated enforcement from removing evidence of abuses.

6. Invest in Persian-language and cultural competency review capacity, since coded threats legible to Persian-speaking audiences will not be identified by systems that lack equivalent expertise.

7. Notify targeted users with specific security warnings where Meta has reason to believe they have been targeted by a state-linked network, since generic notices are insufficient when threats may extend offline.

8. Establish a Mass Atrocity Evidence Preservation Protocol ensuring that content and metadata are preserved for use by international investigative bodies, and address the information blackout asymmetry by lowering enforcement thresholds where state-imposed suppression renders influence operation content disproportionately impactful.

## VII. Conclusion

The Iranian government's information strategy combines the suppression of authentic speech with the artificial manufacture of apparent public support. Internet shutdowns silence civilians while coordinated influence operations fill the resulting vacuum with state narratives, and coded threats raise the personal cost of dissent. The 88-day shutdown transformed Instagram into one of the few channels through which state-produced content could reach an isolated population. In that environment, influence operation content is not merely propaganda. It is an instrument of repression, and Meta's platform is its delivery mechanism.

Meta must treat Iran-linked influence operations not merely as authenticity problems but as potential instruments of human rights abuse, requiring activity-centric detection, Persian-language expertise, protection for targeted users, evidence preservation, and policies designed specifically for periods of mass repression. The IHRDC urges the Board to require this approach. The stakes are measured in the lives, liberty, and safety of civilians who depend on the integrity of the information environment to organize, to document, and to survive.