

Visual Framing in the AI Era: Lessons from Manual Approaches for Computational Methods

Kareem El Damanhoury

Department of Media, Film & Journalism Studies, University of Denver, Denver CO USA

Carol Winkler

Department of Communication, Georgia State University, Atlanta GA USA

Ayse D. Lokmanoglu

Emerging Media Studies, Boston University, Boston MA USA

Keyu Alexander Chen Glanz

Department of Communication, Georgia State University, Atlanta GA USA

Abstract

Computational methods can minimize the time and resources needed to manually code thousands of images. Yet, they also come with challenges, including validation, algorithmic bias, and privacy concerns. Acknowledging that the pictorial turn has now entered a computational phase, this article reports on a manual and automated coding of 7000+ images to better understand online extremist content. Using Rodriguez and Dimitrova's (2011) four-tiered model of visual framing, the study compares manual and OpenAI's ChatGpt4o's coding of Al-Qaeda and ISIS images across the denotative, semiotic, connotative, and ideological levels. AI coding exhibited moderate to strong performance on denotative variables but was weaker in the semiotic and connotative tiers. The study concludes with a discussion of the advantages of human and AI functioning together to better understand visual framing.

Keywords: AI; Visual Framing; Extremism; Photography; ISIS; Al-Qaeda

Introduction

An explosion of visual images characterizes the 21st century public sphere. Individuals capture 57,000 photos every second (Broz, 2023), and daily share over 3 billion images on social media (JasenkaG, 2023) and one million memes on Instagram (Mcgil, 2024). Out of the more than 400 million daily terabytes of online data, video content accounts for more than half (Duarte,

2024). On YouTube alone, users upload more than 500 hours of video every minute (Ceci, 2024). Today, visuals constitute an imperative mode of messaging for virtually any communicator seeking to reach target audiences (Knobloch et al., 2003).

Visuals are also key to strategic, persuasive messaging. Compared to text, visuals appear closer to the truth, rendering them a ready means of proof and authentication (Barthes, 1981; Messaris & Abraham, 2001). They communicate ideological propositions (Edwards & Winkler, 1997), create positive images of political causes (Sontag, 2003), evoke emotional responses (Perlmutter, 1998), constitute cultural memories and meanings (McClancy, 2013; Trachtenberg, 1985), and mobilize supporters (Hariman & Lucaites, 2007; Mattoni & Teune, 2014).

While manual coding methods have helped better understand visual messaging strategies for decades, the sheer number of digital images poses a challenge. By minimizing time and resources needed to manually code images, computational methods are more efficient. Unsupervised learning can help reveal visual themes and categories within big datasets via a bag-of-visual words model (Torres, 2023; H. Zhang & Peng, 2022). Scholars have used such automated approaches to detect clusters in politicians' social media posts (Joo & Steinert-Threlkeld, 2022; Peng, 2021), smartphone screen activities (Muisse et al., 2022), online protest images (H. Zhang & Peng, 2022), ring-wing memes (Lokmanoglu et al., 2023), and migrant photographs (Torres, 2023). Supervised machine learning, by contrast, involves training an AI model using labeled visual data before examining similar imagery. Studies have used this approach to identify protest depictions (H. Zhang & Pan, 2019), image-text mismatches (Ha et al., 2020), differentiations between state and protester violence (Steinert-Threlkeld et al., 2022), and objects and faces including their emotions, gender, age, and visual aesthetics (e.g., Bakhshi and Gilbert, 2015; Dietrich and Ko, 2022; Peng, 2018; Peng and Yingdan, 2023). Nonetheless, computational methods also come with their own fair share of challenges, such as validation, algorithmic bias, and privacy concerns (K. Chen et al., 2024; Y. Chen et al., 2024; Williams et al., 2020; Zou & Schiebinger, 2018).

Acknowledging that W. J. T. Mitchell's (1995) pictorial turn has now entered a computational phase, this article manually examines several thousand images that Islamist extremist groups distributed to provide insights for improving AI's usefulness in understanding online extremist content. After highlighting the usefulness of visual framing practices and discussing key challenges facing computational visual analyses today, we explicate the

study's comparative methodology. Then, we lay out five lessons emergent from the comparison useful for improving visual computational methods. The study concludes with a discussion of how mixed visual methods of AI and manual coding can bring better understandings to the levels of visual framing.

Visual Framing

Rodriguez and Dimitrova's (2011) four-tiered model is one of the most cited works in the literature on visual framing (Walter & Ophir, 2024), as it moves beyond atheoretical or exclusively semiotic approaches to engage with how image selection, cropping, and editing can capture attention, evoke emotions, carry meanings, and influence perceptions (R. Coleman, 2010; Geise & Baden, 2015). Their model disaggregates visual framing into four tiers. The denotative level constitutes meaning by identifying salient frames through who and what an image depicts based on the scene and surrounding texts (Schwalbe, 2006). Denotative frames can be context-specific (Goffman, 1974, 1979) or function as equivalency-based dyads, such as gain versus loss and war versus peace (Cacciatore et al., 2016). The semiotic level suggests social meanings by examining visual grammar and conventions, including camera angles, perceived viewer distance, body posture, eye contact, and facial expressions (Forgas & East, 2008; Hall, 1966; Kress & Leeuwen, 2006). The connotative tier considers abstract and figurative symbols (e.g., metaphors) in the shot that can "combine, compress and communicate social meaning" (Rodriguez & Dimitrova, 2011, p. 96). Finally, the ideological level expands on the visual elements, stylistic choices, and symbols to provide a holistic interpretation that underscores the political, religious, economic and/or demographic underpinnings of the visual constructions. Combined, the four-tiered model allows for a nuanced understanding of visual messaging.

Studies of visual media campaigns by Islamist extremist groups have examined all four levels of the 4-tiered model, albeit with less emphasis on semiotics. At the denotative level, for example, Al-Qaeda's visual campaign encompassed predominantly militant visual frames, such as training, operations, and martyrdom (Center, 2005; Farwell, 2010), highlighting political Islamist connotations associated with objects like flags and swords (A. Coleman, 2006). At the semiotic level, studies of ISIS dissected several pictorial stylistic conventions, including viewer distance, camera angle, eye contact, facial expressions, point-of-view shots, and dynamic versus static imagery (e.g., Impara, 2018; Winkler, El-Damanhoury, et al., 2019). At the connotative level, ISIS and al-Qaeda utilized symbols like AK47s, the monotheism hand

gesture, and depictions of death and dying to convey meaning (Wignell et al., 2017; Winkler et al., 2018). Ideologically, both al-Qaeda and ISIS espoused an extreme Islamist lens steeped in a clash of civilization narrative that promoted religious rule as an alternative to the nation state system (Ciovacco, 2009), but had differing views on the nature and timing of the Caliphate (Kuznar, 2015). Yet, existing literature on al-Qaeda and ISIS's media campaigns, while providing nuanced understandings of the four visual framing tiers, has been mainly limited to manual coding.

Computational Visual Analysis

Computational visual analysis on its own is not yet capable of fully engaging the four-tier model of visual framing. Computational attempts identify some denotative and semiotic elements (e.g., humans, race, age, color, public figures, rifles, umbrellas, facial expressions) (K. Chen et al., 2022; Joo & Steinert-Threlkeld, 2022; Muise et al., 2022; H. Zhang & Peng, 2022). Yet, such analyses stop short of gauging key symbolic and ideological visual components of scenes and their contexts due to what Peng, Lock, and Salah (2024) rightly argue is an automation-theoretical disconnect. This study addresses this gap by comparing manual and automated coding in the online extremism sphere in pursuit of a more effective, hybrid approach.

Traditional computational image analysis typically relies on task-specific computer-vision architectures such as YOLO-style object detectors, Mask R-CNN segmentation models, or transformer-based vision-language models like BLIP-2 or Flamingo. These systems are designed to extract concrete visual features (e.g., objects, faces, scene layouts) and perform discrete tasks with high accuracy when trained on large, labeled datasets. However, they are not built to apply multi-layered interpretive schemas such as Rodriguez and Dimitrova's four-tier model without extensive task-specific fine-tuning. Because our study evaluates GPT-4o in a zero-shot setting—asking a general-purpose vision-language model to follow a human-developed codebook—we position this approach as complementary to, rather than a replacement for, traditional CV pipelines.

AI changes related to the scanning, sampling, and quantization of visual images are rapidly evolving, rendering accurate summations of developments difficult (X. Zhang & Dahu, 2019). Complicating the quickly changing terrain is that machines are now producing most images, often for other machines rather than the human eye (Paglen, 2019). Nonetheless, AI remains data-driven, rather than image-driven, meaning that understanding image-data relationships should remain a priority (Anderson, 2017).

Whether computerized or not, visual cultures influence and are influenced by human biases in both production and consumption of online messaging (Bridle, 2023; Sezen, 2020). Combining quantitative and qualitative analyses of big data can augment visual framing. Dondero, for example, maintains that large-scale quantitatively produced diagrams produced through quantitative and semiotic analysis can assist in identifying “contrasting areas, opposite areas, or superposing of images on the plane of expression” useful for further quantitative and qualitative analysis (Dondero, 2019, p. 140). Such a combined approach preserves the importance of visual context within a specified corpus and across the image components. In short, she advocates for scholars to use her process as a metavisual device for four reasons:

(1) these visualizations are images of images; (2) the parameters used to arrange them are visual; (3) the automatic distribution of the images is visualized spatially in a presentation governed by abscissas and ordinates; (4) the content analysis... remains within the realm of images (filiation, tradition, citation, genre, etc.) and not the abstract realm of verbal description (Dondero, 2019, p. 141).

Here, we agree with Dondero (2019) about the value of computerized quantitative analysis for assisting qualitative results. However, we add that quantitative human coding analysis, combined with statistical assessments, can further strengthen the tracked meaning of results. We begin by asking:

RQ1: How effectively can a vision-language model (GPT-4o) apply a human-developed content-coding schema to Rodriguez and Dimitrova’s four-tiered visual framing analysis?

Large language models and NLP pipelines can efficiently process large text corpora, grouping semantically similar responses and surfacing recurring patterns. Gamielien et al. (2023) find that transformer-based tools can generate highly granular codes across thousands of student reflections, substantially reducing human labor. This aligns with earlier infrastructure-oriented work showing NLP can automatically classify predictable text categories. Similarly, Morgan (2023) reports that ChatGPT performs well when themes are concrete and descriptive, requiring little interpretive inference. In hybrid interfaces, rule-based suggestions can be systematically extended to unseen data (Rietz & Maedche, 2021), increasing agreement rather than replacing human interpretation and underscoring that automation is most reliable for patterned, literal, and structurally evident meaning.

Despite scalability gains, current AI systems appear to underperform when meaning depends on tacit knowledge, ambiguity, or socio-cultural interpretations. Gamielien et al. (2023) note the need for AI researcher oversight when semantic nuance matters. Studies of thematic automation note that disagreement among humans themselves reflects interpretive pluralism (Armstrong et al., 1997; Mackieson et al., 2019) — something models are poorly equipped to resolve. Transformer architectures excel at long-range dependencies (Lakretz et al., 2020), but still primarily attend to textual features rather than framing context, affective tone, or symbolic cues. These limitations indicate that interpretive coding requires judgment beyond probabilistic associations, particularly in indexical, connotative, or historically situated domains.

Across studies, researchers express caution about fully delegating thematic interpretation to automated systems. Marathe and Toyama (2018) report reluctance rooted in opacity, loss of theoretical accountability, and few questioning opportunities (N. Chen et al., 2018). Rietz and Maedche (2021) similarly find researchers use automated suggestions not to accelerate coding, but to reflect on needed codebook refinements. Such reflection aligns with iterative qualitative traditions emphasizing continuous interpretation (e.g., Braun and Clarke, 2006; Saldana, 2021 cited in Gamielien et al., 2023). Thus, for tasks requiring contextual inference, socio-cultural reading, or interpretive framing, human coders continue to outperform computational models. Because visual framing often requires reading symbolism, composition, affect, and implied narratives, we ask:

RQ2: Which dimensions of visual framing remain resistant to automation, and what do these limitations reveal about the strengths of human coding in visual analysis?

Methodology

To assess the effectiveness of GPT-4o for applying a human-developed content-coding schema to visual framing analysis, we began by conducting a human-coded content analysis of 7,292 images from al-Qaeda and ISIS's English and Arabic magazines or newsletters distributed 2009-2020 (see Table 1). For al-Qaeda, the English issues included *Inspire* (1-17) and *Jihad Recollections* (1-4), and the Arabic issues were *al-Masra* (1-57). ISIS's English issues included *Dabiq* (1-15) and *Rumiyah* (1-13), and *al-Naba* (1-229) in Arabic. All items were publicly available through Google, *Jihadology* (Zelin, 2021), or *archive.org*.

Group	Publications	Frequency	Percent	Total / % of group image count
AQ	Jihadi Recollections	442	6.06	3466 (47.5%)
	Inspire	1842	25.26	
	Al Masra	1182	16.21	
ISIS	Dabiq	1391	19.08	3826 (52.5%)
	Rumiyah	273	3.74	
	Al Naba	2162	29.65	
Total across groups		7292	100.0	

Table 1: Image Count in Al-Qaeda and ISIS Online Publications

We utilized 13 expert coders from Egypt, Afghanistan, Turkey, Saudi Arabia, Syria, Poland, Vietnam, and the United States to create, refine, and utilize a visual analysis codebook. Our human coders had doctorates or graduate training in Communication Studies, Psychology, Political Science, and Education. The pilot phase involved three US and Egyptian coders who created coding categories inductively from images in Dabiq’s first issue until intercoder reliability was higher than 0.80 on Cohen’s kappa for all variables. Coders met weekly to identify and resolve discrepancies and cross-cultural differences that produced unacceptable reliability levels. In cases of disagreements, a bias toward the Middle Eastern perspective prevailed in line with the primary target audience. With a reliable codebook, 13 coders received oral and written training and analyzed each image in the dataset. The average intercoder reliability score using Cohen’s kappa across all categories was 0.91 (see Table 2). A third coder resolved discrepancies for statistical analysis.

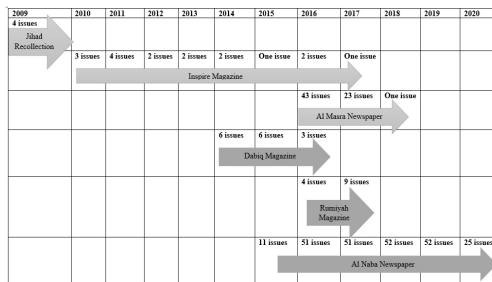


Figure 1: Historical Overview of Study's Al-Qaeda and ISIS Publications

Variable	Description of coding clusters & categories	% Agree	Cohen's κ
<i>Denotative categories</i>			
Military role	Extremist militants; enemy militants; mixed; none	95.9	0.92
Death	About to die; dead bodies; not applicable	94.8	0.93
Flags	ISIS/AQ flag; U.S. flag; MENA country flag; other; mixed; not applicable	97.2	0.93
Humans	1 human; 2-10 humans; > 10 humans; no humans	95.1	0.93
Destruction	Presence of fire, explosions, or other acts of destruction in process; destroyed buildings, bridges, or vehicles; not applicable	97.9	0.92
Leaders	Jihad leaders; Western leaders; Arab state leaders; Asian/Russian leaders; other Muslim leaders; African leaders; mixed; none	97.4	0.91
<i>Semiotic factors</i>			
Viewer position	Viewer looking up at photo subject; looking down; eye level; not applicable	95.6	0.88
Facial expressions	Positive; negative; unclear; not applicable	95.0	0.89
Stance	On knees (not praying); sitting; standing; laying down; praying (on knees or bending over); mixed; not applicable	97.2	0.96
Image position	Foreground; background	97.2	0.86
Viewer distance	Intimate/personal distance (< 4 ft.); social/public distance (> 4 ft.); mixed; unknowable (e.g., infographics, posters, maps, banners, or non-photographic images)	89.4	0.87
Eye contact	Photo subject looks directly at viewer; looks up; looks down; looks away; not looking; not applicable	95.5	0.92
<i>Connotative factors</i>			
Governance	State building (e.g., social services, market, currency, passports, maps, natural landscape); law enforcement; allegiance pledges; media/public information; mixed; not applicable	97.7	0.91
About to die	Certain death; possible death; presumed death; none	97.3	0.94
Religion	Tawhid gesture; reading Qur'an/Qur'anic texts; individual prayer; Hajj; religious iconography or shrines; mixed; none	98.3	0.88

Table 2: Intercoder reliability and description of the manual coding instrument

We sorted our human coding categories into four levels of visual framing. While Rodriguez and Dimitrova (2011) note that the four tiers can overlap, we focused on the level that three coders considered most suited to the images' characteristics. The meaning of denotative elements (or objects that bore an indexical relationship with an individual, thing, or place) involves two interrelated processes. First, the coding process accounts for elements the viewers can see in the shot, rendering them "closer to the truth than other forms of communication" (Messaris & Abraham, 2001, p. 217). Fully gauging the denotative meaning, however, requires generating frames that derive from a second process that looks not only at the textual context (Rodriguez & Dimitrova, 2011), but also into the syntactic relationships between images. Compared to words, images lack an explicit prepositional syntax, or the ability to propagate clear causal relationships, similarities, or other forms of connections (Messaris & Abraham, 2001). Our denotative categories included military fighters, death, humans, leaders, flags, and destruction.

The second level of semiotics (or stylistic, technical, and subject conventions) focuses on "signs and symbols, sign systems, and sign processes" (Moriarty, 2002, p. 20). Visual semiotics fulfills three meta functions: compositional, representative, and interactive (Kress & Leeuwen, 2006; Kress & Leeuwen, 1996), as it emphasizes four types of signification: arbitrary (by convention), memetic (by iconic representations), evidential (by cues and codes), and signaling (by recognition) (Moriarty, 2002). Metaphors or metaphorical thinking (D. Feng & O'Halloran, 2013), and visual metonymics linked to abstract concepts and objects/events are also associated with semiotics (W. Feng, 2017). Accordingly, our semiotic categories included viewer position, image position, viewer distance, eye contact, and facial expressions.

The third framing level of connotation involves visual symbols linked to ideas or concepts associated with individuals, things, or places. Turner insists symbology draws its data from "cultural genres or sub systems of expressive culture. . . as well as narrative genres, such as myth, epic, ballad, the novel and ideological systems [and t]hey would also include non-verbal forms" (Turner, 1979, p. 12). Symbols can allude to power, authority, faith, rituals, and death, among others, to achieve personal or group goals (Turner, 1974). At the iconographical level, visual symbols go beyond the depicted object or person to connote ideas and concepts; they begin to reveal ideological meanings derived from backgrounds and the surrounding context (Leeuwen, 2001; Panofsky, 1955). Our connotative categories included about to die, religion, and state, with the latter disaggregated into state-building,

law enforcement, allegiance pledges, and media propaganda for a more fine-grained analysis (see Appendix A for category meanings; Table A1 for examples).

We removed several manual coding categories from our computational model. We excluded image size and position because the workflow operated on individual images rather than publication layouts. We omitted gender because the overwhelming number of individuals displayed were male, with females only as occasional outliers. We removed age and social infrastructure as neither produced significant results across more than a dozen papers addressing how message strategies intersected with situational factors.

To compare manual content coding with AI visual coding, we built a lightweight, fully reproducible inference pipeline that (1) ingested image files, (2) encoded each image and sent it to GPT-4o together with a structured labeling prompt derived from our codebook, and (3) compiled the model's outputs into a standardized dataset for evaluation against human annotations. The pipeline did not fine-tune model weights; it relied on constrained prompting and schema validation to ensure consistent, interpretable results (see Figure 2).

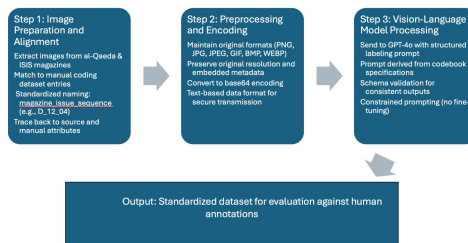


Figure 2: AI Visual Coding Inference Pipeline

Step 1: Image Preparation and Alignment

All images were extracted from the full corpus of al-Qaeda and ISIS publications and matched to their manual coding entries. Each file was saved using a standardized naming convention that included publication, issue number, and image number (e.g., D_12_04), allowing a one-to-one linkage between visual material and its metadata to ensure tracking capacity back to images' sources and manually coded attributes.

Step 2: Preprocessing and Encoding

The repository contained PNG, JPG, JPEG, GIF, BMP, and WEBP formats. Images were maintained at their original resolution without resizing or alteration of embedded metadata to preserve visual detail integrity. Each image was converted into a text-based data format through base64 encoding, which allows secure transmission of visual information as text while retaining the original pixel structure (Lokmanoglu & Walter, 2025). Detailed logs documented each image's processing status to ensure completeness and traceability.

Step 3: Model Inference and Structured Prompting

We provided each encoded image directly to GPT-4o (OpenAI, 2024) along with a structured prompt adapted from the visual framing codebook. The prompt specified the exact coding categories (e.g., military role, human figures, eye contact, state-building, and religious symbolism) and required the model to classify each image according to those predefined labels (see Appendix B). Category definitions were embedded in the prompt to guide consistent decision-making. The model's responses were constrained to a fixed output schema composed of numeric identifiers corresponding to each category.

This study employed a zero-shot prompting design. The model received the structured labeling prompt and visual input simultaneously without prior exposure, calibration subset, or iterative refinement. The objective was not to train or improve GPT-4o's performance but to evaluate how an off-the-shelf vision-language model applied an existing human content-coding schema. Each API call contained a single image and the associated codebook prompt without captions, metadata, or textual context.

To ensure the AI response integrity, we monitored outputs for potential misclassifications or omissions related to sensitive or graphic imagery. System logs were reviewed after each coding run to identify any uncoded images flagged as indeterminate or potentially due to ethical safeguards. No systematic filtering or suppression of graphic content occurred, but the logging process allowed for post hoc verification should future discrepancies arise.

Step 4: Output Aggregation and Reliability Assessment

The model's outputs were compiled into a unified dataset and compared with the manual coding results using standard reliability and performance

metrics, including precision (the proportion of correct positive predictions), recall (the proportion of actual positives correctly identified), and F1-score (a recall-weighted harmonic mean of precision and recall). We also report F2-scores, macro-averaged and per-class metrics, as well as two measures of intercoder reliability between AI and human coders. Reports of percent agreement served as an intuitive measure of alignment.

Following automated content analysis research standards, F1 scores above 0.80 indicated excellent agreement between AI and human coding, scores between 0.70 – 0.80 represented acceptable performance, and scores between 0.60 – 0.70 suggested moderate agreement (Burscher et al., 2014; Chan et al., 2021). For Cohen's κ , values above 0.80 were considered reliable, while values between 0.67-0.80 were deemed acceptable for exploratory research (Krippendorff, 2018). These benchmarks were particularly relevant given that human inter-coder reliability in visual content analysis typically ranges from 0.70-0.85 (Song et al., 2020).

Findings

RQ1 asked how effectively a vision-language model (GPT-4o) could apply a human-developed content-coding schema to Rodriguez and Dimotrova's four tiers of visual framing. The AI coding performance differed in substantial ways across variables and framing tiers (see Table 3 and Appendix B Table B1). Denotative variables showed the strongest alignment between AI and human coding. Humans, Destruction, Leaders, and Flags demonstrated the most consistent performance across metrics. Humans yielded an F1 of 0.80 and a Balanced Accuracy of 0.83. Destruction performed similarly (F1 = 0.79, Balanced Accuracy = 0.83). Leaders and Flags showed moderately strong performance, with F1 values generally ranging from 0.70 to 0.78 and Balanced Accuracy typically above 0.75. Krippendorff's α values for these variables were moderately reliable, ranging from approximately 0.40 to 0.60. As shown in Figure B1, these categories were dominated by absent cases, inflating agreement due to prevalence rather than consistent recognition of positive cases. Weighted F1 values were very close to F1 for denotative variables, suggesting that class imbalance had limited influence on AI performance in this category.

Denotative variables requiring additional contextual interpretation showed more modest alignment. Military Role achieved an F1 of 0.38, a Balanced Accuracy of 0.73, and a Krippendorff's α around 0.30, indicating reliability for distinguishing between combatants and non-combatants below acceptable thresholds. Death showed somewhat higher performance (F1 = 0.52,

Balanced Accuracy = 0.65), although still below that of other denotative variables. For both Military Role and Death, weighted F1 scores were notably higher than F1, indicating that performance was disproportionately driven by the majority “not applicable” class and the model struggled to identify less frequent positive cases.

Semiotic variables requiring interpretations of bodily, expressive, or relational cues showed weaker correspondence. Viewer Position, Viewer Distance, Eye Contact, Stance, and Facial Expression produced F1 values generally 0.60-0.72, Balanced Accuracy values 0.72- 0.76, and Krippendorff’s α around 0.20 to 0.40, falling below the 0.67 acceptable threshold for exploratory research. Across all semiotic variables, Balanced Accuracy consistently exceeded F1, indicating that while the model could distinguish broad classes, it did not reliably identify positive instances. Weighted F1 values were consistently higher than F1, confirming that the predominance of absent codes influenced performance. These semiotic findings suggest that discerning gaze direction, bodily posture, or facial expression requires contextual and relational sensitivity that current image models do not robustly encode.

Connotative variables tied to symbolic or ideological meanings showed the weakest correspondence. State-building performed poorly (F1 \approx 0.24, Balanced Accuracy \approx 0.43, Krippendorff’s $\alpha \approx$ 0.05 to 0.10), indicating minimal reliability. About to Die and Religion displayed comparably low performance, with α values often approaching zero. Weighted F1 values for connotative variables were only slightly higher than F1, suggesting that performance limitations stemmed not only from class imbalance but also from the framing elements’ conceptual and inferential nature. Connotative categories contained very few positive cases, limiting model exposure and contributing to low performance (see Figure B1). While the model detected overt symbolic cues, it struggled with implicit or context-dependent signals of state-building, martyrdom, or religious practice that require interpretive inference beyond visual pattern matching.

RQ2 asked which dimensions of visual framing remain resistant to automation, and what these limitations reveal about the strengths of human coding in visual analysis.

Lesson #1: Denotative Interactions

Despite AI’s stronger performance at the denotative level, inaccuracies and omissions were present. Military roles, for example, showed only modest agreement levels, as manual coders examined taglines and accompanying text to distinguish al-Qaeda and ISIS militants from enemy fighters.

Variable	Number compared	F1	F2	Balanced accuracy	Percent agreement	Krippendorff's α	Cohen's κ
<i>Denotative framing elements</i>							
Humans	6806	0.80	0.80	0.80	82.66	0.76	0.76
Destruction	6787	0.79	0.81	0.82	91.97	0.72	0.72
Military role	5551	0.40	0.36	0.34	72.67	0.51	0.51
Death	7107	0.52	0.45	0.41	65.27	0.06	0.15
Leaders	5029	0.34	0.31	0.29	82.72	0.26	0.26
Flags	6959	0.63	0.63	0.62	89.96	0.64	0.64
<i>Semiotic framing elements</i>							
Distance	6637	0.57	0.54	0.52	58.82	0.43	0.44
Viewer position	5198	0.33	0.31	0.31	75.88	0.36	0.37
Eye contact	5853	0.52	0.53	0.55	69.33	0.54	0.55
Facial expressions	3077	0.55	0.55	0.56	74.36	0.52	0.53
Stance	6173	0.59	0.58	0.58	72.27	0.62	0.62
<i>Connotative framing elements</i>							
State	6283	0.34	0.33	0.32	66.91	0.30	0.30
About to die	7192	0.44	0.39	0.37	94.48	0.11	0.12
Religion	6462	0.45	0.43	0.42	87.84	0.42	0.42

Table 3: F1 and F2 scores, balanced accuracy, percent agreement, Krippendorff's alpha, and Cohen's kappa for each visual framing variable

Without these textual cues, AI required more training on choice of uniforms or human intervention. Additionally, AI could detect ISIS's objects like coins, outdoor markets, and competing currencies, but required researchers to recognize their connotative meaning, such as ISIS's desired frame of economic independence. Similarly, AI could detect bottles of alcohol, drugs, and cigarettes (see Figure 3), but required human coders to recognize them as critical components of ISIS's moral policing apparatus. In short, the addition of human coding can render computational methods less likely to miss key objects or elements in big datasets and more likely to properly assess their functions.

Another key contribution of manual coding to AI processing involves aggregation of denotative elements. Our AI learning approaches generated disparate visual elements, such as militants fighting or training, raised index fingers, swords, maps, doctors treating patients, and sunsets, etc. Human coding, however, captured subtle visual relationships that revealed broader themes of military prowess and state-building, identified how relationships created cohesive narratives, showed how symmetry, repetition or alignment influenced viewer interpretations, and unveiled the visual syntax strategy. For example, human coders identified the interrelationship between images of beheadings, piles of cigarettes, and checkpoints as ISIS's law enforcement apparatus, complete with punishments for alleged spies, the moral policing apparatus, and the access points for determining who could enter the

caliphate. Entman's framing associations were useful to unlock the messaging strategy. The visual law enforcement frame communicated that sins were rampant (problem definition) because of people smuggling contraband, committing treason, and ignoring Islamic rules (causal interpretation), which stained the society (moral evaluation), thus requiring punishments and crackdowns on contraband to ensure community safety (treatment recommendation). An unsupervised learning computational approach on its own would not fully reveal the visual narrative.

Lesson #2: Semiotic Interactions

For the most part, AI performed poorly on visual semiotic elements, rendering human coding highly valuable in this domain. With Krippendorff's α consistently lower than .67 between AI and human coding for viewer distance, eye contact, body stance, and facial expressions, the use of supervised AI was neither efficient nor cost-effective in analyzing the semiotics of al-Qaeda and ISIS's images. Human training and validation, however, could help refine steps for identifying more useful, valid, and efficient processing of visual semiotics patterns. One example is the use of direct eye contact because it conveys dominance (Appendix A). From 2016 to 2018, ISIS used direct eye contact as a frequent visual tactic, but the use of the strategy differed based on whether the depicted subject was a friend or foe and if the language of publication was English or Arabic. Yet, supervised training for AI only produced a .54 Krippendorff's α with manual coding. Intercoder reliability difficulties related to eye contact suggest that a carefully constructed, rule-based definition of what constitutes eye contact and avoidance would be necessary to reduce noise in the AI extraction process.

Human coding could also help narrow categories down to coding op-



Figure 3: Photo from the 10th issue of Dabiq magazine showing ISIS's hisba agents burning cigarettes and alcohol – Released July 2015

tions most useful for understanding visual strategies. For example, semiotic understandings of viewer distance focus on four categories: intimate, personal, public, and social (Jewitt & Oyama, 2008). Each conveys specific meanings associated with standard human interactions (e.g., intimate distances associated with distraught photo subjects; photo subjects shot at a public distance conveying group rather than individual identities). However, human coding and statistical analysis revealed that significant findings primarily appeared only after combining intimate and personal distance (with photo subjects photographed from less than four feet) and comparing them with the combined categories of public and social distance (i.e., greater than four feet). Such findings help avoid overlooking important patterns that could be missed with strict adherence to viewer distance's four standards.

Lesson #3: Connotative Interactions

While computational coding could efficiently identify key symbols, human content expertise helped interpret needed cultural genres and subsystems. The black flag, for example, is a transhistorical object al-Qaeda and ISIS used as a symbol of adherence to Islam and the Prophet Muhammad's path. It typically features the words "No God but Allah" with a white circle beneath carrying the words "Muhammad is Allah's Messenger." Yet, al-Qaeda often deviated from standard black flag depictions, also featuring white and other emblems used in Afghanistan and elsewhere (see Figure 4). Without such insights, computational extractions of flags as denotative elements would miss other symbolic variations essential for understanding the nuance of extremist groups' visual messaging. Another frequent symbol in al-Qaeda and ISIS photographs was militants raising their index fingers. The diverse makeup of our manual coding team, including Muslim researchers, identified the gesture as part of Islamic culture, connoting monotheism and the dedication of deeds to the one God. Culturally informed, human content expertise was instrumental for properly training and validating computational visual analyses to generate insightful media campaign understandings.

When collaborating with AI, another beneficial area for human coders is adding useful insights about tropes and other visual commonplaces. A prime example in al-Qaeda and ISIS's media was the excessive reliance on the about to die visual trope, which appeared in 75 percent of their images. Yet, AI missed many instances of the trope. About to die images assume three forms: presumed (i.e., showing implements of death like weapons and destruction), possible (i.e., showing photo subjects potentially dying without a confirmation their death occurred), or certain (i.e., showing photo

subjects with accompanying text confirming their deaths) (Zelizer, 2010). An unsupervised learning method would likely group all three forms under a military visual frame, hence not distinguishing between the three constructs nor accounting for the groups' emphasized use of the three disaggregated strategies. Instead, breaking down each form into its own core denotative components could facilitate training and validation. Labeled data for supervised learning could account for objects or elements, such as blood, swords, knives, guns, AK-47S, tanks, armored vehicles, rockets, ammunition, fire, smoke, explosions, and sniper crosshairs. Human coding could then complement the computational analysis by grouping the three about to die clusters, each with its own unique viewer interactions (Zelizer, 2010).

Lesson #4: Ideological Interactions

Increasingly, scholars have recognized the linkages between ideology and discourse. Fabiszak et al. define ideology as “systems of beliefs, shared by a social group, with the power to evaluate and explain the social world” (Fabiszak et al., 2021, p. 409). McGee adds that “ideology in practice is a political language, preserved in rhetorical documents, with the capacity to dictate decision and control public belief and behavior” (McGee, 1980, p. 5). Van Dijk (1998) explains that ideologies can influence the human mind's cognitive structures. McGee (1980) posits that a full set of ideological propositions can be summed up in a single term, while Edwards and Winkler (1997) extend such reasoning to a small set of visual images.

Human coding can help AI users distinguish between visual markers of culture and other images not performing ideological functions. One example of how this process could work concerns ideographs. In his study of American culture, McGee (1980) defines a small subset of positive and negative words as ideographs (e.g., freedom, liberty, slavery, and terrorism). They serve as ordinary terms in political discourse, have abstract meanings

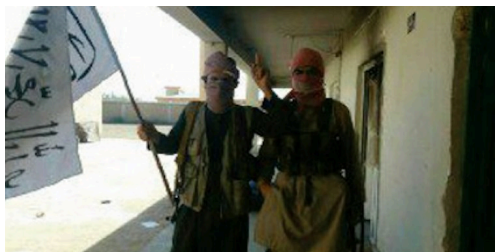


Figure 4: Photo from the 17th issue of Al-Masra newspaper showing a militant holding a white flag with the same text that appears on the black banner – Released July 2016

that allow for collective commitment, warrant the use of power, guide behavior, and have culture-bound meanings. Consideration of the ideograph's characteristics within social groups can assist AI users in narrowing large corpuses of visual images to specific objects that serve as cultural markers. With al-Qaeda and ISIS, for example, one visual ideological code is the group's display of the monotheism gesture, whereby Muslims point their index fingers upwards toward heaven to connote their adherence to Allah. Al-Qaeda and ISIS frequently utilize images of the same gesture to signal potential Muslim recruits sympathetic to their groups' causes.

However, a key function of human coders in the AI training process involves understanding the interactions of visual elements of ideographs within an image. Returning to the monotheism gesture example, AI would be able to scrape all images showing humans pointing their index finger upward. Such an approach, absent human coders' insights, would initially yield many images of Muslims demonstrating their Islamic faith with no affiliation with extremism. AI might also retrieve images of athletes or other individuals raising their index finger to signify success and victory (see Figures 5 & 6). Rather than confound the results with too much noise to produce meaningful results, AI trainers could refine the scraping process by asking for the gesture along with the presence of militants, males, direct eye contact in photo subjects, personal distance, black flags, and number of humans, as each of these variables have a documented relationship with the monotheism gesture in extremism photographs (Winkler, 2022). By considering element constellations rather than single objects or elements, the dataset will become more accurate in discerning ideological frames.



The Inghimasiyin who terrorized the Rafidah in 'Ayn al-Tamr

Figure 5: Photo from the 1st issue of Rumiya magazine showing militants before an attack in Iraq, one of whom (on the right) signaling the monotheism gesture – Released September 2016



Figure 6: Photo showing two indoor soccer players from the Moroccan national team celebrating by signaling the monotheism gesture – Released by Equipe du Maroc/Facebook September 2024

Lesson #5: Image-Context Interactions

Context is a multifaceted resource involving a myriad of forms that can influence interpretations of texts, whether discursive or nondiscursive (Linell, 1998). It both shapes and is shaped by its textual interactions. As McGee notes, “Failing to account for ‘context,’ or reducing ‘context’ to one or two of its parts, means quite simply that one is no longer dealing with discourse as it appears in the world” (McGee, 1990, p. 283).

The complicated interactions between images and contexts in relation to al-Qaeda and ISIS emphasized the need for human coders to supplement AI for meaningful, efficient messaging processing. To begin, human coding aided in the identification of image codes that corresponded to changes in context factors over time. As Appendix C, Table C1 shows, 18 of 30 variables in our manual codebook bore a significant relationship to context variables (e.g., troop withdrawal announcements, online account suspensions, attack lethality, etc.). Those relationships suggest a productive, efficient training regimen for AI, as the context factors are revealing about the groups’ response interactions over time. The remaining coding categories, while potentially useful for assessing image meaning or relationships with other images, did not significantly change in frequency over time, making them less of an AI priority.

Human coding can also help verify the appropriate AI scope for assessing

impact of context variables on strategic image use. The table reveals that the relationships with image strategies vary according to the context factor under consideration. As a result, the outputs of human coding point to context variables that need verification checks prior to premature conclusions that any single context variable is responsible for shifts in the image characteristics. For example, since censorship of militant group online accounts, announcements of anticipated troop withdrawals, and the relative rise in standing of competing ideological groups all relate to significant changes in the use of photo subject distance, users of AI should consider whether each of these context factors overlap during the timespan under evaluation before drawing any conclusions about the potential influence of any single context element.

Human coding can also yield insights regarding appropriate AI disaggregation of context elements in relation to image characteristics over time. For example, our assessment of leader loss based on human coding revealed that the deaths of different levels and types of ISIS leaders resulted in different corresponding changes in the group's image characteristics. Political and military leader deaths corresponded to different image changes. Leaders at different statuses within the media hierarchy corresponded to different changes in visual strategies. Thus, fine-tuned human coding analysis can aid in the development of a more robust, efficient AI system for analyzing the image-context strategies of groups like al-Qaeda and ISIS.

Conclusion

This analysis demonstrates the advantages of having humans and AI functioning together to understand the visual framing of extremist group messaging. Human coding yielded benefits on the retrieval, analysis, and validation of results related to denotative, semiotic, connotative, and ideological framing. It also aided in understanding how AI-human interactions can maximize text-context relationships. Yet, validation checks between the human and AI coding revealed that the percent agreement levels on coding categories varied considerably, suggesting the need for a more robust AI training process, particularly on subjective variables, like facial expressions and perceived distance, and identifying visual constellations linked to connotative and ideological framing.

Examining the extremist visual context provided an ideal opportunity to test and compare manual and computational coding for MENA-based violent groups, but it does not necessarily apply to other types of violent protest visuals. The generalizability of the findings derived from al-Qaeda

and ISIS photographic campaigns should be tested in relation to other forms of political violence (e.g., electoral protests, climate activism, and public vigils). Future studies should determine if the reliability of AI visual framing is transferable to these other settings.

Additionally, variables with inherently nuanced or subjective definitions, such as stance or eye contact, pose significant challenges for consistent annotation. These complexities are reflected in the low recall and F1-scores observed in these categories, as the AI model struggles to align with human coders' interpretations. The limitations of computational methods in capturing subtle cultural or contextual cues further exacerbate these discrepancies, particularly for categories like state, religion, and impending death that rely heavily on contextual understandings. Future studies should examine alternative training protocols to maximize efficient and reliable extraction processes.

Data Availability: Replication materials for this study are hosted on the Open Science Framework (OSF). Due to the presence of violent and potentially harmful imagery, the image data are archived as a restricted-access component on OSF and may be accessed upon request, subject to review and approval: <https://osf.io/r9tjm>, project DOI: 10.17605/OSF.IO/R9TJM. All non-sensitive replication materials—including the coding instrument, model prompts, variable definitions, documentation, and analysis scripts—are publicly available on OSF and are also mirrored on GitHub for ease of access and version control: <https://github.com/aysedeniz09/Visual-Framing-in-the-AI-Era>.

References

- Abdelrahim, Y. (2019). Visual analysis of isis discourse strategies and types in dabiq and rumiyah online magazines. *Visual Communication Quarterly*, 26(2), 63–78. <https://doi.org/10.1080/15551393.2019.1586546>
- Anderson, C. (2017). Data-first manifesto: Shifting priorities in scholarly communications. *Information Services and Use*, 37(3), 335–342. <https://doi.org/10.3233/ISU-170852>
- Armstrong, D., Gosling, A., Weinman, J., & Marteau, T. (1997). The place of inter-rater reliability in qualitative research: An empirical study. *Sociology*, 31(3), 597–606. <https://doi.org/10.1177/0038038597031003015>
- Bakhshi, S., & Gilbert, E. (2015). Red, purple and pink: The colors of diffusion on pinterest. *PLOS ONE*, 10(2), 117–148. <https://doi.org/10.1371/journal.pone.0117148>
- Barr, A., & Herfroy-Mischler, A. (2017). Isil's execution videos: Audience segmentation and terrorist communication in the digital age. *Studies in Conflict and Terrorism*. <https://doi.org/10.1080/1057610X.2017.1361282>

- Barthes, R. (1981). *Camera lucida: Reflections on photography*. Hill; Wang.
- Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), 77–101. <https://doi.org/10.1191/1478088706qp0630a>
- Bridle, J. (2023). New dark age: Technology and the end of the future (Updated) [Verso. (Original work published 2018)].
- Broz, M. (2023, November 24). How many pictures are there (2024): Statistics, trends, and forecasts. <https://photutorial.com/how-many-photos/>
- Burscher, B., Odiijk, D., Vliegenthart, R., Rijke, M., & Vreese, C. (2014). Teaching the computer to code frames in news: Comparing two supervised machine learning approaches to frame analysis. *Communication Methods and Measures*, 8(3), 190–206. <https://doi.org/10.1080/19312458.2014.937527>
- Cacciatore, M., Scheufele, D., & Iyengar, S. (2016). The end of framing as we know it ... and the future of media effects. *Mass Communication and Society*, 19(1), 7–23. <https://doi.org/10.1080/15205436.2015.1068811>
- Carlin, M. (2012). Guns, gold and corporeal inscriptions. *Third Text*, 26(5), 503–514. <https://doi.org/10.1080/09528822.2012.712763>
- Ceci, L. (2024, April 11). Hours of video uploaded to youtube every minute 2007–2022. <https://www.statista.com/statistics/259477/hours-of-video-uploaded-to-youtube-every-minute/>
- Center, I. (2005). Evolution of jihadi video. <https://intelcenter.com/EJV-PUB-v1-0.pdf>
- Chan, C., Bajjalieh, J., Auvil, L., Wessler, H., Althaus, S., Welbers, K., Atteveldt, W., & Jungblut, M. (2021). Four best practices for measuring news sentiment using ‘off-the-shelf’ dictionaries: A large-scale p-hacking experiment. *Computational Communication Research*, 3(1), 1–27. <https://doi.org/10.5117/CCR2021.1.001>
- CHAN
- Chen, K., Duan, Z., & Kim, S. (2024). Uncovering gender stereotypes in controversial science discourse: Evidence from computational text and visual analyses across digital platforms. *Journal of Computer-Mediated Communication*, 29(1), 052. <https://doi.org/10.1093/jcmc/zmad052>
- Chen, K., Kim, S., Gao, Q., & Raschka, S. (2022). Visual framing of science conspiracy videos. *Computational Communication Research*, 4(1), 98–134. <https://doi.org/10.5117/ccr2022.1.003.chen>
- Chen, N., Drouhard, M., Kocielnik, R., Suh, J., & Aragon, C. (2018). Using machine learning to support qualitative coding in social science: Shifting the focus to ambiguity. *ACM Trans. Interact. Intell. Syst.*, 8(2), 1–9 20. <https://doi.org/10.1145/3185515>
- Chen, Y., Zhai, Y., & Sun, S. (2024). The gendered lens of ai: Examining news imagery across digital spaces. *Journal of Computer-Mediated Communication*, 29(1).
- Chouliarakis, L., & Kissas, A. (2018). The communication of terrorism: A typology of isis online death videos. *Critical Studies in Media Communication*, 35(1), 24–39. <https://doi.org/10.1080/15295036.2017.1393096>

- Ciovacco, C. (2009). The contours of al Qaeda's media strategy. *Studies in Conflict and Terrorism*, 32(10), 853–875. <https://doi.org/10.1080/10576100903182377>
- Coleman, A. (2006). The Islamic imagery project. visual motifs in jihadi the Islamic imagery. <https://www.ctc.usma.edu/v2/wp-content/uploads/2010/06/Islamic-Imagery-Project.pdf>
- Coleman, R. (2010). Framing the pictures in our heads: Exploring the framing and agenda-setting effects of visual images. In P. D'Angelo & J. Kuypers (Eds.), *Doing news framing analysis* (pp. 233–261).
- Damanhoury, K. (2022). *Photographic warfare: Isis, Egypt, and the online battle for Sinai*. University of Georgia Press.
- Damanhoury, K., & Winkler, C. (2018). Picturing law and order: A visual framing analysis of Isis's Dabiq magazine. *Arab Media & Society, Winter/Spr*(25), 1–25. <https://doi.org/10.70090/kdcw18vf>
- Dietrich, B., & Ko, H. (2022). Finding fauci. *Computational Communication Research*, 4(1), 135–172. <https://doi.org/10.5117/ccr2022.1.004.diet>
- Dijk, T. (1998). *Ideology: A multidisciplinary approach*. Sage Publications.
- Dondero, M. (2019). Visual semiotics and automatic analysis of images from the cultural analytics lab: How can quantitative and qualitative analysis be combined? *Semiotica*, 2019(230), 121–142. <https://doi.org/10.1515/sem-2018-0104>
- Duarte, F. (2024, June 13). Amount of data created daily. <https://explodingtopics.com/blog/data-generated-per-day>
- Edwards, J., & Winkler, C. (1997). Representative form and the visual ideograph: The Iwo Jima image in editorial cartoons. *Quarterly Journal of Speech*, 83(3), 289–310. <https://doi.org/10.1080/00335639709384187>
- El Karhili, N., Hendry, J., Kackowski, W., El Damanhoury, K., Dicker, A., & Winkler, C. (2021). Islamic/state: Daesh's visual negotiation of institutional positioning. *Journal of Media and Religion*, 20(2), 79–104. <https://doi.org/10.1080/15348423.2021.1930813>
- Fabiszak, M., Buchstaller, I., Brzezińska, A., Alvanides, S., Griese, F., & Schneider, C. (2021). Ideology in the linguistic landscape: Towards a quantitative approach. *Discourse & Society*, 32(4), 405–425. <https://doi.org/10.1177/0957926521992149>
- Farwell, J. (2010). Jihadi video in the 'war of ideas'. *Survival*, 52(6), 127–150. <https://doi.org/10.1080/00396338.2010.540787>
- Feng, D., & O'Halloran, K. (2013). The visual representation of metaphor: A social semiotic approach. *Review of Cognitive Linguistics*, 11(2), 320–335. <https://doi.org/10.1075/rcl.11.2.07>
- Feng, W. (2017). Metonymy and visual representation: Towards a social semiotic framework of visual metonymy. *Visual Communication*, 16(4), 441–466. <https://doi.org/10.1177/1470357217717142>
- Forgas, J., & East, R. (2008). How real is that smile? mood effects on accepting or rejecting the veracity of emotional facial expressions. *Journal of Nonverbal Behavior*, 32(3), 157–170.

- Gamielidien, Y., Case, J., & Katz, A. (2023). Advancing qualitative analysis. *An Exploration of the Potential of Generative AI and NLP in Thematic Coding (SSRN Scholarly Paper, (4487768))*. <https://doi.org/10.2139/ssrn.4487768>
- Geise, S., & Baden, C. (2015). Putting the image back into the frame: Modeling the linkage between visual communication and frame-processing theory. *Communication Theory, 25*(1), 46–69. <https://doi.org/10.1111/comt.12048>
- Glausch, M. (2020). Infographics and their role in the is propaganda machine. *Contemporary Voices: St Andrews Journal of International Relations, 1*(3), 32–50. <https://doi.org/10.15664/jtr.1492>
- Goffman, E. (1974). *Frame analysis: An essay on the organization of experience*. Northeastern University Press.
- Goffman, E. (1979). *Gender and advertisements*. Macmillan.
- Green, L. (2015). *Advertising war: Pictorial publicity, 1914-1918*. Doctoral Dissertation, Manchester Metropolitan University.
- Gregg, H. (2023). The islamic state in africa: The emergence, evolution, and future of the next jihadist battlefield. *Parameters, 53*(3), 159–161. <https://doi.org/10.1080/10220461.2023.2201586>
- Ha, Y., Park, K., Kim, S., Joo, J., & Cha, M. (2020). Automatically detecting image–text mismatch on instagram with deep learning. *Journal of Advertising, 50*(1), 52–62. <https://doi.org/10.1080/00913367.2020.1843091>
- Hall, E. (1966). *The hidden dimension*. Doubleday.
- Hanusch, F. (2013). Sensationalizing death? graphic disaster images in the tabloid and broadsheet press. *European Journal of Communication, 28*(5), 497–513. <https://doi.org/10.1177/0267323113491349>
- Hariman, R., & Lucaites, J. (2007). *No caption needed: Iconic photographs, public culture, and liberal democracy*. University of Chicago Press.
- Heck, A. (2017). Images, visions and narrative identity formation of isis [Retrieved Oct 7, 2025, from]. *Global Discourse, 7*(2-3), 244–259. <https://doi.org/10.1080/23269995.2017.1342490>
- Höijer, B. (2004). The discourse of global compassion: The audience and media reporting of global suffering. *Media, Culture & Society, 26*(4), 513–531. <https://doi.org/10.1177/0163443704044215>
- Impara, E. (2018). A social semiotics analysis of islamic state's use of beheadings: Images of power, masculinity, spectacle and propaganda. *International Journal of Law, Crime and Justice, 53*, 25–45. <https://doi.org/10.1016/j.ijlcrj.2018.02.002>
- JasenskaG. (2023, June 28). How many pictures? photo statistics that will blow you away. <https://www.lightstalking.com/photo-statistics/>
- Jewitt, C., & Oyama, R. (2008). Visual meaning: A social semiotic approach. In T. Leeuwen & C. Jewitt (Eds.), *Handbook of visual analysis* (7th, pp. 134–156). Sage Publications Ltd. <https://doi.org/10.1017/CBO9781107415324.004>
- Joo, J., & Steinert-Threlkeld, Z. (2022). Image as data: Automated content analysis for visual presentations of political actors and events. *Computational Communication Research, 4*(1), 11–67. <https://doi.org/10.5117/CCR2022.1.001.JOO>

- Kaczkowski, W., Winkler, C., El Damanhoury, K., & Luu, Y. (2021). Intersections of the real and the virtual caliphates: The Islamic state's territory and media campaign. *Journal of Global Security Studies*, 6(2). <https://doi.org/10.1093/jogss/ogaa020>
- Knobloch, S., Hastall, M., Zillmann, D., & Callison, C. (2003). Imagery effects on the selective reading of internet newsmagazines. *Communication Research*, 30(1), 3–29. <https://doi.org/10.1177/0093650202239023>
- Kraft, R. (1986). The influence of camera angle on comprehension and retention of pictorial events. *Memory & Cognition*, 15(4), 291–307. <https://doi.org/10.3758/BF03197032>
- Kress, G., & Leeuwen, T. (2006). *Reading images: The grammar of visual design* (2nd). Routledge.
- Kress, G., & Leeuwen, T. (1996). *Reading images: The grammar of visual design* (2nd). Routledge.
- Krippendorff, K. (2018). *Content analysis: An introduction to its methodology* (Fourth). SAGE.
- Kuznar, L. (2015). Daesh's image of the state in their own words.
- Lakretz, Y., Dehaene, S., & King, J.-R. (2020). What limits our capacity to process nested long-range dependencies in sentence comprehension? *Entropy*, 22(4), 446. <https://doi.org/10.3390/e22040446>
- Leeuwen, T. (2001). Semiotics and iconography. In T. Leeuwen & C. Jewitt (Eds.), *The handbook of visual analysis* (pp. 92–118).
- Linell, P. (1998). *Approaching dialogue: Talk, interaction and contexts in dialogical perspectives*. John Benjamins Publishing Company.
- Lokmanoglu, A. (2020). Coin as imagined sovereignty: A rhetorical analysis of coins as a transhistorical artifact and an ideograph in Islamic state's communication. *Studies in Conflict & Terrorism*, 44(1), 52–73.
- Lokmanoglu, A., Winkler, C., Mcminimy, K., & Almahmoud, M. (2022). Troop withdrawal announcements, isis media: Visualizing community and resilience. *International Journal of Communication*, 16, 215–246.
- Lokmanoglu, A., Allaham, M., Abhari, R., Mortenson, C., & Villa Turek, E. (2023). A picture is worth a thousand (s)words: Classification and diffusion of memes on a partisan media platform [Year 3 Award]. *Global Network on Extremism & Terrorism (GNET)*. <https://doi.org/10.18742/pub01>
- Lokmanoglu, A., & Walter, D. (2025). Topic modeling of video and image data: A visual semantic unsupervised approach. *Communication Methods and Measures*. <https://www.tandfonline.com/doi/abs/10.1080/19312458.2025.2549707>
- Mackieson, P., Shlonsky, A., & Connolly, M. (2019). Increasing rigor and reducing bias in qualitative research: A document analysis of parliamentary debates using applied thematic analysis. *Qualitative Social Work*, 18(6), 965–980. <https://doi.org/10.1177/1473325018786996>
- Marathe, M., & Toyama, K. (2018). Semi-automated coding for qualitative research: A user-centered inquiry and initial prototypes. *Proceedings of the 2018 CHI Con-*

- ference on Human Factors in Computing Systems, 1–12. <https://doi.org/10.1145/3173574.3173922>
- Mattoni, A., & Teune, S. (2014). Visions of protest. a media-historic perspective on images in social movements. *Sociology Compass*, 8(6), 876–887.
- McCain, T., Chilberg, J., & Wakshlag, J. (1977). The effect of camera angle on source credibility and attraction. *Journal of Broadcasting*, 21(1), 35–46. <https://doi.org/10.1080/08838157709363815>
- McClancy, K. (2013). The iconography of violence: Television, vietnam, and the soldier hero. *Film & History: An Interdisciplinary Journal*, 43(2), 50–66.
- McGee, M. (1980). The “ideograph”: A link between rhetoric and ideology. *Quarterly Journal of Speech*, 66(1), 1–16. <https://doi.org/10.1080/00335638009383499>
- McGee, M. (1990). Text, context, and the fragmentation of contemporary culture. *Western Journal of Speech Communication: WJSC*, 54(3), 274–289.
- Mcgil, J. (2024, May 23). <https://contentdetector.ai/articles/meme-statistics/>
- McMinimy, K., Winkler, C., Lokmanoglu, A., & Almahmoud, M. (2021). Censoring extremism: Influence of online restriction on official media products of isis. *Terrorism and Political Violence*, 00(00), 1–17. <https://doi.org/10.1080/09546553.2021.1988938>
- Messaris, P., & Abraham, L. (2001). The role of images in framing news stories. In *Framing public life: Perspectives on media and our understanding of the social world* (pp. 215–226). Lawrence Erlbaum Associates Publishers. <https://doi.org/10.4324/9781410605689>
- Mitchell, W. (1995). *Picture theory: Essays on verbal and visual representation*. University of Chicago Press. <https://press.uchicago.edu/ucp/books/book/chicago/P/bo3683962.html>
- Morgan, D. (2023). Exploring the use of artificial intelligence for qualitative data analysis: The case of chatgpt. *International Journal of Qualitative Methods*, 22, 16094069231211248. <https://doi.org/10.1177/16094069231211248>
- Moriarty, S. (2002). The symbiotics of semiotics and visual communication. *Journal of Visual Literacy*, 22(1), 19–28. <https://doi.org/10.1080/23796529.2002.11674579>
- Muise, D., Lu, Y., Pan, J., & Reeves, B. (2022). Selectively localized: Temporal and visual structure of smartphone screen activity across media environments. *Mobile Media & Communication*, 10(3), 487–509. <https://doi.org/10.1177/20501579221080333>
- OpenAI. (2024). Chatgpt-4o api [Large language model]. <https://platform.openai.com/>
- Paglen, T. (2019). Invisible images: Your pictures are looking at you. *Architectural Design*, 89(1), 22–27. <https://doi.org/10.1002/ad.2383>
- Panofsky, E. (1955). *Meaning in the visual arts*. Doubleday Anchor Books.
- Peng, Y. (2018). Same candidates, different faces: Uncovering media bias in visual portrayals of presidential candidates with computer vision. *Journal of Communication*, 68(5), 920–941.
- Peng, Y. (2021). What makes politicians’ instagram posts popular? analyzing social media strategies of candidates and office holders with computer vision. *The*





- International Journal of Press/Politics*, 26(1), 143–166. <https://doi.org/10.1177/1940161220964769>
- Peng, Y., Lock, I., & Ali Salah, A. (2024). Automated visual analysis for the study of social media effects: Opportunities, approaches, and challenges. *Communication Methods and Measures*, 18(2), 163–185.
- Peng, Y., & Yingdan, L. (2023). Computational visual analysis in political communication. In D. Lilleker & A. Veneti (Eds.), *Research handbook in visual politics* (pp. 41–53). Edward Elgar. <https://ssrn.com/abstract=4577025>
- Perlmutter, D. (1998). *Photojournalism and foreign policy: Icons of outrage in international crises*. Praeger.
- Rietz, T., & Maedche, A. (2021). Cody: An ai-based system to semi-automate coding for qualitative research. *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, 1–14. <https://doi.org/10.1145/3411764.3445591>
- Rodriguez, L., & Dimitrova, D. (2011). The levels of visual framing. *Journal of Visual Literacy*, 30(1), 48–65. <https://doi.org/10.1080/23796529.2011.11674684>
- Saldana, J. (2021). *The coding manual for qualitative researchers*. SAGE Publications Ltd.
- Saltman, E., & Smith, M. (2015). Till martyrdom do us part - gender and the isis. http://www.strategicdialogue.org/Till_Martyrdom_Do_Us_Part_Gender_and_the_ISIS_Phenomenon.pdf
- Schwalbe, C. (2006). Remembering our shared past: Visually framing the iraq war on u.s. News Websites. *Journal of Computer-Mediated Communication*, 12(1), 264–289. <https://doi.org/10.1111/j.1083-6101.2006.00325.x>
- Sezen, D. (2020). Without a blink: Machine ways of seeing in contemporary visual culture. *Interactions: Studies in Communication & Culture*, 11(1), 103–107. https://doi.org/10.1386/iscc_00010_7
- Song, H., Tolochko, P., Eberl, J.-M., Eisele, O., Greussing, E., Heidenreich, T., Lind, F., Galyga, S., & Boomgaarden, H. (2020). In validations we trust? the impact of imperfect human annotations as a gold standard on the quality of validation of automated content analysis. *Political Communication*, 37(4), 550–572. <https://doi.org/10.1080/10584609.2020.1723752>
- Steinert-Threlkeld, Z., Chan, A., & Joo, J. (2022). How state and protester violence affect protest dynamics. *The Journal of Politics*, 84(2), 798–813. <https://doi.org/10.1086/715600>
- Stone, E., Sieck, W., Bull, B., Frank Yates, J., Parks, S., & Rush, C. (2003). Foreground:background salience: Explaining the effects of graphical displays on risk avoidance. *Organizational Behavior and Human Decision Processes*, 90(1), 19–36. [https://doi.org/10.1016/S0749-5978\(03\)00003-7](https://doi.org/10.1016/S0749-5978(03)00003-7)
- Tang, D., & Schmeichel, B. (2015). Look me in the eye: Manipulated eye gaze affects dominance mindsets. *Journal of Nonverbal Behavior*, 39(2), 181–194. <https://doi.org/10.1007/s10919-015-0206-8>
- Torres, M. (2023). A framework for the unsupervised and semi-supervised analysis of visual frames. In *Political analysis* (pp. 1–22). <https://doi.org/10.1017/pan.2023.32>

- Trachtenberg, A. (1985). Albums of war: On reading civil war photographs. *Representations*, 9, 1–32.
- Turner, V. (1974). *Dramas, fields, and metaphors: Symbolic action in human society*. Cornell University Press. <https://doi.org/10.2307/3710621>
- Turner, V. (1979). *Process, performance, and pilgrimage: A study in comparative symbolism*. Concept Publishing Company.
- Walter, D., & Ophir, Y. (2024). Meta-theorizing framing in communication research (1992–2022): Toward academic silos or professionalized specialization? *Journal of Communication*, *jqad043*. <https://doi.org/10.1093/joc/jqad043>
- Warrick, J. (2016). Black flags: The rise of isis. *Anchor Books*.
- Wignell, P., Tan, S., & O'Halloran, K. (2017). Under the shade of ak47s: A multimodal approach to violent extremist recruitment strategies for foreign fighters. *Critical Studies on Terrorism*, 10(3), 429–452. <https://doi.org/10.1080/17539153.2017.1319319>
- Williams, N., Casas, A., & Wilkerson, J. (2020). *Images as data for social science research: An introduction to convolutional neural nets for image classification*. Cambridge University Press.
- Wilson, L. (2017). *Understanding the appeal of isis*. Cambridge UP. <https://doi.org/10.17863/CAM.68729>
- Winkler, C., El Damanhoury, K., Dicker, A., & Lemieux, A. (2018). Images of death and dying in isis media: A comparison of english and arabic print publications [Original work published 2019]. *Media, War & Conflict*, 12(3), 248–262. <https://doi.org/10.1177/1750635217746200>
- Winkler, C., El-Damanhoury, K., Dicker, A., Luu, Y., Kaczowski, W., & El-Karhili, N. (2019). Considering the military-media nexus from the perspective of competing groups: The case of isis and al-qaeda in the arabian peninsula. *Dynamics of Asymmetric Conflict*, 13(1), 3–23. <https://doi.org/10.1080/17467586.2019.1630744>
- Winkler, C., McMinimy, K., El Damanhoury, K., & Almahmoud, M. (2021). Shifts in the visual media campaigns of aqap and isis after high death and high publicity attacks. *Behavioral Sciences of Terrorism and Political Aggression*, 13(4), 251–264. <https://doi.org/10.1080/19434472.2020.1759674>
- Winkler, C. (2022). Revisiting representative form in isis media: How emerging collectives reconstitute communities in the 21st century. *Frontiers in Communication*, 7. <https://doi.org/10.3389/fcomm.2022.968302>
- Winkler, C., & Damanhoury, K. (2022). *Proto-state media systems: Al-qaeda and isis as exemplars*. Oxford University Press.
- Winkler, C., Dewick, L., Luu, Y., & Kaczowski, W. (2019). Dynamic/static image use in isis's media campaign: An audience involvement strategy for achieving goals. *Terrorism and Political Violence*, 33(6), 1323–1341. <https://doi.org/10.1080/09546553.2019.1608953>
- Winkler, C., El Damanhoury, K., Dicker, A., & Lemieux, A. (2016). The medium is terrorism: Transformation of the about to die trope in dabiq. *Terrorism and Political Violence*, 31(2), 224–243. <https://doi.org/10.1080/09546553.2016.1211526>







- Winter, C. (2018). Apocalypse, later: A longitudinal study of the islamic state brand. *Critical Studies in Media Communication*, 35(1), 103–121. <https://doi.org/10.1080/15295036.2017.1393094>
- Zelin, A. (2015). Picture or it didn't happen: A snapshot of the islamic state's official media output. *Perspectives on Terrorism*, 9(4), 85–97.
- Zelin, A. (2021). <https://jihadology.net/>
- Zelizer, B. (2010). *About to die: How news images move the public*. Oxford University Press. <https://doi.org/10.1080/10584609.2012.641782>
- Zhang, H., & Pan, J. (2019). Casm: A deep-learning approach for identifying collective action events with text and image data from social media. *Sociological Methodology*, 49(1), 1–57. <https://doi.org/10.1177/0081175019860244>
- Zhang, H., & Peng, Y. (2022). Image clustering: An unsupervised approach to categorize visual data in social science research. *Sociological Methods & Research*, 004912412210826. <https://doi.org/10.1177/00491241221082603>
- Zhang, X., & Dahu, W. (2019). Application of artificial intelligence algorithms in image processing. *Journal of Visual Communication and Image Representation*, 61, 42–49. <https://doi.org/10.1016/j.jvcir.2019.03.004>
- Zou, J., & Schiebinger, L. (2018). Ai can be sexist and racist—it's time to make it fair. *Nature*, 559, 324–326.

Appendix A Manual content coding



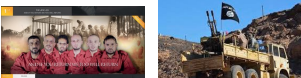

Table A1: Justification for coding instrument categories with examples

Coding categories	Relationship to visual literature	Examples
Denotative framing elements		
Militants	<p>Studies of ISIS and al-Qaeda magazines show the frequent presence of “martyrs” and fighters, with different causes and outcomes for in-group and out-group combatants (Saltman & Smith, 2015; Winkler et al., 2018; Winkler, Dewick, et al., 2019; Winter, 2018).</p>	<div style="display: flex; justify-content: space-around;">   </div> <p><i>Al-Naba</i> (10) <i>Al-Masra</i> (45)</p>
Death	<p>Graphic images of death in wartime highlight the consequences of conflict and who holds the authority to decide who lives or dies (Carlin, 2012; Höijer, 2004). Work on ISIS/AQ magazines shows “about to die” images are more frequent than depictions of dead bodies (Hanusch, 2013; Winkler et al., 2018).</p>	<div style="display: flex; justify-content: space-around;">   </div> <p><i>Dabiq</i> (15) <i>Jihadi Recollections</i> (1)</p>




Continued on next page

Coding categories	Relationship to visual literature	Examples (continued)
Humans	The number of humans in images denotes brotherhood, individual agency, or intended targets (Wilson, 2017; Winkler & Damanhoury, 2022).	  <i>Al-Naba</i> (23) <i>Inspire</i> (3)
Destruction	Explosions and destruction of infrastructure or religious iconography are used to instill fear and acquiescence (Winkler et al., 2018; Winkler, Dewick, et al., 2019).	  <i>Dabiq</i> (11) <i>Al-Masra</i> (4)
Leaders	Images of leaders denote enemies, potential allies, and intra-jihadist conflict following ISIS's break from al-Qaeda (Winkler & Damanhoury, 2022).	  <i>Rumiyah</i> (1) <i>Jihadi Recollections</i> (4)







Continued on next page

Coding categories	Relationship to visual literature	Examples (continued)
Flags	Flags are used to recruit supporters and anchor enmity toward rival states and groups (El Karhili et al., 2021; Warrick, 2016).	 <p><i>Dabiq</i> (13) <i>Al-Masra</i> (12)</p>
Semiotic framing elements		
Viewer position	Camera angle shapes perceived strength and credibility (Kraft, 1986; McCain et al., 1977).	 <p><i>Al-Naba</i> (123) <i>Inspire</i> (2)</p>
Image position	Foregrounding increases salience while backgrounding provides contextual grounding (Stone et al., 2003).	 <p><i>Dabiq</i> (10) <i>Al-Masra</i> (14)</p>
Viewer distance	Intimate versus public distance implies relational closeness or collective anonymity (Jewitt & Oyama, 2008).	 <p><i>Rumiyah</i> (12) <i>Inspire</i> (11)</p>

Continued on next page

Coding categories	Relationship to visual literature	Examples (continued)
Eye contact	Direct gaze communicates dominance while averted gaze suggests reduced confrontation (Tang & Schmeichel, 2015).	  <p data-bbox="778 639 864 667"><i>Dabiq</i> (6)</p> <p data-bbox="908 639 1050 698"><i>Jihadi Recollections</i> (1)</p>
Facial expressions	Positive expressions promote trust whereas negative expressions prompt skepticism (Forgas & East, 2008).	  <p data-bbox="766 999 876 1027"><i>Al-Naba</i> (16)</p> <p data-bbox="933 999 1025 1027"><i>Inspire</i> (4)</p>
Stance	Body posture signals submissiveness or strength and control (Green, 2015; Jewitt & Oyama, 2008).	  <p data-bbox="766 1276 876 1304"><i>Rumiyah</i> (10)</p> <p data-bbox="933 1276 1025 1304"><i>Inspire</i> (15)</p>
Connotative framing elements		

Continued on next page

Coding categories	Relationship to visual literature	Examples (continued)
State-building	Images of governance, services, markets, and territory signal institutional capacity and caliphate aspirations (Damanhoury, 2022; El Karhili et al., 2021; Lokmanoglu, 2020; Winkler, El-Damanhoury, et al., 2019; Zelin, 2015).	  <p data-bbox="766 504 874 528"><i>Al-Naba</i> (2)</p> <p data-bbox="921 504 1034 528"><i>Al-Masra</i> (7)</p>
Law enforcement	Images of capturing and punishing lawbreakers reinforce deterrence and community policing narratives (Barr & Herfroy-Mischler, 2017; Chouliaraki & Kissas, 2018; Damanhoury & Winkler, 2018).	  <p data-bbox="778 938 866 962"><i>Dabiq</i> (4)</p> <p data-bbox="917 938 1043 962"><i>Al-Masra</i> (17)</p>
Allegiance pledges	Allegiance imagery visualizes coalition-building and collective solidarity (Gregg, 2023).	  <p data-bbox="766 1326 874 1350"><i>Al-Naba</i> (12)</p> <p data-bbox="917 1326 1043 1350"><i>Al-Masra</i> (36)</p>

Continued on next page

Coding categories	Relationship to visual literature	Examples (continued)	
Media propaganda	Infographics and posters function as authority strategies and publicity for media products (Abdelrahim, 2019; Glausch, 2020).		
		<i>Dabiq</i> (13)	<i>Inspire</i> (11)
About to die	“About to die” images are especially circulation-prone and vary by certainty and probability (Winkler et al., 2016; Zelizer, 2010).		
		<i>Rumiyah</i> (3)	<i>Inspire</i> (16)
Religion	Religious gestures and iconography signal piety, legitimacy, and ideological commitment (Heck, 2017; Winkler, 2022).		
		<i>Dabiq</i> (7)	<i>Jihadi Recollections</i> (2)

Appendix B Computational analysis

B1.ChatGPT-4o prompt

You are labeling a single image. Return ONLY a minified JSON object with integer codes (no extra text).

Conventions:

- 0 = Not applicable (structurally skipped by dependency rules)
- 99 = Unclear / cannot determine
- Mixed means multiple mutually exclusive states apply in the same image.

Counting & visibility rules:

- Count a human if $\geq 25\%$ of head/torso is visible (silhouettes count if clearly human).
- Small group = 2-10 inclusive; Large group = ≥ 11 .
- Eye direction: Up/Down if head pitch is $\sim \geq 15^\circ$ from neutral; otherwise Eye Level.

Dependencies:

- If Humans=4 (No humans), then FacialExpressions, EyeContact, Stance = 0.
- If Death=3 (N/A), then and AboutToDie=0.
- If AboutToDie $\in \{1,2,3\}$, then Death $\neq 3$.
- State = actions/activities

Categories and codes:

Distance (viewer distance):

- 1 Intimate (face/neck close-up)
- 2 Personal (~1.5-4 ft, waist-up)
- 3 Social/Public (>4 ft)
- 4 Mixed
- 5 Unknowable (e.g., infographics, maps)
- 99 Unclear

ViewerPosition:

- 1 Looking up
- 2 Looking down
- 3 Eye level
- 4 Mixed
- 99 Unclear

Humans:

- 1 One
- 2 Small group (2-10)
- 3 Large group (≥ 11)
- 4 No humans
- 99 Unclear

MilitaryRole (relationship of depicted humans to any military/armed group):

- 1 Martyr (post-mortem)
- 2 Armed group member (group A example: IS)
- 3 Armed group member (non-A)
- 4 Future recruits (children in uniform/with weapons)
- 5 Mixed (1-4)
- 6 No military role present
- 99 Unclear

Leaders (visibly depicted or clearly referenced by portrait/statue/backdrop):

- 1 Jihad/Extremist leaders
- 2 Western state leaders
- 3 Arab state leaders
- 4 Asian/Russian state leaders
- 5 Shiite/Tribal/Other Muslim group leaders
- 6 Mixed (1-5)
- 7 No leaders present
- 99 Unclear

FacialExpressions:

- 1 Positive
- 2 Negative
- 3 Mixed
- 4 No humans (structural N/A → use 0 if Humans=4)
- 99 Unclear

EyeContact:

- 1 Looking at viewer
- 2 Looking upward
- 3 Looking downward
- 4 Looking at other person/object
- 5 Eyes closed/not visible
- 6 No humans (structural N/A → use 0 if Humans=4)
- 99 Unclear

Stance:

- 1 On knees (not praying)
- 2 Sitting (default if riding)
- 3 Standing
- 4 Lying down
- 5 Praying
- 6 Mixed (1-5)
- 7 No humans (structural N/A → use 0 if Humans=4)
- 99 Unclear

Death:

- 1 About to die
- 2 Dead
- 3 Not applicable (no death present)
- 99 Unclear

State (actions/activities):

- 1 Social services/infrastructure use (education, health)
- 2 Law enforcement/punishment
- 3 Maps (not infographics)
- 4 Local markets/economy/agriculture
- 5 Passports
- 6 Natural landscape (pristine)
- 7 Pledging allegiance to Group A (e.g., IS)
- 8 Other extremist-state propaganda (banners, hashtags)
- 9 Mixed (1-8, excluding 3 if purely graphical)
- 10 Not applicable (no state actions)
- 99 Unclear

Religion (visible practices/objects):

- 1 One finger pointing upward (tawhid gesture)
- 2 Reading scripture
- 3 Praying
- 4 Hajj (Kaaba)
- 5 Religious iconography/shrines
- 6 Mixed (1-5)
- 7 Not applicable (no religious content)
- 99 Unclear

Flag:

- 1 Extremist group flag (e.g., Group A)
- 2 U.S. flag
- 3 MENA state flag

4 Other flags
 5 Multiple flags
 6 Not applicable (no flags)
 99 Unclear

Destruction:

1 Active (fire/explosion)
 2 Aftermath (destroyed structures/iconography)
 3 Not applicable (no destruction)
 99 Unclear

AboutToDie (ONLY if applicable):

1 Possible death (no confirmed kill)
 2 Certain death (confirmed kill imminent)
 3 Presumed death (weapons/destruction implying lethal risk)
 0 Not applicable
 99 Unclear

Return JSON with this schema:

```
{
  "Distance": int, "ViewerPosition": int, "Humans": int,
  "MilitaryRole": int, "Leaders": int, "FacialExpressions":
  int, "EyeContact": int, "Stance": int,
  "Death": int, "State": int, "Religion": int, "Flag": int,
  "Destruction": int, "AboutToDie": int}
Optionally include "confidence": {<same keys>: float 0-1}
```

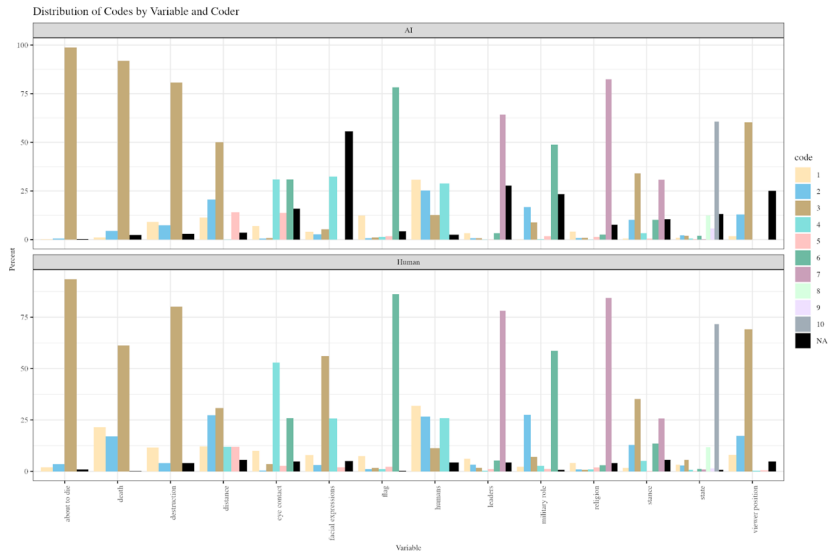


Figure B1: Distribution of codes by variable and coder

Variable	Number compared	Precision	Recall	Weighted precision	Weighted recall	Weighted F1	Weighted F2
<i>Denotative framing elements</i>							
Humans	6806	0.80	0.80	0.83	0.83	0.83	0.83
Destruction	6787	0.77	0.82	0.92	0.92	0.92	0.92
Military role	5551	0.47	0.34	0.73	0.75	0.74	0.75
Death	7107	0.70	0.41	0.65	0.94	0.77	0.86
Leaders	5029	0.43	0.29	0.83	0.86	0.84	0.85
Flag	6959	0.65	0.62	0.90	0.90	0.90	0.90
<i>Semiotic framing elements</i>							
Distance	6637	0.61	0.52	0.59	0.73	0.65	0.70
Viewer position	5198	0.35	0.31	0.76	0.82	0.79	0.80
Eye contact	5853	0.49	0.55	0.69	0.76	0.72	0.74
Stance	6173	0.60	0.58	0.72	0.76	0.74	0.75
Facial expressions	3077	0.54	0.56	0.74	0.88	0.80	0.85
<i>Connotative framing elements</i>							
State	6283	0.35	0.32	0.67	0.66	0.67	0.66
Religion	6462	0.47	0.42	0.88	0.89	0.89	0.89
About to die	7192	0.54	0.37	0.94	0.99	0.97	0.98

Table B1: Precision, recall, and weighted performance metrics for each variable

Appendix C Image-context relationship

Context Variable	Image Characteristic	Chi-Square
Announcements of Anticipated Oppositional Troop Withdrawals (Lokmanoglu et al., 2022)	Number of Images per Page; Foreground/Background; Viewer Distance; Number of Humans; Stance; Eye Contact; Facial Expressions; State-buildings; Flags; Death; About to Die	$p < .01$
Online censorship of Militant Group Accounts (McMinimy et al., 2021)	Number of Images per Page; Militant/Military Roles; Viewer Distance; Death; About to Die	$p < .01$
Online censorship of Militant Group Accounts (McMinimy et al., 2021)	State-building	$p < .05$
Highly Lethal and Publicized Attacks by Militant Group (Winkler et al., 2021)	Number of Images per Page; Leaders; State-building	$p < .01$
Highly Lethal and Publicized Attacks by Militant Group (Winkler et al., 2021)	Religion; Militant/Military Roles	$p < .05$
Militant Group Periods of Growth & Decline (Damanhoury & Winkler, 2018; Kaczkowski et al., 2021)	Number of Images per Page; Militant/Military Roles; State-building	$p < .01$
Death of Militant Media and Political Leaders (Winkler, El-Damanhoury, et al., 2019)	Militant/Military Roles; Death; About to Die; State-building	$p < .01$
Relative Standing Rise of Similar Competing Ideological Groups (Winkler, El-Damanhoury, et al., 2019)	Foreground/Background; Militant/Military Roles; Leaders; Viewer Distance; State-building	$p < .01$

Table C1: Context Relationships that Correlate with Increased or Decreased Levels of Image Types