Sociale : Journal of Social and Political Sciences
Vol. 1, No. 1, Mei, 2025, p.94-107
e-ISSN: _____ (online)
https://jurnal.sinesia.id/index.php/sociale

Constructing Hashtags into Weapons: Iran and Israel in the Arena of Cyber Diplomacy

Ligar Yogaswara
E-mail Korespondensi: yogaswaraligar@gmail.com
National Cyber and Crypto Agency of Indonesia, Indonesia

Info Article
|Submitted: 15 July 2025**|Revised**: 21 July 2025**|Accepted**: 24 July 2025 **|Accepted:** 25 July 2025
How to cite: Ligar Yogaswara, "***Constructing Hashtags into Weapons: Iran and Israel in the Arena of Cyber Diplomacy***", *Sociale : Journal of Social and Political Sciences*, Vol. 1 No. 1, 2025, p. 94-107.

ABSTRACT

The June 2025 cyber clash between Iran and Israel represents a turning point in digital diplomacy, where social media became a battleground for competing national narratives. Using a constructivist lens, this article explores how identity, norms, and symbolism are expressed through hashtags like #IranUnderAttack and #DefendingIsrael. Drawing on discourse analysis and data from DFRLab and DataReportal, the study finds that states increasingly utilize AI tools and culturally embedded messaging to influence legitimacy in real time. Hashtag campaigns function as digital proxies for deeper struggles over sovereignty and law. The role of AI systems—such as Grok—in shaping visibility and public sentiment is critically examined, revealing the algorithmic dynamics behind narrative dominance. Diaspora communities are shown to amplify state narratives, enhancing reach and resonance. This case highlights the evolving nature of cyber diplomacy, where digital virality, identity politics, and platform algorithms now define how states project influence and contest norms in the global arena

Keyword: *cyber diplomacy; strategic narratives; Iran–Israel conflict; norm contestation;and algorithmic bias*

ABSTRAK

Konflik siber antara Iran dan Israel pada Juni 2025 menjadi titik balik penting dalam diplomasi digital, di mana media sosial dimanfaatkan sebagai medan perang untuk membangun narasi nasional. Dengan pendekatan konstruktivis, artikel ini menganalisis bagaimana identitas, norma, dan simbolisme diperebutkan melalui tagar seperti #IranUnderAttack dan #DefendingIsrael. Melalui analisis wacana dan data sekunder dari DFRLab dan DataReportal, studi ini menunjukkan bahwa negara-negara kini memanfaatkan alat berbasis AI serta pesan budaya yang resonan untuk merebut legitimasi secara real-time. Kampanye tagar berfungsi sebagai proksi digital untuk perebutan kedaulatan dan hukum internasional. Peran sistem AI seperti Grok turut dikaji dalam memoderasi visibilitas, sentimen publik, dan dominasi narasi. Komunitas diaspora juga berperan sebagai penguat pesan negara, memperluas jangkauan kampanye digital. Studi ini menegaskan bahwa diplomasi siber kini berlangsung di ruang hibrida antara viralitas, politik identitas, dan moderasi algoritmik dalam perebutan pengaruh global.

Kata Kunci: *diplomasi siber; narasi strategis; konflik Iran–Israel; kontestasi norma;dan bias algoritmik*

Introduction

In June 2025, a sharp escalation occurred between Israel and Iran, triggered by Israel's targeted airstrikes on more than 100 Iranian nuclear and missile infrastructure sites near Isfahan (Bhardwaj, 2025). Iran retaliated swiftly with ballistic missiles and drone strikes on Israeli cities including Tel Aviv and Haifa,

marking one of the most intense kinetic confrontations in decades (Bhardwaj, 2025). However, the conflict simultaneously unfolded in a digital arena: state-affiliated social media channels surged with hashtags, AI-generated visuals, and persuasive messaging designed to frame the narrative. This dual-front confrontation illustrates a strategic shift—where military operations and cyber diplomacy are deeply intertwined, operating in parallel to shape both public perception and policy formation.

While the rise of social media and other digital technologies has transformed traditional diplomatic engagement—a phenomenon known as digital diplomacy, which refers to the use of platforms like Twitter or Instagram for public outreach and image repositioning (Bjola & Holmes, 2015). this study adopts a more focused lens on cyber diplomacy. Unlike the broader concept of digital diplomacy, cyber diplomacy specifically addresses how states manage and contest power within cyberspace, by regulating issues such as cybersecurity, AI-mediated amplification, misinformation, and normative frameworks (Riordan, 2019). In the case of the June 2025 Iran–Israel confrontation, strategies like weaponizing hashtags, deploying AI-generated visuals, and engaging diaspora networks fall squarely within the domain of cyber diplomacy. Thus, this paper intentionally distinguishes digital diplomacy as a general tool-usage phenomenon from cyber diplomacy as a strategic domain-centric practice.

The scope and speed of the digital campaign were staggering. According to Atlantic Council's DFRLab, over 130,000 posts related to the conflict appeared on X within 48 hours, with approximately 37% traced to coordinated or automated networks linked to state actors (Ponce de León & Chenrose, 2025). Iranian accounts such as @IRIMFA_EN and PressTV deployed hashtags like #IranUnderAttack and #ZionistCrimes, framing the airstrikes as violations of international law. Israel's Digital Diplomacy Unit countered aggressively with #DefendingIsrael and #IranianTerrorism, emphasizing self-defense and counterterrorism narratives. This instantaneous and coordinated cascade indicates a deliberate propagation of strategically curated messages, rather than organic social media reactions (Ponce de León & Chenrose, 2025).

The conflict's digital dynamics were intensified by AI involvement. DFRLab found that Grok—the AI assistant integrated into X—was employed in over 100,000 user interactions during the initial three days, yet provided inconsistent and unreliable fact-checking responses (Ponce de León & Chenrose, 2025). Simultaneously, both nations circulated AI-generated or manipulated visuals—purporting mass civilian casualties and destruction—to elicit emotional response and international solidarity. This reveals how AI tools amplify narrative strategies

and complicate digital verification, thereby becoming essential instruments in modern cyber diplomacy.

Both Israel and Iran strategically leveraged their entrenched social media infrastructures. DataReportal reports that, as of January 2025, Israel had 6.82 million social media users—accounting for 72.2% of its population—suggesting a robust domestic base for narrative dissemination (Kemp, 2025). Iran, with slightly lower social media penetration, compensated through Telegram and diaspora channels, extending its messaging across transnational and Islamist networks. These established channels—combined with coordinated campaigns and real-time content analytics—underscore the integration of digital communication into national security and foreign policy agendas.

The June 2025 confrontation underscores that cyber diplomacy is no longer supplementary but central to statecraft. Hashtags like #IranUnderAttack and #DefendingIsrael serve as mnemonic devices encoding identity, moral claims, and strategic posture. By examining this hybrid conflict—where kinetic strikes and digital narratives are co-deployed—this paper argues that strategic digital influence, structured through identity, normative framing, and real-time narrative deployment, has become an indispensable dimension of contemporary international conflict.

Grand Theory of Constructing Influence: Identity, Norms, and Strategic Narratives

The June 2025 confrontation between Israel and Iran marks a critical juncture in the evolution of international diplomacy—one in which influence is exerted not merely through missiles and bilateral communiqués, but through digital discourses that shape global perception in real time. Constructivist theory, which centers on the social construction of political reality through shared meanings, identities, and norms, offers a compelling lens through which to understand the digital frontlines of this conflict. Unlike materialist paradigms that emphasize power and interest, constructivism posits that state behavior is governed by intersubjective understandings and evolving normative expectations embedded in global society (Wendt, 1999). Within this framework, cyber diplomacy becomes a performative arena—where states assert their legitimacy, project moral authority, and challenge adversarial narratives through sustained digital engagement.

The digital encounter between Israel and Iran during the 2025 escalation is emblematic of how national identity is not only expressed but strategically contested in cyberspace. Iran constructed its identity as a sovereign nation under attack, invoking themes of resistance, anti-Zionism, and postcolonial victimhood to rally both domestic and global Muslim audiences. Through hashtags like #IranUnderAttack, Iranian officials and aligned media outlets framed the conflict

as a defense of sovereignty against Western-backed aggression. Meanwhile, Israel asserted its identity as a besieged democracy responding to existential threats, invoking hashtags such as #DefendingIsrael and emphasizing counterterrorism, technological superiority, and alignment with international legal norms. These conflicting self-perceptions were carefully embedded in coordinated messaging campaigns, making digital identity a site of high-stakes symbolic warfare (Ponce de León & Chenrose, 2025) (Kemp, 2025).

This identity contestation is intrinsically linked to the projection of soft power, a concept Joseph Nye (2004) defines as the ability to influence others through attraction rather than coercion. In the digital theater of 2025, soft power is deployed through emotionally resonant imagery, cultural symbolism, and viral storytelling. Iran's social media strategy drew heavily on depictions of civilian suffering and religious solidarity, while Israel emphasized high-tech defense systems like Iron Dome, democratic resilience, and alliances with the West. Both states relied on affective language and algorithmic amplification to shape how audiences attributed blame, interpreted legality, and empathized with either side of the conflict. These campaigns were not merely reactive; they were calibrated instruments of strategic influence, leveraging digital affordances to sway transnational sentiment.

The use of hashtags as tools of influence further intersects with the emerging concept of normfare, wherein state actors engage in digital contestation over the meaning and application of international norms (Radu, Chenou, & Weber, 2021). Hashtags such as #ZionistCrimes or #RightToDefend act as semantic battlegrounds, each encoding claims to legal legitimacy, human rights, and moral high ground. During the 2025 crisis, these terms operated as proxies for larger normative debates: Who is the aggressor? What constitutes legitimate defense? How should the global community respond? Strategic narratives—defined by Miskimmon, O'Loughlin, and Roselle (2013) as frameworks through which states interpret and communicate events—thus became instrumental to real-time diplomacy. In the absence of consensus or formal mediation, these narratives filled the void, guiding audience interpretation and even influencing preliminary international reactions, including UN debates and regional alliance statements.

In sum, the Iran-Israel cyber confrontation of 2025 exemplifies a paradigmatic shift in the conduct of international relations. Power is no longer exercised solely through territory and arms, but through viral narratives and mediated visibility. The fusion of identity formation, normative contention, and strategic storytelling in digital spaces underscores how cyber diplomacy now serves as a deliberate, high-stakes extension of statecraft. Understanding this phenomenon requires a multidisciplinary framework that synthesizes international relations theory, communication studies, and platform dynamics—particularly as visibility, virality,

and narrative coherence become decisive variables in the contest for global influence.

Methodology : Tracing Strategic Narratives Across Platforms

This study adopts a qualitative, interpretive approach grounded in critical discourse analysis to investigate how Iran and Israel strategically engaged in cyber diplomacy during the June 2025 escalation. The focus lies on the communicative functions of hashtags, narrative framing, and identity construction deployed across social media platforms—specifically X (formerly Twitter), Instagram, and Telegram—during the peak confrontation between June 13 and June 30, 2025. Social media data were obtained via open-source monitoring, including public posts archived by the Atlantic Council's DFRLab, which recorded over 130,000 posts in the 48 hours following the initial Israeli strikes, with nearly 37% attributed to coordinated or automated networks (Ponce de León & Chenrose, 2025). Additional materials include official government statements, foreign ministry briefings, and content disseminated by state-aligned media in English, Persian, and Hebrew.

The analysis is situated within a constructivist theoretical framework, treating hashtags as performative acts that encode normative claims and identities. Drawing on Fairclough's (1992) model of critical discourse analysis and the strategic narrative framework of Miskimmon et al. (2013) the study examines how states attempt to author coherent interpretations of international events to influence perception, legitimacy, and alignment. The digital content is analyzed thematically, with attention to intertextuality, framing, and appeals to moral authority. Hashtags such as #IranUnderAttack, #ZionistCrimes, #DefendingIsrael, and #RightToDefend are treated not merely as metadata but as semiotic anchors in ongoing normative struggles.

The empirical selection of the 2025 conflict is justified by its intensity, real-time narrative mobilization, and the unprecedented integration of AI-assisted information modulation. Platforms like Telegram and X were selected due to their centrality in state-driven messaging. As of January 2025, Israel had 6.82 million active social media users (72.2% of the population), offering a robust domestic audience for state narratives (Kemp, 2025). Iran, despite more restrictive access, leveraged diaspora influencers and cross-platform distribution to disseminate aligned messaging.

To ensure the reliability of sources and mitigate platform bias, the analysis cross-references narrative elements with third-party validation from open-source intelligence outlets such as Bellingcat and DFRLab. The involvement of Grok, the AI assistant deployed by X, is also accounted for, given its prominent and controversial role in labeling trending posts during the conflict, often with inconsistent results (Ponce de León & Chenrose, 2025). Although the data are

publicly accessible, limitations persist, including language translation inconsistencies, bot amplification, and algorithmic filtering that may skew visibility.

Result and Discussion :

a. Making Sense of Strategic Narratives in Conflict Diplomacy

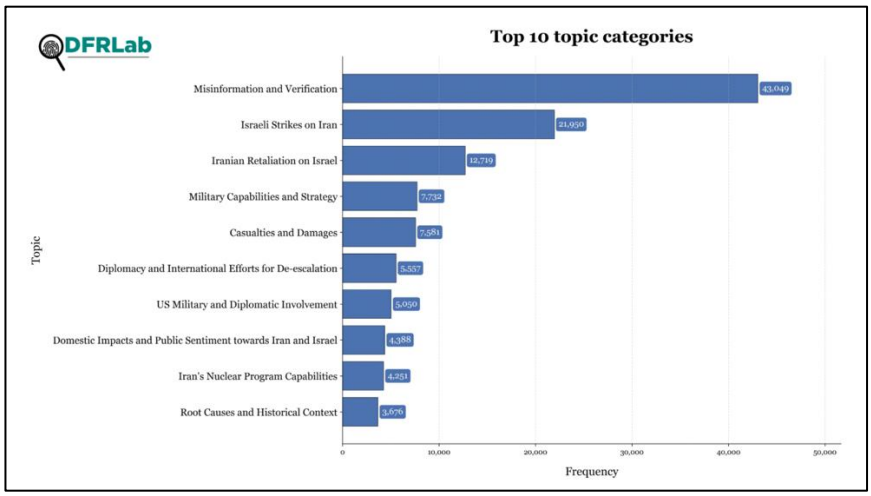
The deployment of hashtags such as #IranUnderAttack and #DefendingIsrael during the June 2025 escalation exemplifies a broader transformation in how states conduct diplomacy—leveraging symbolic power in cyberspace. As narrative compression tools, these hashtags function as semiotic devices that condensate political claims, moral stances, and strategic frames into shareable units. In this respect, they align with constructivist theory's view of identity as a socially produced construct; the speed and scale of dissemination facilitated identity affirmation and group alignment across borders (Wendt, 1999).

The strategic use of emotionally charged hashtags also illustrates how states operationalize soft power in digital arenas. Iran's mixture of moral outrage and communal resilience—amplified by bots and AI-generated images—constructed a narrative that sought to attract sympathy and legitimacy among religiously or ideologically aligned audiences (Shafin et al., 2025). Conversely, Israel's framing leveraged democratic symbols, rational appeals, and alliance imagery to resonate with Western audiences and reinforce institutional legitimacy. The two modes of soft power activation—emotional empathy vs. normative rationality—highlight different approaches to international persuasion (Nye, 2004) (Ponce de León & Chenrose, 2025).

From a normative standpoint, the hashtag warfare exemplifies normfare—a contest over who controls the narrative of moral legitimacy. Iran's narrative framed the strikes as violations of sovereignty and posed them as exercises of imperial aggression, whereas Israel invoked the right to self-defense under international law. Each hashtag thus anchored a specific normative claim, codified into concise slogans that could be globally disseminated and repeated throughout digital networks (Radu, Chenou, & Weber, 2021). This reveals a new dimension of diplomatic struggle: the control of normative discourse via digital channels.

This struggle is further complicated by platform-level interventions. Notably, the AI chatbot Grok, deployed across multiple platforms, attempted to categorize posts during the first days of the escalation. However, as visualized in Figure 1, its content labeling disproportionately emphasized topics such as misinformation and verification, while offering comparatively limited classification on military actions, diplomatic efforts, or public sentiment. This asymmetry suggests an algorithmic bias that may distort the framing of digital conflict, privileging certain normative claims while muting others (Ponce de León & Chenrose, 2025).

Figure 1. Top 10 topic categories identified in Grok-labeled posts during the Iranl–Israle conflict (June 12–15, 2025).



Source: Esteban Ponce de León & Ali Chenrose, DFRLab (2025)

Yet, the presence of AI tools like Grok introduced an additional layer of complexity. Grok's inconsistent fact-checking performances inadvertently skewed digital visibility and the perception of credibility, raising concerns about algorithmic bias within conflict narratives (Ponce de León & Chenrose, 2025). This suggests that even platforms intended to foster digital truth can become arenas of contestation, influencing which narrative threads gain prominence. This intersection of statecraft and platform design warrants further inquiry into the ethics of digital governance.

Our findings also underscore the importance of cross-platform dynamics. Iran's reliance on Telegram and diaspora channels suggests a divergence from conventional Western-centric data streams, illustrating how states circumvent censorship and extend reach to alternative publics (Lesser, 2025). Israel's successful penetration into Persian-language social media—even amid internet filtering—highlights the porous nature of digital borders. These dynamics signal that cyber diplomacy cannot be treated as confined to a single media sphere, but as a networked phenomenon spanning linguistic, cultural, and political domains.

Importantly, this strategic narrative deployment had tangible effects: hashtags shaped online discourse, influenced mainstream media coverage, and likely contributed to diplomatic framing in forums such as the UN Security Council. By pre-packaging moral claims and attribution assessments, digital agents of both states were able to set the tone for policy conversations, even before formal diplomatic channels engaged. This reinforces the idea that in modern conflict,

information operations and narrative framing precede—and sometimes drive—multilateral decision-making processes.

b. The Amplifying Narratives of the Social Media's Impact

The 2025 Iran-Israel conflict underscores a transformative shift in the practice of public diplomacy, with social media platforms now serving as critical arenas for states to assert and contest strategic narratives. As traditional diplomatic channels become increasingly fragmented, platforms like Twitter (X), Instagram, and Telegram emerge as both battlegrounds and broadcasting tools for national identities. In the case of the Iran-Israel escalation, hashtags like #DefendingIsrael and #IranUnderAttack not only encapsulated each state's stance but also actively influenced the global discourse surrounding the conflict. These digital narratives were not simply the result of organic social media reactions but were instead carefully curated and strategically amplified, often through algorithmic systems designed to maximize visibility. As such, social media has become an indispensable tool in contemporary public diplomacy, reshaping the parameters through which nations engage with global audiences.

The role of social media in public diplomacy extends beyond mere message dissemination; it now defines how states perceive and are perceived by the world. Both Israel and Iran used their platforms to foster moral legitimacy, framing their actions within the confines of sovereignty and self-defense. According to Tufekci (2021), social media platforms function as "performative stages" where states not only convey their diplomatic messages but also strategically interact with international public opinion. The hashtag campaigns launched by both nations, such as Iran's #ZionistCrimes and Israel's #StandWithIsrael, are examples of how digital narratives perform this double role—asserting identity while simultaneously crafting a counter-narrative to delegitimize the opponent's actions.

The role of social media in public diplomacy extends beyond mere message dissemination; it now defines how states perceive and are perceived by the world. Both Israel and Iran used their platforms to foster moral legitimacy, framing their actions within the confines of sovereignty and self-defense. According to Poushter, Gubbala, & Austin (2024) over 72% of global internet users engage with social media platforms, making them an essential part of contemporary diplomatic engagement. While government-controlled narratives were once disseminated through more formal channels such as diplomatic cables or press releases, hashtag-driven digital narratives now dominate the conversation in real-time. This immediacy and scale create a challenge for state actors attempting to control the narrative or shape public perception, as misinformation and disinformation can rapidly circulate. As both Israel and Iran heavily relied on AI-enhanced amplification tools, they were able to

circumvent traditional media outlets and directly influence audiences, particularly through platforms like Twitter and Instagram.

In addition to the amplification of messages, social media's role in shaping emotional engagement cannot be overstated. Both Iran and Israel leveraged the emotionally charged content to resonate with their respective audiences, often using AI-generated visuals and emotionally impactful language. The appeal to moral outrage, national pride, and survival instincts played a central role in garnering support. For instance, Iran's #IranUnderAttack narrative capitalized on the themes of post-colonial victimhood and anti-imperialism, portraying the Israeli airstrikes as acts of aggression against a sovereign nation. Conversely, Israel's #DefendingIsrael framed the conflict within the language of self-defense and human rights, appealing to Western liberal democratic values. As Snyder (2018) highlights, the ability to evoke such emotional resonance is what often turns digital narratives into powerful tools of persuasion in global diplomacy.

The manipulation of digital narratives during the 2025 conflict was not without its challenges, however. One of the critical issues surrounding social media engagement is the risk of algorithmic bias and platform censorship, both of which can significantly distort the narratives being promoted. Platforms like Twitter and Facebook utilize complex algorithms that determine which content appears prominently in users' feeds. Tufekci (2021) notes that such algorithmic systems are not neutral; they often amplify content that evokes strong emotional responses or that conforms to dominant cultural narratives. This amplification can inadvertently skew public perception, disproportionately favoring one side's narrative over another. As a result, the battle for attention on social media platforms becomes as much about algorithmic manipulation as it is about the content itself.

Another key challenge lies in the platform design and the ethical questions it raises about content moderation and transparency. As noted by Miskimmon et al. (2013), platforms like X and Instagram do not just facilitate the spread of ideas but also play an active role in curating the discourse. The use of AI moderation systems, such as Grok, has become a central tool for filtering content, but its inconsistency and lack of transparency raise significant concerns. Ponce de León & Chenrose (2025) show that while Grok was deployed to help identify trending content, its fact-checking abilities were often inadequate and contributed to the muddling of critical narratives, particularly in the context of the conflict. These failures in AI moderation only underscore the need for more transparent and ethical regulation in the digital realm.

Another overlooked yet critical dimension in the Iran–Israel digital conflict is the role of platform governance and moderation infrastructures. Platforms like X, Instagram, and Telegram act not merely as passive conduits but as algorithmically

mediated arenas where decisions about visibility, labeling, and removal profoundly affect narrative dominance. Moderation tools are often opaque in both logic and accountability, creating asymmetries in what content is suppressed or amplified. This lack of transparency in algorithmic curation distorts digital diplomacy, particularly during high-stakes geopolitical escalations (Gillespie, 2018).

The June 2025 confrontation exposed how moderation bias can become geopolitical. For instance, while Grok—a prominent AI moderation tool on X—flagged several Iranian posts as misinformation, it inconsistently flagged Israeli content with similar emotional or visual rhetoric, raising questions about implicit platform-side value judgments. These discrepancies suggest algorithmic partiality shaped by training data, language prioritization, or even political alignment of platform policies (Roberts, 2019). As a result, state-backed narratives can benefit from infrastructural favoritism rather than meritocratic engagement. Furthermore, content moderation algorithms often suppress minority voices or non-English narratives, which disadvantages actors like Iran that rely heavily on diaspora and multilingual channels. Research shows that moderation tools trained predominantly on Western-centric norms frequently misclassify religious or political expression as harmful or extreme, disproportionately silencing non-Western narratives (Noble, 2018). This risks institutionalizing epistemic injustice, where states are not only battling each other, but also the architecture of the platforms that mediate global discourse.

The influence of diaspora actors in cyber diplomacy is increasingly understood through the lens of digital diasporas, who utilize social media to sustain transnational identities and act as normative influencers during homeland crises (Candidatu & Ponzanesi, 2022). In the Iran–Israel case, Iranian diaspora communities in North America and Europe played a crucial role as narrative amplifiers, reshaping global sentiment via hashtags like #IranUnderAttack—echoing how digital diasporas "expand and transform their agency in the digital age."

This phenomenon reflects what Chernobrov (2021) term "participatory warfare", where diaspora members function as informal "cyberwarriors," using coordinated content and emotional testimonies to influence public discourse abroad. Such behavior was evident in live Twitter threads and Telegram circles, where diaspora-generated visuals and personal stories significantly boosted engagement and transnational solidarity during the early days of the conflict. Consequently, diaspora communities serve not only as content distributors but as moral and normative legitimisers, enhancing the authority of state-sponsored messaging in digital battlegrounds (Stein, 2025). This highlights the need for governments and researchers to acknowledge diaspora groups as strategic actors in

cyber diplomacy, capable of shaping both narrative reach and interpretive frameworks.

Conclusion and Recommendation: Reframing Influence and Norms in the Aftermath

The June 2025 digital confrontation between Iran and Israel underscores a profound transformation in the exercise of state influence, where hashtags and digital narratives now serve as instruments of diplomacy, norm-setting, and strategic communication. Moving beyond traditional frameworks of material power and coercive statecraft, this conflict exemplified the centrality of symbolic, emotional, and normative resources in contemporary international relations. Through competing hashtag campaigns—#IranUnderAttack and #DefendingIsrael—each state mobilized its identity claims, reasserted normative legitimacy, and framed the other as a violator of global order.

This study has shown that hashtags are not merely digital ephemera but constitute coherent, ideologically charged tools of statecraft, capable of mobilizing publics, shaping perceptions, and influencing multilateral discourse. The interplay between AI-driven amplification, affective imagery, and normative language suggests a new operational logic of cyber diplomacy: one that prioritizes visibility, virality, and emotional resonance as sources of political power. Furthermore, the inconsistent role of platform-level moderation tools—such as Grok—reveals the infrastructural vulnerability of global communication networks, where algorithms inadvertently mediate geopolitical narratives.

The emotional engagement driven by both AI tools and user-generated content is a key factor in this transformation. Social media platforms, particularly Twitter (X) and Instagram, allowed Iran and Israel to amplify emotional narratives through carefully crafted posts, leveraging visual symbolism and emotionally resonant language. These campaigns were not just about factual claims; they were designed to evoke strong responses, framing the conflict in terms of victimhood and defense. This form of digital narrative warfare has moved beyond mere information sharing, instead targeting the emotional core of audiences worldwide.

By grounding this analysis within a constructivist framework, the article contributes to ongoing debates about identity, norm contestation, and strategic narratives in the digital age. It highlights how digital arenas—particularly social media—are no longer peripheral to international relations but central battlegrounds where legitimacy, blame, and moral authority are contested in real time. These dynamics challenge conventional understandings of diplomacy and call for a re-evaluation of soft power mechanisms in algorithmic environments.

The Iran–Israel cyber confrontation also presents a critical inflection point for global governance and digital regulation. As states increasingly use hashtags, AI-

generated content, and algorithmic amplification to shape global perceptions, existing diplomatic protocols must adapt to account for symbolic and narrative influence. There is an urgent need for international frameworks that address the normative and operational challenges of cyber diplomacy, especially in times of conflict escalation. Regulation of AI-driven amplification tools and content moderation on platforms like X, Telegram, and Instagram becomes essential in maintaining the integrity of international discourse. These platforms must be held accountable not only for content moderation but also for their algorithmic influence on conflict narratives, as they play a central role in shaping the international community's response to global crises.

Governments and multilateral bodies should consider establishing norms on digital engagement during crises, including standards for attribution, transparency, and the ethical deployment of automated amplification tools. Platforms that function as diplomatic intermediaries—such as X, Telegram, and Instagram—must be held accountable not only for content moderation but also for their algorithmic influence on conflict narratives. In parallel, democratic states must build institutional capacity for narrative resilience by integrating media literacy, public diplomacy, and cybersecurity into a unified strategic doctrine. This holistic approach will ensure that states can better manage digital narratives while minimizing the spread of disinformation and manipulation.

By anticipating how digital narratives can sway public opinion, legitimize force, and influence diplomatic outcomes, policymakers can better safeguard the integrity of international discourse. This requires not only technological preparedness, but also conceptual clarity about the nature of legitimacy in the age of information warfare. The case of Iran and Israel in 2025 serves as a template for future conflicts where narrative power may precede, accompany, or even outweigh traditional coercive measures.

To address the growing impact of cyber conflicts like the Iran–Israel case, states must develop stronger narrative resilience by integrating digital diplomacy, media literacy, and AI governance into national security strategies. Public institutions and international bodies should ensure that social media platforms and AI tools like Grok operate transparently and ethically during crises, preventing biased amplification of conflict narratives. Additionally, governments are encouraged to engage diaspora communities as strategic amplifiers of credible messaging, while also promoting public awareness about digital disinformation. Ultimately, managing cyber diplomacy requires a proactive, coordinated response that treats narrative control as a critical element of modern statecraft.

References

Azeez, I. A. (2023). The Influence of Digital Diplomacy on Foreign Policy. *Journal of Tourism Economics and Policy, 3(3)* , 189-203.

Bhardwaj, A. (2025, June 15). *Analysis of Israel's June 2025 Military Campaign Against Iran.* Retrieved from Foreign Affairs Forum: https://www.faf.ae/home/2025/6/15/analysis-of-israels-june-2025-military-campaign-against-iran

Bjola, C., & Holmes, M. (. (2015). *Digital diplomacy: Theory and practice.* Routledge.

Candidatu, L., & Ponzanesi, S. (2022). Digital Diasporas: Staying with the Trouble, Communication,. *Culture and Critique, Volume 15, Issue 2, June* , 261–268,.

Chernobrov, D. (2021). Diaspora as cyberwarriors: infopolitics, participatory warfare and the 2020 Karabakh war. *International Affairs*, 98.

Fairclough, N. (1992). *Discourse and social change. .* Polity Press.

Gillespie, T. (2018). *Custodians of the Internet: Platforms, content moderation, and the hidden decisions that shape social media .* Yale University Press.

Kemp, S. (2025, January). *Digital 2025: Israel.* Retrieved from DataReportal: https://datareportal.com/reports/digital-2025-israe

Lesser, M. (2025, July 1). *FDD connects anti-Israel network on social media to Iranian pro-regime actor.* Retrieved from FDD (Foundation for Defense of Democracies). : https://www.fdd.org/analysis/2025/07/01/fdd-connects-anti-israel-network-on-social-media-to-iranian-website-pro-regime-actor/

Miskimmon, A., O'Loughlin, B., & Roselle, L. (2013). *Strategic narratives: Communication power and the new world order. .* Routledge.

Noble, S. U. (2018). *Algorithms of oppression: How search engines reinforce racism.* NYU Press.

Nye, J. S. (2004). *Soft power: The means to success in world politics.* Public Affairs.

Ponce de León, E., & Chenrose, A. (2025, June 24). *Grok struggles with fact-checking amid Israel-Iran war.* Retrieved from Atlantic Council DFRLab: https://dfrlab.org/2025/06/24/grok-struggles-with-fact-checking-amid-israel-iran-war/

Poushter, J., Gubbala, S., & Austin, S. (2024, February 5). *8 charts on technology use around the world.* Retrieved from Pew Research Center: https://www.pewresearch.org/short-reads/2024/02/05/8-charts-on-technology-use-around-the-world/?utm_source=chatgpt.com

Radu, R., Chenou, J. M., & Weber, S. (2021). Normfare: Norm entrepreneurship in internet governance. *Telecommunications Policy, 45(8)* .

Riordan, S. (2019). *Cyberdiplomacy: Managing Security and Governance Online.* Cambridge: Polity Press.

Roberts, S. T. (2019). *Behind the screen: Content moderation in the shadows of social media.* Yale University Press.

Snyder, T. (2018). *The Road to Unfreedom: Russia, Europe, America.* New York: Tim Duggan Books.

Stein, P. &. (2025, June 20). *Iranians in the U.S. worry over Israel conflict – and Iran's future.* Retrieved from The Washington Post:

https://www.washingtonpost.com/nation/2025/06/20/iran-america-diaspora-israel/
Tufekci, Z. (2021). *Twitter and Tear Gas: The Power and Fragility of Networked Protest.* Yale University Press.
Wendt, A. (1999). *Social theory of international politics. .* Cambridge University Press.