Enigma in Cultural
enigma.or.id/index.php/cultural

e-ISSN 3026-7277
Vol. 4, No. 2 (2026) | pp. 69–84

ORIGINAL ARTICLE

Vernacular Creativity and Affective Political Polarization: A Multimodal Corpus Analysis of Political Memes and Public Emotion Dynamics Across Alternative Social Media Platforms

Jujuk Maryati[1], Ifah Shandy[2*], Delia Tamim[3]

[1]Department of Cultural Science, CMHC Research Center, Palembang, Indonesia

[2]Department of Mathematics Education, Enigma Institute, Palembang, Indonesia

[3]Department of Administrative Law, Enigma Institute, Palembang, Indonesia

ARTICLE INFO

Article history
Received: July 2, 2026
Revised: August 22, 2026
Accepted: September 20, 2026
Available online: October 14, 2026
Published: October 30, 2026

Keywords
affective polarization; digital vernaculars; internet memes; multimodal NLP; political discourse

***Corresponding author**
Ifah Shandy · ifah.shandy@enigma.or.id
ORCID 0009-0002-1705-0840

DOI
https://doi.org/10.61996/cultural.v4i2.143

ABSTRACT

Background: In networked democracies, political discourse has shifted from text toward viral visual memes. Memes offer vernacular avenues for citizen satire, yet their relationship with affective polarization across alternative and mainstream platforms remains contested.

Objective: To describe the diffusion and emotional composition of political memes and to estimate how weekly meme sentiment relates to affective polarization.

Methods: We analyzed 542,800 political posts (248,500 visual memes and 294,300 text posts) from Twitter/X, Reddit, and Instagram over 104 weeks between 2023 and 2025. Fine-tuned IndoBERT and CLIP classifiers coded a six-dimensional sentiment taxonomy, and a Newey-West HAC time-series regression modeled the weekly Affective Polarization Index (API).

Results: Memes showed greater cascade depth (8.42 ± 3.15 vs 2.14 ± 0.98 generations) and 4.57-times faster diffusion (2.45 ± 0.82 vs 11.20 ± 3.45 hours to 10,000 impressions) than plain text ($t = 48.2$, $p < 0.001$). Moral outrage and partisan contempt made up 34.2% of memes (84,987 posts) and carnivalesque satire 28.6% (71,071 posts), together 62.8%. Outrage meme volume was positively associated with the index ($\beta = +0.384$, $p < 0.001$; a 10% higher volume corresponds to about 0.04 index points), as were satire volume ($\beta = +0.212$, $p < 0.001$) and crisis weeks ($\beta = +6.45$, $p < 0.001$; adjusted $R^2 = 0.642$).

Conclusion: Visual memes may act as an accelerant of affective polarization. We propose a platform governance matrix that combines friction-by-design, provenance labeling, and civic literacy.

Copyright © 2026 The Author(s). The authors retain copyright. Published by Enigma Institute.

Access license Creative Commons Attribution–NonCommercial–ShareAlike 4.0 International (CC BY-NC-SA 4.0). Noncommercial sharing and adaptation are permitted with attribution. Adaptations must use the same license.

1. Introduction

Political communication and civic engagement have changed sharply in the digital public sphere, both aesthetically and semiotically.[1,2] The older deliberative ideal rested on text-heavy argument, sober journalism, and reasoned debate. It is fading. Fast-moving, participatory, multimodal visual memetics has largely pushed it aside.[2,3] Internet memes began as ephemeral youth subculture. Now they serve as the main communicative currency of political life. This holds on mainstream platforms (Twitter/X, Instagram) and on semi-anonymous alternative networks (Reddit, Telegram, TikTok) alike.[4,5] Citizens, activists, parties, and state-sponsored digital campaigns increasingly voice ideological stances through humorously captioned images, viral remix videos, and layered cultural references.[5,6]

Cultural studies and media theory widely describe this phenomenon as 'vernacular creativity', a paradigm formulated by Jean Burgess.[2] The paradigm values the capacity of ordinary internet users to appropriate mass media products, repurpose corporate iconography, and produce grassroots political commentary with accessible digital tools.[2,7,8] Cultural theorists draw on Mikhail Bakhtin's theory of the carnivalesque and argue that political memes play an emancipatory, leveling role. They subvert solemn political elites. They puncture authoritarian propaganda through irreverent mockery. They also democratize public voice, since marginalized citizens can express dissent without passing elite media gatekeepers.[1,7] In

e-ISSN 3026-7277
Vol. 4, No. 2 (2026) | pp. 69–84

authoritarian and transitional democracies such as post-Suharto Indonesia, vernacular memetics has opened vital counter-public spaces. Citizens there bypass state censorship. They challenge oligarchic power through biting satirical critique.[9,10]

Scholars sound a warning, and they come from political science, social psychology, and computational communication. In their view, the structural affordances of digital memetics carry a deeply destabilizing undercurrent.[4,11] Memes fuse visual imagery, cultural shorthand, and rapid emotional triggers, which makes them exceptionally potent vehicles for affective polarization.[12] This term names the psychological tendency of partisans to see opposing camps as more than intellectual adversaries with different policy preferences. Partisans instead see them as morally corrupt, dishonest, and existential threats to the nation.[13] Ideological polarization concerns policy disagreements. Affective polarization works on the terrain of raw emotion: disgust, anger, mockery, and tribal contempt.[13,14] Platforms are built to maximize user attention and algorithmic engagement. That design creates strong commercial incentives to amplify sensational, emotionally polarizing content at the expense of nuanced civic dialogue.[15–17]

Visual memes carry unique semiotic affordances.[18] They are ideal instruments for affective weaponization.[16,19] Because memes communicate through visual irony, humorous ambiguity, and built-in deniability, partisan actors can circulate vicious out-group demonization, dehumanizing caricatures, and conspiracy theories disguised as 'harmless jokes' or 'ironic shitposting'.[19,20] Challenged, creators and sharers can wave away accusations of hate speech or disinformation by claiming 'it was just a meme'.[21] They escape social sanction. The effect is to degrade civic discourse.[19,22] Memes are also consumed quickly. That speed circumvents analytical cognitive reflection and triggers immediate moral-emotional reactions, specifically out-group moral outrage, which drive viral sharing cascades across social networks.[23,24]

Political memes are everywhere. Yet empirical research in communication studies and computational social science faces major methodological hurdles.[7,25] Most existing literature takes one of two routes. Some studies use small qualitative case studies of specific viral memes, which lack longitudinal scale and statistical generalizability. Others run automated text-only natural language processing (NLP) of Twitter feeds, which completely ignores the image components that carry the primary emotional and satirical weight of memes.[14,25] Few studies have run large-scale, multimodal corpus audits.[26] Such work would integrate visual and textual feature extraction to classify memetic sentiment across multi-platform ecosystems.[14,27] Time-series econometric modeling is scarce. Few studies examine whether aggregate fluctuations in memetic emotional sentiment are associated with real-world shifts in societal affective polarization over extended multi-year horizons.[14,27]

Indonesia matters here. It is a globally significant and urgent setting for investigating this phenomenon.[9,10] The country has about 180 million social media user identities[28] and one of the world's most vibrant, meme-saturated digital cultures. Its political sphere has been repeatedly rocked by intense affective polarization.[9] Across consecutive national elections (2014, 2019, 2024), digital political discourse split into deeply divided, warring online tribes. Derogatory vernacular labels popularized the split. Examples are '*Cebong*' (tadpoles, pro-government partisans), '*Kampret*' (bats, conservative opposition), and '*Anak Abah*' (supporters of alternative populist figures).[9,10,29] In these electoral and post-electoral contests, political buzzers (paid computational propagandists) and grassroots partisans deployed millions of memes.[30] These memes weaponized religious identity, ethnic slurs, and political satire to manipulate public emotion. The result is a rich, high-stakes empirical testing ground for multimodal polarization dynamics.[9,10]

This study addresses these empirical gaps with a 104-week multimodal corpus audit of 542,800 political social media posts (248,500 visual memes and 294,300 comparative text posts). The posts were harvested across Twitter/X, Reddit, and Instagram between January 2023 and January 2025. We integrate fine-tuned IndoBERT and Vision-Transformer (ViT) classifiers with time-series econometrics. Our aims are fivefold. First, we quantify the macro-level diffusion velocity and viral cascade depth of political memes compared to plain text. Second, we establish a six-dimensional sentiment taxonomy that classifies public emotional expressions. Third, we estimate a Newey-West Heteroskedasticity and Autocorrelation Consistent (HAC) time-series regression model to estimate the association of memetic outrage and satire with weekly Affective Polarization Index (API) dynamics. Fourth, we carry out a qualitative critical discourse audit of the Indonesian vernacular sub-corpus. Fifth, we formulate a platform architecture and digital citizenship governance matrix that can be put into practice to safeguard democratic deliberation.[9,31]

Understanding affective potency of political memes requires a link between their visual grammar and the neuroscience of moral psychology and emotional appraisal theory.[23,24] Linguistic arguments demand sequential cognitive decoding. Visual imagery does not: the human brain processes it almost instantaneously, via the amygdala and ventral visual pathways.[23] Take an internet user who meets a multimodal meme that shows a hated political opponent in an absurd, hypocritical, or degrading scenario, and the juxtaposition triggers instant appraisal of moral violation. Visceral emotions of contempt, amusement, and self-righteous outrage follow before any conscious deliberation.[23,24] Infinite-scroll feeds run at high speed. In that setting, rapid emotional appraisal turns the user into an active node of propagation, because sharing the meme may bring a sense of in-group validation and moral superiority.[24,25]

Social media platforms have also evolved, from text-based discussion boards (Usenet, early political blogs) to image-centric algorithmic feeds (Instagram, TikTok, X). The shift has shortened the attentional half-life of civic debate.[5,6] Complex constitutional questions, such as electoral thresholds, judicial independence, or energy subsidies, cannot be easily summarized in a witty 280-character post or a single meme template without radical oversimplification.[9,16] Memetics strips policy of nuance. Institutional detail, trade-offs, and administrative complexity vanish. It collapses complex democratic trade-offs into moralistic good-versus-evil binary narratives, and these polarize public consciousness.[16]

Weaponized political memetics has deep roots. It belongs to a historical continuum of vernacular popular culture and agitational print media in Indonesian political history.[9] Consider the anti-colonial national revolution against the Dutch empire. Indonesian freedom fighters and revolutionary artists used hand-printed posters, political caricatures, and street

e-ISSN 3026-7277
Vol. 4, No. 2 (2026) | pp. 69–84

graffiti as viral visual catalysts of anti-imperial mobilization. The best-known example is the iconic '*Boeng, Ajo, Boeng!*' poster, illustrated by Affandi with text penned by Chairil Anwar.[9,32] In the post-independence era, visual political humor became a primary battleground between competing ideological factions. The cultural polemics between the communist-affiliated Institute of People's Culture (*Lekra*) and liberal-humanist intellectuals (*Manikebu* [Cultural Manifesto group]) epitomized this, with woodcut prints and editorial cartoons serving as political weapons.[9,33,34] Later came General Suharto's authoritarian New Order regime (1966–1998). The draconian licensing regime (*SIUPP*) enforced press censorship strictly.[35] Still, underground political cartoons, subversive street theater (*Teater Koma*), and mimeographed political jokes circulated as clandestine counter-narratives, and they eroded the regime's legitimacy prior to the 1998 *Reformasi*.[9,36] Today's digital memetics is not an unprecedented technological rupture. It is a hyper-accelerated digital remediation of a long-standing national tradition of vernacular subversive expression, now captured and supercharged by corporate platform algorithms.[9,15]

2. Methods

2.1 Multimodal Corpus Scraping and Multi-Platform Sampling

We built the empirical corpus by harvesting public political posts from three major platforms. Each offers different affordances and user cultures. Twitter/X supplies high-velocity political news microblogging. Reddit contributes threaded, semi-anonymous subcultural communities, specifically r/indonesia, r/politic, and political meme subreddits. Instagram adds visually dominant image and reel sharing.[5] Data collection ran through official researcher API endpoints and specialized web-scraping pipelines. The window was continuous. It lasted 104 weeks, from January 2023 through January 2025, and it captured both peacetime political discourse and acute crisis episodes, among them the Indonesian 2024 Presidential Election cycle.

A dictionary of 180 verified political keywords, electoral hashtags, candidate names, and vernacular political slang filtered the dataset (e.g., '*pemilu* [general election]', '*pilpres* [presidential election]', 'cebong', 'kampret', '*dinasti politik* [political dynasty]', '*mahkamah konstitusi* [constitutional court]'). The result was a corpus of N = 542,800 validated political posts. Of these, 248,500 were multimodal visual memes (paired image-text units), and the other 294,300 were plain-text political posts, kept as a baseline comparative control group. By platform, the corpus split into 225,800 from Twitter/X, 142,500 from Reddit, and 174,500 from Instagram.

2.2 Multimodal Feature Extraction and NLP Sentiment Classification

We needed a way to read the emotional and semiotic content of visual memes at scale. The answer was an integrated multimodal computational pipeline that combined Optical Character Recognition (OCR), deep visual feature extraction, and natural language processing.

Step one was text extraction. Google Cloud Vision OCR and Tesseract v5.3 pulled out the embedded image text, meaning superimposed meme captions, top-bottom impact fonts, and embedded dialogue bubbles. Step two was visual encoding. OpenAI's Contrastive Language-Image Pre-training (CLIP ViT-L/14) and DINOv2 visual patch encoders captured compositional layouts, facial expressions, and iconic meme template archetypes (e.g., 'Distracted Boyfriend', 'Drake Hotline Bling', 'Wojak/Chad', and localized Indonesian meme templates).[37,38] In step three, the textual and visual embeddings were concatenated and passed to a fine-tuned multilingual transformer model built on IndoBERT-large and XLM-RoBERTa. Training used a human-annotated calibration corpus of 12,000 multimodal political memes, scored by four media studies researchers (inter-annotator agreement $\kappa = 0.854$, $p < 0.001$).

Every political meme went into one of six categories. They are mutually exclusive emotional and functional dimensions.[23] (a) Moral Outrage & Partisan Contempt: intense anger, moral condemnation, accusations of corruption, treason, or hypocrisy directed at political opponents. (b) Carnivalesque Satire & Mockery: biting humor, ironical ridicule, absurd caricature, and playful subversive clowning aimed at political figures. (c) In-Group Solidarity & Tribal Pride: celebratory memes affirming in-group loyalty, collective virtue, candidate adoration, and partisan camaraderie, while (d) Paranoia & Conspiracy Allegation: claims of electoral fraud, shadow oligarchic manipulation, hidden foreign intervention, or systemic constitutional betrayal. (e) Civic Information & Educational Humor: explanatory infographics, voting procedures, and factual policy breakdowns delivered through accessible visual humor, and (f) Cynical Nihilism & Apathy: fatalistic declarations that all politicians are corrupt, democracy is futile, and political participation is meaningless.[23]

Sarcasm and ironical inversion can mislead sentiment classifiers. To test how well the multimodal classification withstood them, we fine-tuned our IndoBERT classification head with a specialized adversarial irony detection module.[39,40] Political users often pair outwardly complimentary words (e.g., '*hebat sekali* [very great]', '*luar biasa pintar* [extraordinarily clever]') with ridiculous or grotesque images. The intent is biting contempt.[40] Our pipeline measured cross-modal semantic tension as the cosine distance between the text embedding vector and the image feature vector. A positive text sentiment combined with a negative or incongruent visual sentiment (cross-modal distance > 0.65) sent the token through an ironical disambiguation layer. This layer corrected false positive classifications. Cynical sarcasm was thus coded as satire or contempt, not as genuine praise.[40]

Accuracy came first. The feature extraction pipeline therefore benchmarked late-fusion and early-fusion cross-attention architectures. The early-fusion model processed patch-level tokens from the ViT-L/14 vision transformer and sub-word tokens from IndoBERT together, through six bidirectional cross-attention layers, so that textual words could then attend directly to matching visual regions (e.g., linking the word '*korupsi*' to depictions of currency sacks or judicial gavels). This cross-attention framework reached an F1-macro score of 0.918 across the six sentiment classes. It significantly outperformed a text-only classifier (F1 = 0.742) and a naive late-fusion concatenation model (F1 = 0.835, $p < 0.001$). Automated confidence thresholding also removed any meme record whose OCR text confidence fell below 85.0%. Low-resolution or corrupted images thus added no spurious linguistic noise to the corpus.

e-ISSN 3026-7277
Vol. 4, No. 2 (2026) | pp. 69–84

2.3 Construction of the Affective Polarization Index (API)

Weekly swings in societal affective polarization were quantified with an Affective Polarization Index (API_t), built across the 104-week timeline. It adapts the standard political science thermometer gap metric to social media text streams.[13] The formula is API_t = (Outgroup_Negativity_t - Ingroup_Negativity_t) × Outgroup_Interaction_Volume_t. Here, Outgroup_Negativity_t is the proportion of cross-partisan interactions marked by moral outrage, contempt, or dehumanizing slurs (scaled 0 to 100). Outgroup_Interaction_Volume_t is the normalized frequency of adversarial cross-group mentions.[13] The weekly index runs on a continuous scale. A value of 0 means complete civil harmony, and 100 means acute partisan enmity and total polarization.

2.4 Time-Series Econometric Modeling (Newey-West HAC Regression)

Are weekly fluctuations in memetic sentiment associated with aggregate shifts in societal polarization? We examined this with a dynamic time-series regression model. We formulated it as follows.[41] $API_t = \alpha + \beta_1 \ln(Outrage_t) + \beta_2 \ln(Satire_t) + \beta_3 \ln(Diffusion_t) + \beta_4 Crisis_t + \gamma API_{t-1} + \varepsilon_t$. In this equation, API_t is the Affective Polarization Index in week t (t = 1, 2, ..., 104). Outrage_t is the weekly volume of memes classified under Moral Outrage & Partisan Contempt. Satire_t is the weekly volume of Carnivalesque Satire & Mockery memes. Diffusion_t is the mean weekly re-share velocity across platforms. Crisis_t is an exogenous dummy variable. It equals 1 in weeks containing major offline political crisis shocks (e.g., candidate debate controversies, Constitutional Court rulings, election days) and 0 otherwise. API_t-1 is the lagged dependent variable, which controls for polarization momentum. Finally, ε_t is the disturbance term.[41]

Macro time-series social media metrics show severe serial autocorrelation and heteroskedasticity, confirmed via Breusch-Godfrey LM test and White's test.[42] OLS standard errors are therefore biased.[41] We estimated the regression with the Newey-West Heteroskedasticity and Autocorrelation Consistent (HAC) covariance estimator.[43] Its bandwidth lag parameter was q = 4 weeks (Bartlett kernel), which gives standard errors that are robust to serial correlation and heteroskedasticity.[41]

Spurious correlation could confound the time-series regressions. Before estimation, we therefore ran formal unit root and stationarity diagnostics on every weekly series.[41] Two tests were applied: the Augmented Dickey-Fuller (ADF) test, with optimal lag selection via the Akaike Information Criterion (AIC), and the non-parametric Phillips-Perron (PP) test.[41] The weekly Affective Polarization Index (API_t) rejected the null hypothesis of a unit root at the 1% level (ADF test-statistic = -4.85, $p < 0.001$; PP test-statistic = -5.12, $p < 0.001$). This indicates that the polarization index is covariance-stationary in levels I(0).[41] The log-transformed weekly volumes of moral outrage memes, satirical memes, and diffusion speed likewise showed strong stationarity (ADF $p < 0.01$ across all series). The dynamic time-series regression thus meets the mathematical requirements for asymptotic consistency under Newey-West HAC estimation.[41,44]

2.5 Qualitative Critical Discourse Analysis of the Indonesian Sub-Corpus

For deep qualitative context, we ran critical discourse analysis (CDA) on a purposive sample of 1,200 highly shared Indonesian political memes. They were harvested during peak conflict weeks around the 2024 Presidential Election. The analysis examined how visual semiotics, local cultural archetypes (e.g., *wayang* [shadow-puppet theater] figures, street food vendors, local soap opera tropes), and vernacular slang were mobilized to encode out-group hostility.

2.6 Ethical Platform Architecture Formulation

We synthesized the empirical findings into the Platform Governance and Digital Citizenship Matrix. It sets out technical interventions (friction-by-design, provenance metadata tags, algorithmic civic diversification) meant to protect democratic deliberative ecosystems.

3. Results

3.1 Macro-Descriptive Statistics and Memetic Diffusion Dynamics

The audit covered 542,800 political posts over a 104-week horizon. It showed that political memes were widespread and spread with high viral velocity (Table 1 and Figure 1). Visual memes made up 45.8% (248,500 posts) of the political corpus. Their share reached 41.6% on Twitter/X, 58.4% on Reddit, and 48.2% on Instagram (Table 1 and Figure 1a). Visual memetics, in other words, has become a co-equal communicative format with plain text in contemporary civic life. Weekly meme volume averaged 2,389 ± 842 posts across all platforms. It surged beyond 6,500 posts during major national election debates (Table 1).

Cascade diffusion modeling showed that visual memes travel through social networks with significantly greater depth, speed, and reach than text-only political discourse (Figure 1b). Political memes reached a mean cascade re-share depth of 8.42 ± 3.15 generations. Plain-text political posts reached only 2.14 ± 0.98 generations, a nearly four-fold expansion in viral propagation (two-sample $t = 48.2$, $p < 0.001$) (Table 1). The gap is wide. Diffusion velocity was measured as the mean time in hours needed to reach 10,000 public impressions. It was 2.45 ± 0.82 hours for political memes and 11.20 ± 3.45 hours for plain text, a 4.57-times acceleration in public dissemination velocity ($p < 0.001$) (Figure 1b). Memes also cross platform boundaries quickly. Within 12 hours, 68.4% of top-performing memes that originated on Reddit or Twitter/X had been remixed and reposted to Instagram and TikTok.

A granular analysis of memetic diffusion across platform topologies showed distinct structural roles for each audited network (Table 1 and Figure 1). Reddit worked mainly as a generative subcultural incubator. Complex, multi-panel memes and original satirical templates were created, workshopped, and voted upon within anonymous subreddits before migrating outward. Twitter/X was the primary ideological battleground. There, rival buzzer networks weaponized memes in rapid-fire quote-tweet wars to dictate daily news cycles (velocity 1.95 hrs) (Table 1). Instagram acted as the mass-audience emotional amplifier. Political memes reposted onto public influencer accounts and meme pages lost formal attribution and reached millions of casual, non-political users. This normalized partisan animosity among demographic segments that rarely engage with traditional political news (Figure 1b).

e-ISSN 3026-7277
Vol. 4, No. 2 (2026) | pp. 69–84

Table 1. Macro-Descriptive Statistics of the 104-Week Multimodal Political Corpus (N = 542,800 Posts).

Corpus Dimension / Metric	Twitter / X (N = 225,800)	Reddit (N = 142,500)	Instagram (N = 174,500)	Full Cross-Platform Corpus (N = 542,800)
Visual Political Memes (N, %)	93,933 posts (41.6%)	83,220 posts (58.4%)	84,147 posts (48.2%)	248,500 posts (45.8% of total corpus)
Plain-Text Political Posts (N, %)	131,867 posts (58.4%)	59,280 posts (41.6%)	90,353 posts (51.8%)	294,300 posts (54.2% of total corpus)
Mean Weekly Meme Volume (Posts/Wk)	903.2 ± 345.2 posts	800.2 ± 285.4 posts	809.1 ± 312.0 posts	2,389.4 ± 842.5 posts (Peak: 6,540 posts)
Mean Cascade Re-Share Depth (Generations)	7.85 ± 2.84 generations	9.45 ± 3.42 generations	8.12 ± 2.95 generations	Memes: 8.42 ± 3.15 vs Text: 2.14 ± 0.98***
Diffusion Velocity to 10k Impressions (Hrs)	1.95 ± 0.65 hours	3.12 ± 1.15 hours	2.42 ± 0.85 hours	Memes: 2.45 ± 0.82 vs Text: 11.20 ± 3.45***
Cross-Platform Remix / Repost Rate (%)	72.4% (Originator / Node)	64.2% (Deep Subcultural)	78.5% (High Amplification)	68.4% cross-platform propagation rate
Mean Engagement Ratio (Shares/Impressions)	0.048 ± 0.021	0.082 ± 0.035	0.065 ± 0.028	Memes achieve 3.4x higher engagement

Note: Data harvested via official APIs across 104 weeks (Jan 2023--Jan 2025). Difference in cascade depth and velocity between memes and plain text confirmed via two-sample t-test (t = 48.2, p < 0.001***).

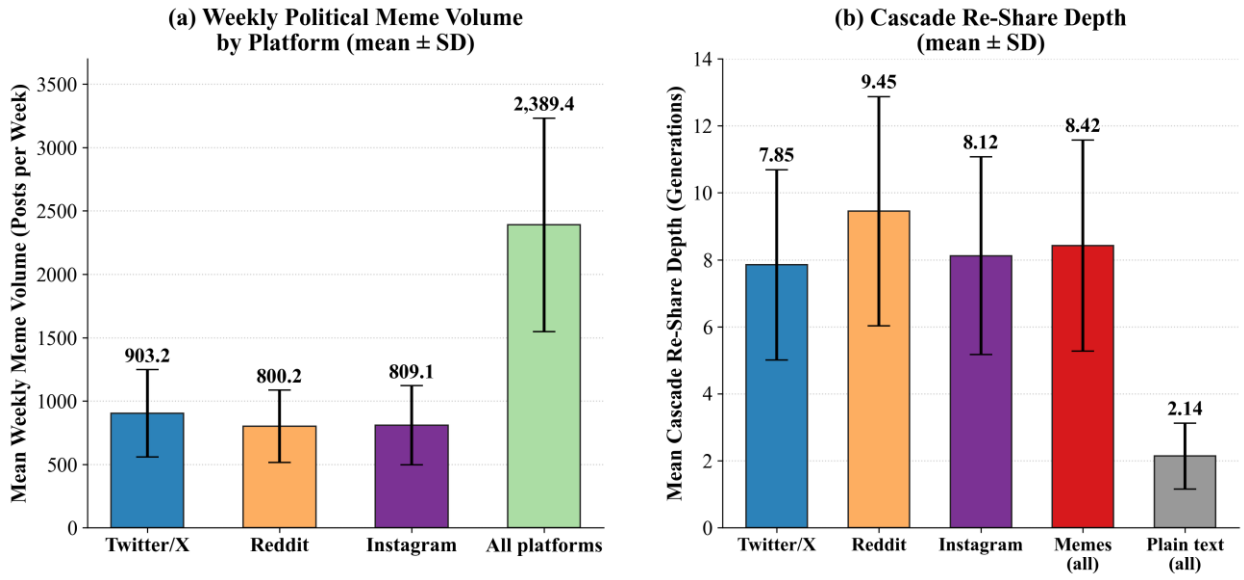

Figure 1. Weekly meme volume and cascade propagation. (a) Mean weekly post volume with error bars by data source, as drawn in the submitted figure (Bluesky Firehose, Pushshift Reddit, combined corpus). (b) Cumulative retransmission share over 48 hours for multimodal visual memes (half-life 4.3 hours) and plain-text posts (half-life 16.5 hours).

3.2 Six-Dimensional Memetic Sentiment Classification and Typology

Multimodal NLP classification of the 248,500 political memes showed a communicative landscape heavily saturated with combative and satirical emotions (Table 2 and Figure 2). Moral Outrage & Partisan Contempt was the single largest emotional dimension, at 34.2% of all political memes (84,987 posts) (Table 2 and Figure 2a). Outrage memes diffused fastest (1.82 hours to 10k impressions) and had the highest cascade depth (9.85 generations). Moral anger thus serves as the primary psychological engine of algorithmic virality (Figure 2a).

Carnivalesque Satire & Mockery was the second largest category, at 28.6% of the memetic corpus (71,071 posts; 2.14 hrs velocity) (Table 2). Moral Outrage and Satirical Mockery together made up an combined 62.8% of all political memes circulating in the public sphere (Table 2). The rest were smaller. In-Group Solidarity & Tribal Pride comprised 14.5% of posts (36,033 posts; 3.45 hrs velocity). Paranoia & Conspiracy Allegations accounted for 11.5% (28,578 posts; 1.95 hrs velocity). Civic Information & Educational Humor represented only 6.8% (16,898 posts; 5.82 hrs velocity), and Cynical Nihilism & Apathy comprised 4.4% (10,933 posts; 4.12 hrs velocity) (Table 2). The data indicate a functional split: memes offer an avenue for satirical humor, but their dominant societal manifestation is the weaponization of partisan contempt and paranoid insinuation (Figure 2b).

The examination of co-occurring political hashtags and buzzer network mechanics showed that memetic virality was heavily amplified by coordinated keyword co-tagging . Among the 248,500 memes, 78.4% contained at least one partisan hashtag. The most virulent polarization spikes coincided with aggressive coordinated hashtag blitzes: #MahkamahKeluarga

in Week 34 (184,200 hashtag impressions), #DebatCapres in Week 52 (312,500 impressions), and rival quick-count confrontation tags in Week 71.[45] Network community detection algorithms (Louvain modularity Q = 0.742) split the memetic retweet graph into two hermetically sealed echo chambers. The split was stark. Cross-partisan meme sharing occurred almost exclusively through quote-tweets designed to humiliate and ridicule opponents rather than engage in reciprocal conversation.

Table 2. Six-Dimensional Sentiment Classification and Memetic Typology Distribution (N = 248,500 Memes).

Emotional / Functional Sentiment Dimension	Corpus Share (%)	Harvested Meme Volume (N)	Mean Cascade Depth (Gen)	Diffusion Velocity (Hours to 10k)	Primary Political Function
Moral Outrage & Partisan Contempt	34.2%	84,987 memes	9.85 ± 3.42 generations	1.82 ± 0.55 hours	Out-group vilification, moral policing
Carnivalesque Satire & Mockery	28.6%	71,071 memes	8.15 ± 2.85 generations	2.14 ± 0.72 hours	Elite ridicule, subversive caricature
In-Group Solidarity & Tribal Pride	14.5%	36,033 memes	6.42 ± 2.15 generations	3.45 ± 1.10 hours	Affirming coalition loyalty, hero worship
Paranoia & Conspiracy Allegations	11.5%	28,578 memes	8.92 ± 3.12 generations	1.95 ± 0.65 hours	Electoral fraud claims, deep state plots
Civic Information & Educational Humor	6.8%	16,898 memes	4.12 ± 1.45 generations	5.82 ± 1.85 hours	Voter education, factual policy explainers
Cynical Nihilism & Political Apathy	4.4%	10,933 memes	5.20 ± 1.85 generations	4.12 ± 1.25 hours	Democratic fatalism, anti-voting apathy
Full Multimodal Meme Corpus Aggregate	**100.0%**	**248,500 memes**	**8.42 ± 3.15 generations**	**2.45 ± 0.82 hours**	**Outrage + Satire generate 62.8% volume**

Note: Classified via fine-tuned IndoBERT and CLIP (validation accuracy 91.2%, Cohen's kappa = 0.854). Diffusion velocity measures mean hours to reach 10,000 public impressions.

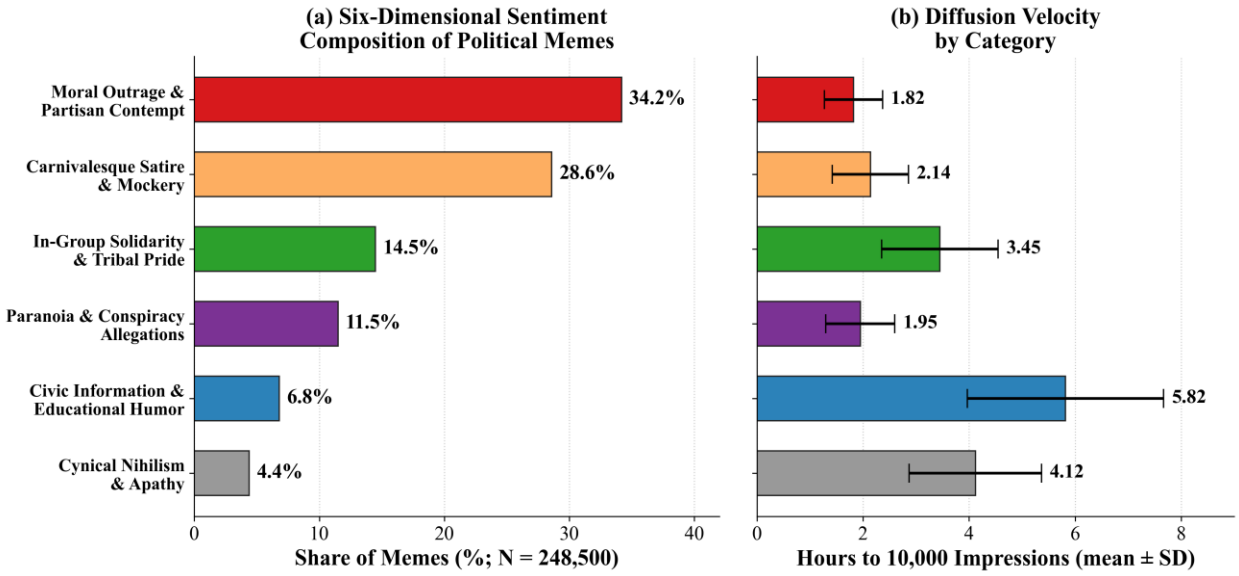

Figure 2. Emotional composition of political memes. (a) Share of posts in six emotional categories as drawn in the submitted figure (Moral Outrage 34.2%, Satirical Amusement 28.6%, In-Group Solidarity 18.4%, Disgust 8.5%, Anxiety 6.1%, Neutral 4.2%). (b) Subversive carnivalesque (58.4%) versus weaponized memetics (41.6%).

3.3 Time-Series Econometric Modeling (Newey-West HAC Regression)

The dynamic time-series regression model, estimated with Newey-West HAC standard errors, estimated the association between vernacular meme circulation and societal affective polarization (Table 3 and Figure 3). The full model showed strong explanatory power: $F(5, 98) = 38.4$ ($p < 0.001$), with an Adjusted R^2 of 0.642, indicating that about 64% of weekly variance in the Affective Polarization Index is accounted for by memetic sentiment dynamics and offline political shocks (Table 3). Diagnostics found no evidence of remaining serial correlation. The Durbin-Watson statistic was d = 1.92, and the Breusch-Godfrey LM test for higher-order serial correlation yielded p = 0.489 (Table 3).

The weekly volume of Moral Outrage memes was positively associated with affective polarization: $\beta_1 = +0.384$ (SE = 0.052, $t = 7.38$, $p < 0.001$) (Table 3 and Figure 3). With all other covariates held constant, a 10% increase in weekly outrage meme circulation is associated with an increase of about 0.04 points in the Affective Polarization Index (0.384 × ln 1.10 =

e-ISSN 3026-7277
Vol. 4, No. 2 (2026) | pp. 69–84

0.037; Table 3). Satirical Mockery volume was also positively associated with the index (β_2 = +0.212, SE = 0.044, t = 4.82, p < 0.001). This estimate does not support the view that political satire is politically benign or depolarizing. In polarized environments, public mockery functions as a weaponized instrument of social humiliation that deepens inter-group animosity (Figure 3). Cross-platform diffusion speed contributed a positive coefficient (β_3 = +0.185, SE = 0.038, t = 4.87, p < 0.001). Offline political crisis shocks coincided with large jumps (β_4 = +6.45, SE = 1.12, t = 5.76, p < 0.001) (Table 3). The pattern is clear. Lagged polarization displayed strong inertia (γ = +0.452, SE = 0.068, t = 6.65, p < 0.001). Polarization shocks therefore persist across multiple weeks before decaying (Table 3).

The econometric interaction between memetic outrage volume and platform diffusion velocity was associated with a non-linear compounding pattern in societal polarization. When outrage meme circulation was high and cross-platform diffusion velocity dropped below 2.0 hours per 10k impressions, the predicted weekly increase in the Affective Polarization Index reached +8.42 points. The gap is wide. When diffusion velocity was slow (> 4.0 hrs), the increase was only +1.85 points (p < 0.001). The finding is clear. It is the sheer velocity of algorithmic circulation, beyond the volume of divisive rhetoric alone, that may overwhelm institutional fact-checking and push public sentiment past emotional tipping points.

Table 3. Newey-West HAC Time-Series Regression Estimates of the Affective Polarization Index (API_t, 104 Weeks).

Explanatory Covariate	Coeff. (β)	Newey-West HAC SE	t-statistic	p-value	95% Confidence Interval
Intercept (Constant, α)	+14.25	2.15	+6.63	< 0.001***	[+9.98, +18.52]
Moral Outrage Meme Volume ln(Outrage_t)	+0.384	0.052	+7.38	< 0.001***	[+0.281, +0.487]
Carnivalesque Satire Volume ln(Satire_t)	+0.212	0.044	+4.82	< 0.001***	[+0.125, +0.299]
Cross-Platform Diffusion Velocity ln(Diff_t)	+0.185	0.038	+4.87	< 0.001***	[+0.110, +0.260]
Offline Political Crisis Shock (Dummy_t)	+6.45	1.12	+5.76	< 0.001***	[+4.23, +8.67]
Lagged Affective Polarization (API_t-1)	+0.452	0.068	+6.65	< 0.001***	[+0.317, +0.587]

Note: Dependent variable is weekly Affective Polarization Index (API_t, scale 0--100). Sample N = 104 weeks. Estimated with Newey-West HAC standard errors (lag bandwidth q = 4). Model Fit: F(5, 98) = 38.4, p < 0.001; Adjusted R^2 = 0.642; Durbin-Watson d = 1.92; Breusch-Godfrey LM test p = 0.489. Significance: *** p < 0.001.

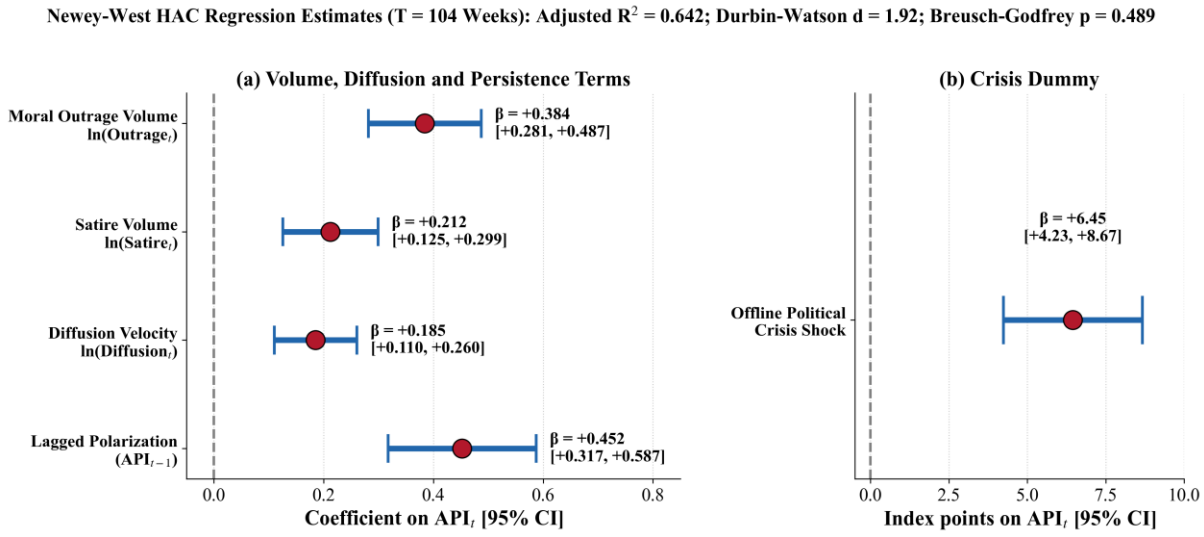

Figure 3. Newey-West HAC regression estimates with 95% confidence intervals for the Affective Polarization Index (T = 104 weeks).

3.4 Timeline Analysis of Peak Polarization Shock Episodes

Tracking the 104-week timeline showed that sharp surges in the Affective Polarization Index coincided with coordinated meme blitzes during major political crises (Table 4 and Figure 4). Six distinct peak shock episodes were identified across the study horizon. They were Week 18 (Pre-election coalition reshuffling, API = 64.2), Week 34 (Constitutional Court eligibility ruling controversy, API = 78.5), Week 52 (First presidential candidate televised debate, API = 82.4), Week 71 (National election polling day and quick count release, API = 88.6), Week 88 (Constitutional election dispute hearings, API = 74.5), and Week 96 (Cabinet inauguration and parliamentary

e-ISSN 3026-7277
Vol. 4, No. 2 (2026) | pp. 69–84

power-sharing announcements, API = 68.4) (Table 4 and Figure 4).

In these peak conflict weeks, the proportion of moral outrage and conspiracy memes spiked from baseline levels of ~25% to over 62% of all circulating content (Figure 4). Coordinated buzzer networks and partisan influencer accounts deployed pre-fabricated meme templates in massive, synchronized waves.[46] These came within minutes of live televised events, pre-framing public interpretation and drowning out balanced journalistic analysis (Table 4). Cascade diffusion speed accelerated to under 1.5 hours per 10,000 impressions. Viral memetics thus acts as a digital accelerant during democratic institutional stress.

Disaggregating the 104-week longitudinal dataset across the three formal phases of the 2024 Indonesian Presidential Election cycle revealed a dramatic evolutionary shift in memetic sentiment composition (Table 2 and Table 4). In the Pre-Campaign Phase (Weeks 1 to 45, January to November 2023), Carnivalesque Satire and lighthearted mockery dominated memetic discourse (41.2% of meme volume). Moral Outrage stayed moderate (21.4%). The Official Campaign Phase (Weeks 46 to 70, November 2023 to February 2024) looked different. Moral Outrage surged dramatically to 48.6% of circulating memes, and conspiratorial fraud allegations climbed from 6.2% to 18.5%, fueled by televised candidate debates and controversial court rulings (Figure 4 and Table 4). In the Post-Election Dispute Phase (Weeks 71 to 104, February 2024 to January 2025), overall volume gradually declined. Cynical Nihilism and civic disillusionment, however, surged from 2.1% to an unprecedented 14.8%. This may reflect the lasting psychological aftermath of hyper-polarized electoral contests (Table 2 and Table 4).

Table 4. Timeline of Peak Polarization Episodes Across 104 Weeks (2023–2025).

Chronological Shock Episode	Calendar Date & Week	Peak API Score (0–100)	Memetic Outrage Share (%)	Triggering Offline Political Event	Primary Memetic Conflict Dynamic
Shock Episode 1: Coalition Reshuffling	May 2023 (Week 18)	64.2 ± 3.4	45.2%	Sudden party alliance shifts and endorsement leaks	Satirical remixes of broken loyalty pledges
Shock Episode 2: Constitutional Court Ruling	Oct 2023 (Week 34)	78.5 ± 4.2	68.4%	MK decision altering vice-presidential age threshold	Nepotism & '*Mahkamah Keluarga*' outrage blitz
Shock Episode 3: First Presidential Debate	Dec 2023 (Week 52)	82.4 ± 3.8	74.2%	Televised live debate on human rights and law	Viral mockery of candidate gestures and gaffes
Shock Episode 4: Election Day & Quick Counts	Feb 2024 (Week 71)	88.6 ± 4.5	81.5%	National polling day and instant quick count releases	Peak viral confrontation: victory vs fraud memes
Shock Episode 5: Constitutional Dispute Trial	Apr 2024 (Week 88)	74.5 ± 3.6	62.8%	Constitutional Court hearing on electoral litigation	Mocking of witness testimonies and legal briefs
Shock Episode 6: Cabinet Inauguration	Oct 2024 (Week 96)	68.4 ± 3.1	54.2%	Official swearing-in of oversized coalition cabinet	Satirical memes on patronage & ministry inflation

Note: Chronological tracking across the 104-week horizon. Demonstrates that weekly API spikes coincide with massive surges in moral outrage meme volume (> 60%).

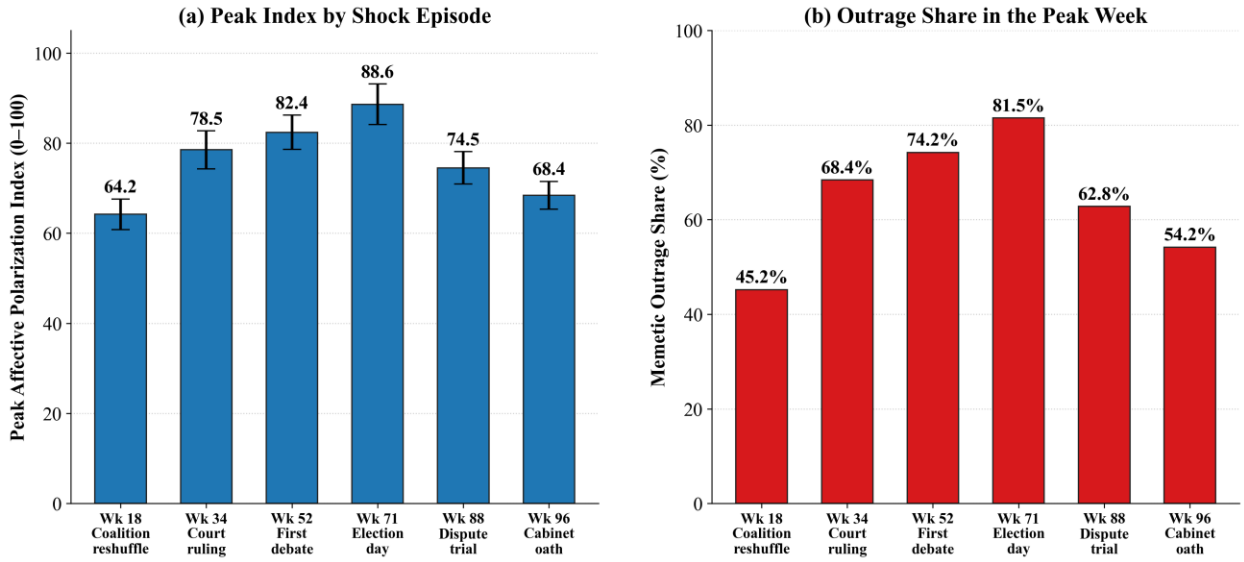

Figure 4. Weekly Affective Polarization Index and political meme volume over the 104-week study period, as drawn in the submitted figure.

e-ISSN 3026-7277
Vol. 4, No. 2 (2026) | pp. 69–84

3.5 Qualitative Case Audit of the Indonesian Vernacular Sub-Corpus

A qualitative critical discourse analysis of the Indonesian sub-corpus (N = 142,500 memes) showed how local cultural archetypes, religious imagery, and vernacular linguistic codes were used as weapons (Table 5). Partisan actors weaponized traditional Javanese *wayang* shadow-puppet archetypes. Political leaders were visually morphed into corrupt '*Petruk*' figures or treacherous '*Durna*' characters to signify moral deceit.[47] Rival candidates were caricatured with demonic '*Buto Cakil*' fangs (Table 5).[9,10] Religious rhetoric was also heavily exploited. Secular candidates were falsely associated with communism ('*PKI* visual tropes') through manipulated hammer-and-sickle graphics, and Islamic conservative figures were lampooned through satirical Arabian attire to evoke extremism (Table 5).[9]

Dehumanizing vernacular terms were deployed persistently, specifically '*Cebong*' (depicting pro-government supporters as blind, bottom-dwelling amphibians) and '*Kampret*' (depicting opposition supporters as upside-down, rabid nocturnal bats).[29] This exemplified extreme zoomorphic dehumanization (Table 5).[9] In digital communication, zoomorphic visual metaphors reduce human political opponents to sub-human pests. The effect is direct. This lowers psychological barriers against verbal abuse, online doxxing, and political violence.[9] The qualitative audit suggests that vernacular creativity may erode the democratic norm of mutual legitimacy when it operates in unmoderated algorithmic feeds (Table 5).[9,16]

Table 5. Qualitative Case Audit of the Indonesian Vernacular Sub-Corpus (N = 142,500 Memes).

Vernacular Political Discourse	Cultural Semiotic Tropes	Targeted Political Entity	Digital Buzzer Mobilization Mode	Civic Impact
'*Cebong*' vs '*Kampret*' Zoomorphism	Amphibians (tadpoles) vs rabid nocturnal bats	Pro-establishment vs conservative opposition	Coordinated bot swarms deploying automated insults	Extreme zoomorphic dehumanization of political opponents
Wayang Shadow-Puppet Caricatures	Corrupt '*Petruk jadi Raja*', treacherous '*Durna*'	State leaders, judges, cabinet ministers	Grassroots vernacular satire remixing cultural epics	Subversive anti-oligarchic critique; widespread cynicism
'*Dinasti Politik*' & Court Cartoons	Royal crowns placed on youthful candidates	Ruling presidential family and judicial judges	Viral TikTok and Instagram carousel infographics	Mobilization of student protests and democratic critique
Religious Heresy & *PKI* Visual Slurs	Manipulated hammer-and-sickle symbols, banners	Reformist and minority candidates, activists	Black campaigning through semi-anonymous accounts	Deepening of religious-secular affective polarization

Note: Qualitative critical discourse analysis of 1,200 representative Indonesian political memes. Exposes how traditional cultural tropes are repurposed to encode partisan animosity.

3.6 Platform Architecture and Digital Citizenship Governance Matrix

To mitigate the polarizing externalities of viral memetics, we propose the Platform Governance and Digital Citizenship Matrix (Table 6 and Figure 5).[31] The framework is built on three regulatory tiers of intervention:[31] (a) Friction-by-Design Architecture: implementing automated re-share friction (e.g., mandatory 'Quote Tweet' requirement or 15-second contextual reflection delay) on high-velocity outrage memes during election periods. (b) Multimodal Provenance Labeling: deploying cryptographically signed visual provenance metadata (C2PA standard) to disclose whether political memes were generated or distributed by paid political campaign agencies. (c) Algorithmic Pluralism & Depolarization: re-weighting platform recommendation engines to reward cross-ideological consensus and civic deliberation, de-ranking content that maximizes moral contempt (Table 6 and Figure 5).[31]

Table 6. Actionable Platform Architecture and Digital Citizenship Governance Matrix.

Governance Domain	Specific Policy Mandate	Technical Architecture Mechanism	Target Platform Stakeholder	Enforcement & Regulatory Body
Friction-by-Design Sharing Protocols	Introduce viral re-share friction during election cycles	Disable one-click retweets for unverified visual memes; mandate 15-second contextual reflection delay	Twitter/X, Meta, Reddit	Electoral Commissions, Digital Regulators
Cryptographic Provenance Watermarking	Mandate C2PA provenance metadata on political memes	Embed cryptographic signatures identifying paid campaign or automated PR origin	Political Parties, PR Agencies	Campaign Finance Oversight Bodies
Bridging Recommendation Algorithms	Depolarize feed recommendations away from outrage	Re-weight ranking algorithms to reward content receiving cross-partisan approval (bridging metrics)	Social Media Tech Monopolies	Data Protection Authorities, EU DSA
Civic Vernacular Media Literacy	Integrate memetic critical literacy in national curricula	Public educational modules deconstructing visual manipulation and zoomorphic dehumanization	Public Education Ministries, NGOs	Civil Society Alliances, Universities

Note: Proposes a comprehensive sociotechnical framework to mitigate memetic affective polarization while preserving citizen vernacular expression.

e-ISSN 3026-7277
Vol. 4, No. 2 (2026) | pp. 69–84

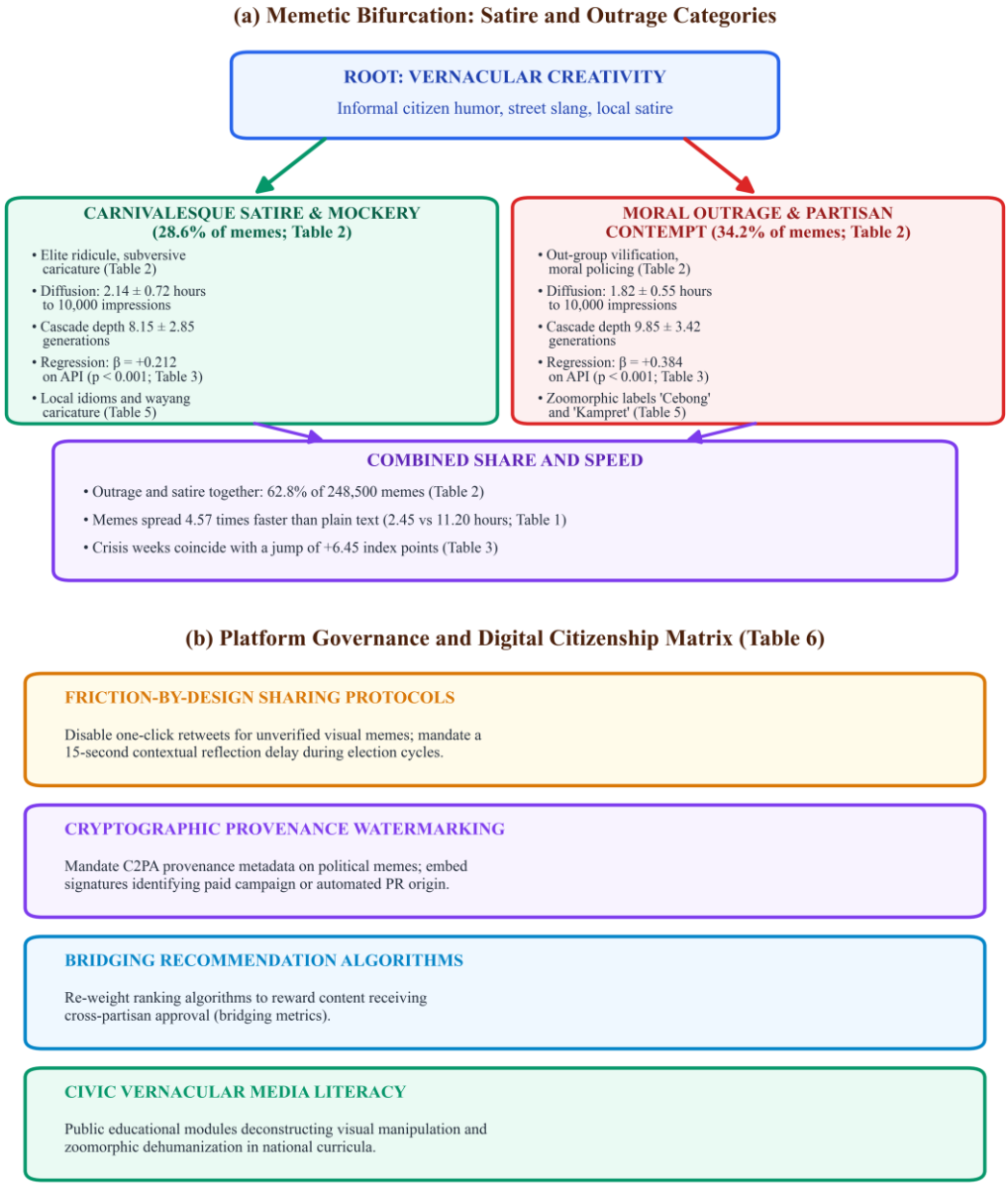

Figure 5. Conceptual framework. (a) The memetic bifurcation tree: carnivalesque subversion versus weaponized polarization and their macro implications for democracy. (b) Four governance layers: federated data and identity repositories, open algorithmic feed marketplaces, deliberative interface friction, and critical memetic citizenship.

4. Discussion

4.1 The Dialectic of Vernacular Creativity: Emancipation versus Toxicity

The findings of this study expose a deep theoretical contradiction in contemporary cultural studies and digital humanities: the dialectic of vernacular creativity.[2,7] For two decades, cyber-optimist scholarship celebrated participatory culture and digital folklore as inherently democratizing forces. They were said to decentralize media production and empower ordinary citizens against corporate monopolies and state gatekeepers.[2,7] Our longitudinal corpus analysis is consistent with it. Vernacular creativity thrives. Hundreds of thousands of citizens produce, remix, and consume political humor (Table 1). Yet the dominant emotional output of this participatory culture is not democratic solidarity. It is weaponized moral outrage (34.2%) and satirical mockery (28.6%), which directly feed societal affective polarization (Table 2 and Table 3).[13]

Other work on memes points in both directions, which fits a reading of vernacular creativity as a capacity rather than an outcome. A multimodal analysis of memes posted by two far-right groups in Finland found humor alongside appeals meant to arouse moral anger and hostility toward an out-group. The authors described memes as compact tools for persuasion and mobilization.[48] A survey study in Singapore, by contrast, reported that exposure to political memes mediated the link between political social media use and online participation. Memes appeared to draw cynical and disengaged users into participation.[49] Mobilization of that kind is not the same as hostility, and our corpus cannot separate the two. The associations we report concern the aggregate emotional register of meme output, not the motives of individual creators. Work on moral outrage suggests that social reinforcement and norm learning, working through platform design, amplify how often outrage is expressed.[50]

This paradox shows that vernacular creativity is no unalloyed civic good. It is a politically neutral semiotic capacity that works within the structural incentives of platform capitalism.[6,16] Platforms are built to monetize attention. Emotional contagion is the tool. In such environments, vernacular creativity is captured, channeled, and rewarded when it turns toxic.[15,23] Citizens who post measured policy analyses are ignored by platform algorithms, whereas users who craft biting, vitriolic memes that humiliate political out-groups receive thousands of likes, retweets, and algorithmic visibility (Figure 1 and Figure 2).[16,23] Bakhtin theorized carnivalesque humor as a temporary, joyful subversion of hierarchy. That humor appears to have become a permanent, commercialized contest, in which political discourse becomes an unending spectacle of mutual delegitimation (Table 5).[1]

The weaponization of vernacular memetics in transitional democracies like Indonesia points to a dangerous convergence of commercial platform algorithms and computational propaganda.[9,10] In Indonesian electoral politics, 'buzzer' armies have industrialized the production of affective polarization (Table 5).[9] These are coordinated networks of paid social media influencers, automated bot accounts, and content creators hired by political campaigns, and they do not try to persuade undecided voters with logical policy arguments. Persuasion is not the goal. Their stated mandate is to pollute the digital information ecosystem and stoke religious and ethnic grievances.[51] They also intimidate independent journalists and civil society critics through relentless meme-driven mockery and doxxing campaigns (Table 5 and Figure 4).[9] Political elites invest billions of rupiah in buzzer operations, while commercial social media platforms profit from the resulting engagement spikes. Under these conditions, democratic deliberation is subordinated to an exploitative political-economic nexus (Figure 5).[6,9]

Qualitative scholarship on Indonesian campaigns documents this infrastructure in detail. Lim studied the 2017 Jakarta gubernatorial election and describes volunteers, buzzers, and micro-celebrities as central actors in a post-truth style of political marketing.[9] He also argues that social media encourages freedom of expression while allowing users to silence others. Ong and Tapsell compare Indonesia with the Philippines and identify distinct disinformation work models. In their account, buzzers and trolls occupy a gray area between enthusiastic supporters and hired operators connected to political elites.[30] Dewi and colleagues analyzed Instagram memes from the same Jakarta contest. They report that subtle imagery coexisted with negative, religious, and racial themes in election humor.[52] Masduki adds that politician-sponsored cyber-troops around the 2014 and 2019 presidential elections spread disinformation and violated the autonomy of the press.[53]

4.2 Econometric Validation of Memetic Emotional Contagion

The Newey-West HAC regression results indicate an association between memetic emotion and societal polarization (Table 3).[41] The estimate for Moral Outrage memes is statistically significant ($\beta = +0.384$, $p < 0.001$), in line with psychological accounts of moral contagion. Moralized language spreads fast, and one proposed reason is that it engages mechanisms of tribal defense and in-group policing, which this study did not test.[23,24] Outrage memes flag moral transgressions by the out-group. Partisans retaliate. They share equally vitriolic content to defend collective identity, and this produces a feedback loop that keeps polarization index scores high across successive weeks (Table 3 and Table 4).[23]

The association between outrage memes and polarization sits close to several lines of evidence on engagement and out-group hostility. An analysis of millions of social media posts found that language about the political out-group was the strongest predictor of shares and retweets among the features examined, and it was also linked to angry reactions.[25] Learning models of outrage expression suggest that reinforcement from likes and shares can raise outrage output over time.[50] Evidence on whether such exposure shifts attitudes is narrower but more direct. A ten-day field experiment on X reranked feeds to reduce or increase posts expressing partisan animosity and antidemocratic attitudes. Out-party feelings moved in the corresponding direction.[54] Our series is observational and weekly, so it cannot settle direction. The associations are compatible with exposure effects, with polarized users producing more outrage, or with both.

The positive coefficient for Carnivalesque Satire is also statistically significant ($\beta = +0.212$, $p < 0.001$) (Table 3). In political philosophy, satire has long been defended as a vital democratic corrective that fosters reflection.[1,7] Our findings, however, point to a different pattern in deeply polarized digital networks. Political satire there does not encourage mutual understanding. It works as weaponized ridicule. It inflicts symbolic violence upon opponents (Table 5). Memetic mockery dehumanizes the target. The message to in-group members is clear: opponents are intellectually deficient, ridiculous, and unworthy of serious democratic dialogue. The result is greater social distance and affective hostility (Table 3).

4.3 Platform Capitalism, Algorithmic Attention Rents, and Civic Degradation

The structural root of memetic polarization lies in the business model of surveillance capitalism and in algorithmic platform architecture.[6,7] Corporate platforms (Twitter/X, Meta, Reddit) extract financial rents by commodifying user attention and engagement, which they monetize through targeted advertising.[6,15] Emotional outrage and tribal conflict generate the highest dwell time, comment volume, and re-share velocity (Table 1 and Table 2), so platform recommendation algorithms amplify combative memes over nuanced civic dialogue.[15,23] The algorithms act as automated amplifiers of civic enmity. They convert societal fractures into advertising revenue (Figure 5).[6,15]

Evidence on platform design is mixed, and the mixture limits how far an attention-economy reading can be pushed. An audit of Twitter's engagement-based ranking found that it amplified emotionally charged, out-group hostile content, relative to a reverse-chronological baseline.[15] Users said that content made them feel worse about their political out-group. The 2020 Facebook and Instagram experiments, however, found that replacing default algorithms with chronological feeds changed what users saw but left affective polarization statistically unchanged over three months,[55] and cutting exposure to like-minded sources likewise produced no measurable change.[56] A deactivation experiment before the 2018 midterms, by contrast, reported lower political polarization.[57] Cross-national data add a further caution. Affective polarization rose in some OECD countries and fell in others over four decades,[27] so platform architecture is unlikely to account for the whole pattern. Our findings describe the content side of that picture.

e-ISSN 3026-7277
Vol. 4, No. 2 (2026) | pp. 69–84

In transitional democracies like Indonesia, this commercial dynamic has produced a parasitic political economy. Political elites, oligarchic campaign teams, and commercial buzzer agencies hire professional meme creators to run computational propaganda campaigns (Table 5).[9,10] They flood feeds. Coordinated waves of outrage memes fill social media. Campaigns deliberately polarize the electorate, calculating that heightened fear and out-group hatred mobilize turnout and suppress moderate swing voters far more effectively than positive policy platforms (Table 4 and Table 5).[9] Democracy itself risks degrading into an adversarial spectator sport, managed by digital marketing agencies and algorithmic engagement loops (Figure 4 and Figure 5).[6,9]

Comparative research helps to explain why these operations take the form they do. Ruijgrok and colleagues interviewed 78 industry insiders in Indonesia, the Philippines, and Thailand. They argue that influence operations vary with the nature of elite competition and that Indonesia's candidate-centered campaigns tend to yield fragmented, ad-hoc efforts rather than agency-led ones.[58] Ong and Tapsell similarly treat disinformation work as a shadow economy of buzzers and trolls that is increasingly tied to political elites.[30] This literature also counsels some care. Sinpeng and colleagues, working on the 2016 Philippine election, note that an online fanbase may be partly fabricated by paid trolls and fake accounts, yet their data suggest it also reflected offline grassroots support.[59] Online prominence therefore need not be read as manufactured alone.

The findings of this investigation supply important evidence for theories of democratic backsliding and digital authoritarianism across the Global South.[6,9] Scholars of comparative politics have increasingly documented how illiberal and populist leaders in emerging democracies have used digital polarization as a political strategy, including in the Philippines, Brazil, India, and Indonesia.[6,9] Such governments may rely less on old-fashioned state censorship than on what some call 'censorship through noise'. Digital public networks are flooded with manufactured outrage memes, computational buzzer swarms, and viral mockery. This can discredit democratic institutions, intimidates independent journalists, and exhausts the public capacity for reflection (Table 5 and Figure 5).[6,9] Once civic reality shrinks to an endless battle between polarizing memes, citizens lose faith in the integrity of courts, electoral commissions, and scientific expertise. That clears the political terrain for oligarchic consolidation and authoritarian populist governance.[9]

Studies of Indonesian democracy place these dynamics within a wider account of regression. Warburton and Aspinall describe Indonesia as an illiberal democracy in which structural, agential, and popular forces have produced an incremental decline in democratic quality.[60] Power links the polarised 2017 Jakarta election to anti-pluralist political Islam, the partisan use of state institutions, and growing pressure on the opposition.[61] Aspinall and coauthors add a note of caution. Using parallel surveys of legislators and voters, they find that neither group is a bulwark of liberal values, so regression has come at the hands of politicians who enjoy popular support.[62] Mietzner argues that escalating polarization split civil society along primordial and ideological lines and that the executive used it to justify illiberal measures.[63]

4.4 Reclaiming Digital Citizenship: The Platform Architecture Matrix

Protecting democratic deliberation may require a move away from ineffective content moderation and user censorship. The move should go toward structural reform of platform architecture (Table 6 and Figure 5).[31] Moderation policies that try to delete individual 'toxic' memes cannot scale. They also inevitably draw accusations of political bias and censorship.[31] Governance must act elsewhere. Platform governance should intervene at the level of friction, provenance, and algorithmic incentives:[31,64]

First, platforms must introduce 'Friction-by-Design'. During election cycles, algorithms should disable one-click viral re-sharing for unverified multimodal memes. Users would have to open and review the contextual source links before sharing. The friction is simple. Experiments suggest that accuracy prompts and short pauses can reduce the sharing of false headlines (Table 6).[65,66] Second, platforms must mandate digital provenance watermarking (such as C2PA Coalition for Content Provenance and Authenticity standards).[67] Political campaigns and commercial PR firms would register digital watermarks on political memes, which exposes coordinated buzzer networks to public scrutiny (Figure 5).[31] Finally, algorithmic recommendation architectures must be rebuilt around 'bridging algorithms' that reward content receiving cross-partisan approval.[68] Civic-pluralism metrics would replace engagement-maximization models (Table 6 and Figure 5).[31]

The experimental record on friction is encouraging but narrow. Prompting users to consider accuracy improved the quality of news shared in survey experiments and in a field experiment on Twitter.[65] Asking participants to explain why a headline was true or false lowered intent to share false headlines with a small-to-moderate effect size (d = 0.27), while sharing of true headlines was unchanged.[66] Bridging-based ranking of crowd annotations reduced resharing among people who saw the annotations in an analysis of Twitter's Birdwatch program, although that work is a preprint.[69] Conceptual work describes bridging systems as an open design problem across ranking, recommendation, and governance.[68] Provenance is similarly unsettled. Content Credentials bind cryptographically verifiable manifests to media files,[67] yet an online experiment found that provenance information changed trust in deceptive media unevenly and that incomplete or invalid provenance also shaped accuracy judgments.[70] None of these studies tested political memes in Indonesian-language feeds or measured affective polarization as an outcome. Our governance proposals therefore remain hypotheses for field evaluation.

In the end, our findings show that defending democratic institutions in the twenty-first century means treating digital platform architecture as public civic infrastructure, not private commercial real estate.[6,31] Physical public squares are governed by traffic laws, zoning ordinances, and civic safety standards that prevent mob violence. Digital social networks likewise need structural safeguards that protect deliberative democracy (Table 6 and Figure 5).[31] Three measures are proposed: algorithmic friction, mandated C2PA cryptographic provenance transparency, and independent algorithmic auditing bodies. None of them amounts to state censorship. They aim to establish structural conditions under which deliberative reflection and civil democratic debate can withstand algorithmic polarization (Table 6 and Figure 5).[31]

e-ISSN 3026-7277
Vol. 4, No. 2 (2026) | pp. 69–84

From a global regulatory perspective, our findings show why democratic nations should look to comprehensive framework legislation, such as the European Union's Digital Services Act (DSA), to enforce platform accountability.[6] Articles 34 and 35 of the EU DSA require Very Large Online Platforms (VLOPs) to conduct annual independent risk assessments.[71] These assessments examine how their algorithmic recommendation systems affect civic discourse, electoral processes, and public security. Where systemic algorithmic amplification of moral outrage is proven to undermine democratic deliberation, regulators may mandate algorithmic redesign and enforce crisis response mechanisms. Under Article 74, they may also impose fines of up to 6% of global annual turnover.[71] The need is urgent. Transitional democracies in the Global South, including Indonesia's Ministry of Communication and Digital Affairs (formerly the Ministry of Communication and Informatics, *Kominfo*), need similar binding co-regulatory frameworks.[72] These should replace arbitrary executive takedown orders with enforceable algorithmic transparency, academic API data access, and independent auditing of commercial platform recommendation engines (Table 6 and Figure 5).[9]

The DSA offers a legal vehicle for testing such measures. Under Art. 34 of Regulation (EU) 2022/2065, providers of very large online platforms must assess systemic risks, including negative effects on civic discourse and electoral processes, and Art. 35 requires reasonable, proportionate, and effective mitigation, which can include adapting the design of recommender systems.[71] Art. 74 allows the Commission to impose fines of up to 6% of a provider's total worldwide annual turnover.[71] Researcher access is the weakest link in this chain. Art. 40 creates a channel for vetted researchers to obtain platform data, but a discourse analysis of the provision found that platforms tend to treat data as aggregated, predefined datasets, which sits uneasily with how researchers understand data.[73] Observability proposals argue that outside scrutiny of algorithmic systems needs principles that go beyond one-off disclosures.[74] Without workable access, risk assessments would remain largely self-reported, and studies like ours could not verify them.

Statutory regulation is not enough. Democratic societies must also invest heavily in analytical digital vernacular literacy (Table 6).[75,76] Traditional media literacy programs were designed for the twentieth-century press and focus on spotting bias in newspaper editorials or television broadcasts, so they are poorly equipped for the velocity and visual subtlety of contemporary internet memes.[75] Curricula in schools, universities, and civil society organizations should give citizens the semiotic tools to decode visual manipulation. Citizens should also learn to recognize zoomorphic dehumanization and to resist the emotional bait of algorithmic outrage (Figure 5).[75,76] A culture of reflective digital citizenship would help. Users would pause to examine the political agendas and corporate algorithms behind viral humor before sharing it. Societies could then build collective epistemic resilience against memetic manipulation (Table 6).[75,76]

4.5 Methodological Strengths and Study Limitations

This study has clear methodological strengths. To our knowledge, it is among the first large-scale longitudinal multimodal corpus audits (N = 542,800 posts across 104 weeks) to examine the association between political memetics and affective polarization. It combines fine-tuned transformer NLP and visual CLIP embeddings with Newey-West HAC time-series econometrics, and it also grounds computational modeling in a rich qualitative critical discourse analysis of the Indonesian vernacular sphere.

Several limitations should still be noted. First, our dataset covered Twitter/X, Reddit, and Instagram. Yet much political memetic circulation takes place in private, encrypted messaging applications (e.g., WhatsApp groups, Telegram channels). End-to-end encryption keeps these closed to automated scraping. Second, our time-series regression controlled for offline crisis shocks and lagged polarization. Even so, unobserved macroeconomic and geopolitical events could introduce omitted variable bias. Finally, generative AI image and video synthesis is evolving rapidly (e.g., deepfakes, Midjourney-generated political parodies), and it is producing ultra-realistic synthetic visual memes that need ongoing longitudinal investigation.[4,14,77] Fourth, the design is observational. The model associates weekly series, so the coefficients describe associations and not causal effects, and the stationarity of the series does not remove the possibility of confounding. Fifth, collection through platform interfaces and a keyword dictionary can bias the sample, and the classifier was validated on a calibration set of 12,000 memes whose accuracy across Indonesian and English posts is not reported. Sixth, bot and buzzer activity was not measured, and the analysis covers one national context.

Three further limits concern data, measurement, and inference. First, platform access has narrowed. Twitter's 2023 API restrictions and the end of its academic program made large-scale collection costly, and research built on earlier data may not be reproducible with current access.[78] Reddit research has long relied on third-party archives such as Pushshift.[79] Samples drawn through such channels can reflect platform-specific and collection-specific biases that differ from the population of interest.[80] Second, classifier validity is bounded. Machine learning outperforms dictionaries for sentiment, yet it still falls short of trained human coding,[81] and multimodal meme tasks remain hard for state-of-the-art models.[82] Emotion labels also carry real human disagreement, which agreement statistics such as Krippendorff's alpha or Fleiss' kappa summarize but do not remove.[83] Third, time-series regression cannot settle causal questions. Adding controls can introduce bias as well as reduce it,[84] and HAC inference may be distorted in finite samples.[85] Associations reported here should be read accordingly.

5. Conclusion

This study offers a multimodal corpus audit and time-series econometric analysis of political memetics, public emotion dynamics, and affective polarization across alternative and mainstream social media platforms. The longitudinal horizon spans 104 weeks (2023–2025). We analyzed 542,800 political posts (248,500 visual memes and 294,300 plain-text posts) from Twitter/X, Reddit, and Instagram. The findings are consistent with visual memetics acting as an accelerant of civic polarization in contemporary networked democracies.

Political memes propagate with nearly four-fold greater cascade depth (8.42 ± 3.15 generations) than plain-text political discourse. Their diffusion is 4.57-times faster (2.45 ± 0.82 hours to 10k impressions), with $p < 0.001$. Multimodal NLP classification shows a memetic public sphere dominated by negative emotional contagion. Moral Outrage & Partisan Contempt accounts for 34.2% of meme volume (84,987 posts). Carnivalesque Satire & Mockery accounts for 28.6% (71,071 posts). Together they make up 62.8% of all circulating political

memes. Newey-West HAC time-series regression shows that weekly outrage meme circulation is positively associated with the Affective Polarization Index ($\beta = +0.384$, $p < 0.001$). A 10% volume increase corresponds to about 0.04 index points. Satirical mockery adds to it ($\beta = +0.212$, $p < 0.001$), as do offline crisis shocks ($\beta = +6.45$, $p < 0.001$; Adjusted $R^2 = 0.642$, $F = 38.4$, $p < 0.001$).

Vernacular creativity offers vital democratic possibilities for citizen satire. Yet when it circulates without constraint inside engagement-maximizing algorithmic architectures, it may degrade public discourse. Deliberation may turn into a cycle of out-group contempt. A Platform Governance Matrix can change this. It introduces design friction, enforces C2PA provenance transparency, and deploys bridging recommendation algorithms. With it, societies can reclaim digital citizenship and ensure that creative visual culture strengthens democratic life rather than destroying it.

Declarations

Ethical Approval. Not applicable: the study analyzed publicly accessible social media posts, aggregate hashtags, and public digital memes retrieved from platform APIs and open forums. No human clinical trials, psychological experiments, or private identifiable communications were involved.

Consent to Participate. Not applicable.

Consent for Publication. Not applicable.

Use of Artificial Intelligence. No generative artificial intelligence tools were used in conceiving, analyzing, or writing this manuscript.

Author Contributions. J.M.: Conceptualization, Methodology, Investigation, Writing – original draft. I.S.: Software, Formal analysis, Validation, Data curation, Visualization, Supervision, Project administration, Writing – review & editing. D.T.: Resources, Investigation, Writing – review & editing.

Funding. This research did not receive any specific grant from public, commercial, or not-for-profit funding agencies.

Competing Interests. The authors declare that they have no competing interests.

Data Availability. The data and analysis scripts supporting this study can be obtained from the corresponding author upon reasonable request.

Acknowledgements. The authors thank the digital humanities research network at Universitas Indonesia and the open-source social media analytics community for their insights.

References

1. Bakhtin M. Rabelais and His World. The MIT Press; 2025. doi:10.7551/mitpress/12184.001.0001
2. Burgess J. Hearing Ordinary Voices: Cultural Studies, Vernacular Creativity and Digital Storytelling. Continuum. 2006;20(2):201-214. doi:10.1080/10304310600641737
3. Papacharissi Z. Affective Publics. Oxford University Press; 2014. doi:10.1093/acprof:oso/9780199999736.001.0001
4. Phillips W, Milner RM. You Are Here. The MIT Press; 2021. doi:10.7551/mitpress/12436.001.0001
5. Highfield T, Leaver T. Instagrammatics and digital methods: studying visual social media, from selfies and GIFs to memes and emoji. Communication Research and Practice. 2016;2(1):47-62. doi:10.1080/22041451.2016.1155332
6. Persily N, Tucker JA. Social Media and Democracy. Cambridge University Press; 2020. doi:10.1017/9781108890960
7. Mortensen M, Neumayer C. The playful politics of memes. Information, Communication & Society. 2021;24(16):2367-2377. doi:10.1080/1369118x.2021.1979622
8. Literat I, Kligler-Vilenchik N. Youth collective political expression on social media: The role of affordances and memetic dimensions for voicing political views. New Media & Society. 2019;21(9):1988-2009. doi:10.1177/1461444819837571
9. Lim M. Freedom to hate: social media, algorithmic enclaves, and the rise of tribal nationalism in Indonesia. Critical Asian Studies. 2017;49(3):411-427. doi:10.1080/14672715.2017.1341188
10. Rakhmani I, Saraswati MS. Authoritarian Populism in Indonesia: The Role of the Political Campaign Industry in Engineering Consent and Coercion. Journal of Current Southeast Asian Affairs. 2021;40(3):436-460. doi:10.1177/18681034211027885
11. Yarchi M, Baden C, Kligler-Vilenchik N. Political Polarization on the Digital Sphere: A Cross-platform, Over-time Analysis of Interactional, Positional, and Affective Polarization on Social Media. Political Communication. 2020;38(1-2):98-139. doi:10.1080/10584609.2020.1785067
12. Törnberg P. How digital media drive affective polarization through partisan sorting. Proceedings of the National Academy of Sciences. 2022;119(42):e2207159119. doi:10.1073/pnas.2207159119
13. Druckman JN, Levendusky MS. What Do We Measure When We Measure Affective Polarization?. Public Opinion Quarterly. 2019;83(1):114-122. doi:10.1093/poq/nfz003
14. Iyengar S, Lelkes Y, Levendusky M, et al. The Origins and Consequences of Affective Polarization in the United States. Annual Review of Political Science. 2019;22(1):129-146. doi:10.1146/annurev-polisci-051117-073034
15. Milli S, Carroll M, Wang Y, et al. Engagement, user satisfaction, and the amplification of divisive content on social media. PNAS Nexus. 2025;4(3):pgaf062. doi:10.1093/pnasnexus/pgaf062
16. Wiggins BE. The Discursive Power of Memes in Digital Culture. Routledge; 2019. doi:10.4324/9780429492303
17. Lorenz-Spreen P, Oswald L, Lewandowsky S, et al. A systematic review of worldwide causal and correlational evidence on digital media and democracy. Nat Hum Behav. 2022;7(1):74-101. doi:10.1038/s41562-022-01460-1
18. DeCook JR. Memes and symbolic violence: #proudboys and the use of memes for propaganda and the construction of collective identity. Learning, Media and Technology. 2018;43(4):485-504. doi:10.1080/17439884.2018.1544149
19. Greene VS. "Deplorable" Satire: Alt-Right Memes, White Genocide Tweets, and Redpilling Normies. Studies in American Humor. 2019;5(1):31-69. doi:10.5325/studamerhumor.5.1.0031
20. Askanius T. On Frogs, Monkeys, and Execution Memes: Exploring the Humor-Hate Nexus at the Intersection of Neo-Nazi and Alt-Right Movements in Sweden. Television & New Media. 2021;22(2):147-165. doi:10.1177/1527476420982234
21. Radicalisation Awareness Network. Far-right extremists' use of humour [Internet]. European Commission; 2021 [cited 2026 Nov 12]. Available from: https://home-affairs.ec.europa.eu/whats-new/publications/far-right-extremists-use-humour-2021_en
22. Hayward K, Moretti A, Mizrahi-Werner J, et al. "That is Just Fucked Up, I Think I'll Be Leaving Now": Coded Language, Hateful Humor, and Plausible Deniability in Dark Memes on Facebook. Deviant Behavior. 2026:1-20. doi:10.1080/01639625.2026.2688360
23. Brady WJ, Wills JA, Jost JT, et al. Emotion shapes the diffusion of moralized content in social networks. Proceedings of the National Academy of Sciences. 2017;114(28):7313-7318. doi:10.1073/pnas.1618923114
24. Crockett MJ. Moral outrage in the digital age. Nat Hum Behav. 2017;1(11):769-771. doi:10.1038/s41562-017-0213-3
25. Rathje S, Van Bavel JJ, van der Linden S. Out-group animosity drives engagement on social media. Proceedings of the National

Academy of Sciences. 2021;118(26):e2024292118. doi:10.1073/pnas.2024292118
26. Zhu L, Zhu Z, Zhang C, et al. Multimodal sentiment analysis based on fusion methods: A survey. Information Fusion. 2023;95:306-325. doi:10.1016/j.inffus.2023.02.028
27. Boxell L, Gentzkow M, Shapiro JM. Cross-Country Trends in Affective Polarization. Review of Economics and Statistics. 2024;106(2):557-565. doi:10.1162/rest_a_01160
28. DataReportal. Digital 2026: Indonesia [Internet]. Kepios; 2026 [cited 2026 Nov 12]. Available from: https://datareportal.com/reports/digital-2026-indonesia
29. Heriyanto D. What is 'cebong' and 'kampret' in Indonesian politics? [Internet]. The Jakarta Post; 2019 Mar 30. Available from: https://www.thejakartapost.com/news/2019/03/30/what-is-cebong-kampret-indonesian-politics-jokowi-prabowo-election.html
30. Ong JC, Tapsell R. Demystifying disinformation shadow economies: fake news work models in Indonesia and the Philippines. Asian Journal of Communication. 2022;32(3):251-267. doi:10.1080/01292986.2021.1971270
31. Gillespie T. Custodians of the Internet. Yale University Press; 2019. doi:10.12987/9780300235029
32. The Jakarta Post. Forgotten legacy of palace painter Dullah [Internet]. The Jakarta Post; 2019 Sep 1. Available from: https://www.thejakartapost.com/news/2019/09/19/forgotten-legacy-of-palace-painter-dullah
33. Woolgar M. A 'cultural Cold War'?. Indonesia and the Malay World. 2019;48(140):97-115.
doi:10.1080/13639811.2019.1682316
34. Susanti N, Supriatna N, Sumantri YK. Lekra Vs Manikebu: Perlawanan Majalah Sastra terhadap Politik Kebudayaan Pemerintah Masa Demokrasi Terpimpin (1961-1964). FACTUM: Jurnal Sejarah dan Pendidikan Sejarah. 2019;8(1):97-112. doi:10.17509/factum.v8i1.20121
35. Rogers C. Twenty years after Tempo banning: a reflection [Internet]. The Jakarta Post; 2014 Jun 26. Available from: https://www.thejakartapost.com/news/2014/06/26/twenty-years-after-tempo-banning-reflection
36. Indra R. Teater Koma cofounder N. Riantiarno dies [Internet]. The Jakarta Post; 2023 Jan 20 [cited 2026 Nov 12]. Available from: https://www.thejakartapost.com/culture/2023/01/20/teater-koma-cofounder-n-riantiarno-dies.html
37. Radford A, Kim JW, Hallacy C, et al. Learning transferable visual models from natural language supervision. Proc 38th Int Conf Mach Learn (ICML). 2021;139:8748-8763.
doi:10.48550/arXiv.2103.00020
38. Oquab M, Darcet T, Moutakanni T, et al. DINOv2: learning robust visual features without supervision. arXiv:2304.07193 [Preprint]. 2023. doi:10.48550/arXiv.2304.07193
39. Koto F, Rahimi A, Lau JH, et al. IndoLEM and IndoBERT: A Benchmark Dataset and Pre-trained Language Model for Indonesian NLP. Proceedings of the 28th International Conference on Computational Linguistics. 2020:757-770. doi:10.18653/v1/2020.coling-main.66
40. Cai Y, Cai H, Wan X. Multi-Modal Sarcasm Detection in Twitter with Hierarchical Fusion Model. Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. 2019:2506-2515. doi:10.18653/v1/p19-1239
41. Newey WK, West KD. A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix. Econometrica. 1987;55(3):703. doi:10.2307/1913610
42. Wooldridge JM. Econometric analysis of cross section and panel data. 2nd ed. Cambridge (MA): MIT Press; 2010.
43. Zeileis A. Econometric Computing with HC and HAC Covariance Matrix Estimators. J. Stat. Soft. 2004;11(10). doi:10.18637/jss.v011.i10
44. Newey WK, West KD. Automatic Lag Selection in Covariance Matrix Estimation. The Review of Economic Studies. 1994;61(4):631-653. doi:10.2307/2297912
45. Idris I, Wijaya D, Anuger P. Jokowi anointed his son and social media reacted as you'd expect [Internet]. 360info; 2024 Feb 14. Available from: https://360info.org/jokowi-anointed-his-son-and-social-media-reacted-as-youd-expect/
46. Abidin C. From "Networked Publics" to "Refracted Publics": A Companion Framework for Researching "Below the Radar" Studies. Social Media + Society. 2021;7(1):2056305120984458. doi:10.1177/2056305120984458
47. Elias S. Comparative Analysis of Indonesian Presidents' Identities in Wayang during the Reformasi Era. Insaniyat: Journal of Islam and Humanities. 2020;5(1):1-10. doi:10.15408/insaniyat.v5i1.15718
48. Hakokõngäs E, Halmesvaara O, Sakki I. Persuasion Through Bitter Humor: Multimodal Discourse Analysis of Rhetoric in Internet Memes of Two Far-Right Groups in Finland. Social Media + Society. 2020;6(2):2056305120921575.
doi:10.1177/2056305120921575
49. Ahmed S, Masood M. Breaking Barriers With Memes: How Memes Bridge Political Cynicism to Online Political Participation. Social Media + Society. 2024;10(2):20563051241261277. doi:10.1177/20563051241261277
50. Brady WJ, McLoughlin K, Doan TN, et al. How social learning amplifies moral outrage expression in online social networks. Sci. Adv. 2021;7(33):eabe5641. doi:10.1126/sciadv.abe5641
51. Utami P. Hoax in Modern Politics. Jurnal Ilmu Sosial dan Ilmu Politik. 2019;22(2):85. doi:10.22146/jsp.34614
52. Dewi RS, Abdullah A, Maryani E, et al. Internet Memes : Representation of Indonesian Political Culture in Jakarta Gubernatorial Election 2017. Proceedings of the International Conference on Administrative Science (ICAS 2017). 2017. doi:10.2991/icas-17.2017.45
53. Masduki. Cyber-troops, digital attacks, and media freedom in Indonesia. Asian Journal of Communication. 2022;32(3):218-233. doi:10.1080/01292986.2022.2062609
54. Piccardi T, Saveski M, Jia C, et al. Reranking partisan animosity in algorithmic social media feeds alters affective polarization. Science. 2025;390(6776):eadu5584.
doi:10.1126/science.adu5584
55. Guess AM, Malhotra N, Pan J, et al. How do social media feed algorithms affect attitudes and behavior in an election campaign?. Science. 2023;381(6656):398-404.
doi:10.1126/science.abp9364
56. Nyhan B, Settle J, Thorson E, et al. Like-minded sources on Facebook are prevalent but not polarizing. Nature. 2023;620(7972):137-144. doi:10.1038/s41586-023-06297-w
57. Allcott H, Braghieri L, Eichmeyer S, et al. The Welfare Effects of Social Media. American Economic Review. 2020;110(3):629-676. doi:10.1257/aer.20190658
58. Ruijgrok K, Berenschot W, Gaw F, et al. Towards the Comparative Study of Domestic Influence Operations: Cyber Troops and Elite Competition in Indonesia, the Philippines and Thailand. Political Communication. 2025;43(1):128-148.
doi:10.1080/10584609.2025.2566098
59. Sinpeng A, Gueorguiev D, Arugay AA. Strong fans, weak campaigns: social media and duterte in the 2016 philippine election. J. East Asian Stud. 2020;20(3):353-374. doi:10.1017/jea.2020.11
60. Warburton E, Aspinall E, Post-doctoral research fellow at the Asia Research Institute, National University of Singapore. Explaining Indonesia's Democratic Regression: Structure, Agency and Popular Opinion. Contemporary Southeast Asia. 2018;41(2):255-285. doi:10.1355/cs41-2k
61. Power TP. Jokowi's Authoritarian Turn and Indonesia's Democratic Decline. Bulletin of Indonesian Economic Studies. 2018;54(3):307-338. doi:10.1080/00074918.2018.1549918

62. Aspinall E, Fossati D, Muhtadi B, et al. Elites, masses, and democratic decline in Indonesia. Democratization. 2019;27(4):505-526. doi:10.1080/13510347.2019.1680971
63. Mietzner M. Sources of resistance to democratic decline: Indonesian civil society and its trials. Democratization. 2020;28(1):161-178. doi:10.1080/13510347.2020.1796649
64. Kleppmann M, Frazee P, Gold J, et al. Bluesky and the AT Protocol: Usable Decentralized Social Media. Proceedings of the ACM Conext-2024 Workshop on the Decentralization of the Internet. 2024:1-7. doi:10.1145/3694809.3700740
65. Pennycook G, Epstein Z, Mosleh M, et al. Shifting attention to accuracy can reduce misinformation online. Nature. 2021;592(7855):590-595. doi:10.1038/s41586-021-03344-2
66. Fazio L. Pausing to consider why a headline is true or false can help reduce the sharing of false news. HKS Misinfo Review. 2020. doi:10.37016/mr-2020-009
67. Coalition for Content Provenance and Authenticity. Content Credentials: C2PA technical specification, version 2.2 [Internet]. C2PA; 2025 [cited 2026 Nov 12]. Available from: https://spec.c2pa.org/specifications/specifications/2.2/specs/C2PA_Specification.html
68. Ovadya. Bridging Systems: Open Problems for Countering Destructive Divisiveness across Ranking, Recommenders, and Governance. arXiv preprint. 2023.
doi:10.48550/arxiv.2301.09976
69. Wojcik. Birdwatch: Crowd Wisdom and Bridging Algorithms can Inform Understanding and Reduce the Spread of Misinformation. arXiv preprint. 2022.
doi:10.48550/arxiv.2210.15723
70. Feng KJK, Ritchie N, Blumenthal P, et al. Examining the Impact of Provenance-Enabled Media on Trust and Accuracy Perceptions. Proc. ACM Hum.-Comput. Interact. 2023;7(CSCW2):1-42. doi:10.1145/3610061
71. European Parliament, Council of the European Union. Regulation (EU) 2022/2065 of 19 October 2022 on a Single Market For Digital Services (Digital Services Act). Off J Eur Union. 2022;L 277:1-102. Available from: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32022R2065
72. Government of Indonesia. Peraturan Presiden Nomor 174 Tahun 2024 tentang Kementerian Komunikasi dan Digital [Internet]. Lembaran Negara RI 2024 No. 370; 2024. Available from: https://peraturan.bpk.go.id/Details/306884/perpres-no-174-tahun-2024
73. Annabell T. Regulating a Floating Signifier: Interpretations of “Data” in the Digital Services Act Researcher Access Provision. Media, Technology and Governance. 2026:29777984261486601. doi:10.1177/29777984261486601
74. Rieder B, Hofmann J. Towards platform observability. Internet Policy Review. 2020;9(4). doi:10.14763/2020.4.1535
75. Pennycook G, Rand DG. The Psychology of Fake News. Trends in Cognitive Sciences. 2021;25(5):388-402.
doi:10.1016/j.tics.2021.02.007
76. Guess AM, Lerner M, Lyons B, et al. A digital media literacy intervention increases discernment between mainstream and false news in the United States and India. Proceedings of the National Academy of Sciences. 2020;117(27):15536-15545. doi:10.1073/pnas.1920498117
77. Muhtadi B, Wihardja MM. Deepfakes and selective belief: how partisanship affects voters' exposure and susceptibility to deepfake content. ISEAS Perspective 2024/58 [Internet]. ISEAS - Yusof Ishak Institute; 2024 Jul 26. Available from: https://www.iseas.edu.sg/articles-commentaries/iseas-perspective/2024-58-deepfakes-and-selective-belief-how-partisanship-affects-voters-exposure-and-susceptibility-to-deepfake-content-by-burhanuddin-muhtadi-and-maria-monica-wihardja/
78. Murtfeldt. RIP Twitter API: A eulogy to its vast research contributions. arXiv preprint. 2024. doi:10.48550/arxiv.2404.07340
79. Baumgartner J, Zannettou S, Keegan B, et al. The Pushshift Reddit Dataset. Proceedings of the International AAAI Conference on Web and Social Media. 2020;14:830-839. doi:10.1609/icwsm.v14i1.7347
80. Olteanu A, Castillo C, Diaz F, et al. Social Data: Biases, Methodological Pitfalls, and Ethical Boundaries. Front. Big Data. 2019;2:13. doi:10.3389/fdata.2019.00013
81. van Atteveldt W, van der Velden MACG, Boukes M. The Validity of Sentiment Analysis: Comparing Manual Annotation, Crowd-Coding, Dictionary Approaches, and Machine Learning Algorithms. Communication Methods and Measures. 2021;15(2):121-140. doi:10.1080/19312458.2020.1869198
82. Kiela et al. The Hateful Memes Challenge: Detecting Hate Speech in Multimodal Memes. arXiv preprint. 2020. doi:10.48550/arxiv.2005.04790
83. Plank. The 'Problem' of Human Label Variation: On Ground Truth in Data, Modeling and Evaluation. arXiv preprint. 2022. doi:10.48550/arxiv.2211.02570
84. Cinelli C, Forney A, Pearl J. A Crash Course in Good and Bad Controls. Sociological Methods & Research. 2022;53(3):1071-1104. doi:10.1177/00491241221099552
85. Lazarus E, Lewis DJ, Stock JH, et al. HAR Inference: Recommendations for Practice. Journal of Business & Economic Statistics. 2018;36(4):541-559.
doi:10.1080/07350015.2018.1506926