Legopolitics: Toying with the Enemy

Ludic Framing in Pro-Iranian Wartime Propaganda

Cantay Caliskan
University of Rochester

Hari Acharya
University of Rochester

Abstract

During the 2026 war between Iran and the United States and Israel, pro-Iranian messaging adopted AI-generated LEGO-style animation. We conceptualize this format as *outward-facing ludic propaganda*: wartime communication oriented toward an adversary's public through the register of play. We argue that this orientation encourages personalized blame and references to the adversary's domestic grievances, while playful representation selectively obscures suffering and accommodates fantastical humiliation. Using `legopolitics`, an open-source multimodal Python pipeline, we analyze 363 videos comprising 7,288 scenes and compare their content with Iranian English-language state media. Iranian actors occupy the agent role in 65.6 percent of depicted confrontations (95 percent CI 59.7–71.6), significantly exceeding parity. Among blame attributions directed at either named U.S. leaders or the American public, 93.1 percent target leaders (95 percent CI 89.6–96.5), overwhelmingly Trump. Within videos and comparable target types, Iranian-inflicted violence is depicted with 0.77 standard deviations less victim suffering than adversary-inflicted violence ($p < .001$). Enemy-humiliation scenes score 0.078 points higher in fantasticality on a 0–3 scale, controlling for violence ($p = .002$). Compared with state media, video text contains more references to U.S. domestic politics and popular culture, with an adjusted difference of 9.7 percentage points ($p < .001$). Three further expectations are not supported: American civilians are not exculpated more than Israeli civilians once shown, humor predicts views between channels but not within them, and sacred legitimation, rarer in the genre as a whole, is more common in its regime-linked core than in state media. The paper provides a systematic quantitative account of LEGO-style propaganda, combining scene-level narrative measurement, comparative media analysis, and reusable software to show how playful wartime communication distributes agency, blame, and suffering.

Keywords: lego; propaganda; computer vision; political communication; computational social science.

"Every joke is a tiny revolution."

—George Orwell, *Funny, but not Vulgar* (1945)

1 Introduction

A few weeks into the 2026 war between Iran and the United States and Israel, a conspicuous strand of pro-Iranian messaging took the form of toy bricks. English-language, AI-generated animations portrayed Donald Trump as petulant and ineffective, staged fantastical Iranian victories, and turned the Strait of Hormuz into a cartoon toll booth. The videos attracted millions of views, and copies continued circulating after YouTube suspended a leading producer's channel on March 27 (Agence France-Presse, 2026). That producer, Explosive Media, presented itself as independent, although a representative subsequently acknowledged in a BBC interview, reproduced by *MARKETING Magazine Asia*, that the Iranian government was a customer (Kaur, 2026). Commentary quickly framed the phenomenon as a contest over who was *winning* the propaganda war (Gault, 2026). Its visibility poses a more specific question: how does propaganda change when it addresses an adversary's public through the language of play?

Research on authoritarian propaganda explains how communication legitimates incumbents, organizes beliefs about regime strength, and disciplines public expression (Wedeen, 2015; Huang, 2018; Carter and Carter, 2023). Research on soft propaganda further demonstrates that emotionally engaging entertainment can mobilize anti-foreign sentiment (Mattingly and Yao, 2022). Research on external state media also shows that governments can promote favorable judgments of authoritarian competence among international audiences (Mattingly et al., 2025). The puzzle here concerns a particular configuration of audience, purpose, and form. Messages aligned with Iran's wartime position repeatedly organize attention around adversary leaders, domestic controversies in the United States, and comic representations of violence. Their relationship to Iranian traditions of martyrdom and sacred commemoration is therefore an empirical question, rather than something established by their political alignment (Rolston, 2020; Saramifar, 2019).

We conceptualize this configuration as *outward-facing ludic propaganda*: political communication that addresses an adversary's public through a playful mode of presentation. Its defining features are the public it addresses and the register through which it presents conflict. The concept can encompass material produced by state-linked organizations, sympathetic creators, and later imitators; each requires separate provenance evidence. Our argument is that addressing an adversary's public creates incentives to discredit its leadership and interpret the war through its domestic grievances, while the ludic register provides distinctive ways to represent power, suffering, and triumph. The resulting observable implications concern enemy construction, the distribution of violence and fantasy, and differences from conventional state-media coverage.

The audience argument builds on longstanding accounts of psychological warfare. Lasswell distinguishes propaganda directed toward persons, groups, and institutions, while Linebarger identifies attempts to separate populations from their leaders as a possible wartime strategy (Lasswell, 1927; Linebarger, 1954). Research on American war support gives these distinctions political relevance: judgments about success and the justification for military action matter, while partisan elites shape how citizens interpret conflict (Gelpi et al., 2009; Berinsky, 2007). An adversary seeking to exploit

these vulnerabilities has reasons to portray opposing leaders as ineffective, assign them responsibility for aggression, and distinguish their decisions from the interests of ordinary citizens. We therefore predict Iranian control over depicted confrontations, personalized blame directed at U.S. leaders, and more frequent explicit exculpation of American than Israeli civilians. These are separate comparisons; the presence of a villainous leader alone does not establish a leader–people split.

The medium argument concerns how the same political claims are rendered. Toy figures permit destruction to be presented through plastic bodies, miniature settings, and comic action. The relevant expectation is selective sanitization: for comparable target types, Iranian-inflicted violence should foreground victims' suffering less often than adversary-inflicted violence. This comparison draws on accounts of moral justification and the minimization of harmful consequences (Bandura, 1999). We also predict that scenes of Iranian victory or adversary humiliation will be more fantastical than other scenes within the same video. Humor research provides a rationale for investigating this organization of fantasy, while also showing why reduced argument scrutiny cannot be equated with greater persuasion (Young, 2008; Nabi et al., 2007). Whether play actually lowers the credibility costs of implausible claims remains a distinct audience-effect proposition.

A further prediction concerns attention: more humorous videos should receive more views among comparable posts observed at a common age. Research connecting emotional arousal and social transmission motivates this expectation (Berger and Milkman, 2012). The comparison with Iranian English-language state media adds two register predictions: a lower proportion of political justifications invoking religious authority or martyrdom, and a higher proportion of narrative units drawing on U.S. domestic politics or popular culture. Distinguishing content, form, and stance makes it possible to investigate continuity in political claims alongside changes in their expression (Shifman, 2013). These predictions allow the evidence to challenge the argument through reversed or precisely negligible contrasts, while keeping exposure, cultural orientation, and persuasion analytically distinct.

We examine these expectations using 363 LEGO-style videos posted on YouTube, TikTok, and X between March and June 2026, organized into 7,288 coded narrative scenes. We introduce `legopolitics`, an open-source Python pipeline that converts brick-built video into linked observations on scenes, actors, relationships, events, symbols, speech, and frames (Caliskan, 2026). A visual layer comprising detector outputs and vision-language annotations across 96,530 frames supports checks of scene coding against image evidence. We compare the videos with a pilot corpus of war-related articles from five Iranian English-language state outlets. The analysis distinguishes material traceable to Iranian pro-government producers or state outlets from that circulated by other Iran-aligned accounts and by commercial or satirical adopters. Because the article pilot is not time-matched to the videos, the cross-register comparisons are organized around common event topics and account for the different genres and platforms represented in the corpora.

The eight hypotheses are tested with regression designs that hold the producer constant wherever the hypothesis permits: video fixed effects for within-video contrasts, channel fixed effects for circulation, and event-topic fixed effects for the register comparison. Four are supported. Iranian actors control the outcome of two-thirds of depicted confrontations, and nearly all of those that

resolve a conflict; blame for U.S. aggression is assigned to named leaders rather than to the American public in 93 percent of attributions; Iranian-inflicted violence is shown with far less victim suffering than adversary-inflicted violence against the same kinds of target; and the videos refer to U.S. domestic politics and popular culture about twice as often as state media does, with the largest gap in grievances addressed to Americans. One is partially supported: scenes that humiliate the enemy are more fantastical than the other scenes of the same video, but victory scenes are not once their violence is accounted for. Three are not: American civilians are not exculpated more than Israeli civilians once civilians appear, the selectivity operating instead through who is shown at all; humor predicts views between channels but not among a channel's own posts; and sacred legitimation, rarer in the genre as a whole, is more common in its regime-linked core than in the state-media articles.

The provenance comparison delimits the argument. Representational dominance and the sacralized frame belong to the videos closest to the state and fade among imitators, while the orientation toward the American audience, its leaders, its grievances, and its culture, is present in every group of producers. The core-to-periphery differences are consistent with political messages being reworked during circulation, although producer composition can also account for part of the observed variation. They establish the importance of provenance without identifying a causal effect of diffusion by themselves.

The paper makes three contributions. Substantively, it provides a scene-level analysis of an AI-generated wartime propaganda genre and a comparison with conventional Iranian English-language coverage. Theoretically, it specifies how communication addressed to an adversary's public combines personalized blame, selective visibility of suffering, and playful representations of violence and triumph, and it identifies where those expectations fail: civilian exculpation is not selective by nationality, play does not reward humor within a channel, and the sacred frame survives in the regime-linked core. Methodologically, `legopolitics` connects audiovisual observations with interpretable narrative units, making the source evidence underlying political classifications inspectable. The study evaluates representations and circulation; changes in viewers' credibility judgments or support for war require additional audience evidence. The following sections review the relevant literature, develop the research questions and hypotheses, describe the data and descriptive patterns, present the empirical methods and results, and discuss their implications.

2 Literature Review

2.1 Computer vision, political images, and film

Political analysis of film begins with the recognition that moving images organize political meaning through composition, sequencing, and perspective. Shapiro's *Cinematic Geopolitics* examines how feature films and documentaries can contest geopolitical imaginaries and representations of state violence (Shapiro, 2009). This interpretive tradition identifies politically consequential features that computational analysis must preserve: who acts, whose experience structures the narrative, and how audiovisual form makes particular political worlds intelligible.

The images-as-data literature supplies tools for studying these representations systematically.

Webb Williams et al. (2020) introduce convolutional neural networks for image classification, while Torres and Cantú (2022) explain their implementation and limitations through handwritten electoral tallies. These are foundations for visual measurement rather than studies of cinema. Joo and Steinert-Threlkeld (2022) extend the approach to political communication through analyses of politicians' Facebook photographs and Twitter images of protests. Their distinction between general-purpose labels and classifiers trained for substantive research questions is especially relevant when generic object categories fail to capture political responsibility or vulnerability.

Moving-image research within political science already goes beyond isolated photographs. Tarr et al. (2023) combine audio and image features to code campaign advertisements and evaluate their measurements against human-coded Wesleyan Media Project data. Their results support automation for many variables without establishing equal performance across every interpretive category. More recently, Breuer et al. (2025) process 9,707 presidential campaign advertisements through a pipeline of video preparation, transcription, storyboarding, and summarization, evaluating outputs against human alternatives. These studies establish precedents for multimodal political-video pipelines and locate the methodological challenge in the measurement and validation of substantive constructs.

Film and digital-humanities research addresses the additional problem of narrative sequence. Arnold and Tilton's distant-viewing framework treats computational metadata as an interpretation of visual materials, linking large-scale pattern detection to close viewing (Arnold and Tilton, 2019). Their television analysis follows characters, shots, and scenes rather than treating every frame as an independent cultural object. In computer vision, MovieGraphs represents social situations in film clips through annotated characters, relationships, interactions, attributes, and motivations (Vicol et al., 2018). Its human-annotated graphs offer a model for relational representation; they do not establish that political motives can be inferred automatically without validation.

Together, these traditions motivate an analysis that connects visual detection to narrative relations. A leader's presence does not establish blame, and an explosion does not identify whose suffering the story foregrounds. In LEGO-style animation, stylized bodies also make transfer from photographic training data a measurement question. The present study builds on these precedents by connecting auditable image evidence with scene-level measures of agency, responsibility, suffering, and fantasy through `legopolitics`. Attention to temporal context and comparisons across producers are necessary to turn scalable video processing into evidence about political representation.

2.2 Propaganda, political control, and foreign audiences

Political-science research treats propaganda as strategic communication whose operation depends on institutions, audiences, and political objectives. Lasswell's account emphasizes the organization of collective attitudes through significant symbols (Lasswell, 1927). Wartime propaganda can strengthen domestic solidarity, cultivate allies, or weaken an adversary. Linebarger's account of psychological warfare explicitly considers separating hostile populations from their leadership (Linebarger, 1954). Enemy-centered communication and appeals to an adversary's citizens consequently have longstanding theoretical foundations.

Contemporary comparative research explains why propaganda varies across political settings. Carter and Carter (2023) connect propaganda strategies to autocratic institutions and the constraints rulers face. This perspective cautions against treating authoritarian communication as uniformly self-glorifying. The relevant question is which claims and styles serve particular political relationships. A claim that humiliates foreign leaders may simultaneously advertise domestic competence, define national identity, and solicit sympathy abroad.

A second tradition distinguishes persuasion from the organization of public compliance. Wedeen's analysis of Syrian political symbolism shows how extravagant representations of leadership can structure public performances even without sincere belief (Wedeen, 2015). Huang's Chinese experiments likewise show that hard propaganda can communicate state strength and discourage dissent while worsening evaluations of the regime (Huang, 2018). These findings establish that literal credibility is not always propaganda's decisive resource. They do not, however, show that intimidation mechanisms carry over to foreign audiences voluntarily watching comic videos.

Propaganda also governs attention and attribution. King et al. (2017) find that fabricated Chinese social-media posts primarily pursue distraction and cheerleading rather than sustained argument with critics. Studying Russian state television, Rozenas and Stukal (2019) show how unfavorable economic developments are attributed to external factors while favorable developments are credited to domestic politicians. These studies explain why counting positive references to the regime alone can miss politically useful communication. The allocation of blame and the displacement of attention are substantive outcomes.

Research on effects further rejects a universal model of audience response. Adena et al. (2015) show that radio's political consequences changed with control of broadcasting in interwar Germany and that responses to antisemitic propaganda depended on prior local predispositions. Mattingly and Yao's experiments demonstrate that emotionally engaging soft propaganda can increase anger and anti-foreign sentiment in China (Mattingly and Yao, 2022). A nineteen-country experiment also finds that Chinese state-media messages can improve international evaluations of China's political-economic model (Mattingly et al., 2025). Entertainment and foreign targeting therefore belong within established propaganda research, although their effects remain conditional.

The present study connects these arguments at the level of observable content. Addressing an adversary's public supplies a rationale for personalized blame and appeals to domestic grievances; playful form supplies possible resources for depicting violence and triumph. This synthesis motivates directional comparisons without assuming that content reveals producer intentions or proves persuasion. Views measure circulation, and narrative exculpation measures representation. Establishing demoralization, belief change, or reduced support for war requires evidence about audiences beyond the videos themselves.

2.3 Animation, toy aesthetics, and state messaging

Governments' use of animation predates digital platforms and generative artificial intelligence. Keller's archival study of Disney's collaboration with the U.S. Office of Inter-American Affairs documents a wartime program combining entertainment, education, distribution, and audience evaluation in Latin America (Keller, 2025). Cartoons served foreign-policy objectives through familiar characters and accessible stories, with government agencies, commercial studios, and local intermediaries participating in their circulation. Animation was therefore already a vehicle for outward-facing state communication. Keller also shows that officials' evaluations conflated popularity and reach with political effect, a problem that persists when online visibility becomes evidence of propaganda success.

Contemporary Chinese research shows how animated personification connects geopolitical narratives to popular culture. Guan and Hu's analysis of *Year Hare Affair* finds that animalized nation-states organize distinctions between national self and foreign others while affirming aspirations to national strength (Guan and Hu, 2020). Yet the series also revises aspects of official patriotic narratives. Their account of negotiation between relatively independent animation and state hegemony is useful for analyzing a genre whose producers, sponsors, and imitators need not

share identical organizational relationships.

Playful identification also involves audiences and platforms. Wong et al. (2021) document how state-run social-media accounts adopted fan practices around personified construction vehicles during China's COVID-19 hospital-building campaign. Their concept of fandom governance identifies a route from affection for familiar objects to attachment to the nation. Sun's study of *Year, Hare, Affair* further shows how related online practices acquire competing interpretations across Bilibili and Weibo (Sun, 2025). Participatory circulation can support official patriotic readings while also generating contested social meanings. These studies establish neither uniform reception nor a straightforward chain of state command.

There is also a direct LEGO-style precedent. In April 2020, Xinhua distributed *Once Upon a Virus*, an English-language animation contrasting China's depicted warnings about COVID-19 with contradictory American responses (Xinhua, 2020). Hagström and Gustafsson discuss its brick-like national personifications within the wider Sino-American narrative contest (Hagström and Gustafsson, 2021). The example establishes an earlier use of toy aesthetics in state-distributed international messaging. Here, the relevant evidence is the use of toy figures to stage a national competence contrast. Their broader analysis also shows how strategic narratives encounter existing interpretive frameworks and limited uptake, qualifying assumptions that a state controls the meanings its stories acquire abroad.

These precedents locate the Iranian videos within an established repertoire of animated diplomacy, personification, and participatory nationalism. They leave open the specific distribution of blame, suffering, fantasy, and sacred justification across a wartime corpus. Humor research suggests that jokes can reduce argument scrutiny while also encouraging message discounting (Young, 2008; Nabi et al., 2007); neither consequence follows automatically from depicting toys. Likewise, Iranian research on martyrdom imagery motivates measuring sacred legitimation rather than assuming that play eliminates it (Rolston, 2020). The contribution is to test these representational differences systematically, compare animation with state-media coverage of the same events, and distinguish documented state connections from stylistic imitation.

3 Research Questions and Hypotheses

LEGO-style propaganda combines familiar wartime claims with toys, parody, and fantasy. Following Shifman's distinction among content, form, and stance, we examine how this playful, or *ludic*, register organizes representations of the conflict (Shifman, 2013). Three research questions connect its portrayal of adversaries, treatment of violence, and differences from Iranian English-language state media.

RQ1. Who acts, who is blamed, and who deserves sympathy?

Propaganda organizes attitudes toward political actors, while framing assigns responsibility and evaluates their conduct (Lasswell, 1927; Entman, 1993). Outward-facing propaganda also has a strategic reason to distinguish an adversary's public from its leaders: citizens may be addressed as potential opponents of their government's war. Linebarger's account of psychological warfare identifies this separation as one propaganda approach (Linebarger, 1954). We therefore expect narratives that favor Iranian agency and concentrate blame on U.S. leaders. H1c extends this

argument by proposing that the invitation to identify with civilians is stronger for Americans than Israelis; this asymmetry is a case-specific expectation, rather than a general implication of the theory.

- **H1a — Representational dominance:** In confrontations featuring Iranian and adversary actors, Iranian actors are portrayed as controlling the outcome more frequently than adversary actors.
- **H1b — Personalized blame:** Among attributions of responsibility for U.S. aggression, named U.S. political leaders are assigned responsibility more frequently than the undifferentiated American public.
- **H1c — Selective civilian exculpation:** When civilian publics are represented or discussed, American civilians are more likely than Israeli civilians to be explicitly separated from responsibility for their leaders' actions.

These hypotheses distinguish narrative control, blame, and exculpation. Civilian absence is recorded separately from how civilians are portrayed.

RQ2. What does the ludic register do to violence, triumph, and circulation?

Bandura's account of moral disengagement suggests that minimizing consequences can make violence easier to justify (Bandura, 1999). We expect the videos to downplay suffering caused by Iran relative to suffering caused by its adversaries. Humor may also accommodate implausible triumphs: it can reduce critical scrutiny, although audiences may discount a joke's political relevance (Young, 2008; Nabi et al., 2007). This motivates testing where fantasy appears, without assuming that viewers believe its claims. Finally, research linking emotional arousal to sharing motivates a separate expectation about circulation (Berger and Milkman, 2012).

- **H2a — Selective sanitization:** For comparable target types, scenes of Iranian-inflicted violence are less likely to foreground victims' suffering than scenes of U.S.- or Israeli-inflicted violence.
- **H2b — Fantastical triumph:** Within the same video, scenes depicting Iranian victory or adversary humiliation have higher fantasticality scores than other scenes.
- **H2c — Attention association:** Among comparable posts, higher independently coded humor scores are associated with higher view counts measured at a common post age.

Fantasticality concerns departures from ordinary physical or causal constraints beyond the shared toy aesthetic. H2b tests its concentration in particular scenes; H2c concerns exposure. Neither establishes persuasion or a reduction in credibility costs.

RQ3. How does the playful register differ from Iranian state-media coverage?

Studies of Iranian murals and war photography document the importance of martyrdom and sacred narratives in particular visual traditions (Rolston, 2020; Saramifar, 2019). We propose that

playful animation places less emphasis on these sources of legitimacy and more on U.S. controversies and popular culture. These are proposed register differences: the literature establishes the relevance of sacred and cultural references, but does not establish their relative prevalence in these corpora.

- **H3a — Reduced sacred legitimation:** Among political-justification claims about comparable events, the proportion invoking religious authority or martyrdom is lower in LEGO-style videos than in English-language Iranian state-media coverage.
- **H3b — U.S.-oriented cultural reference:** Among comparable narrative units, the proportion referring to U.S. domestic politics or popular culture is higher in LEGO-style videos than in English-language Iranian state-media coverage.

Comparisons use common event topics, with differences in producers, genres, and platforms limiting causal interpretation. Cultural references indicate content orientation rather than audience response. Section 5 specifies the measures, comparisons, and clustering. Contrasts opposite to a prediction count against it when precisely estimated; estimates spanning both directions remain inconclusive.

4 Data and Descriptive Analysis

4.1 Collecting and screening the videos

The corpus consists of every LEGO-style political video we could locate on YouTube, TikTok and X that was presented as Iranian, or as being about Iran, in the period surrounding the 2026 war. Collection yielded 549 candidate videos with 74.9 hours of footage: 536 from YouTube, nine from TikTok, and four from X. Twelve of them, nine Chinese and three Russian LEGO-style videos, were collected as reference points. A two-stage screening protocol separated the object of study—videos that are themselves brick-built or LEGO-style animations—from the commentary ecosystem around them. Each candidate was first triaged from twelve frames sampled across its full length, its metadata, and its first 90 seconds of transcribed speech. The triage record stores a category, a decision, the coder's confidence, whether the video is likely AI-generated, the estimated share of LEGO-style footage, the main language, and a one-line topic summary. Videos were retained when the video itself was a LEGO-style animation or a compilation of such animations; re-uploads of the same animation by other accounts were retained and coded separately. Videos were excluded when they were news reports about the LEGO videos (48), reaction videos or podcasts (63), commentary that merely showed clips (7), non-LEGO animation (8), or unrelated content (55). Screening removed 181 videos and 47.9 hours of footage.

This left 368 coded videos. Our main analysis sample comprises the 363 videos uploaded between the start of the war on 28 February 2026 and the end of collection in June 2026; the remaining five are three LEGO-style videos from the COVID-19 period (two copies of the Chinese state news agency Xinhua's "Once Upon a Virus" and a Taiwanese explainer) and two items from the June 2025

Israel–Iran war, which we use only as historical reference points. The 363 videos were posted by 136 channels or accounts (353 on YouTube, seven on TikTok, three on X) and run to 26.9 hours of footage. At the time of collection they had accumulated 30.1 million views, 1.2 million likes and 139,000 comments (like counts are visible for 315 videos and comment counts for 251). Table 1 describes the videos.

Table 1: Video-level descriptive statistics (2026 war sample, $n = 363$)

Variable	N	Mean	Median	10th pct.	90th pct.	Total
Length (seconds)	363	267	225	112	510	96,747
Views	360	83,491	2,896	29	175,154	30,056,939
Likes	315	3,872	253	3	9,270	1,219,704
Comments	251	554	56	1	1,200	139,104
Scenes	363	20.1	19	12	30	7,288
Camera shots (estimated)	363	80.0	66	25	131	29,023
Character records	363	15.5	14	9	23	5,617
Directed actor relationships	363	12.1	11	7	17	4,378
Events	363	9.2	9	6	13	3,329
Symbol records	363	20.4	20	12	30	7,409
Speech and on-screen text segments	363	44.8	38	18	74	16,270
Words transcribed (Whisper)	358	354	272	2	819	126,767
Distinct speakers (diarization)	358	1.9	2	1	3	–
Frames (1 per second)	363	266	225	112	510	96,530
Share of frames with a brick figure	363	0.45	0.44	0.22	0.64	–
OCR characters per frame	363	22.6	16.6	5.0	45.9	–

Note. Views, likes, and comments as displayed at collection. Scene-level records are from the scene-coding layer; words and speakers from the Whisper transcripts; frames, brick figures and OCR from the frame layer.

Attention and production are both concentrated, attention far more so. The median video has 2,896 views, but the mean is 83,491, and the ten most viewed videos account for 39 percent of all views. The ten most active channels posted 35 percent of the videos, the largest ("CNN IRAN LEGO") posted 41, and 76 channels posted only one.

Because re-uploads were coded individually, we also identify content clusters—copies of the same animation—by comparing the overlap of four-word sequences in their transcribed speech, lyrics and on-screen text. The 363 videos correspond to 331 distinct animations; 15 animations were posted more than once, the most widely copied eight times. As a check, we also report the main H1a estimate with each animation weighted equally.

Figure 1 shows what the videos look like. It reproduces one still from each of fifteen distinct animations, five from each provenance group described in Section 4.4, ordered by upload date. Nearly all are rendered by generative video models as glossy three-dimensional brick figures, lit, and framed like feature animation. The settings range from the Oval Office and Pentagon briefing rooms to Iranian command centers, Gulf seascapes and American kitchens. Formats range from rap diss tracks and heavy-metal songs to mock news bulletins and short dramas. Recurring set pieces include caricatures of Trump and members of his cabinet, flag-draped coffins, oil-price boards and household bills, Iranian missiles, and Iranian civilians amid ruins.



(a) *Iran Trolls Trump About Epstein*
Shizzil's Shorts
11 Mar 2026 · 4k views

(b) *The Strait of Hormuz Will Remain Closed*
Noman Johan
22 Mar 2026 · 8k views

(c) *Come Closer*
A Life in Progress
30 Mar 2026 · 997k views

(d) *Until the Blood Is Out*
Rock ID Official
3 Apr 2026 · 818k views

(e) *Another LEGO Animation Mocking Trump*
APT
9 Apr 2026 · 2.6M views

(A) Regime-linked producers: 51 videos, 33 channels, 11.6M views

(f) *Specially for Pete Hegseth*
Al IRAN
2 Apr 2026 · 1.2M views

(g) *Minor Damage Only*
A Life in Progress
6 Apr 2026 · 131k views

(h) *We Love Americans, But Not Their Leaders*
Sophmara
22 Apr 2026 · 985k views

(i) *Why Are You Paying So Much in Taxes?*
Creedlow
22 Apr 2026 · 169k views

(j) *Goodbye T.A.V. (Metal Version)*
NadaNita
12 May 2026 · 117k views

(B) Iran-aligned accounts: 65 videos, 40 channels, 8.7M views

(k) *Iran Just Answered USA Blockade in the Gulf*
EmpireArmsLego
13 Apr 2026 · 265k views

(l) *Trump as Jesus?*
Sophmara
15 Apr 2026 · 738k views

(m) *Kash Patel Passed Out Drunk?!*
BNN
18 Apr 2026 · 770k views

(n) *LEGO Rap Exposes Trump's Lies About Obama*
Prose in Plastic
11 May 2026 · 694k views

(o) *$29 Billion & "I Don't Think About You"*
Lego Records
14 May 2026 · 198k views

(C) Trend adopters: 247 videos, 95 channels, 9.8M views

Figure 1: Selected stills from fifteen distinct animations in the corpus, by provenance (rows) and upload date (left to right). Titles are shortened from the upload title or on-screen title; channel is the uploading account. Regime-linked videos are mostly re-uploads of animations credited to Explosive Media (*Akhbar-e Enfejari*) or other Iranian producers.

4.2 Measurement: the legopolitics pipeline

We converted each video into structured data with `legopolitics`, an open-source Python package for the multimodal analysis of brick-built political video (version 0.1.1, available on PyPI under an MIT license) (Caliskan, 2026). The package separates the analysis into linked units—videos, scenes, frames, figures, audio, and interpretation—and keeps automated visual observation distinct from political interpretation. Three layers of measurement are used in this paper. Figure 2 shows how their tables link, and Appendix B lists every variable.

The first is an **audio and text layer**. Speech and lyrics were transcribed with Whisper large-v3-turbo after voice-activity detection, segments were diarized by speaker, and non-speech audio was tagged in five-second windows with an AudioSet classifier. The transcripts contain 126,767 words from 358 videos, a median of 272 words per video, with an average of 1.9 distinct voices. Fifty-nine videos, mostly instrumental or carried by on-screen text, have fewer than 20 transcribed words. During coding, every transcript was checked against the on-screen captions and corrected where the two disagreed (Whisper renders "P. Hegseth" as "Pete Hexeth", for example; Appendix C).

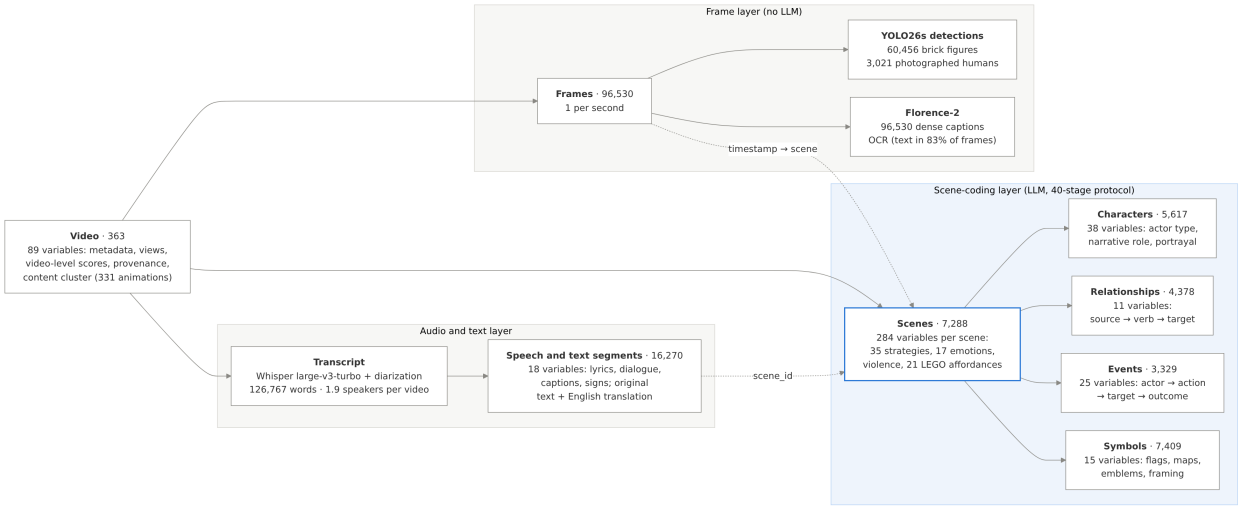

Figure 2: The data model. Boxes are tables, with the number of records in the 2026 war sample and the number of variables in the codebook. Solid arrows link tables by video and scene identifiers; dotted arrows show how text segments and frames are assigned to scenes by timestamp. The frame layer is processed separately from the scene-coding LLM; Florence-2 is itself a vision-language model.

On-screen text was transcribed during coding and, independently, by optical character recognition. The resulting text layer has 16,270 segments: 7,568 spoken or sung lines and 8,702 pieces of on-screen or other text, chiefly written labels (3,321), subtitles (1,424), signs (1,115), banners (656), titles (590) and logos (425). Ninety-seven percent of the segments are in English, 170 in Persian, and 46 in Arabic, each stored with an English translation. Ninety-five percent of the videos are predominantly in English; music is present in 99 percent of coded runtime, and rap or hip-hop accounts for 73 percent of it.

The second is a **scene-coding layer**. Each video was segmented into narrative scenes—coherent units of setting, actors, and action rather than camera shots—and each scene was coded on a 40-stage protocol covering actors and their narrative roles, directed relationships, events, 35 propaganda strategies, 12 protagonist and 12 antagonist attributes, 17 emotions, violence, and its graphic realism, symbols, political topics, cinematic style, and 21 medium-specific affordances of the LEGO register. Coding was performed by a large language model (Claude Opus 5.5) running unattended under a fixed protocol: it viewed every contact sheet (one frame per second for videos up to ten minutes), read the full transcript and the audio tags, and was instructed to separate raw observation from interpretation, record confidence for every identification, and treat absence (0) and indeterminacy (missing) as distinct. The coder saw sampled frames and transcripts, not continuous video, and was given the project's research questions but not the hypotheses tested here (Appendix D). A deterministic quality-control step rejected any coding that left timeline gaps, overlapping scenes, or references to undefined actors and symbols. The codebook defines 570 variables in 18 linked tables, 284 of them at the scene level (Appendix B). Most interpretive codes are ordinal scores from 0 (absent) to 3 (dominant), 127 variables in all; 73 are binary indicators and 25 are counts. Table 2

summarizes the resulting data. The 363 videos yield 7,288 scenes (7,054 substantive), 5,617 character records, 3,329 events, 4,378 directed actor relationships, 7,409 symbol records and 16,270 transcribed speech and text segments. The average video has 20 scenes and an estimated 80 camera shots, a cut every 3.3 seconds (Table 1).

The third is an **independent frame layer**, processed separately from the scene-coding LLM. All 96,530 frames sampled at one per second from the 363 videos were passed through a custom YOLO26s detector trained to distinguish brick figures from photographed humans, and through Florence-2 for dense captions and OCR. At a confidence threshold of 0.5 the detector found 60,456 brick figures and 3,021 photographed humans. Brick figures are detected in 45 percent of frames and photographed humans in 3 percent, which confirms how little live-action footage the videos contain. Florence-2 produced a caption for every frame, and its OCR found text in 83 percent of frames. We use this layer to validate the scene codes (Section 5.5) and to provide coder-independent evidence for H2a.

Table 2: The LEGO corpus at each unit of analysis (2026 war sample)

Unit	Count	Variables	Source layer
Candidate videos screened	549	15	Triage
Videos coded (2026 war sample)	363	89	Scene coding
Distinct animations (content clusters)	331	–	Text overlap
Channels or accounts	136	–	Metadata
Scenes (substantive)	7,288 (7,054)	284	Scene coding
Character records	5,617	38	Scene coding
Directed actor relationships	4,378	11	Scene coding
Events	3,329	25	Scene coding
Symbol records	7,409	15	Scene coding
Speech and on-screen text segments	16,270	18	Audio and text
Frames (1 per second)	96,530	19	Frame layer (YOLO26s, Florence-2)
Brick-figure and human detections (confidence ≥ 0.5)	63,477	37	Frame layer (YOLO26s)

Note. Variables are counted from the project codebook for the scene-coding tables and from the output files for the triage and frame layers (Appendix B).

4.3 Anatomy of a coded video

Figure 3 shows one video in all three layers. "Lego video from IRAN Specially for Pete Hegseth" was posted by the Iran-aligned channel AI IRAN on 2 April 2026 (panel (f) of Figure 1). It had 1.2 million views, 41,799 likes and 4,100 comments when collected, the most of any Iran-aligned video. It is a 131-second rap diss track. The coder divided it into 17 scenes and recorded 11 characters, 13 directed relationships, 10 events, 28 symbols, and 29 speech and text segments. The frame layer adds 130 frames with their detections, captions and OCR.

The timeline shows the video's two halves. The first 70 seconds ridicule and demonize the U.S. defense secretary: scenes 3 to 8 score 3 on adversary delegitimation and at most 2 on self-legitimation,

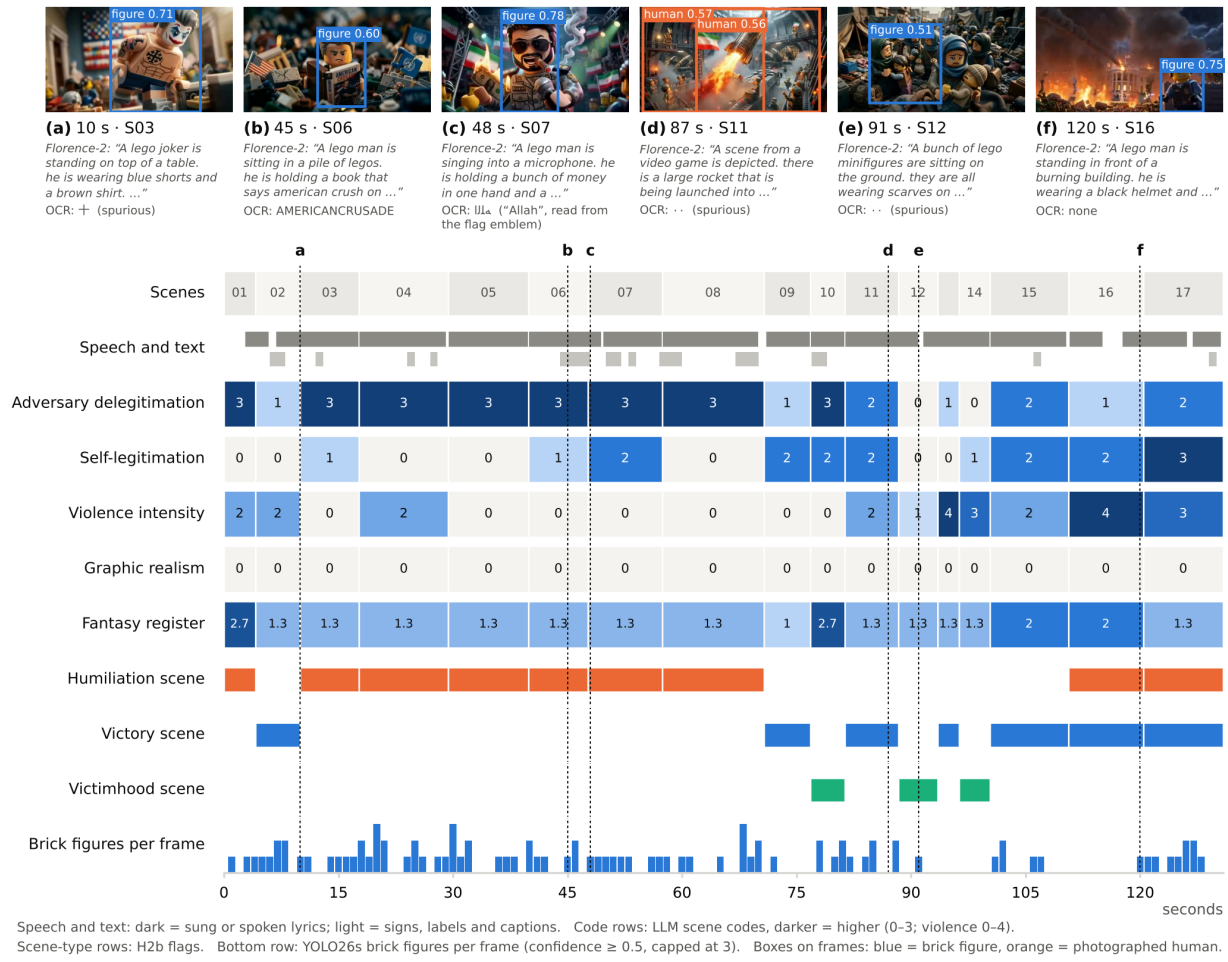

Figure 3: One video across the three layers: "Lego video from IRAN Specially for Pete Hegseth" (AI IRAN, 2 April 2026). Top: six frames from the frame layer with YOLO26s detections at confidence ≥ 0.5, the Florence-2 caption (truncated) and OCR output. Bottom: the scene-coding layer, scene by scene; the speech and text layer; and the number of brick figures detected in each frame. Letters mark the frames shown above.

and only one of them contains any violence. The second half turns to Iranian power. Self-legitimation reaches 2 or 3 in most scenes, and violence intensity reaches 4 as missiles strike the White House (frame f). Three short victimhood scenes interrupt the second half, among them Iranian civilians in a displacement camp (frame e) and Tehran under air attack. Graphic realism is 0 in every scene, including the four scenes coded at violence intensity 3 or 4.

The frame layer reads the same video independently, and imperfectly. YOLO26s finds brick figures in most of the dialogue scenes and few in the bombardment scenes, which are dominated by aircraft, missiles, and fire. OCR reads the title of the book in frame b ("AMERICAN CRUSADE") and the word "Allah" from the emblem of an Iranian flag in frame c. It also returns spurious characters, Florence-2 describes the Iranian flag in frame d as Indian, and the detector mistakes the missile launch in the same frame for two photographed humans. Errors of this kind are why we use the frame

layer for aggregate validation rather than to measure interpretive content. Appendix C reproduces the underlying records for this video: the video summary, a scene, a character, a relationship, an event, a symbol, a text segment beside the raw transcript, and a frame with its detections.

4.4 Who made the videos?

Attribution matters for interpreting propaganda, and it is rarely clean. We coded provenance into three categories (Table 3) from the scene coder's free-text producer notes, which record the watermarks, credits, end cards, channel descriptions, and on-screen handles seen in each video. Keyword rules assign the category, and 18 manual overrides correct notes that negate an attribution (for example, "not identifiable as state media"). **Regime-linked** videos carry an affirmative attribution to Iranian pro-government producers (Explosive Media, also styled *Akhbar-e Enfejari*; PersiaBoi; Afand) or to Iranian state outlets and accounts (Tasnim, IRIB, an Iranian embassy). Most are re-uploads by third parties, since YouTube terminated Explosive Media's channel on 27 March 2026 (Agence France-Presse, 2026). **Iran-aligned independent** videos come from channels that adopt a pro-Iranian voice or brand without an attributable regime-linked producer (for example "AI IRAN", which solicits cryptocurrency donations). **Trend adopters** are commercial, satirical or content-farm channels with no Iranian alignment cue, many of which use "Iran LEGO" as a genre label; the largest, "CNN IRAN LEGO", posted 41 videos, many of them about U.S. domestic politics.

Table 3: Provenance of the 2026 LEGO corpus

Provenance	Videos	Channels	Views (millions)	Share of views	Median views	Median length (s)
Regime-linked	51	33	11.6	38.7%	20,441	136
Iran-aligned independent	65	40	8.7	28.8%	9,392	164
Trend adopter	247	95	9.8	32.4%	1,544	251

Channels sum to more than 136 because 29 channels posted videos in more than one category. Regime-linked videos are a seventh of the corpus but draw the largest share of views. Their median video is seen thirteen times as often as a trend adopter's, and they are re-uploaded more often: an animation with a regime-linked origin is posted about 1.6 times as many times as other animations (incidence-rate ratio 1.57 in a Poisson model with channel-clustered errors, $p = .007$).

4.5 The timing of the genre

Figure 4 shows the genre passing from regime-linked producers to everyone else within two weeks of its peak. Regime-linked videos dominate the first month of the war and peak in the week of 30 March. From mid-April, after Explosive Media's YouTube channel was terminated and the 7 April ceasefire was announced, trend adopters account for most uploads, and regime-linked uploads fall to about one per week. A second surge of trend-adopter content follows renewed U.S.–Iranian strikes

in early June. Topics track the war. Keyword measures on titles, lyrics, and on-screen text show ceasefire and negotiation themes rising from 12 percent of videos in the week of 30 March to 52 percent in the week of the 7 April ceasefire. The Strait of Hormuz and the U.S. blockade appear in about 30 percent of videos through mid-April. References to Jeffrey Epstein are most common early: they appear in all five videos posted in the first two weeks of March and in 64 percent of the eleven posted in the week of 23 March.

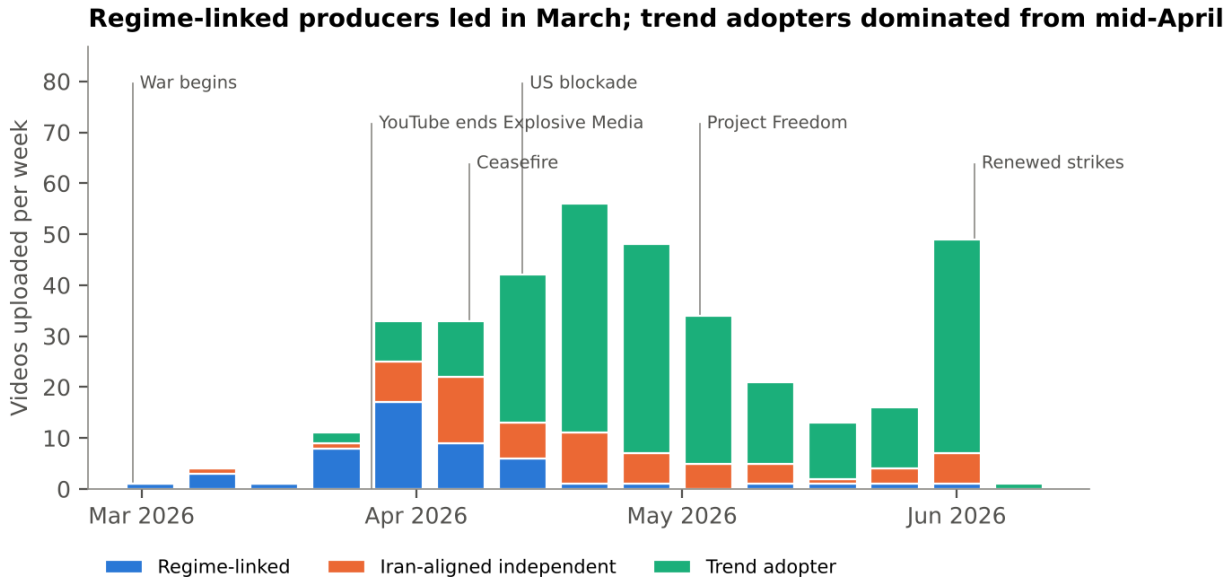

Figure 4: Weekly uploads of LEGO-style videos by provenance, March–June 2026 ($n = 363$). Vertical lines mark war events; dates from Wikipedia, "2026 Iran war ceasefire", Annahar and AFP.

4.6 What the videos contain

Three descriptive regularities frame the hypothesis tests. First, the dominant strategy is ridicule: enemy ridicule is the highest-scoring of 35 strategies (mean 0.82 on the 0–3 scale across all substantive runtime) and the single dominant strategy in 34 percent of videos, followed by victimhood (0.60), moral superiority (0.56) and enemy demonization (0.50). Taking the first emotion the coder listed as dominant, the most common are tension (21 percent of videos), contempt (19 percent) and amusement (12 percent). Second, the corpus is overwhelmingly synthetic: the coder judged generative-AI production central in 358 of 363 videos (99 percent), and the toy aesthetic is clearly present or central in 99 percent of substantive scenes. Third, the genre is musical and narrative rather than documentary. Videos average 20 scenes, with a median scene length of ten seconds, and narration is present in only 16 percent of runtime, because most meaning travels through lyrics, characters, and action. Like music videos, they repeat themselves: 42 percent reuse footage shown earlier in the same video, typically under a chorus, and repeated footage makes up 12 percent of runtime on average.

The character, relationship, event, and symbol tables show who and what populates these stories. Among the 5,617 character records, civilians are the most common actor type (852), followed by soldiers (725), political leaders (644) and government officials (587). Where the coder identified a nationality, Americans outnumber Iranians (607 to 399), and Israelis are rare (49). Trump himself appears as a character in 74 percent of videos, Netanyahu in 22 percent, and Ali or Mojtaba Khamenei in 8 percent. The most frequent directed relationships are attacks (517), condemns (447), threatens (409), victimizes (355) and mocks (325). The most frequent event actions are attacking (219), launching a missile (180), destroying (171), threatening (163) and bombing (152). The symbol table is dominated by national flags and by objects that stand for American politics and the costs of war. A U.S. flag appears in 70 percent of videos, an Iranian flag in 54 percent and an Israeli flag in 26 percent. Money or dollar amounts appear in 39 percent, oil or fuel in 30 percent, the White House in 25 percent, references to Epstein in 18 percent, coffins in 14 percent and MAGA caps in 13 percent.

Violence is common but rarely bloody. Twenty-two percent of substantive scenes contain violence, 11 percent an explosion and 7 percent a missile attack, but only 3 percent depict death and 2 percent injury. When violence appears, the coder most often framed it as tragic (348 scenes), villainous (315) or spectacular (305), and less often as comic (204), retaliatory (191) or heroic (149).

Provenance shapes these profiles. Military superiority is the most common dominant strategy among regime-linked videos (27 percent), followed by enemy ridicule (24 percent). Among Iran-aligned independents (40 percent) and trend adopters (34 percent), enemy ridicule leads. Regime-linked videos also devote far more runtime to violence (56 percent, against 16 percent for trend adopters).

4.7 The comparison corpus: Iranian English-language state media

The comparison corpus contains 506 English-language articles from five Iranian state or state-aligned outlets: Press TV (109), Tehran Times (113), the Islamic Republic News Agency (IRNA, 88), Mehr News (130) and Iran Press (66). Articles were discovered from publisher sitemaps and homepages and parsed to full text, 160,183 words in all. Each record stores the canonical URL, title, lead, body, publication time, section, tags, language checks, and a hash of the source page (Appendix C). The collection is a pilot, and its timing does not match the videos. It clusters on the first days of the June 2025 war (135 articles), the first day of the 2026 war (100 articles from 28 February), and September 2026 (213 articles). It contains only seven articles from March to June 2026, when the videos were posted. We therefore treat it as a baseline for the register of Iranian state communication in English during wartime, not as a time-matched control. Section 6.3 shows that the contrasts hold when the comparison is restricted to the 80 war-related articles published during the 2026 war, 74 of them on its first day. After keeping war-related articles of at least 30 words, 347 articles remain (Mehr 100, Press TV 78, Tehran Times 73, IRNA 64, Iran Press 32), with a median length of 236 words; 332 of them have a parsed body of more than 200 characters and at least three war-keyword matches, the filter used when the articles are split into sentences for the register comparison (Section 5.4).

5 Empirical Methods

We estimate linear probability and logit models for binary outcomes, linear models with fixed effects for within-unit comparisons, and count models for count outcomes. Standard errors are clustered by channel, video, or outlet, as specified for each analysis. For each hypothesis, we define the unit of analysis, estimating equation, statistical test, and robustness checks. We report two-sided p-values and 95 percent confidence intervals. A result supports a directional hypothesis when the estimate has the predicted sign and its confidence interval excludes the relevant null value. Section 5.1 presents the shared definitions and inferential procedures; Sections 5.2–5.4 describe the hypothesis-specific analyses; and Section 5.5 compares scene codes with measurements from independently processed frames.

Figure 5 summarizes how the two corpora are transformed into linked measurements and hypothesis-specific comparisons.

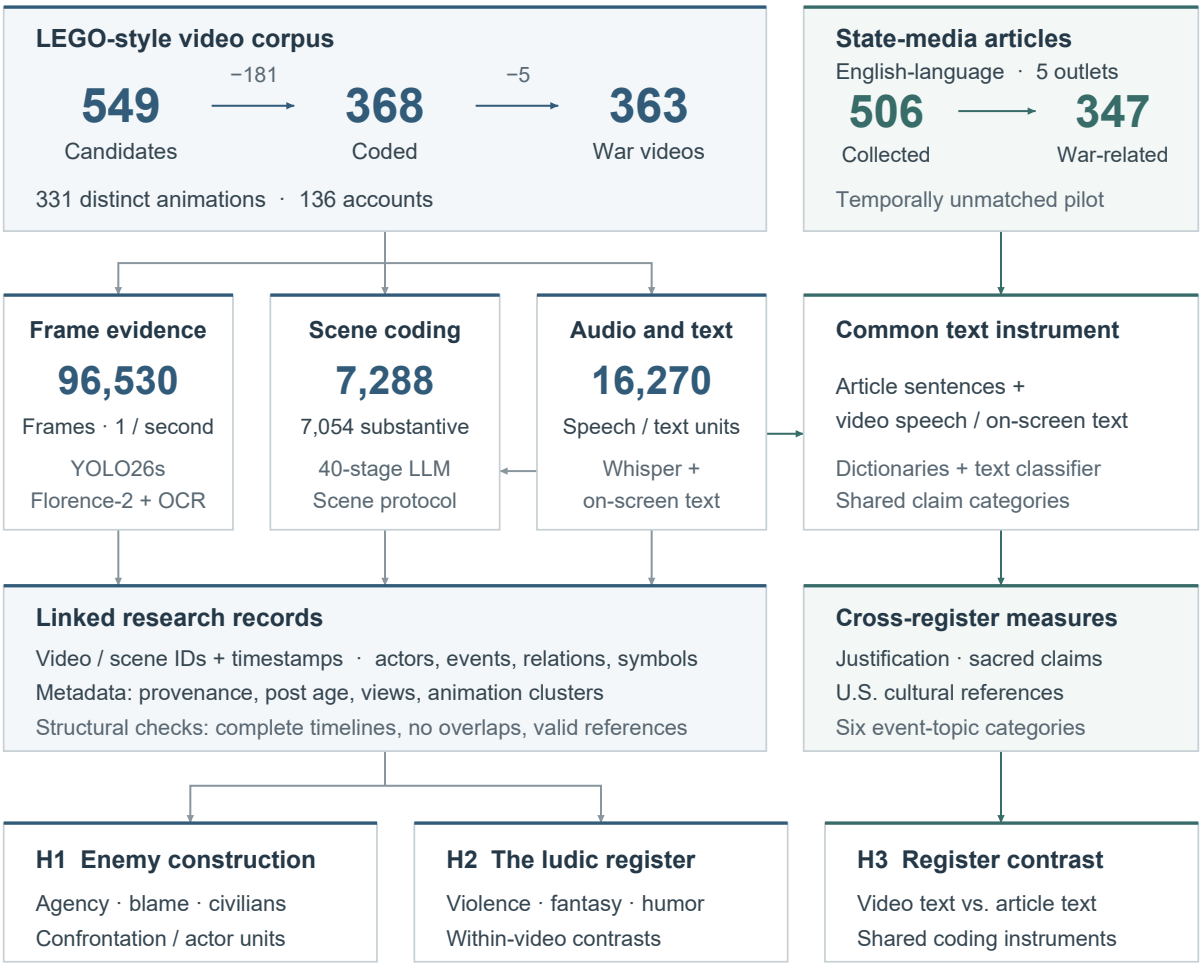

Figure 5: Study pipeline linking multimodal video data and state-media text to hypothesis tests. Arrows indicate processing and analysis dependencies.

5.1 Preliminaries common to all tests

The scene coder recorded a country or identity for every character, event, and symbol. We map these into three sides: *Iran* (Iran, Iranian, the IRGC and Iranian leaders, soldiers and civilians); *adversary* (the United States and Israel, their leaders, militaries, institutions, and publics); and *other* (third countries, international organizations, unclear). Inferred and implied attributions ("USA (implied)") are kept with a flag, and every test is repeated without them. Within the adversary side we distinguish the United States from Israel, and within each country we distinguish tiers of actor: *named leaders* (political leaders and government officials identified by name), *unnamed leaders or officials*, the *state or government* as an institution ("the Pentagon", "the administration"), the *military* (soldiers, generals, the armed forces, and security institutions), *civilians* (civilians, workers, families, children, journalists, and protesters), and the *undifferentiated public or nation*, which covers actors recorded as "the United States", "America", "Americans" or a national symbolic figure such as Uncle Sam without reference to a leader, institution or individual.

Videos posted by the same channel share producers, templates, and audiences, and content clusters contain copies of the same animation. Unless stated otherwise, standard errors are clustered by channel ($G = 136$). Where a test is restricted to a provenance group, the number of clusters falls to 33–95, and we report wild-cluster bootstrap p-values (Rademacher weights, 9,999 draws) alongside the analytic ones (Cameron et al., 2008). Every estimate is also reported with each video weighted by the inverse of its content-cluster size, so that an animation re-uploaded eight times counts once.

Each model is estimated on the full 2026 sample and separately within the three provenance groups (regime-linked, Iran-aligned independent, trend adopter). Because the hypotheses concern pro-Iranian narratives, the Iran-aligned, and regime-linked groups are the theoretically central samples; the trend-adopter group shows whether a pattern survives diffusion to producers without an Iranian alignment cue. All outcome variables are language-model codes.

5.2 H1: Agency, responsibility, and vulnerability

5.2.1 H1a: Representational dominance

We examine directed confrontations between Iranian and adversary actors. The main sample comes from relationships coded as `attacks`, `defeats`, `threatens`, `victimizes`, `commands`, `protects`, `rescues` or `mocks`, each of which assigns agency to a source actor. Relationships coded as `competes_with`, `cooperates_with`, `represents` or `condemns` enter only the broader robustness specification. Events with an Iranian actor and an adversary target, or the reverse, provide a separate sample.

For confrontation i in video v, $\text{AGENT}_{iv} = 1$ when the Iranian actor is the source and 0 when the adversary actor is. The baseline linear probability model is

$$\text{AGENT}_{iv} = \pi + \varepsilon_{iv}, \tag{1}$$

where $\hat{\pi}$ estimates the Iranian agency share. H1a predicts $\pi > 0.5$, assessed with a Wald test against parity and standard errors clustered by channel. We then add verb fixed effects $\lambda_{k(i)}$, an indicator for violence, the scene's position in the video's runtime, and provenance indicators:

$$\text{AGENT}_{iv} = \pi + \lambda_{k(i)} + \beta_1 \text{Violent}_{iv} + \beta_2 \text{Position}_{iv} + \boldsymbol{\gamma}' \text{Prov}_v + \varepsilon_{iv}. \tag{2}$$

A fractional logit relates each video's Iranian agency share, s_v, to provenance, providing predicted shares for each producer group.

To assess sensitivity to the definition of a confrontation, we repeat the test using only relationships that resolve one (`defeats` and `rescues`) and events ending in destruction, surrender, flight or capture. A broader specification includes every directed relationship between the two sides. Removing `mocks` separately tests whether the pattern depends on ridicule. We also weight confrontations by the product of the participants' screen-prominence scores, each ranging from 1 to 3. To assess dependence on re-uploads, we randomly select one video per content cluster and average the estimate across 1,000 selections.

5.2.2 H1b: Personalized blame

For H1b, we identify attributions of responsibility for U.S. aggression. These include relationships with a U.S. source coded as `attacks`, `threatens`, `victimizes` or `commands` an attack; events in which a U.S. actor attacks, bombs, launches a missile, destroys or threatens a target; and events whose U.S. actor is coded as the cause of the conflict. We distinguish five categories of responsible actor: named leaders, including Trump, Vance, Hegseth and Rubio; unnamed leaders or officials; the state or government; the military; and the undifferentiated public or nation.

The main test compares named leaders with the public. Among eligible attributions j in video v, $\text{LEADER}_{jv} = 1$ for a named leader and 0 for the public or nation. We estimate

$$\text{LEADER}_{jv} = \pi + \varepsilon_{jv}, \tag{3}$$

with H1b predicting $\pi > 0.5$. To account for differences in the kinds of acts attributed to leaders and publics, the conditional model adds act-type fixed effects $\lambda_{k(j)}$, violence intensity and provenance:

$$\text{LEADER}_{jv} = \pi + \lambda_{k(j)} + \beta_1 \text{Violence}_{jv} + \boldsymbol{\gamma}' \text{Prov}_v + \varepsilon_{jv}. \tag{4}$$

A multinomial logit estimates the attribution share for each of the five actor categories. We also model the event coder's perpetrator-named flag, NAMED_{jv}, with the same covariates. This provides an alternative measure that does not rely on our actor categories.

We repeat the main comparison after removing Trump to distinguish a broader emphasis on leaders from personalization in one figure. We then examine speech and on-screen text segments coded as enemy demonization, enemy ridicule, accusation, blame or threat and directed at a U.S. target. Targets are assigned to the same five categories using the linked character record, and

equation (3) is re-estimated at the segment level. This checks whether the pattern also appears in the videos' verbal and written content.

5.2.3 H1c: Selective civilian exculpation

The H1c analysis separates the decision to show civilians from their portrayal once shown. We first construct a video–nationality panel with 363×2 observations, one each for American and Israeli civilians in every video. We then compare civilian character records, including workers, family members, children, journalists, and protesters of either nationality.

In the first stage, $\text{REP}_{vn} = 1$ if video v contains at least one civilian of nationality n. In the second, $\text{EXC}_{cv} = 1$ when civilian c has a non-culpable role (victim, protagonist, ally, protector, rescuer or martyr) and a sympathetic, positive or neutral portrayal. It equals 0 for antagonists, foreign or domestic enemies and villains, or for characters portrayed negatively, ridiculed or demonized. Unclear or ambiguous records are excluded from the main test and assigned 0 in a sensitivity check. A subsidiary outcome, HARM_{cv}, records whether a civilian is targeted by a harmful relationship (`victimizes`, `attacks` or `threatens`) originating from a leader, official or military actor of the civilian's own country.

Representation is estimated with a logit:

$$\Pr(\text{REP}_{vn} = 1) = \Lambda\big(\alpha + \delta\,\text{American}_n + \boldsymbol{\gamma}'\text{Prov}_v\big)\,, \tag{5}$$

where δ is the American–Israeli difference in the log odds of civilian representation. For represented civilians, we estimate

$$\text{EXC}_{cv} = \alpha + \beta\,\text{American}_c + \boldsymbol{\theta}'X_{cv} + \mu_v + \varepsilon_{cv}, \tag{6}$$

where X_{cv} includes screen prominence, scene count and armed status, and μ_v denotes video fixed effects. H1c predicts $\beta > 0$. The within-video estimate uses only videos containing civilians of both nationalities. We also report a pooled specification with provenance and upload-week fixed effects, and a pooled model with HARM_{cv} as the outcome. Standard errors are clustered by video for the within-video comparison and by channel for the pooled models. Given the scarcity of Israeli civilian records, we supplement these estimates with Fisher's exact test, a channel-bootstrap interval for the raw nationality difference, and the minimum detectable effect at 80 percent power.

Selection into representation may affect these comparisons. Following Manski (1990), we calculate bounds at the video–nationality level by treating unrepresented publics as either exculpated or not exculpated. The stronger version of H1c requires the American rate to exceed even the Israeli upper bound. We also estimate a two-step selection correction: a probit of representation on nationality and provenance supplies an inverse Mills ratio for equation (6), with provenance serving as the exclusion restriction.

Finally, we repeat the portrayal analysis for soldiers and for leaders and officials. The leader–people argument predicts a smaller nationality gap for soldiers and none for leaders. A gap

extending to American leaders would instead suggest a broader preference for American characters. Dropping characters whose nationality is coded as implied or inferred checks sensitivity to nationality assignment.

5.3 H2: Violence, triumph, and circulation in the ludic register

5.3.1 H2a: Selective sanitization

We compare the portrayal of victim suffering in scenes of Iranian and adversary violence. Perpetrators and targets are identified through character records or descriptions. Perpetrators are assigned to the Iranian, adversary, other or unidentified category; targets are classified as civilians, soldiers, leaders, and officials, military assets, or cities, and infrastructure. Scenes with unidentified or unseen perpetrators are excluded from the main comparison and examined separately.

For scene s in video v, the suffering index S_{sv} combines six components: graphic realism (0–3), injury and death indicators (0/1), grief and sympathy scores (each 0–3), and the victimhood score (0–3). We standardize each component, average them, and standardize the resulting index across violent scenes. Coefficients are therefore expressed in standard deviations of suffering. The same rule applies to both sides, and each component is also analyzed separately.

The main specification compares scenes within videos while adjusting for target type:

$$S_{sv} = \beta \, \text{IranPerp}_{sv} + \lambda_{t(s)} + \theta_1 \text{Intensity}_{sv} + \theta_2 \ln \text{Length}_{sv} + \mu_v + \varepsilon_{sv}, \tag{7}$$

where $\text{IranPerp}_{sv} = 1$ for Iranian perpetrators and 0 for adversary perpetrators, $\lambda_{t(s)}$ denotes target-type fixed effects, and μ_v denotes video fixed effects. The remaining covariates are violence intensity (0–4) and log scene length. H2a predicts $\beta < 0$. Identification comes from videos containing both kinds of violence; a pooled specification replaces video effects with provenance and upload-week effects. Interactions between perpetrator side and target type estimate differences for specific targets, particularly civilians. Standard errors are clustered by video in the within-video model and by channel in the pooled model.

We also use coarsened exact matching on target type, violence intensity, battle versus single strike, and the tercile of scene position in the video (Iacus et al., 2012). The matched comparison uses standard errors clustered by video. Two further sensitivity analyses assign all unidentified perpetrators to the Iranian side and then to the adversary side.

A separate measure uses the share of Florence-2 frame captions mentioning blood, wounds, bodies, crying, mourning or funerals. We substitute this share for the suffering index in equation (7). Mentions of explosions, fire or smoke provide a negative-control outcome for destruction. A suffering gap without a corresponding destruction gap would be consistent with selective sanitization rather than a difference in the scale of depicted violence.

5.3.2 H2b: Fantastical triumph

The H2b analysis uses 7,054 substantive scenes. Fantasticality, F_{sv}, averages the realism-to-fantasy code (0 = predominantly realistic, 3 = overt fantasy or impossible representation) and the fantasy-affordance code (0 = absent, 3 = central). Both concern departures from ordinary physical or causal constraints beyond the shared toy aesthetic. We exclude the exaggeration code because it overlaps with ridicule ($r = 0.53$); the retained measures have smaller correlations with humor and ridicule ($r = 0.14$ and 0.20).

Victory scenes have a narrative function of victory, punishment, retaliation or enemy defeat, or score at least 2 on inevitability of victory or military superiority. Humiliation scenes score at least 2 on enemy ridicule, antagonist ridiculousness or antagonist weakness. Since these categories overlap, both indicators enter the same model:

$$F_{sv} = \beta_V \,\text{Victory}_{sv} + \beta_H \,\text{Humiliation}_{sv} + \theta_1 \ln \text{Length}_{sv} + \theta_2 \text{Intensity}_{sv} + \mu_v + \varepsilon_{sv}, \tag{8}$$

where H2b predicts $\beta_V > 0$ and $\beta_H > 0$. Video fixed effects compare scenes within the same production, and standard errors are clustered by video. We report models with and without violence intensity to distinguish the scene-type associations from the spectacle of battle. A third indicator identifies victimhood scenes, defined by a victimization narrative function or a victimhood score of at least 2. Selective realism predicts a negative coefficient for this comparison.

We assess sensitivity to measurement by analyzing the two fantasticality components separately, estimating a conditional ordered logit for the realism-to-fantasy code, and replacing the composite score with its within-video rank. We also vary the scene indicators, using strict narrative-function definitions and broader thresholds of at least 1, and dropping contributing items in turn. Removing ridicule from the humiliation indicator checks whether the association depends on that coding item.

5.3.3 H2c: Attention association

View counts are available for 360 of the 363 videos. Each is observed once, at collection, rather than at a common post age. We adjust for the time between upload and collection using log post age and upload-week fixed effects. These adjustments account for differences in exposure time but do not reproduce views measured at a fixed age.

The humor score H_v is the duration-weighted mean of humor scores (0–3) over substantive scenes, standardized across videos. The protocol defines humor from scene content. Engagement counts were also included in the preprocessing metadata, so the coding should not be treated as blinded to views. Ridicule, amusement and entertainment orientation are secondary predictors.

The pooled benchmark is

$$\begin{aligned} \ln(\text{Views}_v + 1) = \beta H_v + \theta_1 \ln \text{Length}_v + \theta_2 \ln \text{Age}_v \\ + \theta_3 \text{Violent}_v + \theta_4 \text{SL}_v + \omega_{w(v)} + \pi_{p(v)} + \varepsilon_v. \end{aligned} \tag{9}$$

Here, Violent_v is the violent share of runtime, SL_v is the self-legitimation index, and $\omega_{w(v)}$ and $\pi_{p(v)}$ are upload-week and platform fixed effects. Provenance is added as a control, and standard errors are clustered by channel. H2c predicts $\beta > 0$. The primary specification replaces week and platform effects with channel fixed effects, comparing posts from the same account and absorbing stable account characteristics. We report the additional R^2 contributed by the content measures beyond the fixed effects.

Alternative specifications use a negative binomial model for view counts with log post age as an offset, and likes, and comments as outcomes. We also restrict the sample to the earliest dated posts within each content cluster and estimate a Poisson model for the number of re-uploads as another measure of circulation.

Finally, we replace the coded humor score with two measures produced separately: the share of five-second audio windows tagged as laughter or comedy by the AudioSet classifier, and the frequency of ridicule terms in speech and on-screen text. Comparing these estimates assesses sensitivity to how humor is measured.

5.4 H3: Register contrast with Iranian state media

H3 compares the content of LEGO-style videos with Iranian English-language state media using common measures applied to speech and on-screen text segments and article sentences. Excluding logos, watermarks, and empty segments leaves 15,845 of the videos' 16,270 text segments. The article sample contains 5,865 sentences from the titles, leads and bodies of 332 war-related articles with parsed bodies. Video segments are shorter than article sentences, with medians of 7 and 21 words. All models control for unit length, and a sensitivity analysis restricts both corpora to units of 5–40 words. Each unit retains its document's channel or outlet identifier, along with provenance or publication date.

The article corpus is not time-matched to the videos (Section 4.7). We therefore adjust for six broad event topics assigned by keyword rules: the opening U.S.–Israeli strikes; Iranian retaliation; Hormuz and the blockade; casualties and civilian harm; ceasefire and negotiations; and other or none. Topic fixed effects compare references within these categories but do not establish that the texts concern the same event or period. Units classified as other remain in the main analysis and are excluded in a sensitivity check.

The instruments distinguish three kinds of content. Political-justification claims justify, legitimate or condemn a party's conduct, including claims about deserved retaliation, aggression, rightful victory or suffering as a grievance. Sacred legitimation invokes religious authority, divine sanction, sacred history or martyrdom. U.S.-oriented references concern domestic politics, such as elections, parties, taxes, prices, and scandals, or popular culture, including music, film, television, celebrities, memes, and sports.

The dictionary classifications cover every unit and form the main specification. A language-model classifier applies the same categories to a random sample of 1,200 units, stratified equally between

video segments and article sentences. It receives the text without source identifiers under a fixed prompt. We use this sample to compare classifications, assess dictionary recall relative to the classifier, and re-estimate the register contrasts.

5.4.1 H3a: Reduced sacred legitimation

H3a restricts the comparison to political-justification claims. For claim u in document d, $\text{SACRED}_{ud} = 1$ if the claim invokes religious authority or martyrdom. We estimate

$$\Pr(\text{SACRED}_{ud} = 1) = \Lambda\big(\alpha + \beta\,\text{LEGO}_d + \lambda_{e(u)} + \theta \ln \text{Words}_u\big)\,, \tag{10}$$

where $\text{LEGO}_d = 1$ for video segments and 0 for article sentences, $\lambda_{e(u)}$ denotes event-topic fixed effects, and $\ln \text{Words}_u$ adjusts for unit length. H3a predicts $\beta < 0$. We report average marginal effects as percentage-point differences and cluster standard errors by channel for videos and outlet for articles.

Only five article outlets contribute clusters, limiting the reliability of cluster-based inference (MacKinnon and Webb, 2018). As a sensitivity analysis, we permute register labels across documents within event topic using 2,000 draws. This procedure assumes that documents are exchangeable across registers conditional on topic; it does not remove dependence among documents from the same outlet. A second specification interacts provenance with LEGO to estimate separate contrasts for regime-linked, Iran-aligned, and trend-adopter videos. These comparisons account for provenance differences without constituting a comparison of the same producer across formats.

We assess sensitivity to measurement by substituting the classifier's labels on its sample, separating religious authority from martyrdom, and removing the justification-claim filter. A restricted comparison uses the 80 articles published during the 2026 war and videos from the corresponding period. At the document level, we estimate a Poisson model of sacred-term counts with a log word-count offset, event-topic fixed effects, and standard errors clustered by channel or outlet.

5.4.2 H3b: U.S.-oriented cultural reference

H3b includes all segments and sentences because cultural references need not express a justification claim. The outcome, USREF_{ud}, equals 1 when a unit refers to U.S. domestic politics or popular culture. The model is

$$\Pr(\text{USREF}_{ud} = 1) = \Lambda\big(\alpha + \beta\,\text{LEGO}_d + \lambda_{e(u)} + \theta \ln \text{Words}_u\big)\,, \tag{11}$$

with H3b predicting $\beta > 0$. Clustering, permutation procedures and provenance interactions follow the H3a specification.

We distinguish domestic-political from popular-cultural references and divide the former into grievance-framed and other references. Grievance framing requires a reference to co-occur with a blame or victimhood function. This separates appeals about taxes, prices or scandals from reports of

congressional votes or presidential statements. The theoretical expectation concerns grievance-framed and popular-cultural references most directly.

To assess dependence on prominent figures, we remove Epstein and Trump from the outcome in turn, then exclude references to all named U.S. leaders. We also estimate the model using the classifier's labels. Sensitivity to aggregation is assessed by repeating the analysis with scenes and paragraphs and with a document-level fractional logit, where the outcome is the share of units containing a U.S.-oriented reference and the control is the document's modal event topic. The 5–40-word restriction provides a further comparison among units of similar length.

5.5 Validating the scene codes

Language-model coding of video is new, and two concerns follow: that the model sees things that are not there, and that its interpretation reflects our hypotheses. The protocol addresses the second concern in part. The coder recorded raw observation before interpretation and scored all 35 strategies symmetrically, and it was not given the hypotheses tested here. It was given the project's broad research questions, however, one of which asks whether LEGO allows intense violence with little graphic suffering. That is a reason to rely on the coder-independent frame layer wherever a hypothesis concerns violence, which the H2a robustness check does.

We address the first concern by comparing the scene codes with the independent frame layer, which was produced by different models with no access to the codes (Table 4). For concrete visual content, agreement is high. Scenes the coder marked as containing flags, missiles, explosions or soldiers are well separated from scenes it did not, using only Florence-2 captions (AUC 0.80–0.89). The coder's character counts track YOLO26s figure counts (AUC 0.86), and its counts of on-screen text elements track OCR character counts (AUC 0.82). Agreement is weak for two codes. Crowd size barely separates on caption mentions of crowds (AUC 0.59). Graphic realism separates only weakly on gore captions (AUC 0.60), because both are rare: 2 percent of scenes have any graphic realism, and 0.3 percent of frames have a gore caption. The frame layer therefore corroborates that gore is rare, not the coder's finer gradations of realism.

Table 4: Convergent validity of the scene codes against the independent frame layer

Scene code (LLM)	Frame measure (separate models)	Scenes	Spearman ρ	AUC (code > 0)
Flag count	Florence-2 caption mentions a flag	6,629	0.71	0.89
Explosion present	Caption mentions explosion, fire or smoke	7,208	0.50	0.87
Character count	YOLO26s brick figures per frame	3,085	0.50	0.86
Missile count	Caption mentions a missile or rocket	6,912	0.66	0.84
Text elements	OCR characters per frame	7,114	0.57	0.82
Soldier count	Caption mentions soldiers or uniforms	6,400	0.41	0.80
Crowd size	Caption mentions a crowd or group	7,220	0.40	0.59
Graphic realism	Caption mentions blood, wounds or corpses	7,224	0.24	0.60

Table 5: Summary of the empirical tests

Hypothesis	Unit of analysis	Outcome	Test	Robustness checks
H1a Representational dominance	Directed Iran–adversary confrontation	Iranian side is the agent	$\pi > 0.5$, eq. (1); verb FE, eq. (2)	Narrow and broad control; prominence weights, one video per cluster
H1b Personalized blame	Attribution of responsibility for U.S. aggression	Named leader vs. public; perpetrator named	$\pi > 0.5$, eq. (3); act-type FE, eq. (4); multinomial	Excluding Trump; text-layer replication
H1c Selective civilian exculpation	Video–nationality pair; civilian character	Represented; exculpated; harmed by own leaders	δ, eq. (5); $\beta > 0$ with video FE, eq. (6)	Manski bounds and selection correction; placebo on soldiers and leaders
H2a Selective sanitization	Violent scene	Suffering index	$\beta < 0$ on Iranian perpetrator, eq. (7)	CEM and bracketing; caption-based suffering with destruction as negative control
H2b Fantastical triumph	Substantive scene	Fantasticality	$\beta_V, \beta_H > 0$ within video, eq. (8)	Components and ordered logit; flag definitions
H2c Attention association	Video	ln views at collection	$\beta > 0$ on humor with channel FE, eq. (9)	Negative binomial, likes, first uploads; AudioSet and lexicon humor
H3a Reduced sacred legitimation	Justification claim (segment; sentence)	Religious authority or martyrdom	$\beta < 0$ on LEGO, eq. (10); permutation inference	Classifier sample, split categories, no filter; 2026 articles, document-level rates
H3b U.S.-oriented cultural reference	Narrative unit (segment; sentence)	U.S. domestic politics or popular culture	$\beta > 0$ on LEGO, eq. (11)	Dropping Epstein, Trump, named leaders; scene and paragraph grain, length-matched units

6 Empirical Results

6.1 H1: Agency, responsibility, and vulnerability

Figure 6 summarizes the three tests. The videos give Iran control over confrontations (H1a, supported), and they blame U.S. aggression on named leaders rather than on the American public (H1b, supported as stated, but the personalization is almost entirely Trump). They do not, however, exculpate American civilians more than Israeli civilians once civilians appear (H1c, not supported); the selectivity lies elsewhere, in who is shown at all and in how soldiers are cast.

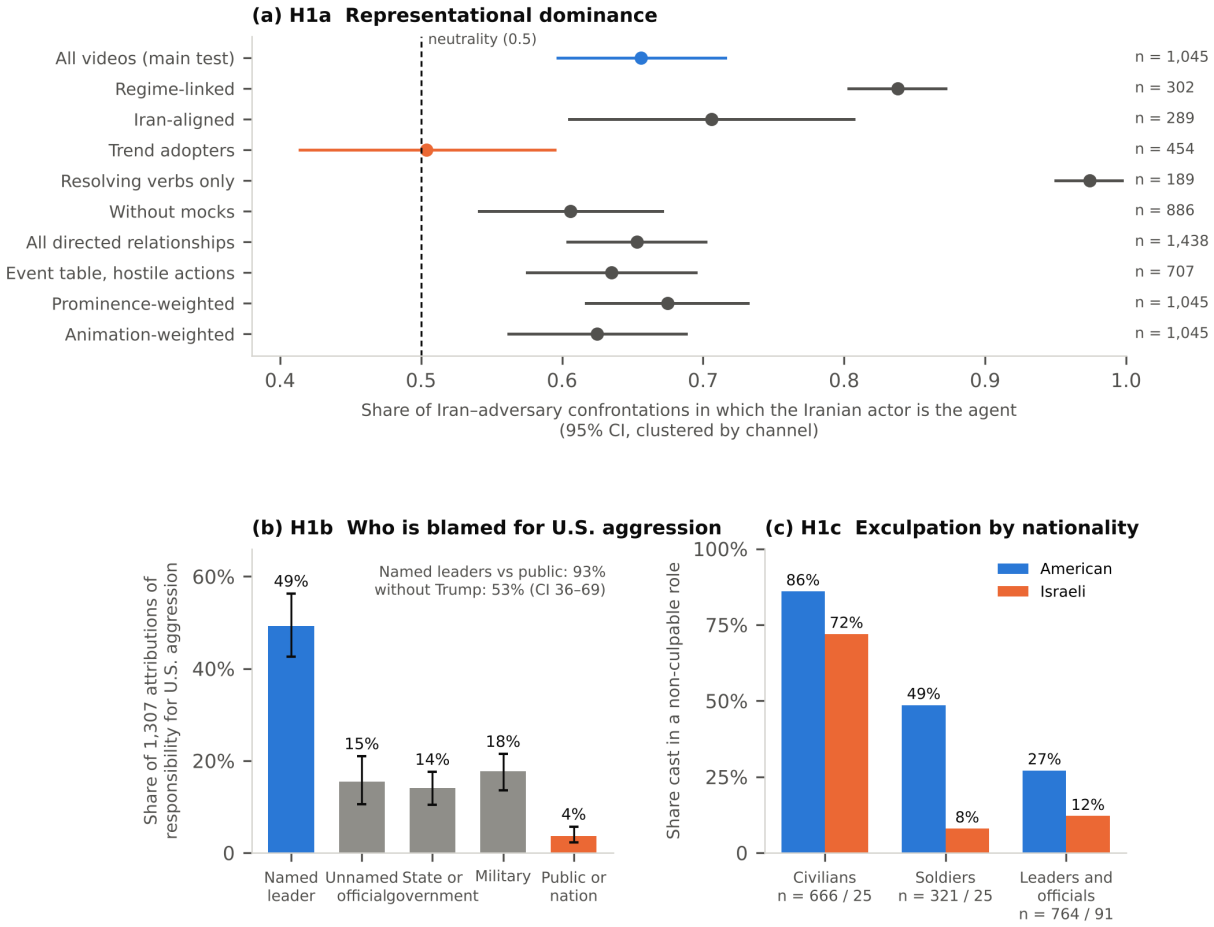

Figure 6: Tests of the H1 hypotheses. (a) Share of directed Iran–adversary confrontations in which the Iranian actor is the agent, with 95 percent confidence intervals clustered by channel; the dashed line is representational neutrality. (b) Share of 1,307 attributions of responsibility for U.S. aggression by the tier of the U.S. actor blamed, with channel-cluster bootstrap intervals. (c) Share of American and Israeli characters cast in a non-culpable role, for civilians (the H1c test) and for soldiers and leaders (the placebo tiers); n is American / Israeli character records.

6.1.1 H1a: Representational dominance

Across the 2026 sample, 1,045 directed relationships place an Iranian-side actor against an adversary-side actor with a verb that assigns control of the outcome. In 65.6 percent of them the Iranian actor is the agent (95 percent CI 59.7–71.6; $p < .001$ against the neutral value of 0.5, wild-cluster bootstrap $p < .001$; Table 6). The share is stable across specifications. With verb, violence, scene-position, and provenance controls (equation (2)) the predicted share is unchanged at 0.656; scene position is the one covariate with a large coefficient (0.31, SE 0.04): Iranian agency rises over the course of a video, so that confrontations near the end are the ones Iran wins. Weighting each animation once gives 0.625 (CI 0.562–0.688), and the mean over 1,000 draws of one video per content cluster is 0.625. The event table gives 0.635 (CI 0.575–0.695). At the level of the video, 239 videos contain at least

Table 6: H1a: share of Iran–adversary confrontations with Iranian agency

Sample or specification	n	Share	95% CI	p vs. 0.5
Main test: relationships with control verbs	1,045	0.656	0.597–0.716	$< .001$
Regime-linked	302	0.838	0.803–0.872	$< .001$
Iran-aligned independent	289	0.706	0.605–0.807	$< .001$
Trend adopter	454	0.504	0.414–0.595	.92
Identified sides only (no implied attributions)	896	0.662	0.601–0.723	$< .001$
Animation-weighted	1,045	0.625	0.562–0.688	$< .001$
Event table, hostile actions	707	0.635	0.575–0.695	$< .001$
R1 Narrow: resolving verbs (defeats, rescues)	189	0.974	0.950–0.997	$< .001$
R1 Narrow: events with a resolving outcome	322	0.745	0.677–0.814	$< .001$
R1 Broad: all directed relationships	1,438	0.653	0.604–0.702	$< .001$
R1 Main test without `mocks`	886	0.606	0.541–0.671	.001
R2 Prominence-weighted	1,045	0.675	0.617–0.732	$< .001$
R2 One video per content cluster (1,000 draws)	863	0.625	0.620–0.630	—
Iranian agency by verb (main test)				
attacks / threatens / victimizes	357 / 263 / 62	0.62 / 0.45 / 0.07		
defeats / mocks	186 / 159	0.97 / 0.94		

Note. Linear probability models with an intercept only; standard errors clustered by channel. The one-video-per-cluster row reports the mean and the 2.5th and 97.5th percentiles over 1,000 random selections.

one such confrontation, and in 56.5 percent of them Iran is the agent in a majority.

The first robustness check shows that dominance is strongest where it matters most. Restricted to verbs that resolve a confrontation (`defeats`, `rescues`), the Iranian share is 97.4 percent (CI 95.0–99.7, $n = 189$); events whose coded outcome records destruction, surrender, flight or capture give 74.5 percent (CI 67.7–81.4). Broadening to every directed relationship between the two sides leaves the share at 65.3 percent, and dropping `mocks` lowers it to 60.6 percent (CI 54.1–67.1), still well above neutrality. The verb-specific shares in Table 6 show the structure: Iran is the agent of 62 percent of attacks and 97 percent of defeats, the adversary of 94 percent of victimizations and a slight majority of threats. The videos let the adversary strike first and threaten, and give Iran the last word. The second check, weighting by the screen prominence of the two actors, raises the share to 67.5 percent.

Provenance moderates the result sharply. Iranian agency is 83.8 percent in regime-linked videos (CI 80.3–87.2) and 70.6 percent in Iran-aligned videos (CI 60.5–80.7), but 50.4 percent among trend adopters (CI 41.4–59.5; wild-bootstrap $p = .94$). The fractional logit at the video level gives the same ordering (0.84, 0.71, 0.50). Representational dominance is a property of the pro-Iranian core of the genre, not of the genre as such; commercial and satirical imitators depict the two sides as equally capable of winning.

6.1.2 H1b: Personalized blame

We identify 1,307 attributions of responsibility for U.S. aggression with a U.S.-side agent in one of five tiers. Named leaders receive 49 percent of them (bootstrap CI 43–56), the military 18 percent,

unnamed officials 15 percent, the state or government 14 percent, and the undifferentiated public or nation 4 percent (CI 2–6). In the direct test, among the 691 attributions whose agent is either a named leader or the public, 93.1 percent name a leader (CI 89.6–96.5; $p < .001$, wild-bootstrap $p < .001$). The share is 93.6, 88.1 and 94.3 percent in Iran-aligned, regime-linked and trend-adopter videos, 91.0 percent in the relationship table and 95.2 percent in the event table, and the act-type and provenance controls of equation (4) leave the predicted share at 0.931. The coder's own flag tells the same story: the perpetrator is named in 72.2 percent of U.S.-aggression events (CI 65.1–79.3). The text layer replicates the visual result: among 2,229 speech and text segments with a blaming function and a U.S.-side target that is a named leader or the public, 94.3 percent target a leader (CI 89.6–99.1).

The first robustness check changes the interpretation. Removing Trump leaves 101 attributions, and among them named leaders receive 52.5 percent (CI 36.1–68.8; $p = .77$ against parity). Trump accounts for 590 of the 643 attributions to named leaders; Vance, Hegseth, Rubio, and the rest together draw about as much blame as "America" does. H1b is therefore supported in the form stated, blame is assigned to leaders rather than to the public, but the leader-centered frame is a Trump-centered frame. What the videos avoid is not personalization in one figure, which they embrace, but any suggestion that the American nation is the aggressor: the public is cast as the perpetrator in one attribution in twenty-seven, less often than the military or the unnamed state.

6.1.3 H1c: Selective civilian exculpation

The first stage shows selectivity in representation. American civilians appear in 59.2 percent of videos and Israeli civilians in 5.8 percent (logit coefficient on American 3.17, SE 0.38); only five videos show both. The character table contains 722 American and 30 Israeli civilian records, of which 666, and 25 have a role and portrayal that can be scored. The second stage does not support the hypothesis. American civilians are cast in a non-culpable role 86.0 percent of the time and Israeli civilians 72.0 percent, a raw gap of 14 points (Fisher's exact $p = .074$; channel-bootstrap interval for the difference -0.03 to 0.38). The pooled linear probability model with prominence, scene count, provenance, and upload-week fixed effects (equation (6)) gives $\hat{\beta} = -0.007$ (SE 0.082, $p = .93$): the raw gap is explained by which channels and weeks show Israeli civilians at all. The video-fixed-effects model is identified by five videos and 17 characters and gives $\hat{\beta} = -0.16$ (SE 0.14). The subsidiary measure agrees: 13.2 percent of American civilians and 6.7 percent of Israeli civilians are shown harmed by their own leaders, state or military (pooled $\hat{\beta} = 0.044$, SE 0.065). The minimum detectable effect at 80 percent power is 20 points for the exculpation outcome, so the test can rule out a large gap but not a moderate one. Coding ambiguous characters as culpable raises the raw gap (79.4 vs. 60.0 percent, Fisher $p = .02$) but not the adjusted estimate ($\hat{\beta} = 0.011$, SE 0.10); dropping implied and inferred nationalities gives $\hat{\beta} = -0.083$ (SE 0.082).

The robustness checks sharpen the null. The Manski bounds at the video–nationality level run from 0.54 to 0.96 for American civilians and from 0.04 to 0.98 for Israeli civilians: with Israeli civilians represented in 19 videos, the unrepresented cases dominate, and no conclusion about exculpation

Table 7: H1b and H1c: main estimates and robustness checks

	n	Estimate	95% CI or SE
H1b: share of attributions naming a leader (vs. the public)			
Main test	691	0.931	0.896–0.965
Regime-linked / Iran-aligned / Trend adopter	118 / 203 / 370	0.881 / 0.936 / 0.943	
Relationships only / events only	356 / 335	0.910 / 0.952	
Conditional model, predicted share (eq. (4))	691	0.931	
Perpetrator named (coder flag), U.S.-aggression events	600	0.722	0.651–0.793
R1 Excluding Trump	101	0.525	0.361–0.688
R2 Text layer, blaming segments	2,229	0.943	0.896–0.991
H1c: exculpation of civilians, coefficient on American			
Representation, logit δ (eq. (5))	726	3.17	(0.38)
Raw rates, American / Israeli	666 / 25	0.860 / 0.720	Fisher $p = .074$
Pooled LPM, provenance and week FE	691	−0.007	(0.082)
Video fixed effects (5 videos)	17	−0.164	(0.142)
Harmed by own leaders, pooled LPM	752	0.044	(0.065)
Ambiguous coded as culpable, pooled LPM	752	0.011	(0.100)
Identified nationality only, pooled LPM	575	−0.083	(0.082)
R1 Manski bounds, American / Israeli	363 / 363	0.54–0.96 / 0.04–0.98	
R1 Selection-corrected LPM	691	0.868	(1.083)
R2 Placebo: soldiers, pooled LPM	346	0.146	(0.085)
R2 Placebo: leaders, video FE (77 videos)	242	0.081	(0.036)

Note. Standard errors clustered by channel (pooled models) or by video (fixed-effects models). Minimum detectable effect at 80 percent power for the civilian exculpation test: 0.20.

survives the assumption that producers chose whom to show. The selection-corrected estimate is imprecise ($\hat{\beta} = 0.87$, SE 1.08; inverse Mills ratio 0.65, SE 0.80), because representation is predicted by nationality far better than by provenance. The placebo tiers show where the nationality gap actually lies. American soldiers are cast as non-culpable 48.6 percent of the time and Israeli soldiers 8.0 percent (Fisher $p < .001$; pooled $\hat{\beta} = 0.146$, SE 0.085); American leaders 27.1 percent and Israeli leaders 12.1 percent (video-fixed-effects $\hat{\beta} = 0.081$, SE 0.036, $p = .027$, across 77 videos). The gap that H1c predicts for civilians appears instead for soldiers and, more weakly, for leaders. Civilians of either nationality are exculpated at high rates; it is the Israeli soldier who is cast as an enemy where the American soldier is cast as a victim or a bystander.

6.2 H2: Violence, triumph, and circulation in the ludic register

Figure 7 summarizes the three tests. Iranian-inflicted violence is shown with far less suffering than adversary-inflicted violence against the same kinds of target (H2a, supported). Scenes that humiliate the enemy are more fantastical than the other scenes of the same video, and scenes of victimhood are more realistic, but victory scenes are not more fantastical once their violence is accounted for (H2b, supported for humiliation, not for victory). Humor is associated with views across channels but not within them (H2c, not supported among comparable posts).

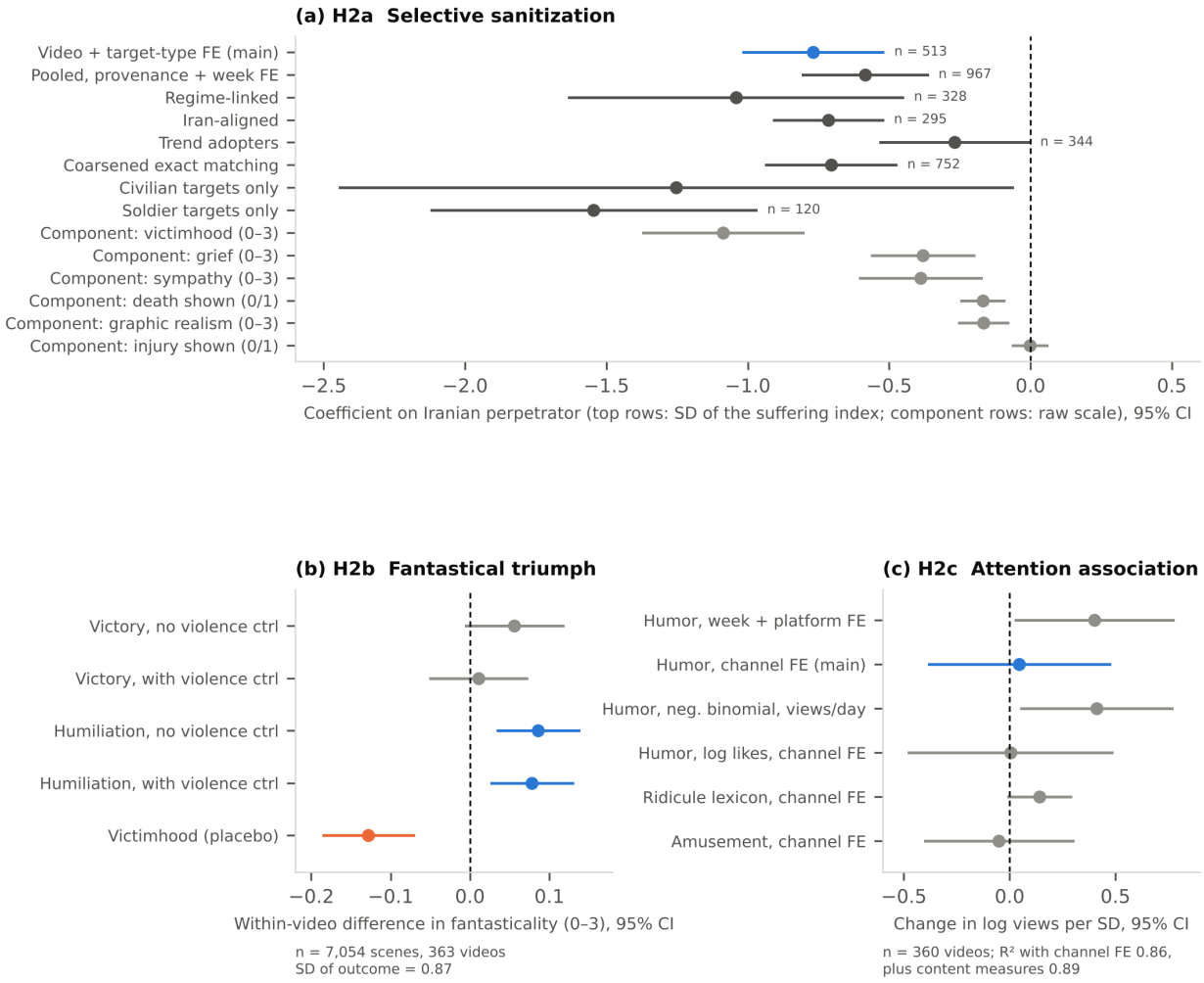

Figure 7: Tests of the H2 hypotheses. (a) Coefficient on an Iranian perpetrator in regressions of victim suffering among violent scenes with an identified perpetrator; the top rows use the standardized suffering index, the component rows its six elements on their own scales. (b) Within-video coefficients on scene type in regressions of fantasticality with video fixed effects. (c) Coefficients on standardized humor and alternative measures in circulation models; outcomes and estimators are identified by row. All intervals are 95 percent; standard errors clustered by video for fixed-effects models and by channel otherwise.

6.2.1 H2a: Selective sanitization

Of the 1,557 violent scenes, 1,556 carry a usable perpetrator code: 539 have an Iranian-side perpetrator, 428 an adversary-side perpetrator, 585 an unidentified or unseen one, and 4 both. The target types are unspecified (481), cities and infrastructure (287), military assets (263), civilians (239), soldiers (155) and leaders (131). The raw contrast is large on every component of the suffering index (Table 8). Scenes of adversary violence average 1.54 on the victimhood code, 0.54 on grief and 0.70 on sympathy, and show a death in 18.5 percent of cases; scenes of Iranian violence average 0.20, 0.05 and 0.07, and show a death in 8.9 percent. In the main model with video and target-type fixed effects (equation (7); 513 scenes in 73 videos that contain both kinds of violence), an Iranian perpetrator lowers the suffering index by 0.77 standard deviations (SE 0.13, $p < .001$). The pooled model with provenance and upload-week fixed effects gives -0.59 (SE 0.11), the animation-weighted version -0.44 (SE 0.10). The effect is present in every provenance group: -1.04 (SE 0.30) in regime-linked videos, -0.72 (SE 0.10) in Iran-aligned videos and -0.27 (SE 0.13, $p = .046$) among trend adopters.

The components show what is being withheld. Within videos, Iranian violence scores 1.09 points lower on victimhood (SE 0.14), 0.38 lower on grief and 0.39 lower on sympathy, and is 17 points less likely to show a death (SE 4); graphic realism is 0.17 points lower (SE 0.04), and injury does not differ (0.00, SE 0.03). The gap is largest where the hypothesis says it should be: among scenes with civilian targets the Iranian-perpetrator coefficient is -1.25 (SE 0.61; 153 scenes, of which only 17 have an Iranian perpetrator), and among scenes with soldier targets -1.55 (SE 0.29). For leaders, military assets and cities the gaps are small (0.18, -0.17, -0.24). The imbalance in the civilian cell is itself a finding: Iranian violence is rarely aimed at civilians at all (17 of 539 scenes, against 136 of 428 for the adversary), so sanitization operates first through target selection and second through the treatment of the few civilian targets that remain.

The first robustness check supports the regression estimates. Coarsened exact matching on target type, violence intensity, battle versus single strike, and position in the video retains 752 scenes in 58 strata and gives -0.71 (SE 0.12). Assigning the 585 unidentified perpetrators to the Iranian side gives -0.36 (SE 0.08), and assigning them to the adversary -0.39 (SE 0.08), so the estimate is bounded well away from zero whatever the unseen perpetrators are. The second check, the coder-independent caption measure, is uninformative rather than contradictory: Florence-2 captions mention blood, wounds, bodies or mourning in 0.6 percent of frames from adversary-violence scenes and 0.2 percent from Iranian-violence scenes (pooled coefficient -0.003, SE 0.002, $p = .11$). Gore captions are too rare in a corpus without gore to separate the two sides. The negative control behaves as predicted: captions mention explosions, fire or smoke equally often (coefficient 0.02, SE 0.03), so Iranian violence is not shown as smaller, only as less painful. The caption check therefore corroborates the absence of bodies on both sides; the asymmetry in suffering rests on the coder's grief, sympathy, victimhood, and death codes.

Table 8: H2a: victim suffering by perpetrator among violent scenes

	n	Iranian perpetrator	SE	p
Suffering index (standardized)				
Main: video and target-type FE	513	−0.769	0.126	< .001
Pooled: provenance and week FE	967	−0.585	0.112	< .001
Animation-weighted	967	−0.438	0.102	< .001
Regime-linked / Iran-aligned / Trend adopter	328 / 295 / 344	−1.042 / −0.715 / −0.268	0.30 / 0.10 / 0.13	
By target type (pooled): civilians / soldiers / leaders	153 / 120 / 75	−1.253 / −1.545 / 0.184	0.61 / 0.29 / 0.20	
military assets / cities and infrastructure	205 / 183	−0.174 / −0.239	0.10 / 0.08	
R1 Coarsened exact matching (58 strata)	752	−0.705	0.117	< .001
R1 Unidentified perpetrators as Iranian / as adversary	1,552	−0.363 / −0.389	0.075 / 0.077	< .001
Components, video and target-type FE (raw scales)				
Victimhood (0–3)	513	−1.087	0.144	< .001
Grief (0–3)	513	−0.381	0.092	< .001
Sympathy (0–3)	513	−0.389	0.109	< .001
Death shown (0/1)	513	−0.169	0.038	< .001
Graphic realism (0–3)	513	−0.167	0.044	< .001
Injury shown (0/1)	513	−0.002	0.031	.94
R2 Frame-caption measures (pooled)				
Share of frames with a gore or mourning caption	967	−0.003	0.002	.11
Share of frames with an explosion caption (negative control)	967	0.024	0.026	.36

Note. All models control for violence intensity and log scene length. Raw means of the index: 0.48 for adversary violence and −0.39 for Iranian violence.

6.2.2 H2b: Fantastical triumph

Across 7,054 substantive scenes, 18 percent are victory scenes, 33 percent humiliation scenes, and 18 percent victimhood scenes; the fantasticality score has a standard deviation of 0.87. Within videos, humiliation scenes score 0.086 higher on fantasticality than other scenes (SE 0.026, $p < .001$) and 0.078 higher once violence intensity is held constant (SE 0.026, $p = .002$). Victimhood scenes score 0.092 lower (SE 0.028), and 0.128 lower with the violence control (SE 0.029, $p < .001$); the internal placebo has the sign the theory of selective realism predicts. Victory scenes are 0.056 higher without the violence control (SE 0.031, $p = .066$) and 0.011 higher with it (SE 0.031, $p = .72$): victories are fantastical because they are battles, not beyond that. These are within-video differences of about a tenth of a standard deviation, consistent across the robustness checks (Table 9).

The first check replaces the estimator and the outcome. Each component gives the same pattern (humiliation 0.077 and 0.079, victimhood −0.106 and −0.148, victory −0.016 and 0.041, none of the latter significant). The within-video percentile rank gives humiliation +2.3 percentile points and victimhood −5.4. The conditional ordered logit on the realism-to-fantasy code, estimated

Table 9: H2b: fantasticality by scene type, within video

Specification	n	Victory	Humiliation	Victimhood
Main, no violence control	7,054	0.056 (0.031)	0.086 (0.026)	−0.092 (0.028)
Main, with violence intensity	7,054	0.011 (0.031)	0.078 (0.026)	−0.128 (0.029)
Regime-linked	801	0.003 (0.094)	0.217 (0.075)	−0.028 (0.137)
Iran-aligned independent	1,258	−0.025 (0.046)	0.102 (0.049)	−0.134 (0.067)
Trend adopter	4,995	0.033 (0.041)	0.032 (0.033)	−0.139 (0.030)
Animation-weighted	7,054	0.017 (0.032)	0.062 (0.027)	−0.123 (0.030)
R1 Component: realism-to-fantasy code	7,043	−0.016 (0.029)	0.077 (0.024)	−0.106 (0.030)
R1 Component: fantasy affordance	7,054	0.041 (0.041)	0.079 (0.034)	−0.148 (0.035)
R1 Within-video percentile rank	7,054	0.011 (0.011)	0.023 (0.010)	−0.054 (0.012)
R1 Conditional ordered logit (log-odds)	7,043	−0.104 (0.099)	0.380 (0.084)	−0.576 (0.084)
R2 Strict flags	7,054	0.037 (0.035)	0.057 (0.029)	−0.133 (0.029)
R2 Broad flags	7,054	0.043 (0.029)	0.066 (0.022)	−0.124 (0.029)
R2 Humiliation without the ridicule item	7,054	0.012 (0.031)	0.072 (0.027)	−0.131 (0.029)

Note. Outcome is the fantasticality score (0–3, SD 0.87) unless stated; all models include video fixed effects and log scene length, and rows after the first include violence intensity. Standard errors in parentheses, clustered by video (bootstrap over scenes for the conditional ordered logit).

by stacking its three thresholds with video-by-threshold fixed effects, gives log-odds of 0.38 for humiliation (scene-cluster bootstrap SE 0.08) and −0.58 for victimhood (SE 0.08), with victory at −0.10 (SE 0.10). The second check varies the flags. Strict definitions give humiliation 0.057 (SE 0.029, $p = .051$) and broad definitions 0.066 (SE 0.022); dropping the ridicule item from the humiliation flag gives 0.072 (SE 0.027), and dropping any one of its three items leaves the coefficient between 0.062 and 0.076. The victory coefficient never exceeds 0.047 under any definition. By provenance, the humiliation effect is 0.217 (SE 0.075) in regime-linked videos, 0.102 (SE 0.049) in Iran-aligned videos and 0.032 (SE 0.033) among trend adopters; the realism of victimhood comes mainly from trend adopters (−0.139, SE 0.030) and Iran-aligned videos (−0.134, SE 0.067). The producers closest to the state are the ones who most clearly mark mockery as fantasy.

6.2.3 H2c: Attention association

The humor score averages 0.84 on the 0–3 scale (SD 0.84) and is highly correlated with the coder's ridicule and amusement codes ($r = 0.86$ and 0.93) and modestly with the ridicule lexicon ($r = 0.23$). With upload-week, platform, and provenance fixed effects and controls for length, post age, violent share, and self-legitimation, a one-standard-deviation increase in humor is associated with 0.40 more log views (SE 0.19, $p = .036$), about 50 percent more views. In the main specification, which compares videos posted by the same channel, the association is 0.05 (SE 0.22, $p = .83$). Ridicule, amusement and entertainment orientation show the same contrast (0.12, 0.32 and 0.31 with week effects; −0.04, −0.05 and 0.12 with channel effects). Channel fixed effects alone explain 86 percent of the variance in log views; the four content measures add three points. H2c as stated, an association among comparable posts, is not supported. What the data show is an association between channels: accounts that make funnier videos have larger audiences, but a funnier video does not outperform

Table 10: H2c: humor and views

Outcome and specification	Treatment	n	Coefficient	SE	p
Log views, week + platform + provenance FE	Humor	360	0.400	0.190	.036
Log views, channel FE (main)	Humor	360	0.046	0.218	.83
Log views, week FE / channel FE	Ridicule	360	0.123 / −0.038	0.21 / 0.21	.55 / .86
Log views, week FE / channel FE	Amusement	360	0.322 / −0.051	0.17 / 0.18	.056 / .77
Log views, week FE / channel FE	Entertainment	360	0.309 / 0.115	0.16 / 0.17	.056 / .51
R1 Views per day, negative binomial, week FE	Humor	360	0.411	0.182	.024
R1 Log likes, channel FE	Humor	312	0.004	0.245	.99
R1 Log comments, channel FE	Humor	248	0.020	0.269	.94
R1 Log views, first uploads only, week FE	Humor	335	0.384	0.194	.048
R1 Number of re-uploads, Poisson	Humor	335	−0.066	0.295	.82
R2 Log views, channel FE / week FE	Ridicule lexicon	360	0.141 / −0.021	0.08 / 0.14	.060 / .88

R^2: week FE 0.37; week FE + content 0.39; channel FE 0.86; channel FE + content 0.89.

Note. Treatments standardized across videos. All models control for log length, log post age, violent share of runtime, and the self-legitimation index; standard errors clustered by channel.

its channel's other videos.

The robustness checks sharpen rather than change this reading (Table 10). The negative binomial model of views per day of exposure gives 0.41 (SE 0.18, $p = .024$) with week effects, and restricting to the first upload of each animation gives 0.38 (SE 0.19); both are between-channel estimates. Likes and comments with channel effects give 0.00 (SE 0.25) and 0.02 (SE 0.27). The number of re-uploads an animation receives is unrelated to humor (Poisson coefficient −0.07, SE 0.30). The second check uses measures that do not depend on the scene coder. The AudioSet classifier tags laughter or comedy in only 2 percent of videos, so it cannot serve as a measure. The ridicule lexicon gives a within-channel coefficient of 0.14 (SE 0.08, $p = .060$) and a between-channel coefficient of −0.02 (SE 0.14), the reverse of the pattern for the coder's humor score; the two measures agree on nothing except the absence of a robust association. View counts at a common post age would remove the dependence on the log-age control; the between-channel estimate is the one most exposed to that limitation.

6.3 H3: Register contrast with Iranian state media

The comparison corpus holds 15,845 speech and on-screen text segments from the 363 videos (median 7 words) and 5,865 sentences from 332 war-related articles (median 21 words). The event-topic rules assign both corpora mostly to the same three topics, casualties, and civilian harm, ceasefire and negotiations, and Hormuz and the blockade (83 percent of video units and 63 percent of article units), while the opening strikes are discussed far more in the articles (20 percent of units against 1.5 percent), which is the 28 February cluster at work. The dictionaries flag 26 percent of article sentences and 5 percent of video segments as justification claims. On the 1,200-unit sample classified by the source-blind language model, the dictionaries agree with the classifier on 82 percent of units

for claims, 96 percent for sacred legitimation and 89 percent for U.S.-oriented reference (Cohen's κ 0.47, 0.59 and 0.52). The disagreements run in one direction that matters below: the dictionaries miss about half of the claims and half of the U.S. references that the classifier finds (recall 0.47 and 0.52), and they miss them more often in the short, slangy video segments than in article prose. Figure 8 summarizes both tests.

6.3.1 H3a: Reduced sacred legitimation

Among the 2,360 justification claims (838 from videos, 1,522 from articles), 7.5 percent of video claims and 13.5 percent of article claims invoke religious authority or martyrdom. The logit with event-topic fixed effects gives a coefficient on LEGO of -0.52 (SE 0.42, clustered $p = .21$), an average marginal effect of -4.6 percentage points; the document-level randomization test gives $p = .050$ (Table 11). The direction is the predicted one, but the claim-level test does not reject at conventional levels with clustered inference. The robustness checks explain why. Splitting the category, religious authority gives -0.56 (SE 0.40) and martyrdom -0.81 (SE 0.59), each with a lower video rate (5.9 vs. 11.0 and 1.9 vs. 4.1 percent) and neither significant. Removing the claim filter, so that the denominator is all 21,710 units, gives -0.95 (SE 0.35, $p = .006$; permutation $p < .001$): sacred vocabulary appears in 2.2 percent of video units and 9.7 percent of article units. At the document level, the Poisson rate of sacred terms per word is less than half as high in videos (IRR 0.47, CI 0.25–0.86), and at the scene-and-paragraph grain the claim-level contrast is -0.77 (SE 0.39, $p = .050$). Restricting to the 2026-war window gives -0.82 (SE 0.61). On the classifier sample, where the claim filter is applied by the language model, sacred legitimation appears in 6.0 percent of video claims and 7.3 percent of article claims (-0.31, SE 0.59); across all sampled units, 2.5 against 6.2 percent (-0.77, SE 0.66).

The provenance interaction reverses the sign for the regime-linked core. Among regime-linked video claims, 26.5 percent invoke sacred legitimation, twice the article rate (coefficient $+1.74$, SE 0.50, $p < .001$). Iran-aligned videos are at 3.9 percent (-1.23, SE 0.81) and trend adopters at 3.2 percent (-1.82, SE 0.51, $p < .001$). The pooled contrast is therefore a mixture of a regime-linked register that is *more* sacralized than English-language state media and an independent and commercial register that is almost entirely desacralized. H3a is therefore inconclusive as stated: the genre as a whole uses less sacred vocabulary per unit and per word, but not a lower share of sacred justification claims, and the producers closest to the state, whose videos carry the martyrdom and clerical vocabulary that the outlets, writing for a foreign audience, leave out, reverse the predicted sign.

6.3.2 H3b: U.S.-oriented cultural reference

Across all 21,710 units, 13.6 percent of video segments and 6.9 percent of article sentences refer to U.S. domestic politics or popular culture. The logit with event-topic fixed effects gives $+1.20$ (SE 0.20, $p < .001$), an average marginal effect of $+9.7$ percentage points; the permutation p-value is below .001. The result holds in every provenance group (regime-linked 7.9 percent, $+0.75$, SE

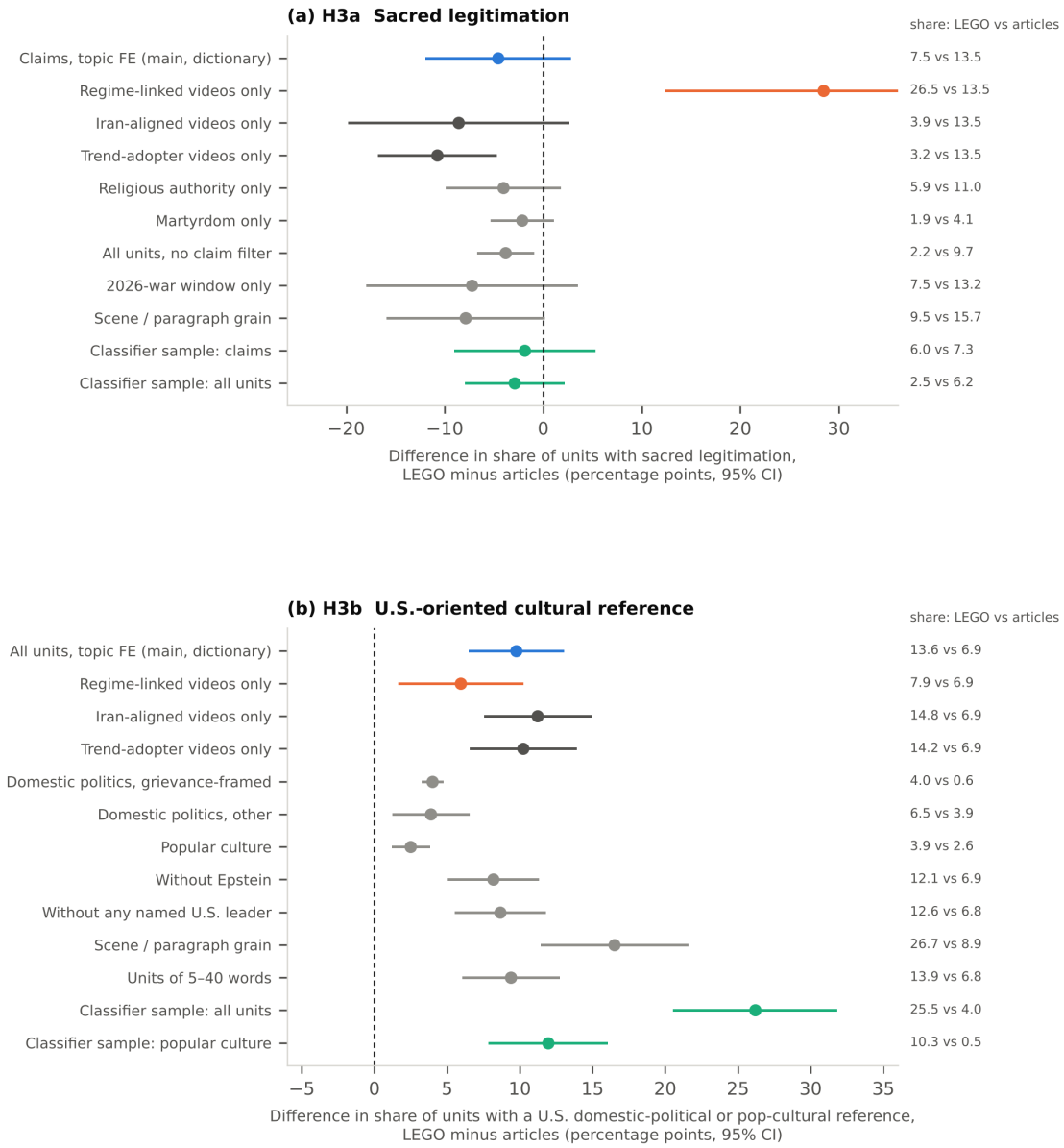

Figure 8: Tests of the H3 hypotheses. Each point is the difference between the LEGO videos and the state-media articles in the share of units with the outcome, from logit models with event-topic fixed effects and log unit length (average marginal effect, in percentage points; 95 percent intervals from standard errors clustered by channel and outlet). Raw shares are given at the right. Gray rows are robustness checks; green rows use the source-blind language-model classifications on a stratified sample of 1,200 units.

0.27; Iran-aligned 14.8 percent, +1.17; trend adopters 14.2 percent, +1.21) and under every check in Table 11. Removing Epstein leaves +1.07; removing Trump +1.21; removing every named U.S. leader +1.11. The scene-and-paragraph grain gives +1.31 (26.7 vs. 8.9 percent), length-matched units +1.10, and the document-level fractional logit predicts shares of 10.2 and 5.7 percent; the Poisson rate of U.S.-oriented terms per word is three times as high in videos (IRR 3.11, CI 2.28–4.25).

The split specifications show where the gap lies. Grievance-framed references to U.S. domestic politics, taxes, prices, the Epstein files, and the costs of the war presented as a complaint, appear in 4.0 percent of video units and 0.6 percent of article units (+2.26, SE 0.19, the largest coefficient in the section). Domestic-political references that are not grievance-framed, which in the articles are mostly reports of what Congress or the administration did, are 6.5 against 3.9 percent (+0.86, SE 0.29). Popular-culture references are 3.9 against 2.6 percent (+0.86, SE 0.21). The classifier, which reads lyrics better than the dictionaries, widens every gap: on the sample, 25.5 percent of video units and 4.0 percent of article units carry a U.S.-oriented reference (+2.67, SE 0.29), 18.2 against 3.7 percent for domestic politics, and 10.3 against 0.5 percent for popular culture (+3.66, SE 0.63). H3b is supported in the strong form the theory specifies: the videos' advantage is largest in grievance and popular-cultural references, not in reporting of U.S. politics as such.

Table 11: H3: register contrasts between LEGO videos and Iranian state media

Specification	Units	LEGO %	Articles %	Coef. on LEGO (SE)	AME (pp)
H3a: sacred legitimation					
Main: justification claims, topic FE	2,360	7.5	13.5	−0.52 (0.42)	−4.6
permutation p (document level, 2,000 draws)				$p = .050$	
Regime-linked / Iran-aligned / Trend adopter	1,673 / 1,676 / 2,055	26.5 / 3.9 / 3.2	13.5	+1.74 (0.50) / −1.23 (0.81) / −1.82 (0.51)	
R1 Religious authority only / martyrdom only	2,360	5.9 / 1.9	11.0 / 4.1	−0.56 (0.40) / −0.81 (0.59)	
R1 No claim filter (all units)	21,710	2.2	9.7	−0.95 (0.35)	−3.8
R1 Classifier sample: claims / all units	304 / 1,200	6.0 / 2.5	7.3 / 6.2	−0.31 (0.59) / −0.77 (0.66)	
R2 2026-war window only	1,172	7.5	13.2	−0.82 (0.61)	
R2 Scene / paragraph grain (claims)	2,124	9.5	15.7	−0.77 (0.39)	
R2 Document-level Poisson rate, IRR (95% CI)	695			0.47 (0.25–0.86)	
H3b: U.S. domestic-political or popular-cultural reference					
Main: all units, topic FE	21,710	13.6	6.9	+1.20 (0.20)	+9.7
permutation p				$p < .001$	
Regime-linked / Iran-aligned / Trend adopter	7,618 / 8,274 / 17,548	7.9 / 14.8 / 14.2	6.9	+0.75 (0.27) / +1.17 (0.19) / +1.21 (0.22)	
Split: grievance-framed domestic politics	21,710	4.0	0.6	+2.26 (0.19)	+4.0
Split: other domestic politics / popular culture	21,710	6.5 / 3.9	3.9 / 2.6	+0.86 (0.29) / +0.86 (0.21)	

Continued on next page

Table 11 continued

Specification	Units	LEGO %	Articles %	Coef. on LEGO (SE)	AME (pp)
R1 Without Epstein / Trump / any named U.S. leader	21,710	12.1 / 12.6 / 12.6	6.9 / 6.0 / 6.8	+1.07 / +1.21 / +1.11	
R1 Classifier sample: all units / popular culture	1,200 / 972	25.5 / 10.3	4.0 / 0.5	+2.67 (0.29) / +3.66 (0.63)	+26.2 / +11.9
R2 Scene / paragraph grain	10,432	26.7	8.9	+1.31 (0.20)	+16.5
R2 Units of 5–40 words	14,889	13.9	6.8	+1.10 (0.20)	+9.4
R2 Document-level fractional logit (predicted shares)	695	10.2	5.7	+0.63 (0.18)	
R2 Document-level Poisson rate, IRR (95% CI)	695			3.11 (2.28–4.25)	

Note. Logit coefficients with event-topic fixed effects and log unit length; standard errors clustered by channel (136) and outlet (5). AME is the average marginal effect in percentage points. Dictionary classifications unless marked as classifier sample.

The register contrast is asymmetric. What the videos add is clear and robust: references to the American audience's own politics, grievances, and culture, at two to six times the state-media rate depending on the instrument, in every provenance group and under every specification (H3b supported). What the videos remove is less clear. Sacred vocabulary is rarer in them per unit and per word, but among the claims that justify or condemn, the gap is small, and imprecise, and the regime-linked videos, the ones a state-media comparison should most resemble, use more sacred legitimation than the outlets do (H3a inconclusive for the genre, reversed in its core). Desacralization in this corpus is what happens as the genre leaves the regime's hands; adaptation to the American audience is present from the start.

Table 12 brings together the findings for all eight hypotheses, distinguishing the primary tests from secondary comparisons and provenance differences.

Table 12: Summary of findings across all hypotheses

Hypothesis	Principal finding and qualification	Assessment
H1a Representational dominance	Iranian actors are the agents in 65.6 percent of eligible confrontations (95 percent CI 59.7–71.6). The pattern is concentrated in regime-linked and Iran-aligned videos; trend adopters are near parity.	Supported
H1b Personalized blame	Named leaders receive 93.1 percent of leader-versus-public blame attributions (95 percent CI 89.6–96.5). Excluding Trump gives an inconclusive 52.5 percent (CI 36.1–68.8): the pattern depends on Trump.	Supported (Trump-driven)
H1c Selective civilian exculpation	The adjusted American–Israeli civilian difference is $\hat{\beta} = -0.007$ (SE 0.082; $p = .93$). Sparse Israeli civilian observations and only five videos with both publics limit the within-video comparison.	Not supported
H2a Selective sanitization	Within videos and target types, an Iranian perpetrator is associated with a suffering index 0.769 SD lower (SE 0.126; $p < .001$). The separate frame-caption check is inconclusive.	Supported
H2b Fantastical triumph	With violence controlled, humiliation scenes are more fantastical ($\hat{\beta} = 0.078$, SE 0.026; $p = .002$). The victory contrast is inconclusive ($\hat{\beta} = 0.011$, SE 0.031; $p = .72$).	Partially supported
H2c Attention association	Within channels, humor's association with log views is $\hat{\beta} = 0.046$ (SE 0.218; $p = .83$). The pooled, between-channel association is positive; views are collection-time snapshots.	Not supported
H3a Reduced sacred legitimation	Among justification claims, the adjusted gap is −4.6 percentage points ($p = .21$). Lower sacred-vocabulary rates per unit and per word coexist with a reversed contrast in regime-linked videos.	Inconclusive (reversed in the core)
H3b U.S.-oriented cultural reference	Video units contain more U.S. domestic-political or popular-cultural references (adjusted gap +9.7 percentage points; $p < .001$). The positive contrast appears in every provenance group.	Supported

Note. Assessments prioritize each hypothesis's primary specification and summarize the reported observational evidence under the current coding and inferential procedures (Tables 6–11). *Not supported* and *inconclusive* do not establish a zero effect. H2c's common-post-age outcome is not observed; H3 uses a temporally unmatched article pilot. SE denotes standard error; SD, standard deviation.

7 Discussion

The videos are built around the adversary. They let Iran win the confrontations they depict, assign the war to one American leader rather than to Americans, strip Iranian violence of visible victims, and speak to the American audience in the vocabulary of its own politics and culture. Those four findings survive every specification we ran, and three of them hold within videos or within channels, where the producer's style and audience are held constant. The evidence also narrows the account in three ways that the theory did not anticipate. The leader–people split operates through representation rather than through differential blame: American civilians are shown in six videos of ten and Israeli civilians in one of seventeen, but once a civilian is shown, nationality does not predict how it is treated, and the nationality gap appears among soldiers instead. Play is reserved for mockery, not for triumph: humiliation is marked as fantasy and suffering as real, but battles are rendered as battles whoever wins them. And the attention advantage of humor is a property of channels, not of videos; within an account, a funnier upload does not outperform its siblings.

The theoretical contribution is to connect the audience orientation of propaganda to its representational form. Addressing an adversary's citizens makes recognizable leaders and domestic grievances useful narrative resources, which links the psychological-warfare tradition to research on the audience conditions of propaganda (Linebarger, 1954; Carter and Carter, 2023). Within a shared toy aesthetic, the selective visibility of suffering gives playful presentation political significance: comparable harm receives different treatment depending on its perpetrator, consistent with the minimization of consequences described by Bandura (1999). These are findings about representation. Whether humor reduces scrutiny or instead encourages viewers to discount a claim requires direct audience evidence (Young, 2008; Nabi et al., 2007).

Provenance separates the two halves of the argument. Representational dominance and the sacred frame belong to the videos closest to the state: Iranian agency falls from 84 percent in regime-linked confrontations to parity among trend adopters, and the regime-linked core invokes martyrdom and religious authority in a quarter of its justification claims, twice the rate of the English-language outlets, while the imitators barely do. The orientation toward the American audience, by contrast, is present in every group of producers. What travels when a state seeds a genre, on this evidence, is the audience strategy rather than the legitimating frame. This motivates a theory of selective adaptation across producers; establishing that diffusion caused these differences requires traced sequences of originals, copies, and edits, which the content clusters identified here make possible.

The dataset combines scale with linked measurement: every estimate can be traced to a scene, a frame, and a timestamp, and the same architecture extends to other conflicts, languages and animation styles through `legopolitics` (Caliskan, 2026). Its scene count is of the same order as MovieGraphs' 7,637 annotated clips, although units, annotation procedures, and validation standards differ (Vicol et al., 2018), and larger political-video archives exist (Breuer et al., 2025). Nested scenes and frames deepen measurement without increasing the number of independent producers, which is why inference throughout is clustered by channel.

Several limits remain consequential. The interpretive codes are language-model judgments; the frame layer corroborates concrete content, not humor, responsibility or suffering. Search-based collection can miss deleted or less discoverable material. The state-media baseline is a temporally unmatched pilot from five outlets, so the register comparison rests on event-topic rather than date matching. Provenance categories do not directly observe state control, and displayed views establish neither unique audiences nor persuasion.

Future research should combine audience experiments with prospective comparative collection. Holding claims constant while varying toy aesthetics, humor, and source disclosure would test effects on credibility, empathy, recall, and war attitudes. Engagement measured at fixed post ages, documented re-upload sequences, and date-matched state-media samples would distinguish content associations from distribution advantages. Continued experimentation with synthetic animation is a plausible expectation: experiments demonstrate the persuasive potential of generated propaganda text, and research on a Russia-affiliated operation documents increased output and topical breadth after AI adoption (Goldstein et al., 2024; Wack et al., 2025). Extending those findings to animation remains an inference; reusable characters and templates may encourage organized producers and imitators, while moderation, audience fatigue, and unequal distribution may constrain their reach. The enduring research question is how political claims change as they become entertainment, circulate across producers, and encounter different publics.

References

Maja Adena, Ruben Enikolopov, Maria Petrova, Veronica Santarosa, and Ekaterina Zhuravskaya. Radio and the rise of the Nazis in prewar Germany. *The Quarterly Journal of Economics*, 130(4): 1885–1939, 2015. doi: 10.1093/qje/qjv030. URL https://doi.org/10.1093/qje/qjv030.

Agence France-Presse. YouTube suspends pro-Iran channel posting Lego-style clips mocking Trump. The Standard, April 2026. URL https://www.thestandard.com.hk/world/article/329464/YouTube-suspends-pro-Iran-channel-posting-Lego-style-clips-mocking-Trump. Published April 16, 2026.

Taylor Arnold and Lauren Tilton. Distant viewing: Analyzing large visual corpora. *Digital Scholarship in the Humanities*, 34(Supplement 1):i3–i16, 2019. doi: 10.1093/llc/fqz013. URL https://doi.org/10.1093/llc/fqz013.

Albert Bandura. Moral disengagement in the perpetration of inhumanities. *Personality and Social Psychology Review*, 3(3):193–209, 1999. doi: 10.1207/s15327957pspr0303_3. URL https://doi.org/10.1207/s15327957pspr0303_3.

Jonah Berger and Katherine L. Milkman. What makes online content viral? *Journal of Marketing Research*, 49(2):192–205, 2012. doi: 10.1509/jmr.10.0353. URL https://doi.org/10.1509/jmr.10.0353.

Adam J. Berinsky. Assuming the costs of war: Events, elites, and American public support for military conflict. *The Journal of Politics*, 69(4):975–997, 2007. doi: 10.1111/j.1468-2508.2007.00602.x. URL https://doi.org/10.1111/j.1468-2508.2007.00602.x.

Adam Breuer, Bryce J. Dietrich, Michael H. Crespin, Matthew Butler, JA Pryse, and Kosuke Imai. Using AI to summarize US presidential campaign TV advertisement videos, 1952–2012. *Scientific Data*, 12:1552, 2025. doi: 10.1038/s41597-025-05558-9. URL https://doi.org/10.1038/s41597-025-05558-9.

Cantay Caliskan. legopolitics: A multimodal Python research package for analyzing political narratives in brick-built videos. Computer software, July 2026. URL https://pypi.org/project/legopolitics/0.1.1/. Public release, version 0.1.1.

A. Colin Cameron, Jonah B. Gelbach, and Douglas L. Miller. Bootstrap-based improvements for inference with clustered errors. *The Review of Economics and Statistics*, 90(3):414–427, 2008. doi: 10.1162/rest.90.3.414. URL https://doi.org/10.1162/rest.90.3.414.

Erin Baggott Carter and Brett L. Carter. *Propaganda in Autocracies: Institutions, Information, and the Politics of Belief*. Cambridge University Press, Cambridge, 2023. doi: 10.1017/9781009271226. URL https://doi.org/10.1017/9781009271226.

Robert M. Entman. Framing: Toward clarification of a fractured paradigm. *Journal of Communication*, 43(4):51–58, 1993. doi: 10.1111/j.1460-2466.1993.tb01304.x. URL https://doi.org/10.1111/j.1460-2466.1993.tb01304.x.

Matthew Gault. Iran is winning the AI slop propaganda war. 404 Media, March 2026. URL https://www.404media.co/iran-is-winning-the-ai-slop-propaganda-war/. Published March 27, 2026.

Christopher Gelpi, Peter D. Feaver, and Jason Reifler. *Paying the Human Costs of War: American Public Opinion and Casualties in Military Conflicts*. Princeton University Press, Princeton, NJ, 2009. doi: 10.1515/9781400830091. URL https://doi.org/10.1515/9781400830091.

Josh A. Goldstein, Jason Chao, Shelby Grossman, Alex Stamos, and Michael Tomz. How persuasive is AI-generated propaganda? *PNAS Nexus*, 3(2):pgae034, 2024. doi: 10.1093/pnasnexus/pgae034. URL https://doi.org/10.1093/pnasnexus/pgae034.

Tianru Guan and Tingting Hu. The conformation and negotiation of nationalism in China's political animations—a case study of Year Hare Affair. *Continuum*, 34(3):417–430, 2020. doi: 10.1080/10304312.2020.1724882. URL https://doi.org/10.1080/10304312.2020.1724882.

Linus Hagström and Karl Gustafsson. The limitations of strategic narratives: The Sino-American struggle over the meaning of COVID-19. *Contemporary Security Policy*, 42(4):415–449, 2021. doi: 10.1080/13523260.2021.1984725. URL https://doi.org/10.1080/13523260.2021.1984725.

Haifeng Huang. The pathology of hard propaganda. *The Journal of Politics*, 80(3):1034–1038, 2018. doi: 10.1086/696863. URL https://doi.org/10.1086/696863.

Stefano M. Iacus, Gary King, and Giuseppe Porro. Causal inference without balance checking: Coarsened exact matching. *Political Analysis*, 20(1):1–24, 2012. doi: 10.1093/pan/mpr013. URL https://doi.org/10.1093/pan/mpr013.

Jungseock Joo and Zachary C. Steinert-Threlkeld. Image as data: Automated content analysis for visual presentations of political actors and events. *Computational Communication Research*, 4(1): 11–67, 2022. doi: 10.5117/CCR2022.1.001.JOO. URL https://doi.org/10.5117/CCR2022.1.001.JOO.

Harvin Kaur. We spoke to the man making viral Lego-style AI videos for Iran: Experts say it's powerful propaganda. MARKETING Magazine Asia, April 2026. URL https://marketingmagazineasia.com/we-spoke-to-the-man-making-viral-lego-style-ai-videos-for-iran-experts-say-its-powerful-propaganda/. Published April 13, 2026; reproduces BBC interview reporting.

Olivier Keller. Saludos Amigos: Disney propaganda for Latin America. *The Americas*, 82(2):219–247, 2025. doi: 10.1017/tam.2025.10076. URL https://doi.org/10.1017/tam.2025.10076.

Gary King, Jennifer Pan, and Margaret E. Roberts. How the Chinese government fabricates social media posts for strategic distraction, not engaged argument. *American Political Science Review*, 111(3):484–501, 2017. doi: 10.1017/S0003055417000144. URL https://doi.org/10.1017/S0003055417000144.

Harold D. Lasswell. The theory of political propaganda. *American Political Science Review*, 21(3): 627–631, 1927. doi: 10.2307/1945515. URL https://doi.org/10.2307/1945515.

Paul M. A. Linebarger. *Psychological Warfare*. Duell, Sloan and Pearce, New York, second edition, 1954. URL https://www.gutenberg.org/files/48612/48612-h/48612-h.htm.

James G. MacKinnon and Matthew D. Webb. The wild bootstrap for few (treated) clusters. *The Econometrics Journal*, 21(2):114–135, 2018. doi: 10.1111/ectj.12107. URL https://doi.org/10.1111/ectj.12107.

Charles F. Manski. Nonparametric bounds on treatment effects. *The American Economic Review*, 80(2):319–323, 1990. URL https://www.jstor.org/stable/2006592. Papers and Proceedings.

Daniel Mattingly, Trevor Incerti, Changwook Ju, Colin Moreshead, Seiki Tanaka, and Hikaru Yamagishi. Chinese state media persuades a global audience that the "China model" is superior: Evidence from a 19-country experiment. *American Journal of Political Science*, 69(3):1029–1046, 2025. doi: 10.1111/ajps.12887. URL https://doi.org/10.1111/ajps.12887.

Daniel C. Mattingly and Elaine Yao. How soft propaganda persuades. *Comparative Political Studies*, 55(9):1569–1594, 2022. doi: 10.1177/00104140211047403. URL https://doi.org/10.1177/00104140211047403.

Robin L. Nabi, Emily Moyer-Gusé, and Sahara Byrne. All joking aside: A serious investigation into the persuasive effect of funny social issue messages. *Communication Monographs*, 74(1):29–54, 2007. doi: 10.1080/03637750701196896. URL https://doi.org/10.1080/03637750701196896.

George Orwell. Funny, but not vulgar. *Leader*, July 1945. URL https://orwell.ru/library/articles/funny/english/e_funny. Published July 28, 1945.

Bill Rolston. When everywhere is Karbala: Murals, martyrdom and propaganda in Iran. *Memory Studies*, 13(1):3–23, 2020. doi: 10.1177/1750698017730870. URL https://doi.org/10.1177/1750698017730870.

Arturas Rozenas and Denis Stukal. How autocrats manipulate economic news: Evidence from Russia's state-controlled television. *The Journal of Politics*, 81(3):982–996, 2019. doi: 10.1086/703208. URL https://doi.org/10.1086/703208.

Younes Saramifar. The pain of others: Framing war photography in Iran. *Ethnos*, 84(3):480–507, 2019. doi: 10.1080/00141844.2017.1415367. URL https://doi.org/10.1080/00141844.2017.1415367.

Michael J. Shapiro. *Cinematic Geopolitics*. Routledge, London and New York, 2009. doi: 10.4324/9780203892008. URL https://www.routledge.com/Cinematic-Geopolitics-1st-Edition/Shapiro/p/book/9780415776356.

Limor Shifman. Memes in a digital world: Reconciling with a conceptual troublemaker. *Journal of Computer-Mediated Communication*, 18(3):362–377, 2013. doi: 10.1111/jcc4.12013. URL https://doi.org/10.1111/jcc4.12013.

Jiarui Sun. Patriotic rabbits or toxic men? Media ideology, entextualization, and enregisterment on Chinese interfaces. *Journal of Linguistic Anthropology*, 35(2):e70007, 2025. doi: 10.1111/jola.70007. URL https://doi.org/10.1111/jola.70007.

Alexander Tarr, June Hwang, and Kosuke Imai. Automated coding of political campaign advertisement videos: An empirical validation study. *Political Analysis*, 31(4):554–574, 2023. doi: 10.1017/pan.2022.26. URL https://doi.org/10.1017/pan.2022.26.

Michelle Torres and Francisco Cantú. Learning to see: Convolutional neural networks for the analysis of social science data. *Political Analysis*, 30(1):113–131, 2022. doi: 10.1017/pan.2021.9. URL https://doi.org/10.1017/pan.2021.9.

Paul Vicol, Makarand Tapaswi, Lluís Castrejón, and Sanja Fidler. MovieGraphs: Towards understanding human-centric situations from videos. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 8581–8590, 2018. URL https://openaccess.thecvf.com/content_cvpr_2018/html/Vicol_MovieGraphs_Towards_Understanding_CVPR_2018_paper.html.

Morgan Wack, Carl Ehrett, Darren Linvill, and Patrick Warren. Generative propaganda: Evidence of AI's impact from a state-backed disinformation campaign. *PNAS Nexus*, 4(4):pgaf083, 2025. doi: 10.1093/pnasnexus/pgaf083. URL https://doi.org/10.1093/pnasnexus/pgaf083.

Nora Webb Williams, Andreu Casas, and John D. Wilkerson. *Images as Data for Social Science Research: An Introduction to Convolutional Neural Nets for Image Classification*. Elements in Quantitative and Computational Methods for the Social Sciences. Cambridge University Press, 2020. doi: 10.1017/9781108860741. URL https://doi.org/10.1017/9781108860741.

Lisa Wedeen. *Ambiguities of Domination: Politics, Rhetoric, and Symbols in Contemporary Syria*. University of Chicago Press, Chicago, 2015. URL https://press.uchicago.edu/ucp/books/book/chicago/A/bo22776830. With a new preface by the author. Originally published in 1999.

Jamie Wong, Crystal Lee, Vesper Keyi Long, Di Wu, and Graham M. Jones. Let's go, baby forklift!: Fandom governance and the political power of cuteness in China. *Social Media + Society*, 7(2):1–18, 2021. doi: 10.1177/20563051211024960. URL https://doi.org/10.1177/20563051211024960.

Xinhua. Once Upon a Virus. English.news.cn, animated video and accompanying text, April 2020. URL https://www.xinhuanet.com/english/2020-04/30/c_139020877.htm. Published April 30, 2020.

Dannagal Goldthwaite Young. The privileged role of the late-night joke: Exploring humor's role in disrupting argument scrutiny. *Media Psychology*, 11(1):119–142, 2008. doi: 10.1080/15213260701837073. URL https://doi.org/10.1080/15213260701837073.

A The `legopolitics` Python Package

`legopolitics` is an open-source Python package for converting brick-built video into linked research records (Caliskan, 2026). This description refers to public release 0.1.1, uploaded to PyPI on July 13, 2026, and checked against its published source archive. The release requires Python 3.10 or later, is distributed under the MIT License, and provides a command-line interface and a Python API. Its architecture combines video processing, optional machine-learning adapters, structured annotation, and export (Figure A1). Appendix D describes how the package and the project-specific protocol were combined to build the corpus analyzed in this paper; the package's defaults, quoted below, are not the study's realized settings unless stated there.

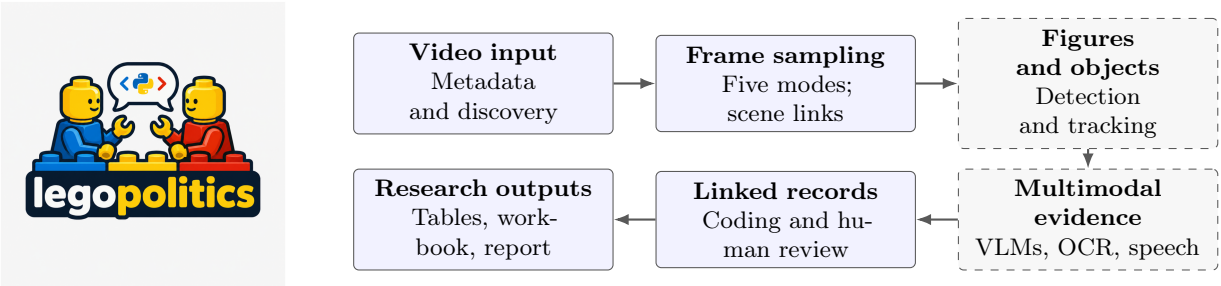

Figure A1: Schematic of the package workflow. Dashed boxes require configured optional adapters and, where applicable, external model weights. The logo is reproduced from the published source distribution; the workflow is drawn from the inspected release.

A.1 Records

The basic observation is not a single description of an entire video. Records distinguish scenes, timestamped frames, detections, tracks, text, audio, and model responses. A detection records a bounding box and class label for an object in a frame; a track links observations of an object through time; codebook records identify a variable and its unit of analysis. This structure makes it possible to inspect the image evidence behind a coding decision and to aggregate observations at the scene or video level. Identifier links (Figure A2) permit comparisons across scenes, frames, objects, and model outputs. Exports include Excel, JSONL, SQLite, optional Parquet tables and HTML, alongside configuration, environment information, prompts, raw responses, and provenance. Deterministic SHA-256 partitioning supports batch assignment, and caching supports resumption.

Table A1: Quantitative structure of release 0.1.1

Component	Count	Interpretation
Recognized video extensions	6	MP4, MKV, MOV, AVI, WebM, M4V
Frame-sampling modes	5	Fixed, scene, hybrid, adaptive, all frames
Example codebook variables	21	Each permits integer values from 0 to 5
Exported data-table types	21	Includes a derived model-provenance table
Excel worksheets	24	Data tables plus README, run metadata and data dictionary
Hardware-specific templates	5	CPU through maximum-capacity GPU profiles
Entries in the ensemble example	4	Two enabled and two disabled (Table A2)

Counts describe the released schemas and configurations, not nonempty outputs or predictive accuracy.

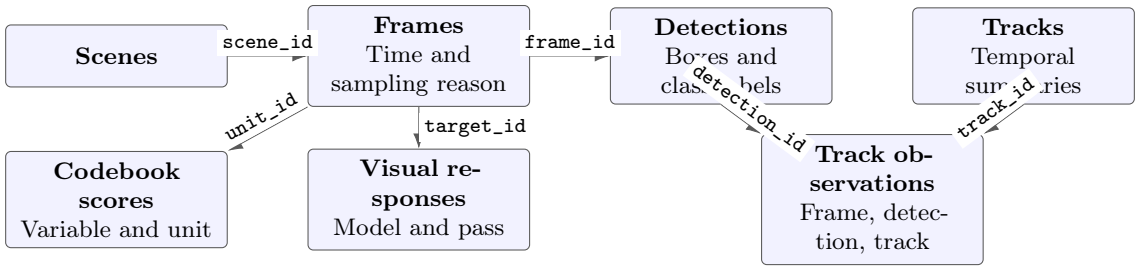

Figure A2: Selected identifier links in the public schemas. Joins are made within a video; response and score records also specify their target or unit type. Arrows indicate joinable identifiers, not causal relationships or guaranteed population of every table.

A.2 Sampling, detection and multimodal models

Frame sampling controls the volume of evidence sent to downstream models. Fixed sampling selects a regular cadence; scene sampling adds scene boundaries and representatives; hybrid sampling combines those candidates; adaptive sampling additionally considers motion; all-frame mode retains the complete sequence (capped at 1,000,000 frames). With source rate r, requested rate f, and N source frames, the regular candidate rule is

$$q = \max\{1, \operatorname{round}(r/f)\}, \qquad N_{\text{regular}} = \lceil N/q \rceil \,, \tag{A1}$$

so that a 60-second video at 30 frames per second (1,800 frames) contributes 60 regular candidates at one frame per second; endpoint and scene rules can add candidates and perceptual-hash deduplication (default distance threshold 4) can remove them. Each selected frame retains its source index, timestamp, scene identifier, and sampling reasons, together with motion and image-quality measures.

The optional YOLO adapter uses supplied detector weights, with default confidence threshold 0.30, intersection-over-union threshold 0.50 and inference size 1,280 pixels; bounding boxes support crops, normalized area, and centrality measures, and tracking stores object observations over time. OCR and speech adapters add timestamped evidence when enabled. For frame-level vision-language analysis the implementation requests, separately, a blind description and a context-conditioned interpretation of each frame and preserves both: with F frames, M models and both passes, this entails $2MF$ requests. Hardware profiles, quantization options and sequential model unloading configure resource use.

Table A2: The distributed Hugging Face ensemble example

Model identifier	Nominal size	Enabled
SmolVLM2-2.2B-Instruct	2.2B	Yes
Qwen2.5-VL-3B-Instruct	3B	Yes
Idefics3-8B-Llama3	8B	No
LLaVA-OneVision-Qwen2-7B	7B	No

Sizes are the nominal designations in the configured model identifiers.

A.3 Political measurement, validation, and reproducibility

The 21-dimension example codebook uses frame-level scores from 0 to 5, including heroism, threat, victimization, victory, humiliation, civilian suffering, and religious legitimacy; entries specify inclusion

and exclusion rules and evidence requirements, and scores are exported only when a parsed contextual response contains the matching variable. Default relations record object co-occurrence and default narrative labels mark the first scene as an introduction; the operational definitions of agency, responsibility, and suffering used in this paper are supplied by the study protocol (Appendix D), not by these defaults. Human-review sampling selects random detections, low-confidence detections, and units with model disagreement (default budgets of 100 per category), and validation utilities compute accuracy, precision, recall, F1, intersection-over-union summaries, percentage agreement and Cohen's kappa. Reproduction requires preserving input identities, package, and model versions, weights, prompts, and configuration, all of which the export records.

B Data Dictionary and Text Dictionaries

B.1 Codebook and tables

The project codebook (corpus_codebook.csv in the replication files) documents 570 variables in 18 tables. For each variable it gives the table, the unit of analysis, a conceptual and an operational definition, the data type, the coding scale and valid values, the missing-value rule, an example, the feasibility of automated coding and whether interpretation is required. Table B1 lists the tables and their variable names; Table B2 groups the 284 scene-level variables into families. Ordinal codes run from 0 (absent) to 3 (dominant or central) unless stated otherwise; violence intensity runs from 0 to 4. Absence (0) and indeterminacy (missing) are distinct.

Table B1: Tables, units of analysis and variables

Table	Unit	Records (2026 sample)	Variables
Triage	Candidate video	549 (all candidates)	15
Video summary	Video	363	89
Scenes	Scene	7,288	284
Characters	Character or collective actor	5,617	38
Actor relationships	Directed relationship	4,378	11
Events	Event	3,329	25
Symbols	Symbol	7,409	15
Text and audio	Speech or on-screen text segment	16,270	18
Propaganda summary	Strategy within video (top five)	1,815	11
Other per-video coding tables	Scene × character, scene × symbol, narrative phase, notes and checks	Per-video workbooks	79
Frames	Frame (1 per second)	96,530	19
YOLO26s detections	Detected object	63,477 at confidence ≥ 0.5	37
Whisper transcripts	Diarized speech segment	358 videos	4
State-media articles	Article	506	44
Analysis variables (this paper)	Video, scene, character, confrontation, text unit	—	—

Triage. category, channel, confidence, decision, duration_seconds, is_ai_generated_likely, lego_style_share_estimate, main_language, reason, source_url, title, topic_summary, upload_date, video_id, view_count

Video summary. video_id, video_key, video_title, source, source_url, platform, channel, channel_url, uploader, upload_date, view_count, like_count, comment_count, country_folder, file_path, triage_category, language, video_duration_seconds, total_scenes, total_substantive_scenes, total_shots_estimated, shots_detected_automatically, mean_scene_duration, median_scene_duration, shortest_scene_duration, longest_scene_duration, primary_country, principal_protagonist, principal_antagonist, dominant_narrative, dominant_emotion, dominant_propaganda_strategy, national_glorification_score, military_glorification_score, religious_legitimation_score, historical_legitimation_score, mythological_legitimation_score, enemy_demonization_score, enemy_dehumanization_score, enemy_ridicule_score, martyrdom_sacrifice_score, military_superiority_score, technological_superiority_score, inevitability_of_victory_score, justification_of_violence_score, leader_glorification_score, victimhood_score, fear_appeal_score, pride_appeal_score, anger_appeal_score, foreign_threat_score, moral_superiority_score, moral_binary_score, temporal_compression_score, sacred_secular_fusion_score, heroic_genealogy_score, enemy_transformation_score, weapon_sacralization_score, military_spectacle_score, population_unanimity_score, mythic_compression_score, enemy_specificity_score, hero_villain_visual_distance_score, max_violence_intensity, mean_violence_graphic_gap_violent_scenes, share_runtime_violent, entertainment_orientation, information_orientation, persuasion_orientation, overall_LEGO_medium_specificity, propaganda_intensity_mean, reused_footage_seconds, reused_footage_share, clip_period_candidate_seconds, clip_boundary_share_on_grid, generative_ai_production, national_flag_absence, producer_type_inferred, opening_emotion, midpoint_emotion, climax_emotion, ending_emotion, dominant_emotional_transition, emotional_turning_point_scenes, notes, asr_language_majority, pipeline_version, coding_model, coded_at

Scenes. See Table B2

Characters. video_id, character_id, character_name, character_description, actor_type, political_identity, country_or_identity, affiliation, nationality_if_identifiable, historical_identity, religious_identity, mythological_

identity, first_scene, last_scene, first_timestamp, last_timestamp, number_scenes_present, on_screen, estimated_screen_prominence, narrative_role, protagonist_antagonist, portrayal, heroic_score, threatening_score, competent_score, righteous_score, violent_score, ridiculous_score, inhuman_monsterized_score, main_actions, associated_symbols, associated_emotions, historical_reference, religious_reference, mythological_reference, propaganda_function, identification_confidence, notes

Actor relationships. video_id, relationship_id, scene_id, source_actor_id, source_actor, relationship, target_actor_id, target_actor, relationship_description, political_function, confidence

Events. video_id, event_id, event_type, start_scene, end_scene, actor_id, actor, action, target_id, target, instrument, location, country, outcome, political_meaning, narrative_function, causal_role, historical_analogy, religious_analogy, mythological_analogy, perpetrator_named, confidence, notes, start_timestamp, end_timestamp

Symbols. video_id, symbol_id, symbol, symbol_type, associated_actor, associated_country, associated_identity, first_scene, last_scene, number_scenes_present, approximate_prominence, framing, symbolic_function, interpretation_confidence, notes

Text and audio. video_id, text_audio_id, scene_id, start_timestamp, end_timestamp, start_seconds, end_seconds, modality, speaker, source, original_language, original_text, English_translation, transcription_confidence, political_function, emotional_function, target, notes

Propaganda summary. video_id, strategy_rank, propaganda_strategy, overall_strength, max_scene_score, share_of_runtime_score_ge2, strongest_scene_1, strongest_scene_2, strongest_scene_3, evidence_summary, interpretation_confidence

Other per-video coding tables. scene_characters (17); scene_symbols (9); narrative_arc (13); narrative_reconstruction (3); comparative_hypotheses (9); automation (7); coding_notes (6); new_variables (8); large_n_recommendations (4); qc_checks (3)

Frames. folder, sec, n_lego_figure, maxarea_lego_figure, n_human, maxarea_human, video_id, caption, ocr, cap_flag, cap_explosion, cap_missile, cap_gore, cap_lego, cap_soldier, cap_crowd, ocr_chars, has_lego, has_human

YOLO26s detections. batch_id, num_batches, timestamp, frame_path, frame_relative_path, video_folder_relative, frame_filename, frame_width, frame_height, detection_id, class_id, class_name, confidence, x1, y1, x2, y2, box_width, box_height, x_center, y_center, x1_norm, y1_norm, x2_norm, y2_norm, x_center_norm, y_center_norm, box_width_norm, box_height_norm, crop_x1, crop_y1, crop_x2, crop_y2, crop_path, crop_relative_path, status, message

Whisper transcripts. start time, end time, speaker label, text

State-media articles. record_id, outlet, outlet_name, article_id, requested_url, canonical_url, title, lead, body, body_chars, parse_status, date_published, date_published_raw, date_modified, date_timezone, publisher_displayed_dates, declared_language, language, language_check, latin_letter_fraction, section, tags, body_selector, fetched_at, http_status, content_type, html_sha256, body_sha256, discovery, media_files_downloaded, full_text, date_published_local, publication_date_local, publisher_date_local, scope_timezone, authors, author, sections, publisher_name, publisher_url, genre, processed_at, publication_date_verified, provenance

Analysis variables (this paper). provenance, main (2026 sample), content_cluster, upload week; side (Iran, adversary, other) and tier (named leader, unnamed official, state, military, civilian, public) for every character, event and symbol; directed confrontations with AGENT; attributions of U.S. aggression with LEADER and NAMED; REP, EXC and HARM for civilian characters; perpetrator side, target type and the six-component suffering index for violent scenes; fantasticality and the victory, humiliation and victimhood flags; duration-weighted humor, ridicule, amusement and entertainment scores and the ridicule-lexicon rate; for H3 units, event topic, claim, sacred (religious authority, martyrdom), usref (domestic politics, popular culture) and grievance flags

Note. The frame table includes caption keyword flags derived for validation (cap_*). Detection records also store crop paths and batch metadata.

Table B2: Scene-level variables by family (284 variables)

Family	**Variables**
Identification, timing and segmentation	14
Setting, actors and narrative	18
Political topics	28
Emotions	21
Propaganda strategies (0–3)	36

Continued on next page

Table B2 continued

Family	Variables
Protagonist and antagonist attributes (0–3)	24
Composite rhetorical devices (0–3)	11
LEGO-medium affordances	23
Violence	19
Counts of characters and objects	28
Content orientation	5
Audio, speech, text and key symbols	17
Cinematic style and production	16
Narrative change markers	23
Coder uncertainty	1

Identification, timing and segmentation. video_id, scene_id, scene_order, start_timestamp, end_timestamp, start_seconds, end_seconds, duration_seconds, substantive_scene, estimated_shot_count, estimated_average_shot_duration, boundary_reason, transition_type, boundary_confidence

Setting, actors and narrative. raw_scene_description, setting, secondary_setting, indoor_outdoor, geographic_location, country, historical_period, time_of_day, realistic_vs_fantasy, main_characters, main_groups, protagonist, antagonist, main_action, event_type, narrative_function, causal_role, scene_narrative_contribution

Political topics. political_topics, countries_actors, pt_nation, pt_government, pt_leader, pt_military, pt_police, pt_political_party, pt_opposition, pt_war, pt_terrorism, pt_foreign_relations, pt_territory, pt_borders, pt_protest, pt_revolution, pt_religion, pt_ethnicity, pt_national_identity, pt_national_history, pt_economic_development, pt_military_technology, pt_technology, pt_security, pt_international_organization, pt_collective_memory, pt_historical_conflict, pt_contemporary_geopolitical_conflict

Emotions. dominant_emotion, emotional_target, emotional_source, emotional_intensity, emo_fear, emo_anger, emo_sadness, emo_grief, emo_pride, emo_hope, emo_happiness, emo_excitement, emo_tension, emo_triumph, emo_contempt, emo_humiliation, emo_sympathy, emo_reverence, emo_amusement, emo_disgust, emo_determination

Propaganda strategies (0–3). national_glorification, leader_glorification, government_glorification, military_glorification, patriotism, national_unity, collective_mobilization, religious_legitimation, historical_legitimation, mythological_legitimation, martyrdom_sacrifice, victimhood, historical_grievance, enemy_demonization, enemy_dehumanization, enemy_ridicule, foreign_threat, domestic_threat, fear_appeal, anger_appeal, pride_appeal, hope_appeal, moral_superiority, military_superiority, technological_superiority, cultural_superiority, government_competence, government_benevolence, inevitability_of_victory, justification_of_violence, justification_of_policy, call_for_resistance, call_for_obedience, national_revival, collective_efficacy, propaganda_intensity_mean

Protagonist and antagonist attributes (0–3). protagonist_heroic, protagonist_brave, protagonist_strong, protagonist_competent, protagonist_protective, protagonist_patriotic, protagonist_righteous, protagonist_disciplined, protagonist_technologically_capable, protagonist_sacrificial, protagonist_unified, protagonist_resilient, antagonist_threatening, antagonist_violent, antagonist_cruel, antagonist_incompetent, antagonist_cowardly, antagonist_ridiculous, antagonist_immoral, antagonist_foreign, antagonist_treacherous, antagonist_chaotic, antagonist_inhuman_or_monsterized, antagonist_weak

Composite rhetorical devices (0–3). temporal_compression, sacred_secular_fusion, heroic_genealogy, enemy_transformation, moral_binary, weapon_sacralization, military_spectacle, population_unanimity, enemy_specificity, mythic_compression, hero_villain_visual_distance

LEGO-medium affordances. toy_aesthetic, visual_simplification, abstraction, miniaturization, fantasy, exaggeration, playfulness, humor, satire, ridicule, violence_without_graphic_realism, emotional_distancing, moral_simplification, hero_villain_visual_polarization, historical_reconstruction, mythological_reconstruction, military_reconstruction, gamification, memetic_quality, mass_uniformity, social_media_shareability, LEGO_medium_specificity, medium_specific_function

Violence. violence_present, threat_of_violence, weapon_present, weapon_used, physical_combat, gunfire, missile_attack, explosion_present, destruction_present, injury_present, death_present, detention_present, violence_intensity, graphic_realism, violence_graphic_gap, violence_framing, violence_perpetrator, violence_target, violence_outcome

Counts of characters and objects. character_count, civilian_count, armed_actor_count, soldier_count, religious_figure_count, political_leader_count, protagonist_count, antagonist_count, child_count, female_character_count, male_character_count, crowd_present, crowd_size_category, approximate_character_count, flag_

count, religious_symbol_count, weapon_count, firearm_count, missile_count, vehicle_count, military_vehicle_count, aircraft_count, ship_count, explosion_count, destroyed_building_count, text_element_count, banner_count, anachronism_count

Content orientation. informational_content, entertainment_content, persuasion_content, emotional_content, primary_scene_function

Audio, speech, text and key symbols. important_symbols, important_text, spoken_content_summary, music_present, music_type, music_mood, vocal_chant, narration, dialogue, major_sound_effects, audio_emotional_function, audio_top_tags, audio_music_share, audio_speech_share, audio_singing_share, audio_explosion_share, audio_gunfire_share

Cinematic style and production. animation_type, dominant_shot_scale, rapid_cutting, close_up_emphasis, low_angle_heroic_framing, high_angle_vulnerability, slow_motion, dramatic_framing, spectacle_level, camera_movement, shots_per_10_seconds, symbol_density, character_density, reused_footage, reused_footage_source, narration_visual_lag_seconds

Narrative change markers. event_begins, event_continues, event_ends, new_character_introduced, character_removed, new_setting, new_country, new_historical_period, new_conflict, new_topic, new_political_message, new_symbol, new_emotional_tone, violence_begins, violence_ends, music_changes, narrative_turning_point, protagonist_change, antagonist_change, time_shift, geographic_shift, transition_from_realism_to_fantasy, transition_from_political_to_mythic

Coder uncertainty. uncertainty_notes

B.2 Dictionaries used in the register comparison (H3)

Table B3 lists the term families behind the four H3 classifications. Matching is case-insensitive and word-bounded; an asterisk marks any word ending; "Islamic" is matched only when not followed by "Republic". A unit is a *justification claim* if it matches the justification family; it is *sacred* if it matches religious authority or martyrdom; it carries a *U.S.-oriented reference* if it matches domestic politics or popular culture; and a domestic-political reference is *grievance-framed* if the same unit also matches the grievance or blame family. The full regular expressions are in `code/28_h3_units.py` of the replication archive, and the source-blind classifier prompt that provides the second instrument is reproduced in Appendix D.

Table B3: Term families for the H3 classifications

Family	Terms
Justification claim	justif*, deserv*, legitimat*, retaliat*, revenge, payback, respond(ed/ing/s) to, response to, punish*, defend*, self-defen*, aggress*, crime*, criminal, condemn*, violat*, illegal, unlawful, war crime, rightful, right to, had no choice, forced to, provok*, unprovoked, terroris*, genocid*, massacre, slaughter, murder*, atrocit*, innocent, guilt*, blame*, responsib*, accountab*, evil, wicked, coward*, liar*, lie(s/d), lying, hypocri*, arrogan*, shame*, victory, we won/win, humiliat*, defeat*, sovereign*, dignity, honour/honor, resist*, stand(s/ing) firm/strong/up, will not bow, never surrender
Religious authority	Allah, God, Islam, Islamic (not "Islamic Republic"), Muslim*, Quran*, prophet*, Imam, Hussain/Husayn, Karbala, Ashura, Mahdi, mosque, shrine, jihad*, ummah, Shia, Shiite, Zainab/Zaynab, Ayatollah, Supreme Leader, cleric*, holy, sacred, divine, blessed, almighty, pray*, faith, heaven, paradise, fatwa, halal, haram, Ramadan, Islamic Revolution
Martyrdom	martyr*, shahid, shaheed, sacrific*

Continued on next page

Table B3 continued

Family	Terms
U.S. domestic politics	Congress, senat*, Democrat*, Republican*, GOP, MAGA, election*, midterm*, voter*, vote(s/d), ballot, campaign, impeach*, taxpayer*, tax(es), gas price*, price at the pump, pump, inflation, grocer*, cost of living, debt, deficit, Epstein, scandal, Fox News, CNN, MSNBC, Capitol Hill, White House, Oval Office, approval rating, poll*, America First, Biden, Obama, Vance, Hegseth, Rubio, Kash Patel, Bondi, Leavitt, Melania, Musk, DOGE, tariff*, Social Security, Medicare, veteran(s), VA, border, deport*, immigra*, Second Amendment, NRA, Supreme Court, No Kings, Pentagon budget, military budget, billion(s) of dollars/in aid, your money/sons/children/taxes, American people/families/citizens/mothers/workers/taxpayers, Ohio, Texas, Florida, California, Wall Street, stock market, Dow, Nasdaq
U.S. popular culture	hip-hop, rap(per/pers), diss track, Drake, Kendrick, Eminem, Taylor Swift, Kanye, Ye, Beyoncé, Jay-Z, Snoop, NFL, Super Bowl, NBA, LeBron, MLB, Hollywood, Netflix, Disney, Marvel, Star Wars, Batman, Superman, Avengers, TikTok, YouTube, Instagram, Twitter, meme*, podcast, Joe Rogan, Kardashian, Elon, Tesla, McDonald's, Burger King, Coca-Cola, Starbucks, Walmart, Amazon, iPhone, Xbox, PlayStation, Fortnite, Minecraft, LEGO, Barbie, wrestl*, WWE, UFC, Halloween, Thanksgiving, Black Friday, Las Vegas, casino, cowboy, Hank, country music, Grammy, Oscar*, Emmy
Grievance and blame (for the grievance-framed split)	tax*, price*, cost*, pay(s/ing), bill(s), billion*, trillion, your money, struggl*, poor, homeless, veteran(s), your sons, coffin*, body bag*, lie(s/d), lying, cheat*, corrupt*, scandal, Epstein, sacrific* your/American, while you, for Israel/a lie, not your war, your children, bleed*, die(s/d) for; and the blame terms liar, blame, fault, coward, loser, clown, puppet, weak, pathetic, shame, humiliat*, victim, suffer, pays, bleed, die, body bag, coffin, sacrific*
Named U.S. leaders (dropped in the "without named leaders" check)	Trump, Donald, Netanyahu, Hegseth, Vance, Rubio, Biden, Obama, Kash Patel, Bondi, Leavitt, Melania, Musk, Witkoff, Kushner

B.3 Descriptive lexicons

Table B4 lists the descriptive lexicons used for the topic timeline in Section 4.5, the ridicule-lexicon measure of humor in the H2c robustness check, and the dictionary counts of Table C9. Counts are normalized per 1,000 words; the regular expressions are in `code/09_h3_text.py`.

Table B4: Descriptive lexicons

Category	Terms
Shia-Islamic vocabulary	allah, islam, islamic (excluding "Islamic Republic"), muslim*, quran*, prophet*, imam, hussain/husein, husayn, karbala, ashura, mahdi, mosque, shrine, jihad*, ummah, shia, shiite, zainab, zaynab

Continued on next page

Table B4 continued

Category	Terms
Generic religious vocabulary	god, pray*, heaven, faith, holy, sacred, divine, blessed, almighty, soul*
Demonic imagery	satan*, devil*, demon*, antichrist, hell, evil
Martyrdom	martyr*, shahid, shaheed, sacrific*
Iranian leaders	khamenei, supreme leader, leader of the (Islamic) revolution, the leader, pezeshkian, araghchi, larijani, ghalibaf, qalibaf, mojtaba, baqaei, irgc commander, president of iran, iranian president, Iran's president, foreign minister
Iranian institutions	irgc, revolutionary guard*, armed forces, islamic republic, majlis, parliament, government of iran, iranian government, foreign ministry, judiciary, minist*, spokesman, spokesperson
Enemy leaders (named)	trump*, netanyahu, bibi, hegseth, vance, rubio, witkoff, kushner, kash patel, leavitt, donald
Enemy institutions	united states, u.s., us forces, washington, pentagon, white house, American government/administration/regime/forces/military, zionist, israeli regime, occupying regime, idf, tel aviv, mossad, cia, nato, imperialis*, global arrogance, the enemy
Ridicule lexicon (H2c)	loser(s), clown(s), coward*, joke*, pathetic, fool(s), foolish, idiot*, crybaby, liar(s), scared, beg*, humiliat*, laugh*, embarrass*, panic*, weak, white flag, bye-bye, clueless, delusion*, crazy, dumb, stupid, tantrum, cry*, shame*
U.S. domestic grievances	gas, gasoline, fuel price, oil price, price at the pump, pump, inflation, grocer*, taxpayer*, tax*, debt, economy, job(s), working class, worker(s), veteran(s), body bag(s), coffin(s), epstein, maga, america first, democrat*, republican*, congress, midterm*, approval, impeach*, no kings, israel first, aipac, lobby, your sons, your children, American people/families/citizens/soldiers/troops/mothers

C Example Records

This appendix reproduces the records behind Figure 3 for one video, "Lego video from IRAN Specially for Pete Hegseth" (IRN_YT_NMQ4rIncFJw), and one record from the state-media corpus. Long text fields are shortened, marked [. . .]; insults in the lyrics are bracketed as in the coded data. Each table shows a selection of fields; the full records are in the replication files.

Table C1: Triage record

Field	Value
video_id	IRN_YT_NMQ4rIncFJw
title	Lego video from IRAN Specially for Pete Hegseth
channel	AI IRAN
upload_date	20260402
duration_seconds	130.868
category	lego_animation
decision	full
confidence	HIGH
is_ai_generated_likely	True
lego_style_share_estimate	0.95
main_language	en
reason	Entire video is LEGO-style animation with rap lyrics; Iranian-voiced political content.
topic_summary	AI-generated LEGO-style rap diss track from the 'AI IRAN' channel insulting US Defense Secretary Pete Hegseth and threatening missile retaliation.

Table C2: Video summary record (selected fields of 89)

Field	Value
video_title	Lego video from IRAN Specially for Pete Hegseth
channel	AI IRAN
upload_date	2026-04-02
view_count / like_count / comment_count	1,197,768 / 41,799 / 4,100
video_duration_seconds	130.9
total_scenes / total_shots_estimated	17 / 79
principal_protagonist	CHAR_002 (Iranian rapper-commander) / ACTOR_IRAN
principal_antagonist	CHAR_001 (Pete Hegseth); CHAR_003 (Trump)
dominant_narrative	An Iranian rapper-commander roasts US Defense Secretary Pete Hegseth as a drunk, abusive, crusader-clown and Trump as a Baal-worshipping 'Epstein' associate, then presents Iranian missiles destroying US forces, bases and the White House as payback for Muslim victims.
dominant_emotion	contempt (first half) / triumph (second half)
dominant_propaganda_strategy	enemy_ridicule
enemy_ridicule_score / enemy_demonization_score / military_superiority_score	1.77 / 1.61 / 1.22
share_runtime_violent / max_violence_intensity	0.55 / 4
generative_ai_production	3
producer_type_inferred	pro-Iran AI-content channel ('AI IRAN', ~52k subscribers, crypto donation link); Iranian-aligned voice, state link not verifiable

Continued on next page

Table C2 continued

Field	Value
provenance (this paper)	Iran-aligned independent

Note. Video-level strategy scores are duration-weighted means of the scene scores (0–3).

Table C3: Scene record, scene S12 (selected fields of 284)

Field	Value
scene_id	S12
start_timestamp	00:01:28.3
end_timestamp	00:01:33.5
boundary_reason	emotional_tone_change
raw_scene_description	A woman cradling a baby among grieving civilians in a damaged street; a crowd of figures stands amid collapsed buildings; women in headscarves hold children in a tent camp; a woman cowers in a dark corner among debris. Lyric: '...you broke the wives, you broke the Muslims... every victim screaming in the dark.'
setting	Damaged street; ruins; displacement camp; dark corner
geographic_location	Iran (implied)
realistic_vs_fantasy	1.0
protagonist	CHAR_007
antagonist	ACTOR_USA
main_action	mourning
event_type	civilian_suffering
narrative_function	victimization
causal_role	justification
dominant_emotion	grief
emo_grief / emo_sympathy / emo_anger	3 / 3 / 1
violence_present / violence_intensity / graphic_realism	1 / 1 / 0
violence_framing	tragic
violence_perpetrator / violence_target / violence_outcome	USA (implied) / civilians / destroyed homes; displacement
victimhood / foreign_threat / justification_of_violence / enemy_ridicule	3 / 2 / 2 / 0
toy_aesthetic / fantasy / violence_without_graphic_realism	3 / 0 / 3
medium_specific_function	Toy civilians convey mass suffering and displacement without bodies or blood.
music_type	hip-hop / rap (diss track)
audio_top_tags	Music; Hip hop music; Reggae; Rapping; Afrobeat
dominant_shot_scale / shots_per_10_seconds	wide/medium / 7.69

Note. Scene S12 corresponds to frame (e) in Figure 3.

Table C4: Character record CHAR_001

Field	Value
character_id	CHAR_001

Continued on next page

Table C4 continued

Field	Value
character_name	Pete Hegseth
character_description	Grey-haired minifigure in suit or uniform; shown with tattoos, clown paint, drunk, crying; named 'P. Hegseth' on a targeting screen
actor_type	government_official
country_or_identity	American
first_scene	S01
last_scene	S17
number_scenes_present	13
estimated_screen_prominence	high
narrative_role	antagonist
protagonist_antagonist	antagonist
portrayal	ridiculed
ridiculous_score	3
threatening_score	1
violent_score	2
heroic_score	0
propaganda_function	enemy_ridicule; enemy_demonization
identification_confidence	HIGH

Table C5: A relationship, an event and a symbol record

Field	Value
Relationship R08	United States → victimizes → Muslim civilians (scene S12); "Destroyed homes; 'you broke the Muslims"'; function: victimhood; confidence: MEDIUM
Event EV04	missile_launch: Iran — launching_missile — Trump's 'crew'; instrument: ballistic missile; location: Iranian silo; outcome: missile in flight; meaning: Retaliation as payback; confidence: HIGH
Symbol SYM_001	Iranian national flag (patches, flags, on missiles); type: national_flags; actor: ACTOR_IRAN; first scene S01, 7 scenes; prominence: high; framing: positive; function: National pride; marks protagonists and their weapons

Table C6: Raw Whisper transcript and coded text segments

Raw transcript (Whisper large-v3-turbo, diarized)	Coded segment (text_audio table)
[00:00:00.910 –> 00:00:09.200] SPEAKER_01: Yo, Pete Hexeth, this one's specially for you. [. . .]	TA001, scene S02, 2.7–5.9 s, spoken, speaker CHAR_002: "P. Hegseth. This one's specially for you." Function: targeted address; target: CHAR_001; confidence: HIGH
[00:00:40.500 –> 00:00:48.940] SPEAKER_01: even guard Biden's inauguration, your own military said, next. Book called American Crusade, dreaming holy war, but reality check.	TA010, scene S06, 39.9–49.5 s, spoken, speaker CHAR_002: "Couldn't even guard Biden's inauguration, your own military said next. Book called American Crusade, dreaming holy war, but reality changed." Function: enemy_ridicule; religious othering; target: CHAR_001; confidence: HIGH
—	TA011, scene S06, 44.0–46.0 s, written_label: "AMERICAN CRUSADE" Function: religious othering; target: CHAR_001

Continued on next page

Table C6 continued

Raw transcript (Whisper large-v3-turbo, diarized)	Coded segment (text_audio table)
—	TA019, scene S09, 71.0–76.8 s, spoken: "You thought our missiles maxed 2,000 kilometers? Cute rookie mistake." Function: technological_superiority; target: ACTOR_USA

Note. Whisper misrenders the defense secretary's name; the coder corrected it against the on-screen label "TARGET: P. HEGSETH" (TA004).

Table C7: Frame record and YOLO26s detection, frame (c) in Figure 3

Field	Value
folder / sec	v_99e2a57721b4_2026-04-02_Lego_video_from_IRAN_Specially_for_Pe / 48
caption (Florence-2)	a lego man is singing into a microphone. he is holding a bunch of money in one hand and a microphone in the other. he has sunglasses on his face and has smoke coming out of his mouth. behind him there are several flags hanging from the ceiling. there are several other lego men standing around him.
ocr (Florence-2)	"Allah" in Arabic script, from the flag emblem
ocr_chars	4
n_lego_figure / maxarea_lego_figure	1 / 0.303
n_human / maxarea_human	0 / 0.000
cap_flag / cap_explosion / cap_missile / cap_gore / cap_soldier / cap_crowd	1 / 1 / 0 / 0 / 0 / 0
Detection: frame_filename / class_name / confidence	frame_48sec.jpg / lego_figure / 0.780
Detection: box x1, y1, x2, y2 (pixels; frame 640 × 360)	201, 8, 399, 360
Detection: normalized center and size	0.469, 0.511; 0.309 × 0.978

Note. cap_explosion is triggered here by the smoke in the caption; caption flags are keyword matches used only in aggregate.

Table C8: State-media article record

Field	Value
record_id	tehran_times:524321
outlet_name	Tehran Times
canonical_url	https://www.tehrantimes.com/news/524321/Iranian-missiles-hit-five-key-US-bases
title	Iranian missiles hit five key US bases
lead	TEHRAN - At least five key U.S. military bases in the region have come under heavy fire from Iranian missiles, as Iran continues to respond to the U.S.-Israeli attacks that targeted Tehran and other Iranian cities Saturday morning.
date_published	2026-02-28T09:53:49Z
section	Politics
tags	['Iran', 'Iranian missiles', 'US bases', 'Israel', 'United States']
language	en

Continued on next page

Table C8 continued

Field	Value
body_chars	1278
parse_status	ok
discovery_method	publisher_dated_sitemap
html_sha256	02e10e2266021c4a…

Table C9: Dictionary counts for the example article and the example video

Category	Article tehran_times:524321	Video IRN_YT_NMQ4rIncFJw
Words	258	350
Shia-Islamic	0	2
Martyrdom	0	0
Iranian leaders	0	0
Iranian institutions	0	0
Enemy leaders named	0	5
Enemy institutions	1	0
Ridicule	0	5
U.S. domestic grievances	0	6
Personalization share	0.00	1.00

Note. Counts from the descriptive lexicons of Table B4. The video document joins its English speech, lyrics and on-screen text.

D Computational Pipeline and Language-Model Coding

This appendix documents how the corpus was produced and analyzed, with particular attention to the two places where large language models enter the study: the scene coder that produced the interpretive codes, and the agentic sessions that implemented and executed the estimation code. Everything described here is in the replication archive; file names refer to it.

D.1 Stages and compute

Table D1 lists the stages in order, with the software used at each and where it ran. Stages 2 to 4 ran locally on one workstation with an NVIDIA GeForce RTX 5080 GPU and an Intel Core Ultra 9 285K processor and took about one month of wall-clock time in total, most of it in frame captioning and detection over 96,530 frames and in transcription. The scene coder (stage 5) ran as an unattended agent in a cloud container, in scheduled batches, between 16 and 22 October 2026. Corpus assembly and the statistical analysis (stages 6 and 7) run on an ordinary laptop in under an hour.

Table D1: Pipeline stages

	Stage	Software and settings	Output
1	Collection and screening	Search on YouTube, TikTok and X for LEGO-style videos presented as Iranian or about Iran; download with platform metadata (title, description, channel, upload date, views, likes, comments); two-stage triage (Appendix D.2).	549 candidates; 368 coded; 363 in the 2026 sample
2	Frame extraction	`ffmpeg`, one frame per second, 640×360, for every video in the 2026 sample.	96,530 frames
3	Frame layer	YOLO26s detector fine-tuned by the authors to separate brick figures from photographed humans (confidence 0.5 for analysis); Florence-2-large-ft for a detailed caption and OCR of every frame; keyword flags derived from captions for validation. GPU batches of frames; logs per batch.	Detection and caption tables (Table B1)
4	Audio layer	Whisper large-v3-turbo after voice-activity detection, with speaker diarization; AudioSet tagging in five-second windows; loudness per second; shot-boundary detection.	Transcripts for 358 videos; audio tags
5	Scene coding	Claude Opus 5.5 (model identifier `claude-opus-5-5`, high reasoning effort) under the 40-stage protocol, run unattended in scheduled batches; deterministic quality control (`build_workbook.py`).	One workbook per video: 18 tables, 570 variables
6	Corpus assembly	`corpus.py`: concatenates the per-video workbooks, checks identifiers, and writes the corpus tables; `01-04` of the analysis code add provenance, content clusters, scene indices, and frame aggregates.	`corpus_*.csv`; `data/*.pkl`
7	Estimation and figures	Python scripts `24_h1_tests.py` to `30_fig_h3.py`, wrapped in `run_empirical_methods.py`; seed 2026.	`out/*.csv`, `out/*.json`, `fig/*.png`

D.2 The scene coder

What the coder received: Each video was first *triaged*: the agent saw twelve frames sampled across the full length, the platform metadata, and the first 90 seconds of transcribed speech, and recorded a category, a decision, a confidence, whether the video is likely AI-generated, the estimated share of LEGO-style footage, the main language, and a one-line topic summary (Table C1). Videos retained at triage were then *fully coded*. For full coding the agent received timestamped contact sheets of sixteen frames each, at one frame per second for videos up to ten minutes (sparser for

longer videos); the Whisper transcript with speaker labels and timestamps; the AudioSet tags per five-second window; detected shot boundaries; and the platform metadata. The coder therefore worked from sampled frames and text, not from continuous video or audio: the protocol instructs it to use the audio tags as evidence for music, explosions and gunfire, and never to invent sound it cannot hear. The agent was given the project's research questions (what stories the videos tell, who the protagonists, and victims are, how violence is represented, what the toy medium permits, including whether it allows intense violence with little graphic suffering) and the full codebook, but not the eight hypotheses of Section 5, which were formulated after the protocol was fixed and tested only after coding was complete.

The protocol: The research prompt (`_pipeline/research_prompt.txt`, about 5,200 words) defines forty stages: scene segmentation by narrative unit rather than shot (stage 1); basic scene information and a raw scene description written as observation only, with evaluative words such as "propaganda", "heroic" or "villain" prohibited (stages 2–3); characters, counts, relationships, and events with controlled vocabularies (stages 4–7); narrative function, causal role, and protagonist and antagonist construction (stages 8–11); political topics, the 35 propaganda strategies, emotions, violence, and object counts (stages 12–16); symbols, text, audio, and cinematic style (stages 17–20); the LEGO-specific affordances and composite devices (stages 21–24); the dataset specifications (stages 25–35); and reflective stages on the medium, comparison with conventional media, recommended variables, and automation (stages 36–40), which are stored but not used in the analysis. Ordinal codes run from 0 to 3 and are scored symmetrically for every actor; an omitted code means "judged absent" and is stored as 0, while an indeterminate code is stored as missing, and the protocol forbids coding uncertainty as absence. A calibration example of a finished coding was supplied with the protocol.

Execution and quality control: A runbook (`_pipeline/RUNBOOK.md`) governed each scheduled run: the agent pulled the next items from a queue, staged the video, and its metadata, ran the preprocessing script, viewed every contact sheet before deciding scene boundaries, wrote the coding as a JSON document, and ran `build_workbook.py`, which rejects any coding whose scenes leave gaps or overlaps in the timeline, whose first scene does not start at zero or whose last scene does not end at the video's duration, or which references a character, symbol, scene or event identifier that is not defined in its own table. The agent corrected its coding until the check passed; the builder then wrote the per-video workbook with the model identifier and coding date recorded in every file. Runs handled at most six videos in 75 minutes; the queue recorded videos that failed twice for attention, and none did. The agent could not ask questions during a run and did not see any other video's coding except the calibration example. Across the 549 candidates, 181 were triaged out, 368 were fully coded, and the ordering of coding was by queue position, not by provenance or topic.

D.3 The source-blind text classifier

The second instrument for the H3 classifications is a language-model classifier applied to a random sample of 1,200 units, 600 video segments and 600 article sentences. The units were shuffled, stripped of every identifier and split into six files of 200 lines containing only an anonymous id and the text. Each file was classified by a separate agent instance of Claude given the prompt below and nothing else; the outputs were joined back to the units by the anonymous id. Agreement with the dictionaries is reported in Section 6.3 (Cohen's κ 0.47 for claims, 0.59 for sacred legitimation and

0.52 for U.S.-oriented reference; dictionary recall against the classifier 0.47, 0.73 and 0.52). The prompt, reproduced verbatim from `data/h3_llm/PROMPT.md`:

> You are classifying short text units about the 2026 Iran–US/Israel war. Read the JSONL file given to you; each line has an "id" and a "text". You know nothing about where a unit comes from; judge the text alone.
>
> For EVERY unit, output one JSON line: {"id": "...", "claim": 0/1, "sacred": 0/1, "dompol": 0/1, "popc": 0/1}
>
> Definitions (apply the same way to every unit):
>
> - claim = 1 if the unit makes a political-justification claim: it justifies, legitimates, condemns or assigns blame for a party's conduct in the conflict (retaliation presented as deserved, aggression condemned, a victory presented as rightful, suffering presented as a grievance, an actor called a criminal, liar, coward, aggressor, terrorist, or similar). Plain description of events, logistics, unrelated news, headings, or lyric fragments with no evaluative content = 0.
> - sacred = 1 if the unit invokes religious authority or sacred legitimation: God/Allah, divine will or blessing, prayer, holy war, Islam or Shia references (Imam, Hussein, Karbala, Ashura, Mahdi, Quran), religious clerics as authorities (Ayatollah, Supreme Leader in a religious role), martyrdom or sacrifice in a religious sense. Mentions of a mosque as a building, or "Islamic Republic" as a state name alone, = 0.
> - dompol = 1 if the unit refers to US domestic politics: elections, parties, Congress, MAGA, polls, scandals (e.g. Epstein files), taxes, prices, inflation, the cost of the war to Americans, domestic political figures in a domestic role, cable news, etc. A US official acting in foreign policy only (e.g. "the US president ordered strikes") = 0.
> - popc = 1 if the unit refers to US popular culture: music genres and artists, films, TV, sports, brands, celebrities, memes, social media platforms, games, consumer culture.
>
> Rules: output exactly 200 lines, one per input id, in input order, and nothing else. Use the Write tool to write them to the output path given, as a .jsonl file. Do not skip units; if a unit is empty or unintelligible, output all zeros.