

Threat Assessment and Mitigation Report: State-Sponsored Hate Speech, Digital Surveillance, and OSINT De-escalation Frameworks in Iran

The Structural Framework of Iran's Digital Repression and Network Architecture

The landscape of information control in the Islamic Republic of Iran has undergone a profound transformation, shifting from localized, reactive kinetic interventions to a continuous, multi-tiered digital repression infrastructure. Historically, the state's efforts during periods of civil unrest, such as the Green Movement of 2009, relied on sporadic filtering and visible security deployments. However, as demonstrated during the nationwide protests that erupted on December 28, 2025, and the subsequent external military escalations of February 28, 2026, the state has institutionalized centralized command structures designed to maintain absolute control over the domestic and international information ecosystems.

This system is overseen by the Supreme Council for Cyberspace, a central policy-making body established in 2012 under the direct oversight of the Supreme Leader. The Council's decisions are operationalized by the National Cyberspace Center, which coordinates enforcement across various state organs, including the Cyber Police Force (FATA). Established in 2011, FATA actively monitors the digital sphere with a mandate to suppress political dissent and prosecute activities deemed contrary to state-defined religious and moral values.

The technical foundation of this system is the National Information Network (NIN), a state-managed domestic internet ecosystem initiated around 2010. The NIN's decentralized architecture allows the state to implement graduated network responses. During periods of heightened domestic tension or external threat, authorities can execute blanket shutdowns of the global internet while keeping domestic services—including banking, state databases, and approved domestic messaging applications—fully online.

This dual capability was demonstrated on January 8, 2026, when the state imposed a nationwide shutdown of internet, phone, and mobile networks amid escalating protests. This communication blackout coincided with the killing of at least 34 people and the arrest of over 2,200 individuals, including students, journalists, and human rights defenders. By cutting off global connectivity, the state prevented the documentation and real-time reporting of human rights violations.

The ultimate operational test of this infrastructure occurred on February 28, 2026, during military strikes launched by the United States and Israel. Authorities reduced international bandwidth connectivity to a mere 1 percent of ordinary levels, effectively isolating 92 million citizens from external news sources.

Incident Date	National Security Context	Network Security Posture	Socio-Economic and Humanitarian Impact	Violations of International Legal Instruments
January 8, 2026	Escalation of protests initiated on December 28, 2025	Complete nationwide shutdown of internet, mobile, and phone networks	34 fatalities; over 2,200 arbitrary arrests; complete interruption of emergency healthcare and legal services	Violation of ICCPR Article 19 (Freedom of Expression) and Article 21 (Peaceful Assembly)
<i>February 28, 2026</i>	US and Israeli military strikes against domestic targets	Global connectivity throttled to 1% of baseline capacity via the National Information Network (NIN)	80% to 90% revenue losses in domestic e-commerce; 15% drop in the Tehran Stock Exchange; total isolation of the civilian population	Breach of proportionality and necessity tests required under international human rights frameworks

During these nationwide blackouts, the state employs a selective connectivity model, granting filtered internet access to approved state actors through special credentials, colloquially referred to as "White SIM cards". This selective system allows state-aligned figures to promote official narratives internationally while independent and dissenting voices remain completely offline. For example, during the February 2026 wartime blackout, Foreign Minister Abbas Araghchi conducted a live, unfiltered Zoom interview with CBS, illustrating how the state utilizes selective connectivity to manage international perceptions while isolating its civilian population. The international watch dog group Article 19, a UK-based human rights organization founded in 1986, has consistently documented these censorship practices. It advocates for the protection of fundamental rights under the Universal Declaration of Human Rights and the International Covenant on Civil and Political Rights (ICCPR). Through its regional monitoring, Article 19 has highlighted how repressive legal frameworks, national security laws, and state-imposed blackouts in Iran, Saudi Arabia, and Egypt are systematically utilized to silence civil society and target vulnerable communities, including women and LGBTQI individuals.

Domestic Platform Migration as a Vector for Surveillance, Moral Policing, and Narrative Monopolization

To reinforce its network boundaries, the state has pursued a policy of forced migration, driving citizens off secure global communication platforms and onto state-controlled domestic alternatives. By maintaining blocks on international services like Instagram, the state has left citizens with no practical choice but to use domestic platforms such as Eitaa, Rubika, Bale, and Shad for daily communication, business, and education. This forced migration integrates state surveillance and censorship directly into the daily lives of the population.

Domestic Platform	Governing Body / Corporate Developer	Core Functionality	Primary Surveillance and Censorship Vectors
Eitaa Messenger	Idea of Today's Civilization Supporters Company	Instant messaging, VoIP, and social networking with over 40 million users	Integrated into the Message Exchange Bus (MXB); lacks encryption, exposes device metadata, and was removed from global app stores
Rubika	Tusca Holding, Mobile Telecommunication Company of Iran (MCI), MTN Irancell	Video-on-demand, social media, and digital wallet services with over 50 million users	Monitored by the MCI Digital Platform Monitoring Center; utilizes intrusive device permissions, conducts gallery scans, and enforces strict moral policing
Bale	State-Backed Financial Institutions	Instant messaging and integrated mobile banking services	Monitored by state banking compliance systems; imposes high advertising costs and strict moral approvals
Shad	Ministry of Education / State IT Developers	National online educational portal	Restricts student interaction; operates a unidirectional audio-visual channel where only instructors can speak during class hours
Zarebin	Ministry of Information and Communications Technology	National search engine portal	Algorithmic deletion of queries regarding politically sensitive topics or items deemed morally unapproved

The structural unification of this domestic app ecosystem is achieved through the Message Exchange Bus (MXB). The MXB is an interoperability protocol designed to connect Eitaa, Bale, Soroush, Rubika, Gap, and iGap, allowing users to exchange files, messages, and calls across platforms without separate accounts. While presented as a technological convenience, the MXB functions as a data centralization tool, consolidating communication data and metadata into unified monitoring streams.

A technical assessment conducted by the security lab of the Open Technology Fund (OTF) revealed that Eitaa, Rubika, and Bale lack basic security protocols and compromise user privacy. These applications require extensive device permissions, and users have reported that platforms like Rubika can access and scan local device galleries. Due to these security vulnerabilities and surveillance links, these apps were removed from Google Play and the Apple App Store, leaving users dependent on local repositories like Bazaar and Myket.

Censorship and behavioral modification are actively enforced within these domestic platforms. Small businesses that rely on Rubika or Bale for marketing face high advertising fees and strict moral policing. For example, Rubika charges business owners up to 63 million tomans (approximately \$359) for a 15-minute promotional slot. However, these promotions must undergo a lengthy, highly conservative approval process. Advertisements are frequently rejected for minor perceived infractions, such as a female business owner's elbow being visible for a few seconds or the promotion of female undergarments.

This moral policing is mirrored in the state-sanctioned search engine Zarebin, which filters search queries to omit commercial listings for female undergarments while returning standard results for male equivalents. In the education sector, the Shad platform is structurally configured to limit student interaction. The application restricts communication so that only the teacher can speak during designated hours, blocking student feedback and questions until late afternoon. Furthermore, these domestic networks serve as channels for political censorship and state propaganda. Accounts that use historical national symbols, such as the pre-1979 Lion and Sun emblem, or profiles featuring portraits of Mohammad Reza Pahlavi, are systematically banned. Concurrently, these platforms are used to circulate coordinated disinformation. During the 2026 military crisis, state-linked channels used Rubika and Bale to spread false reports, including claims of an absolute military victory over the United States and Israel, and the false death of Israeli Prime Minister Benjamin Netanyahu. This controlled information environment aims to project an image of the state as a land of "peace, friendship, and human rights" while suppressing the reality of the domestic crisis.

State-Sponsored Hate Speech and Systematic Persecution of the Baha'i Religious Minority

A critical element of domestic narrative control is the systematic, state-sponsored dehumanization of marginalized groups, most notably the Baha'i community, which represents Iran's largest non-Muslim religious minority. Unrecognized by the constitution of the Islamic Republic, Baha'is are denied fundamental civil rights, access to higher education, and public employment. Following the national protests of January 2026 and the subsequent military escalations in February 2026, the state has intensified its targeting of the Baha'is, treating them as internal enemies of the state to divert public anger and justify authoritarian crackdowns. The current crackdown is characterized by a high degree of state coordination, involving arbitrary detentions, violent home raids, properties being seized, and incommunicado detentions that may amount to enforced disappearances. Between June and November 2025, the Baha'i International Community (BIC) documented over 750 distinct persecutory acts across Iran—a threefold increase compared to the same period in 2024. This wave has continued to escalate into mid-2026, with sweep operations targeting families in Mashhad, Shiraz, Borujerd, Aligudarz, and Semnan.

Baha'i business owners have had their workplaces sealed, and gravesites belonging to the community have been systematically desecrated. This systematic campaign is designed to induce terror, isolate the community, and enforce collective punishment under fabricated charges such as "propagating against Sharia Islam" or "forming groups to act against national security".

To validate these physical crackdowns, the state orchestrates pervasive online and offline hate speech. Since 2017, more than 33,000 pieces of anti-Baha'i content have been broadcast or published across state radio, television, newspapers, and social media platforms. Data collected

during key monitoring periods, such as between November 2020 and April 2021, demonstrated a 36 percent increase in state-sponsored hate content compared to previous quarters, indicating a calculated escalation in propaganda. The state actively utilizes manipulated and desecrated imagery to cultivate public hostility, such as the public distortion of "the Greatest Name," a highly sacred Baha'i symbol equivalent to the Holy Cross in Christianity or the word Allah in Islam. Another common technique involves superimposing false, occultist, or cult-like imagery over photographs of the Shrine of Bahá'u'lláh to frame the peaceful community as a dangerous cult. Because Baha'is are strictly denied the right of reply in any state-controlled or domestic media, these narratives go completely unchallenged within the domestic space, allowing the regime to systematically build a hostile environment that serves to justify ongoing judicial abuses, lengthy prison sentences, and asset confiscation. Baha'i women face a dual layer of vulnerability; they are disproportionately targeted for their participation in civil rights movements and are subjected to harsher judicial sentences due to the intersection of gender-based repression and religious discrimination.

Target Subject	Documented Regional Operations	Judicial Verdict / Alleged Offenses	State-Sponsored Disinformation and Media Vectors
Vafa Kashefi	Mashhad	Detained during January 2026 protests; re-arrested on April 21, 2026, following a court summons	Framed as an internal saboteur and instigator of political instability during civil protests
Sara Sepehri	Shiraz	Arrested on April 9, 2026, and held at an undisclosed location under incommunicado detention	Associated with "deviant cultist activities" and acting against national security
Semnan Mother and Humanitarian Activist	Semnan	Convicted by the Semnan Revolutionary Court of "forming groups to act against national security, educational activities, and propagating against Sharia Islam"	Accused of utilizing humanitarian care for Afghan refugee children as a front for anti-regime propaganda
Mehdi Hajati	Shiraz City Council	Arrested by security agents to serve a one-year prison sentence following an in-absentia trial	Charged with defending a "false faith" after publicly advocating for the civil rights of Baha'i citizens
Baha'i Families and Businesses	Borujerd and Aligudarz	Coordinated raids on February 2, 2026; sealing of five Baha'i homes and three workplaces	Categorized as economic saboteurs and financial agents of foreign intelligence services

The gendered dimension of this persecution is documented in submissions by Article 19 to the United Nations Special Rapporteur on Freedom of Expression. The research shows that women

who speak out against state abuses or report on women's rights face significant risks of gender-based violence, harassment, and intimidation.

This risk is particularly acute before and after protests; research in Latin America and the Middle East shows that women face persecution on their way home from demonstrations, when they are isolated on public transport and easily identified by opponents. Within newsrooms and digital spaces, women journalists face verbal attacks, comments about their clothing, and discrimination based on their sexuality. For Baha'i women, this structural risk is amplified, as the state has accused the Baha'i community of instigating the "Woman, Life, Freedom" protests, resulting in an escalation of targeted judicial and digital attacks.

Computational Natural Language Processing and Farsi Hate Speech Datasets

The automated identification of hate speech and incitement to violence in the Iranian digital ecosystem presents significant computational challenges. Farsi remains a low-resource language (LRL) in the domain of Natural Language Processing (NLP). The development of robust, automated models is historically hindered by a severe scarcity of high-quality, annotated public datasets, combined with the complex linguistic nuances of Farsi, such as code-mixing, colloquialisms, and highly metaphorical or uncivil rhetorical structures.

Dataset Name	Total Sample Volume	Annotation Structure and Depth	Key Attributes and Computational Use Cases
Pars-OFF	10,563 annotated data samples	Three-layered hierarchy based on SemEval-2019 OLID	Level A: Binary (Offensive/Non-Offensive) Level B: Target presence (Targeted/Untargeted) Level C: Target type (Individual/Group/Others)
PHATE	7,000+ manually-annotated Persian tweets	Two-level hierarchical multi-label structure	Top level: Normal or Hate Sub-labels: Violence, Hate, and Vulgar; includes the specific text spans serving as the label's rationale
Mozafari et al. Dataset	6,000 Farsi tweets	Hierarchical annotation based on SemEval-2019	Collected over a two-month window in 2020; sources include Iranian e-commerce and delivery platforms Digikala and Snappfood
POLID	4,988 distinct linguistic entities	Binary Classification	Divides data samples into Offensive ("OFF")

Dataset Name	Total Sample Volume	Annotation Structure and Depth	Key Attributes and Computational Use Cases
			and Not Offensive ("NOT") classes

These datasets highlight the challenge of class imbalance in Farsi hate speech detection. In the Pars-OFF dataset, for instance, only 10 percent of offensive tweets are labeled as untargeted at Level B. To address this imbalance, researchers deploy preprocessing pipelines that utilize oversampling techniques (such as 50 percent, 75 percent, and 100 percent random duplication of the minority class) to prevent model bias and ensure accurate classification.

The choice of machine learning architecture is critical for detection accuracy. Classical machine learning approaches (such as Logistic Regression, Random Forest, and Naive Bayes utilizing TF-IDF representations) struggle with semantic depth, irony, sarcasm, and cross-sentence contexts. To overcome these limitations, fine-tuned, encoder-only transformer models have been introduced. Models such as ParsBERT, which are pre-trained specifically on Farsi corpora, have demonstrated significant performance advantages.

Comparative evaluations show that ParsBERT significantly outperforms decoder-only models, including multiple ChatGPT model families, in identifying Farsi hate speech. This discrepancy arises because generic large language models (LLMs) struggle to parse the subtle, rhetorical forms of Farsi incivility and distinguish them from general profanity. In Farsi digital spaces, uncivil messages often contain no explicit vulgar terms, relying instead on metaphorical constructs to attack individuals or communities.

In addressing these challenges, research shows that changing the language of the prompt (such as prompting in English versus Farsi) has no remarkable effect on ChatGPT's analytical output, pointing to a fundamental limitation in the underlying training representations of general-purpose LLMs when dealing with Farsi. To improve detection, researchers classify Farsi incivility into three distinct categories: Pejorative Speech (PS), Insults, and Threatening Messages (TMs). This taxonomy allows fine-tuned models to distinguish between general profanity and targeted, harmful speech, providing a more robust framework for automated monitoring.

Open-Source Intelligence Frameworks and Actionable De-escalation Methodologies

The conversion of raw social media data and online chatter into actionable intelligence requires a structured Open Source Intelligence (OSINT) framework. An OSINT framework provides a repeatable, disciplined pipeline designed to process vast amounts of unstructured, public data while bypassing the challenges of information overload, fragmented data sources, and accuracy errors. This pipeline operates through the standard stages of the intelligence cycle:

[Preparation] → [Collection] → [Processing] → [Analysis] → [span_166] (start_span) [span_166] (end_span)

To distinguish hate speech from offensive language, OSINT frameworks must employ rigorous diagnostic models. A notable development in this domain is the **xList-Hate** framework. Rather than prompting a language model to make a binary, black-box decision on whether a text constitutes hate speech, xList-Hate decomposes the evaluation into a checklist of ten explicit,

concept-level binary questions grounded in legal and policy standards (such as targeting of protected groups, dehumanization, and incitement to violence).

These diagnostic signals are then aggregated by a lightweight, fully interpretable decision tree.

Let $H(x)$ represent the final classification of a text x as hate speech, determined by a mapping function f applied to the set of ten binary diagnostic factors $q_i \in \{0, 1\}$:

where the function f is represented by a transparent, auditable decision tree. This methodology separates semantic judgment from classification logic, making the output robust to distribution shifts and contextual ambiguity, and highly suitable for human rights monitoring.

At the international level, this diagnostic approach aligns with the **Rabat Plan of Action**, a comprehensive threshold test adopted by the United Nations to evaluate whether speech constitutes incitement to discrimination, hostility, or violence. The Rabat Plan of Action outlines six distinct factors that must be evaluated:

- **Social Context:** The prevailing political, social, and economic climate surrounding the speech.
- **Speaker Status:** The position, influence, and authority of the speaker relative to the target audience.
- **Intent:** The deliberate intent to advocate hatred or incite violence against a protected group.
- **Content and Form:** The style, tone, and specific rhetorical devices employed in the speech.
- **Reach and Extent:** The audience size, frequency of communication, and dissemination channels.
- **Imminence of Harm:** The direct causal likelihood that the speech will result in real-world violence or discrimination.

This framework is highly relevant to international human rights monitoring. Under Article 19(3) of the ICCPR, states may restrict freedom of expression to protect national security, public order, or the rights of others. However, such restrictions must be provided by law that is precise and public, and they must meet strict tests of necessity and proportionality.

The UN Special Rapporteur on Freedom of Expression has highlighted how counter-terrorism and national security laws are frequently misused by states to suppress legitimate dissent and target minority groups. The Special Rapporteur's findings indicate a global surge in hate speech and disinformation, with social media platforms like Facebook, Telegram, and particularly X serving as major vectors during polarized conflicts. While international criminal law prohibits direct and public incitement to genocide, distinguishing between offensive speech and actual incitement requires the rigorous contextual analysis outlined in the Rabat Plan.

The predictive link between digital speech and real-world violence has been empirically demonstrated by pioneering organizations such as **HateLab**. Merging criminology with machine learning, HateLab's methodology tracks hate speech in real-time by mapping "breadcrumb trails" (temporal, geographic, and network dispersion patterns).

Their empirical research has established a direct predictive correlation, demonstrating that localized spikes in online hate speech are reliably followed by physical hate crimes in the same geographic areas. Building on this foundation, the spin-out platform **Nisien.ai** has integrated these capabilities into its **Hero Platform**.

The platform features a multi-tiered defense architecture:

- **Hero Detect:** Uses AI models to map language patterns and identify rising societal tensions or emerging harmful narratives.
- **Hero Defend:** Allows administrators to protect digital environments by hiding or removing identified toxic content.

- **Hero Resolve:** Leverages generative AI to suggest de-escalation responses. Rather than relying on heavy-handed censorship, the platform suggests empathy-inducing counter-narratives grounded in behavioral science to persuade individuals to alter their behavior. This resolution capability has been shown to reduce negative online interactions by up to 65 percent in testing environments, preserving free speech while mitigating real-world harm.

Actionable OSINT Recommendations and Strategic De-escalation Initiatives

To counter state-sponsored hate speech and digital repression in the Iranian information ecosystem, civil society organizations, international human rights monitors, and technology developers should implement targeted, technical de-escalation strategies.

Decentralized OSINT Data Collection Pipelines

Because the Iranian state utilizes the National Information Network to execute global internet shutdowns, monitoring organizations must build decentralized data collection systems. These systems should deploy a distributed network of local proxies and virtual private networks to maintain connectivity during bandwidth throttling events.

OSINT collection pipelines must prioritize the secure ingestion of data from domestic applications such as Eitaa, Rubika, and Bale, which have become the primary channels for forced migration. To protect local collectors, these pipelines should employ containerized, virtualized environments that isolate application data, preventing domestic platforms from accessing local hardware identifiers or scanning device galleries. Collected data should be continuously streamed to secure, off-shore servers for real-time analysis.

Localized NLP Classification Models

To identify and analyze Farsi hate speech, monitoring organizations should deploy fine-tuned, localized encoder models rather than relying on general-purpose LLMs. Automated detection systems should utilize ParsBERT, pre-trained on Farsi corpora, to effectively identify the subtle, non-insulting rhetorical incivility common in Farsi digital spaces.

These models should be integrated with diagnostic frameworks like xList-Hate, decomposing hate speech evaluation into explicit, concept-level binary questions rather than relying on direct black-box classification. Preprocessing pipelines must incorporate oversampling techniques (such as 100 percent random duplication of minority classes) to address severe class imbalances in Farsi datasets, ensuring accurate classification across targeted groups.

AI-Driven Behavioral Counter-Speech

To mitigate online hate speech without relying on heavy-handed censorship, organizations should deploy automated, generative AI assistance tools that suggest empathy-inducing counter-narratives based on Nisien.ai's "Resolve" methodology. These de-escalation tools should be integrated into communication channels used by activists and human rights defenders, providing real-time, behaviorally validated responses to hostile online actors. The suggested counter-speech must be culturally and linguistically tailored to Farsi, directly

addressing the underlying anxieties exploited by state-sponsored propaganda. By utilizing empathy-inducing counter-narratives, monitoring groups can de-escalate digital conflicts and reduce negative online interactions, preserving free speech while mitigating real-world harm.

International Regulatory Pressure and Technical Accountability

International human rights organizations should leverage documented OSINT findings to pressure global digital intermediaries and international regulatory bodies. Organizations must document state-imposed internet blackouts as systematic violations of Articles 19 and 21 of the ICCPR, submitting comprehensive reports to the United Nations Special Rapporteur on Freedom of Expression to support targeted sanctions against the leadership of the Supreme Council for Cyberspace.

Concurrently, civil society should coordinate with global application stores to block the distribution of state-backed domestic apps, such as Eitaa and Rubika, that lack basic security protocols and actively compromise user privacy. By restricting state access to global distribution channels, international actors can increase the political cost of digital repression and protect the digital rights of the civilian population.

Works cited

1. From Coercion to Code: Iran's Digital Security Transformation ..., <https://www.meforum.org/meq/from-coercion-to-code-irans-digital-security-transformation> 2. Iran: Communication blackout puts lives at risk and violates human rights - ARTICLE 19, <https://www.article19.org/resources/iran-communication-blackout-puts-lives-at-risk-and-violates-human-rights/> 3. Internet shutdown pushes Iranians onto distrusted domestic apps ..., <https://www.iranintl.com/en/202605079104> 4. Article 19 (organization) | Law | Research Starters - EBSCO, <https://www.ebsco.com/research-starters/law/article-19-organization> 5. ARTICLE 19's submission to the report on Gender Justice by the UN Special Rapporteur on Freedom of Expression Introduction - ohchr, https://www.ohchr.org/sites/default/files/2021-11/ARTICLE19_3.pdf 6. Middle East and North Africa - ARTICLE 19, <https://www.article19.org/regional-office/middle-east-north-africa/> 7. Eitaa - Wikipedia, <https://en.wikipedia.org/wiki/Eitaa> 8. Rubika - Wikipedia, <https://en.wikipedia.org/wiki/Rubika> 9. VOA Persian: Iranian domestic messaging apps lack security, <https://www.voanews.com/a/voa-persian-iranian-domestic-messaging-apps-lack-security/7906713.html> 10. Iran's Baha'i Community Under Intensified State Attack Amid Escalating Domestic Repression, <https://iranhumanrights.org/2026/05/irans-bahai-community-under-intensified-state-attack-amid-escalating-domestic-repression/> 11. Iran: Weaponizing Justice System to Persecute Bahá'ís - Baha'i International Community, <https://www.bic.org/news/iran-weaponizing-justice-system-persecute-bahais> 12. 3.6.2. Baha'is | European Union Agency for Asylum, <https://www.euaa.europa.eu/country-guidance-iran-2025/362-bahais> 13. Iran - International Religious Freedom Reports: Custom Report Excerpts - United States Department of State, <https://www.state.gov/report/custom/ebc65bdb4a> 14. State-sponsored hate speech and disinformation against the Bahá'í ..., <https://humanrightscommission.house.gov/sites/evo-subsites/humanrightscommission.house.gov/files/documents/2021.06.30%20Hate%20Speech%20Explainer%20Document%20-%20Updat>

ed_1.pdf 15. Spanning the Spectrum of Hatred Detection: A Persian Multi-Label Hate Speech Dataset with Annotator Rationales - AAAI Publications, <https://ojs.aaai.org/index.php/AAAI/article/view/29743/31279> 16. Pars-OFF: A Benchmark for Offensive Language Detection on Farsi Social Media, <https://www.computer.org/csdl/journal/ta/2023/04/09936700/1105rlw1FzG> 17. Old wine in old glasses Comparing computational and qualitative methods in identifying incivility on Persian Twitter during the #MahsaAmini movement - arXiv, <https://arxiv.org/pdf/2602.08688> 18. One_by_zero@ NLU of Devanagari Script Languages 2025: Target Identification for Hate Speech Leveraging Transformer-based Approach - ACL Anthology, <https://aclanthology.org/2025.chipsal-1.38.pdf> 19. (PDF) OSINT-Based Detection of Hate Speech and Offensive Language on Social Media Using Machine Learning - ResearchGate, https://www.researchgate.net/publication/405290751_OSINT-Based_Detection_of_Hate_Speech_and_Offensive_Language_on_Social_Media_Using_Machine_Learning 20. OSINT Framework: What It Is, How It Works, and the Best Tools (2026 Guide) - Bitsight, <https://www.bitsight.com/learn/cti/osint-framework> 21. OSINT Framework Explained - Recorded Future, <https://www.recordedfuture.com/threat-intelligence-101/intelligence-sources-collection/osint-framework> 22. A Checklist-Based Framework for Interpretable and Generalizable Hate Speech Detection - arXiv, <https://arxiv.org/pdf/2602.05874> 23. Global threats to freedom of expression arising from the conflict in Gaza - Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression, Irene Khan (A/79/319) - Question of Palestine - the United Nations, <https://www.un.org/unispal/document/report-special-rapporteur-23aug24/> 24. Using artificial intelligence (AI) to tackle online hate – UKRI, <https://www.ukri.org/who-we-are/how-we-are-doing/research-outcomes-and-impact/esrc/using-artificial-intelligence-ai-to-tackle-online-hate/>