



Digital Autopsy of the 2025 Kenyan By-Elections: A Forensic Analysis of Coordinated Narrative Warfare

1 Attribution & Methodology Note

This report was written and edited by AI Salon and Election Ledger Consortium members. Editorial Team consisted of Mwende Mukwanyaga and Lilian Mutinda, together with three additional contributors who have chosen to remain anonymous. The analysis draws on data and technical insights from Ampitech, members of the AI Salon Consortium and our data gurus.

Patterns of coordination, narrative formation, actor behaviour, and coordinated inauthentic behaviour (CIB) were identified using Sema, a proprietary election-analysis tool developed by the AI Salon Kenya / Election Ledger team. The outputs were further cross-validated on NotebookLM and human editorial team, ensuring consistency across methodologies and analytical approaches. SeedDream, NotebookLM, Nano Banana and Flourish were used to produce selected data visualisations and infographics. All analytical conclusions remain grounded in the original datasets and methods described above.

CONTENT

1. Executive summary
2. Core findings
3. The stakes
4. The choice
5. Introduction to Industrialisation of Influence
6. National & Constituency Overview: Theatres of Narrative Warfare
7. The Actor Ecology: Anatomy of the Seed-Amplifier War Machine
8. Tactical Analysis: The DISARM Playbook in Action
9. Legal and Regulatory Grey Zones: The Enforcement Gap When Law Meets Industrial Manipulation
10. Proposed Recommendations

01

Executive Summary

Executive Summary

This document is a forensic intelligence dossier on the sophisticated information operations observed during the 2025 Kenyan by-elections. The analysis of online discourse between November 01 and December 05, drawn from three structured monitoring datasets totalling 54,998 posts and ≈ 17 million views, reveals not a spontaneous public conversation but a highly coordinated, industrialized system of narrative manufacturing.

The digital public square was effectively privatized by a hostile broadcast operation that treated specific constituencies as laboratories for testing and refining influence tactics, producing a clear operational playbook for the 2027 General Election.

A low-angle, upward-looking shot of several modern skyscrapers with glass facades. The buildings are reflected in each other, creating a complex geometric pattern. The sky is a pale, overcast blue. The overall color palette is dominated by various shades of blue and grey.

02

Core Findings

Industrial-Scale Manufacturing

The post-election information space was overwhelmingly manufactured. In the extended post-election window (Post election 28 Nov- 05 Dec, 14,998 posts), only about 2,850 unique Text strings exist. This means, $\approx 84\%$ of posts in this period belong to duplicate text clusters, where the same sentence or paragraph is repeated across multiple accounts. This is called astroturfing -when coordinated actors make something look like real, grassroots public opinion when it is not.

Across all three phases combined (pre, immediate post, and extended post), scripted clusters dominate key narratives:

A single Mbeere nullification script appears $\approx 2,119$ times across the immediate and extended post windows.

A Malava 2027 fear script appears ≈ 433 times.

A Kasipul violence cluster appears ≈ 741 times.

What appears as "debate" is in reality industrial broadcast: centrally authored scripts repeated thousands of times by an ecosystem of amplifiers and swarms.

Choreographed Narrative Pivots

Specific constituencies were used as "narrative laboratories." In Mbeere North, pre - election chatter was a messy mix of local policy, personality politics, and gossip.

Immediately after results, this was abruptly replaced by a cold, uniform legal-delegitimization script that was completely absent from the pre-election dataset.

The forensic timeline shows a precise narrative ladder in Mbeere:

- Fear of force-discussion of security deployment and "escort units."
- Logistical crisis-"ballot shortage" and mishandling claims.
- Legal inevitability- the nullification script repeated over 2,000 times.

Comparable pivots occur in Malava (from community issues to 2027 apocalypse framing) and Kasipul (from routine contestation to an image of state-sponsored violence centred on Agoro Sare).

These are not organic mood swings; they are pre-scripted, outcome-contingent pivots.

Hierarchical Actor Ecology

Operations were executed by a structured "Seed-Amplifier-Swarm" war machine, not a loose crowd of angry citizens.

Our actor-level inspection of Author Handle across the three datasets reveals:

A concentration of high-volume amplifiers posting dozens to hundreds of times across phases.

Thousands of one-shot or low-frequency accounts, many contributing exactly one instance of a script.

A persistent core of accounts active both before and after the election, forming a standing influence infrastructure, not a one-off outburst.

The geometry is layered:

1. Seed layer; a handful of accounts that originate the nuanced legal, moral, or fear-inducing texts (e.g. the Mbeere nullification phrasing, Malava "imagine 2027" framing, Kasipul rigging/violence lines).

2. Amplifier layer; mid/high-volume accounts that spray those texts across constituencies and hashtags.

3. Swarm base; thousands of low-activity accounts whose main visible function is to repeat a script once.

AI as a Force Multiplier

This election cycle marks the emergence of AI as a structural actor in the information ecosystem.

Within these datasets, the AI-linked persona @grok appears 663 times, making it one of the most prolific authors by count. Its content is recognizably LLM/Q&A-style: answers, explanations, and summaries of ongoing political discourse.

@Grok's role is not primarily to invent disinformation, but to:

1

Ingest chaotic feeds of already-biased content.

2

Summarise and normalise those patterns as "what Kenyans are saying" or "what the post argues."

3

Wrap partisan-tilted conversations in the language of neutral technological analysis.

In doing so, AI becomes a high-reach mediator and launderer of frames, amplifying whichever narratives have already been industrially pushed into the underlying timelines.

Why 2027 Requires a New Regulatory Paradigm

Current Frameworks Are Inadequate:

Content-based regulation chases individual falsehoods while missing the production schedule.

Bot-focused policies miss human-operated astroturfing.

Defamation law doesn't address true-but-misleading lawfare narratives.

AI regulation doesn't exist, allowing narrative laundering to go unchecked.

The 2027 Imperative:

Shift from content to coordination as the signal of harm.

Measure manufacture, not just monitor posts.

Expose production chains, not just debunk individual lies.

Harden institutions against predictable narrative capture.

Teach forensic literacy so citizens can recognize scripts.



03

The Stakes

An abstract graphic on the left side of the slide featuring a network of glowing green nodes connected by thin lines, set against a dark background with a green gradient.

The Stakes

What was piloted in 4 constituencies (54,998 posts) will scale to 290 constituencies in 2027.

Without regulatory reform, the 2027 General Election will be fought in a fully industrialized information environment where coordination is invisible, manipulation is legal, and AI launders it all as consensus.



04

The Choice

The Choice

Option 1

Continue chasing individual falsehoods while the factory runs 24/7.

Option 2

Change the level of analysis, expose the production schedule, regulate coordination, and ensure Kenyans can distinguish between a nation talking to itself and a nation being talked at by a machine.



Implications & Projections for the 2027 General Election

The tactics and infrastructure stress - tested in 2025 form a clear playbook for 2027:

1

A shift from visible copy - paste brigades to LLM-generated paraphrase swarms: thousands of posts that are semantically identical but textually different enough to evade simple duplication filters.

2

Increased reliance on lawfare narratives, treating courts and petitions as the main battlefield.

3

Expansion of temporal projection scripts ("if this is a by - election, imagine 2027") to pre - emptively delegitimise national polls.

4

Continued weaponization of AI explainers as quasi - journalistic narrators of the election; struggles over what AI systems like @grok see, ingest, and present as the "state of public opinion."

5

Crime scenes at scale; incidents in scattered polling stations turned into a continuous montage of democratic failure.

6

Outsourced operations and diaspora proxies; posting activity partly located outside Kenya, complicating enforcement and accountability.

Primary recommendation: Stakeholders must stop treating disinformation as a post - hoc fact - checking problem. The priority for 2027 is proactive, real - time monitoring of coordinated inauthentic behaviour and narrative scripts, exposing the machinery before its output hardens into public common sense.

05

In depth research: The Industrialisation of Influence



Introduction

The 2025 Kenyan by - elections, though limited in seats, were strategically significant. They functioned as a controlled stress test of an influence infrastructure designed to scale into the 2027 General Election.

This report treats the by - elections not as isolated local contests but as theatres in a national experiment on:

1

Narrative scripts and their emotional arcs.

2

Actor networks engineered for amplification.

3

AI mediation as a new layer in the battlespace.

Methodological Basis

The analysis is grounded in forensic examination of three datasets drawn from Twitter/X and open web monitoring systems:

Pre - by - elections

20,000 posts (Pre by-election csv.xlsx- Sheet 1-Mbeere_or_Malava_or_k.csv).

Immediate post - by - elections

20,000 posts (Post byelection csv.xlsx- Sheet 1-Mbeere_or_Malava_or_k.csv).

Extended post - election (28 Nov–5 Dec)

14,998 posts (Post election 28 Nov- 05 Dec.xlsx).

All statistics, counts, and sentiment distributions were computed programmatically (e.g. using Python/pandas), not by manual spreadsheet inspection. For analytical integrity, the Excel(.xlsx) exports are treated as canonical. The DISARM framework is used conceptually to classify tactics, techniques and procedures.

Thesis

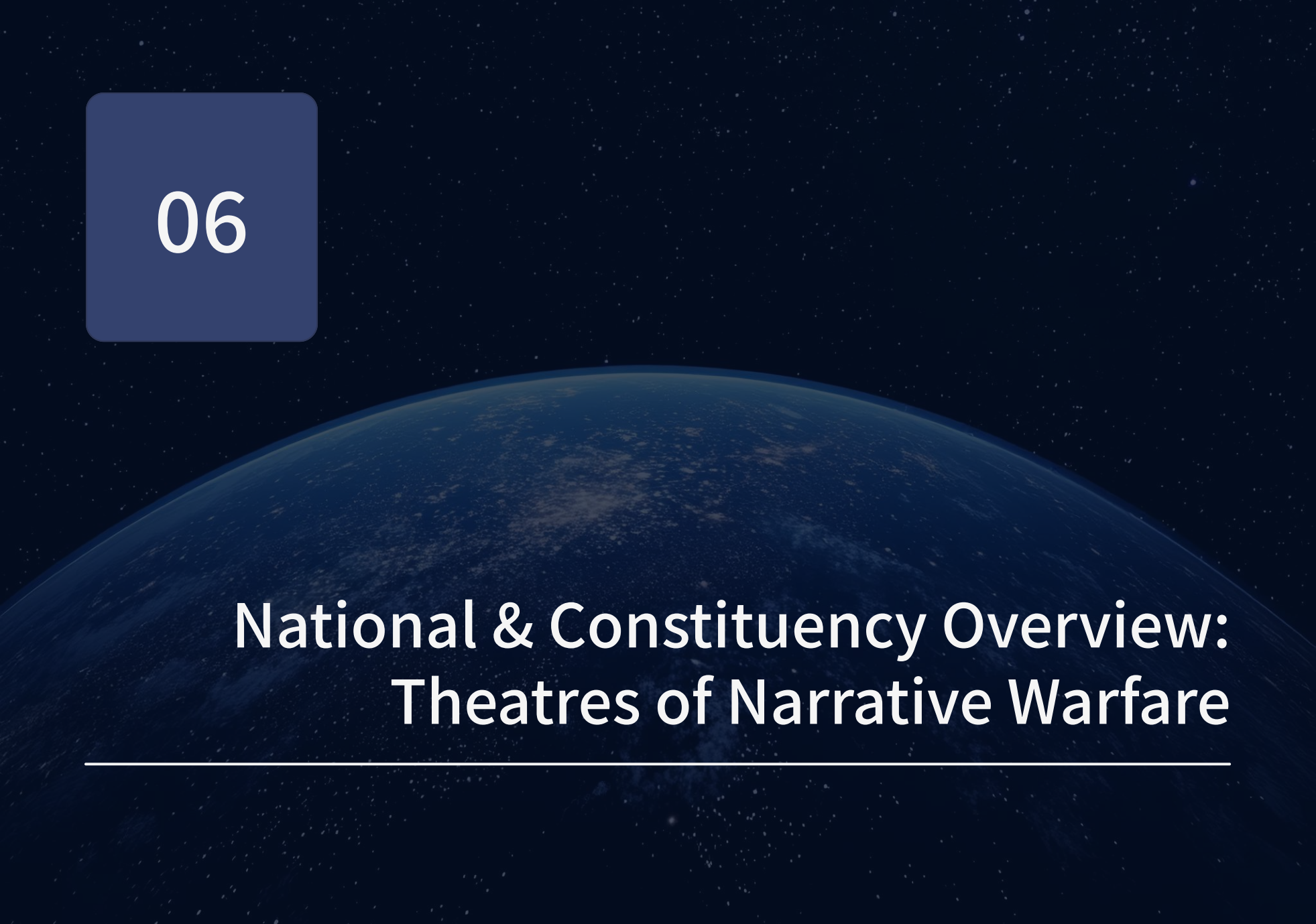
The core thesis is stark:

The digital conversation around the 2025 by - elections was dominated by an industrial broadcast system, not a pluralistic public square.

This system was designed to:

- 1** Dominate the agenda by flooding feeds with a small set of scripts.
- 2** Game algorithmic visibility to hijack trends.
- 3** Pre - emptively frame legal and institutional processes as either salvation or proof of systemic failure.





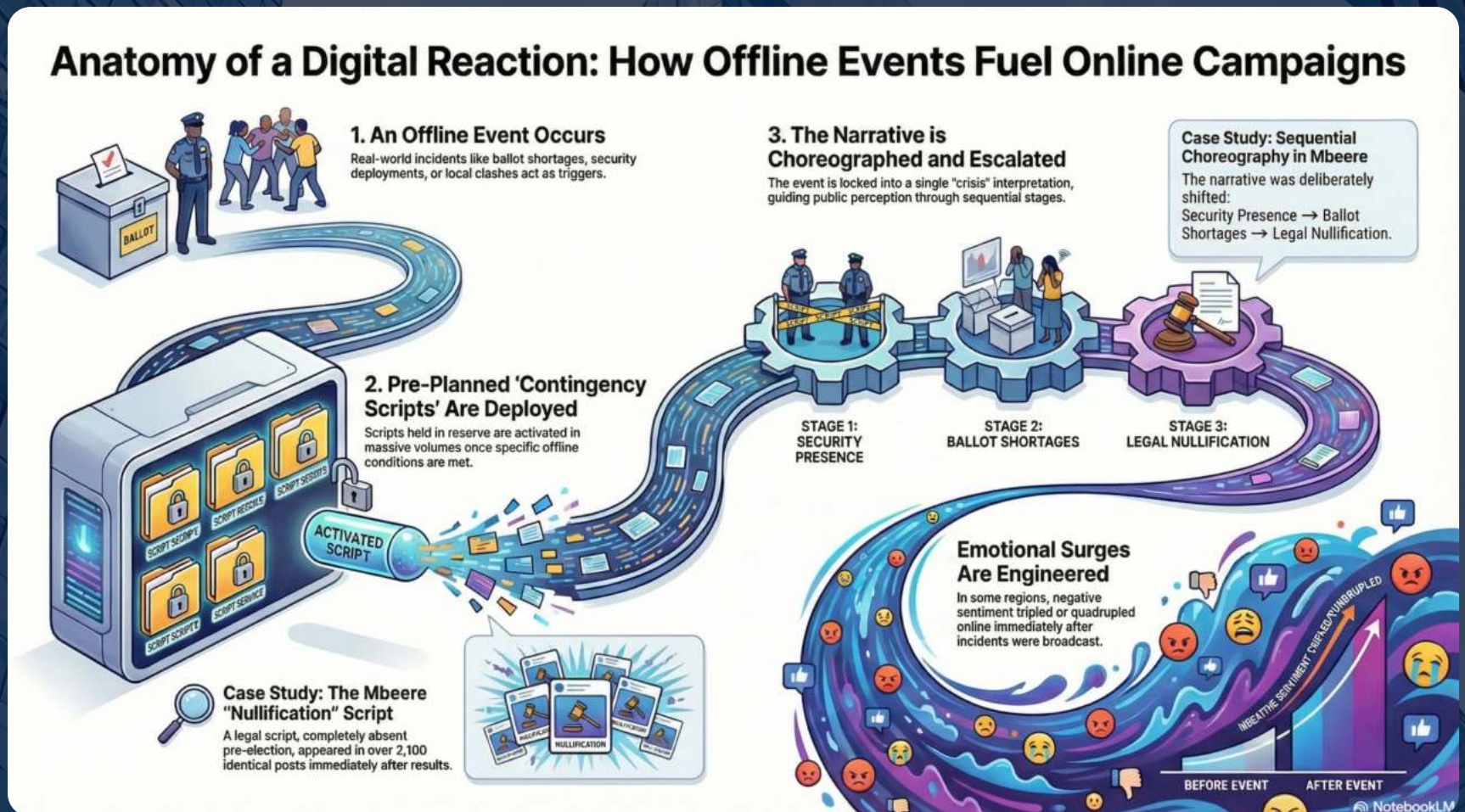
06

National & Constituency Overview: Theatres of Narrative Warfare

National Social Media Landscape

a. Digital Behaviour Around Offline Incidents

Across the three phases, the data shows digital campaigns largely reacting to, and then escalating, offline events:



Spikes in scripted content follow tally announcements and televised incidents, rather than preceding them.

Physical events (ballot shortages, security deployments, clashes) are rapidly absorbed into pre-existing frames, not debated from scratch.

Key patterns:

Contingency scripts; In Mbeere North, the legalistic nullification script is absent pre-election but appears in hundreds of near-identical copies immediately after results. The timing indicates a script held in reserve, triggered only when outcome conditions are met.

Sequential choreography; In Mbeere, content moves in stages aligned with the offline timeline:

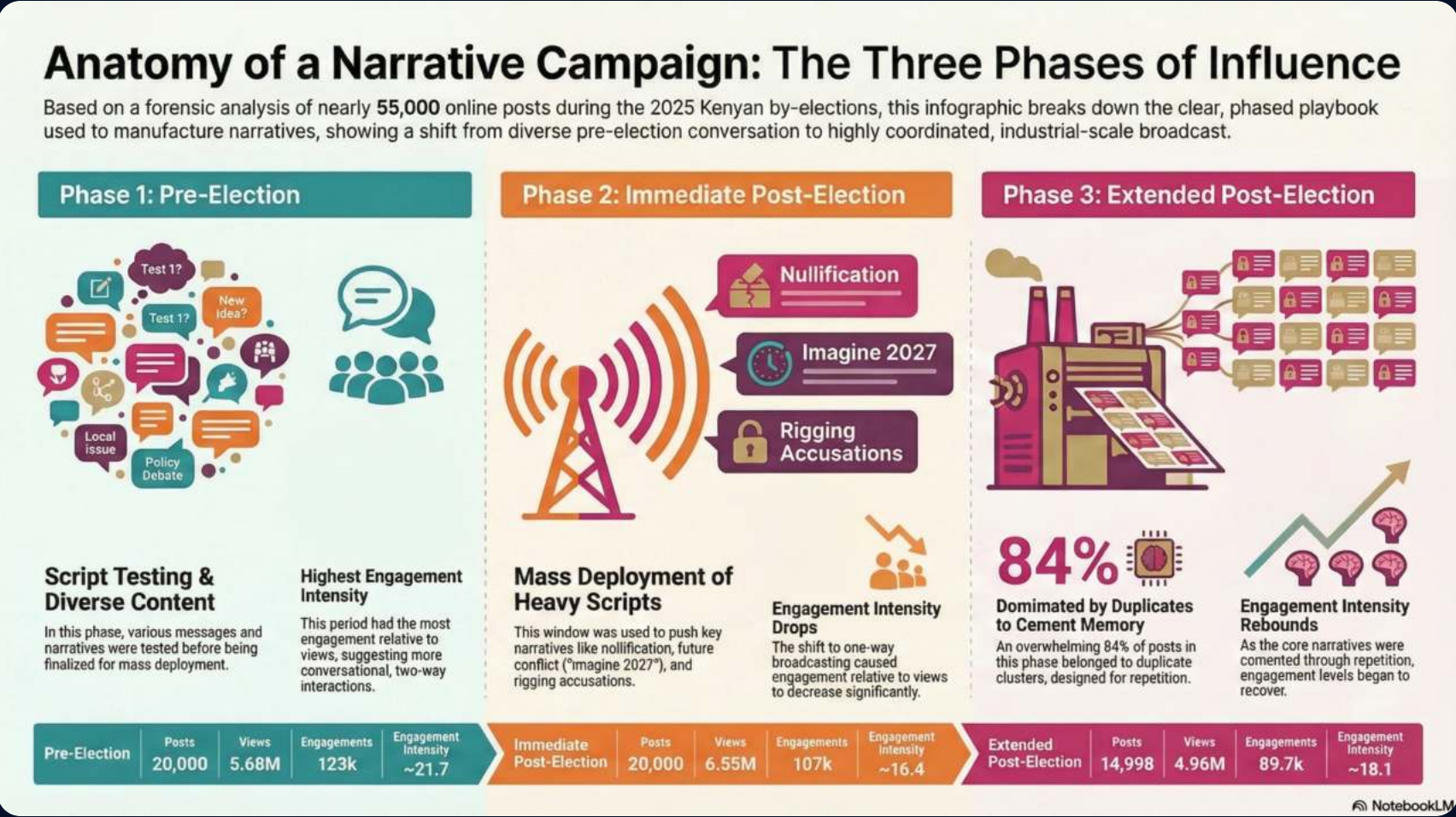
- Security presence → Administrative friction (ballot shortages) → Legal certainty of nullification.

Post-incident emotional surges; In Malava and Kasipul, negative sentiment triples or quadruples immediately after results are known and incidents (e.g. Agoro Sare chaos) are broadcast. The digital apparatus is used to lock those events into a single "crisis" interpretation.

National Social Media Landscape

b. From Conversation to Broadcast

Across phases, the transformation is structural:



Phase	Engagement Intensity
Pre - election	Highest engagement intensity (\~21.7 engagements per 1k views). More diverse content; scripts are being tested, not yet slammed into place.
Immediate post - election	Engagement intensity drops to \~16.4 per 1k views. This window is used to deploy heavy scripts (nullification,"imagine2027", violence/rigging accusations).
Extended post - election	Engagement intensity rebounds to \~18.1 per 1k views. 84% of posts belong to duplicate clusters; this phase is about cementing memory.

The national meta - narratives that cut across constituencies:



State vs. Voter
Elections framed as a struggle between the people and a coercive state:

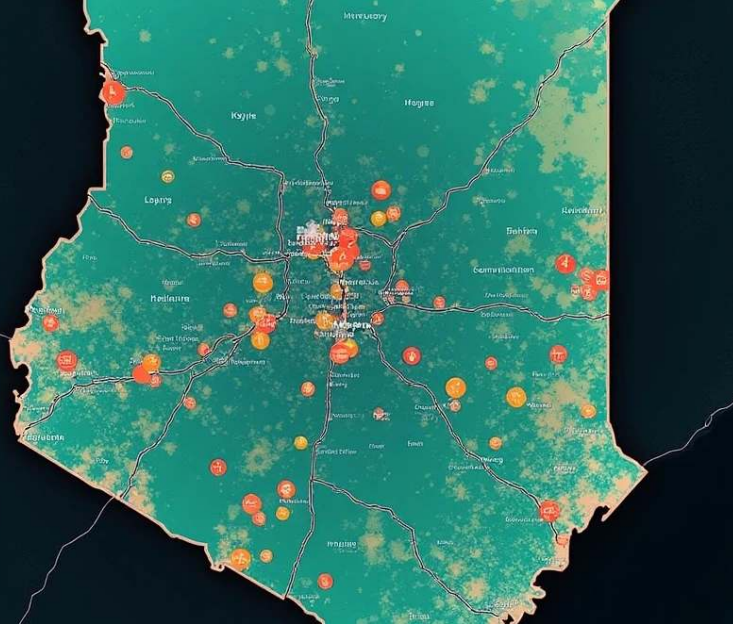
- Repeated allegations of state-sponsored rigging and violence.
- Scripted mentions of security organs as political weapons.



Courts as the Real Battleground
Judicial petitions recast from safeguards into the primary arena where "stolen" elections will be reversed.



2025 as Dress Rehearsal for 2027
Local by - election incidents treated as trailers for an inevitable national crisis.

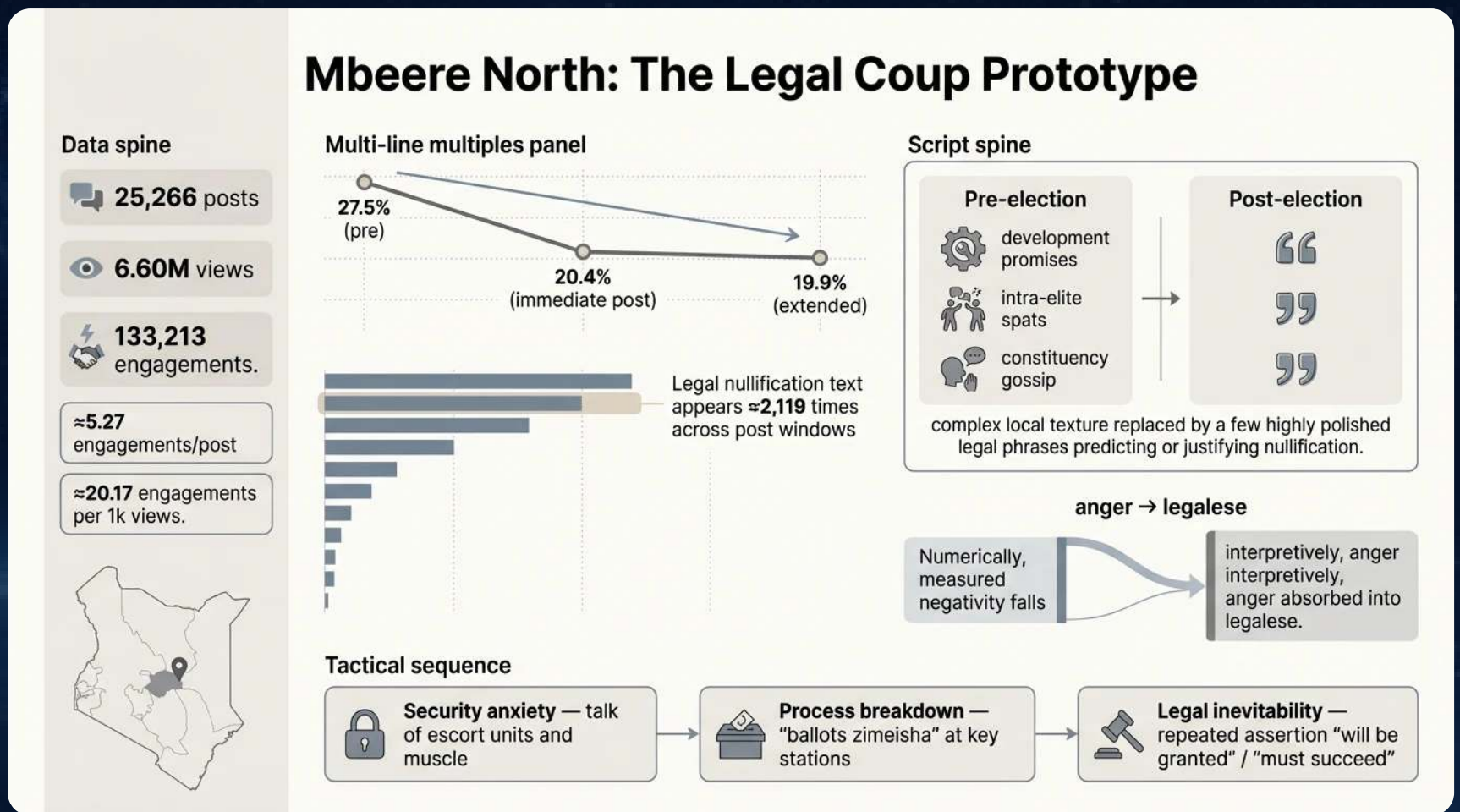


Constituency-Level Deep Dive

Influence operations are not evenly distributed. They are concentrated in specific constituencies, each used for a distinct tactical demonstration.

Constituency-Level Deep Dive

a. Mbeere North: The Legal Coup Prototype



1

Data spine

- 25,266 posts; 6.60M views; 133,213 engagements.
- ≈5.27 engagements/post; ≈20.17 engagements per 1k views.
- Negative sentiment: 27.5% (pre) → 20.4% (immediate post) → 19.9% (extended).

2

Script spine

- Legal nullification text appears ≈2,119 times across post windows.
- Pre - election, discourse is locally textured: development promises, intra - elite spats, constituency gossip. Post - election, this complexity is replaced by a few highly polished legal phrases predicting or justifying nullification.

3

The tactical sequence

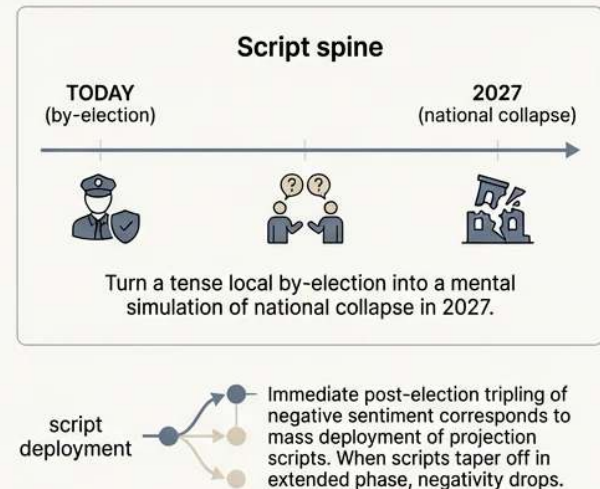
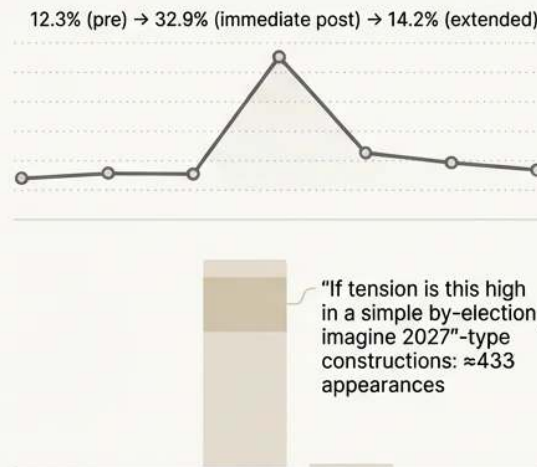
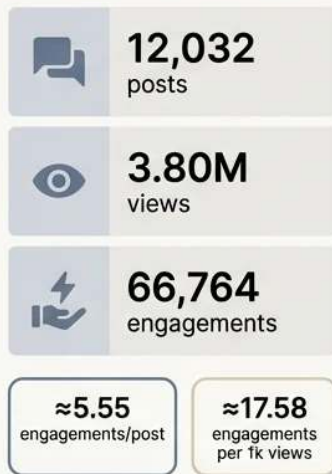
- Security anxiety- talk of escort units and muscle.
- Process breakdown- "ballots zimeisha" at key stations.
- Legal inevitability- repeated assertion that any nullification petition "will be granted" or "must succeed."

Numerically, measured negativity falls; interpretively, anger is being absorbed into legalese. Mbeere becomes a proof - of - concept for lawfare framing: democracy narrated as something that will be resolved in court, according to scripts, before evidence is even discussed.

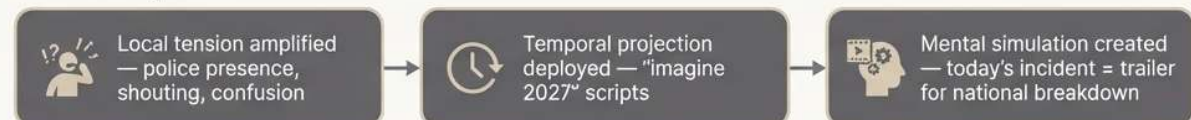
Constituency-Level Deep Dive

b. Malava: The Apocalypse Trailer

Malava: The Apocalypse Trailer



Tactical sequence



Data spine

- 12,032 posts; 3.80M views; 66,764 engagements.
- ≈5.55 engagements/post; ≈17.58 engagements per 1k views.
- Negative sentiment: 12.3% (pre) → 32.9% (immediate post) → 14.2% (extended).

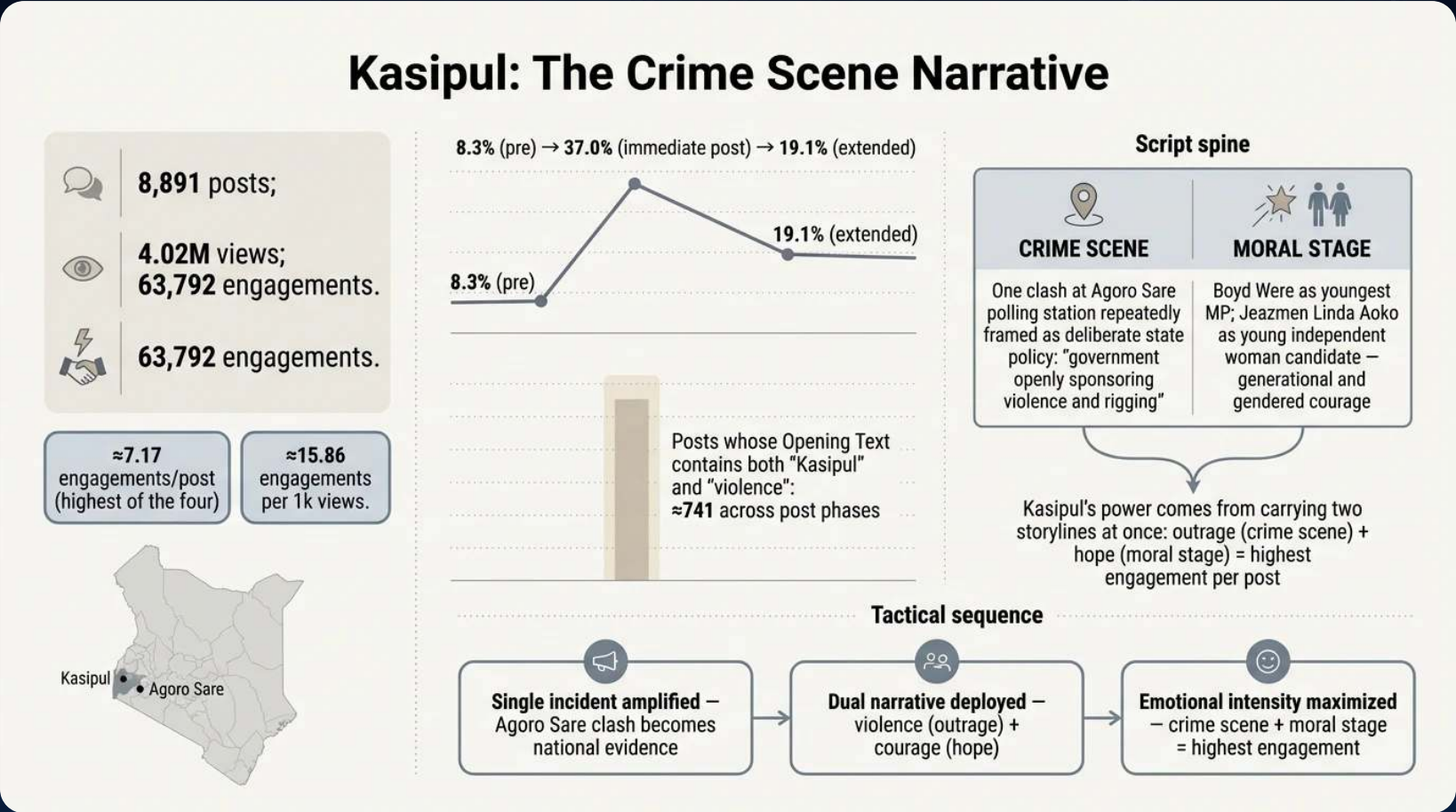
Script spine

- "If tension is this high in a simple by-election, imagine 2027"-type constructions: ≈433 appearances.
- Malava's core function is temporal projection:
 - Turn a tense local by-election into a mental simulation of national collapse in 2027.
 - Instruct audiences to treat today's police presence, shouting, or confusion as the trailer for a wider breakdown.

The immediate post-election tripling of negative sentiment is not a random flare-up; it corresponds to the mass deployment of projection scripts. When those scripts taper off in the extended phase, negativity drops and positivity rises. Malava is used to inject into Kenya's political bloodstream the idea that 2027 is already doomed.

Constituency-Level Deep Dive

c. Kasipul: The Crime Scene Narrative



- 1

Data spine

 - 8,891 posts; 4.02M views; 63,792 engagements.
 - ≈7.17 engagements/post (highest of the four); ≈15.86 engagements per 1k views.
 - Negative sentiment: 8.3% (pre) → 37.0% (immediate post) → 19.1% (extended).
- 2

Script spine

 - Posts whose Opening Text contains both "Kasipul" and "violence": ≈741 across post phases.
 - A single locality, Agoro Sare polling station, becomes national evidence of regime brutality. The central narrative moves as follows:
 - Take one clash and repeatedly frame it as deliberate state policy: "government openly sponsoring violence and rigging the by - election in Kasipul."

Simultaneously, Kasipul hosts narratives of hope and courage:

- 1

Boyd Were as the youngest MP.
- 2

Jeazmen Linda Aoko as a young independent woman candidate whose story is framed as the beginning of a new politics, not a defeat.

Kasipul's power comes from carrying two storylines at once: A crime scene (for outrage, evidence of state violence) and A moral stage for generational and gendered courage. This duality, plus the violence imagery and stories, drives the highest engagement per post of any constituency.

Constituency-Level Deep Dive

d. Magarini: The Momentum Myth

Magarini: The Momentum Myth



4,703
posts



1.63M
views



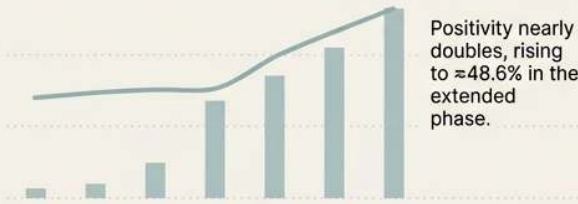
29,296
engagements

≈6.23 engagements/post

≈18.0 engagements per 1k views



9.0% (pre) → 24.7% (immediate post) → 5.2% (extended)



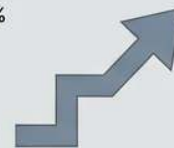
Positivity nearly doubles, rising to ≈48.6% in the extended phase.

Pre Immediate Post Extended

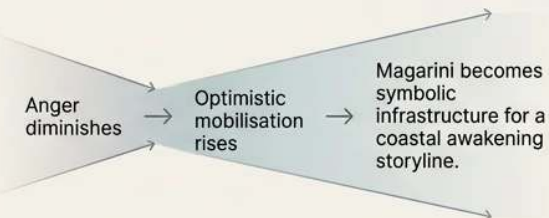
Over time, anger diminishes and is replaced with optimistic mobilisation.

Script spine

Opposition's ≈33% share is spun as proof that "if a fragmented opposition gets 33% here, 50% in Mombasa will be easy".



Magarini is less about present grievances, more about future coastal power.



Data spine

- 4,703 posts; 1.63M views; 29,296 engagements.
- ≈6.23 engagements/post; ≈18.0 engagements per 1k views.
- Negative sentiment: 9.0% (pre) → 24.7% (immediate post) → 5.2% (extended).

Magarini shows how even a loss can be repurposed into a narrative of inevitability:

Opposition's ≈33% share is spun as proof that "if a fragmented opposition gets 33% here, 50% in Mombasa will be easy."

Magarini is less about present grievances, more about future coastal power.

Over time, anger diminishes and is replaced with optimistic mobilisation. Magarini becomes symbolic infrastructure for a coastal awakening storyline.

Comparative Constituency Metrics

Using lenient matching on Keywords across all phases:

Constituency	Posts	Views	Engagement	Eng/Post	Eng/1,000 Views	Negativity Shift (Pre → Immediate Post → Extended)
Mbeere	25,266	6,603,800	133,213	5.27	20.17	27.5% → 20.4% → 19.9% (anger "legalised")
Malava	12,032	3,796,717	66,764	5.55	17.58	12.3% → 32.9% → 14.2% (fear spike, then partial cooling)
Kasipul	8,891	4,023,216	63,792	7.17	15.86	8.3% → 37.0% → 19.1% (crime scene framing)
Magarini	4,703	1,627,489	29,296	6.23	18.00	9.0% → 24.7% → 5.2% (from grievance to momentum)

Each constituency is deployed for a different emotional outcome:

1

Mbeere: legalisation of anger.

2

Malava: projection of dread into 2027.

3

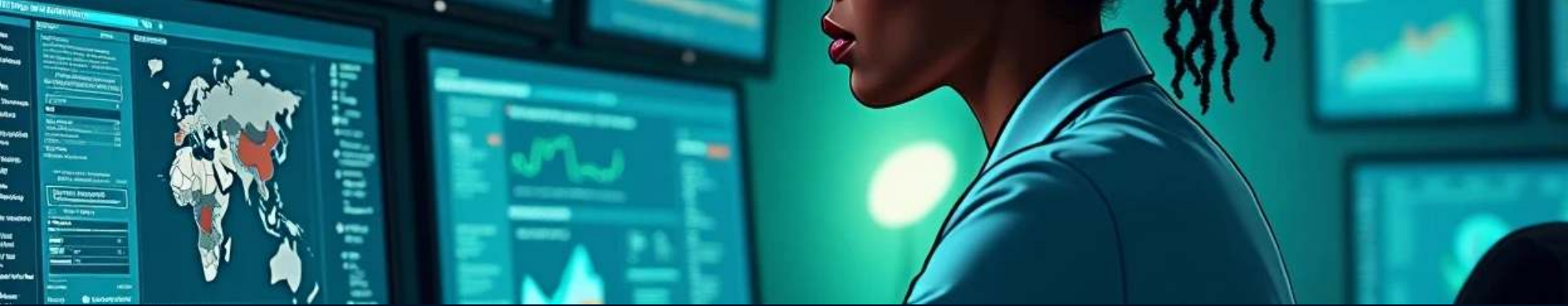
Kasipul: spectacle of violence and moral test.

4

Magarini: proof of future coastal realignment.

07

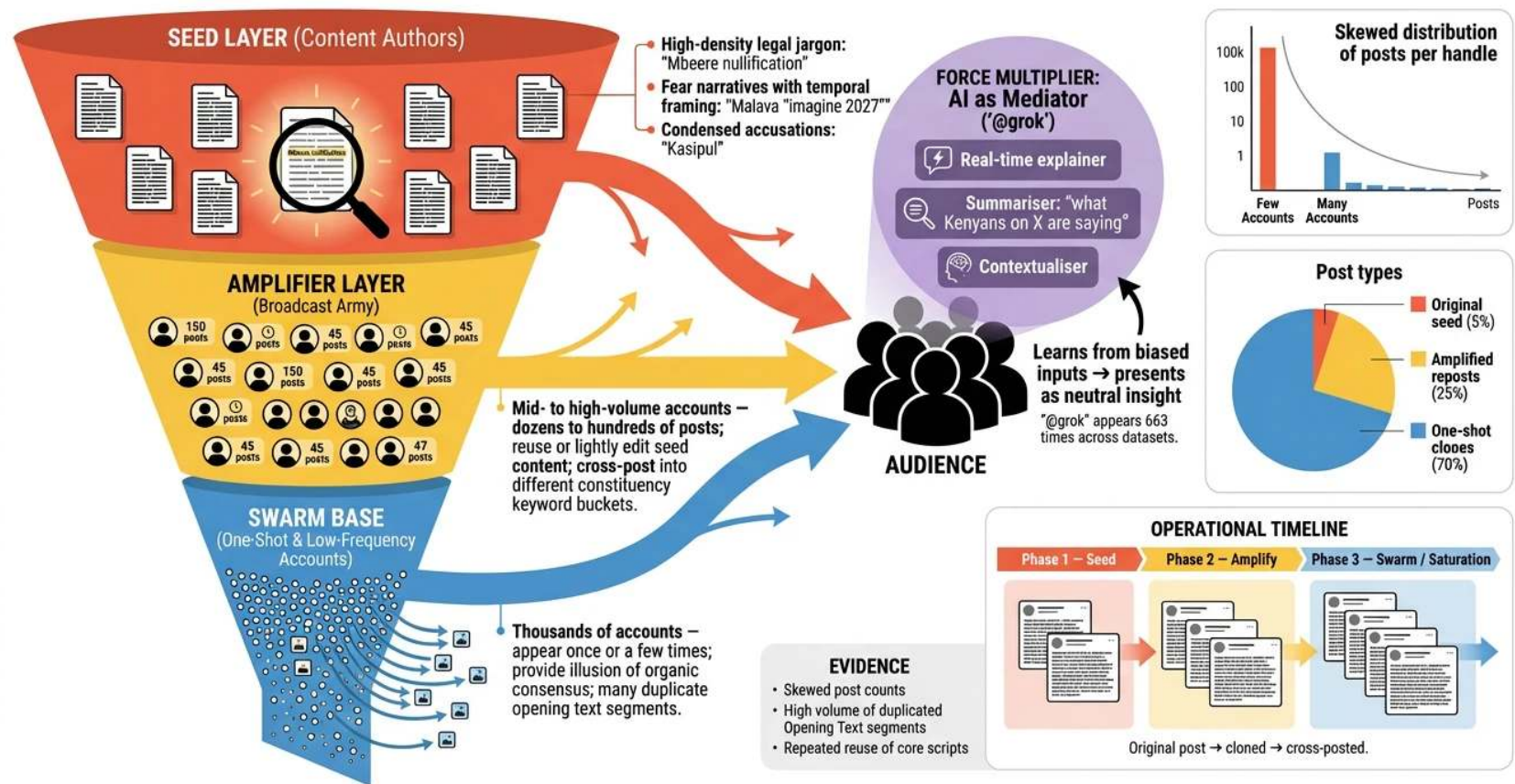
The Actor Ecology: Anatomy of the Seed-Amplifier War Machine



The Actor Ecology: Anatomy of the Seed-Amplifier War Machine

The operations observed in these datasets are structural, not spontaneous. They are executed by an ecosystem with clear functional layers.

DIGITAL INFLUENCE OPERATIONS: THE FUNNEL, THE MULTIPLIER, AND THE EVIDENCE



Layers and Roles

a. Seed Layer (Content Authors)

Crafts the core texts:

High -
density legal
jargon
(Mbeere
nullification).

Fear
narratives
with
temporal
framing
(Malava
"imagine
2027").

Condensed
accusations
of rigging
and state
violence
(Kasipul).

These accounts appear early in a script's lifecycle with original, complex postings that are later cloned.





Layers and Roles

b. Amplifier Layer (Broadcast Army)

Mid - to high - volume accounts that:

1

Post dozens to hundreds of times across the three phases.

2

Reuse or lightly edit seed content.

3

Cross - post scripts into different constituency keyword buckets.

They form the central nervous system of distribution.



Layers and Roles

c. Swarm Base (One-Shot & Low-Frequency Accounts)

Thousands of accounts that:

Appear once or a handful of times.

Mostly post a single instance of a known script.

Provide the illusion of organic consensus: many "different" voices saying the same thing.

Evidence of this pattern comes from the skewed distribution of post counts per Author Handle and the sheer volume of duplicated Opening Text segments.

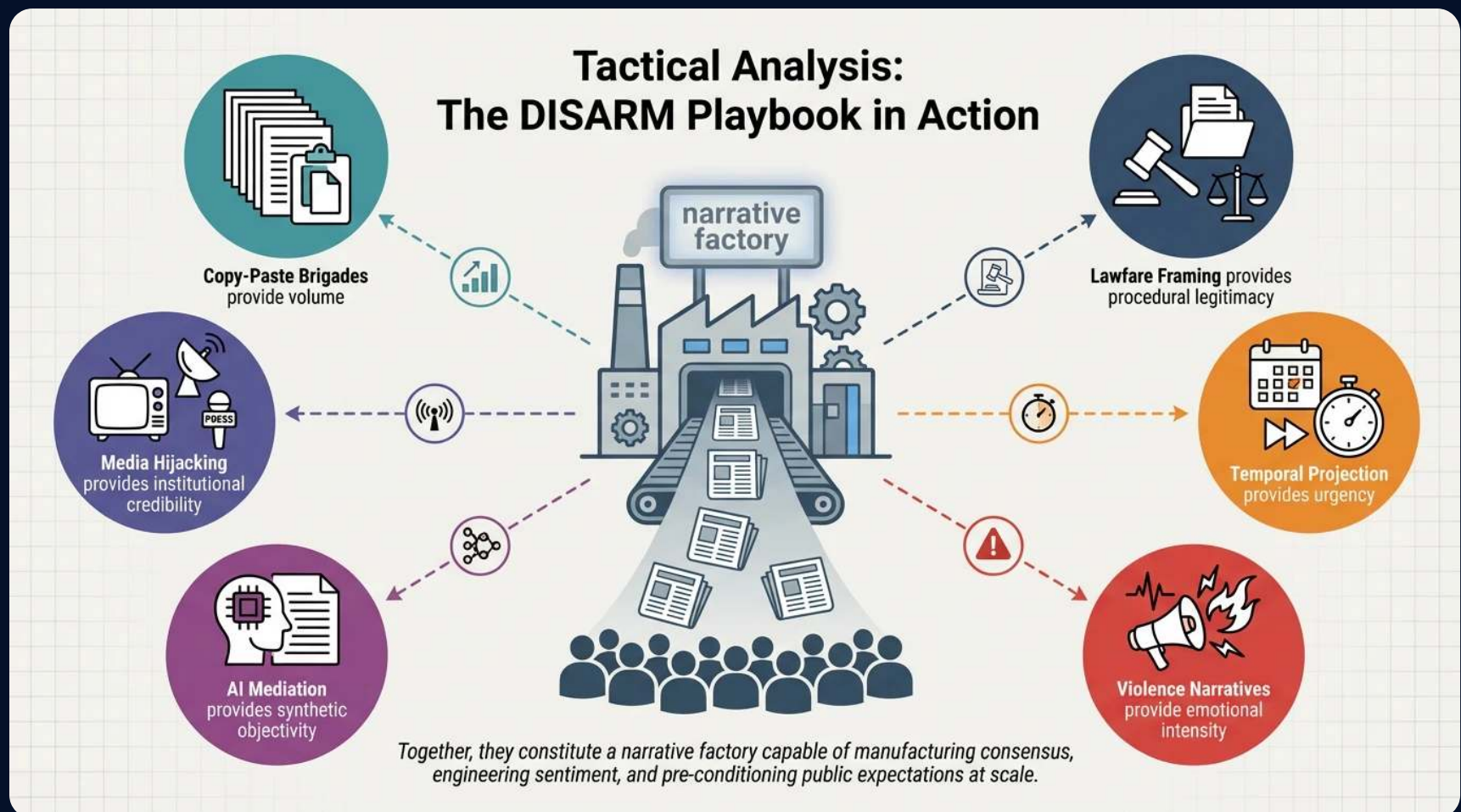
08

Tactical Analysis: The DISARM Playbook in Action



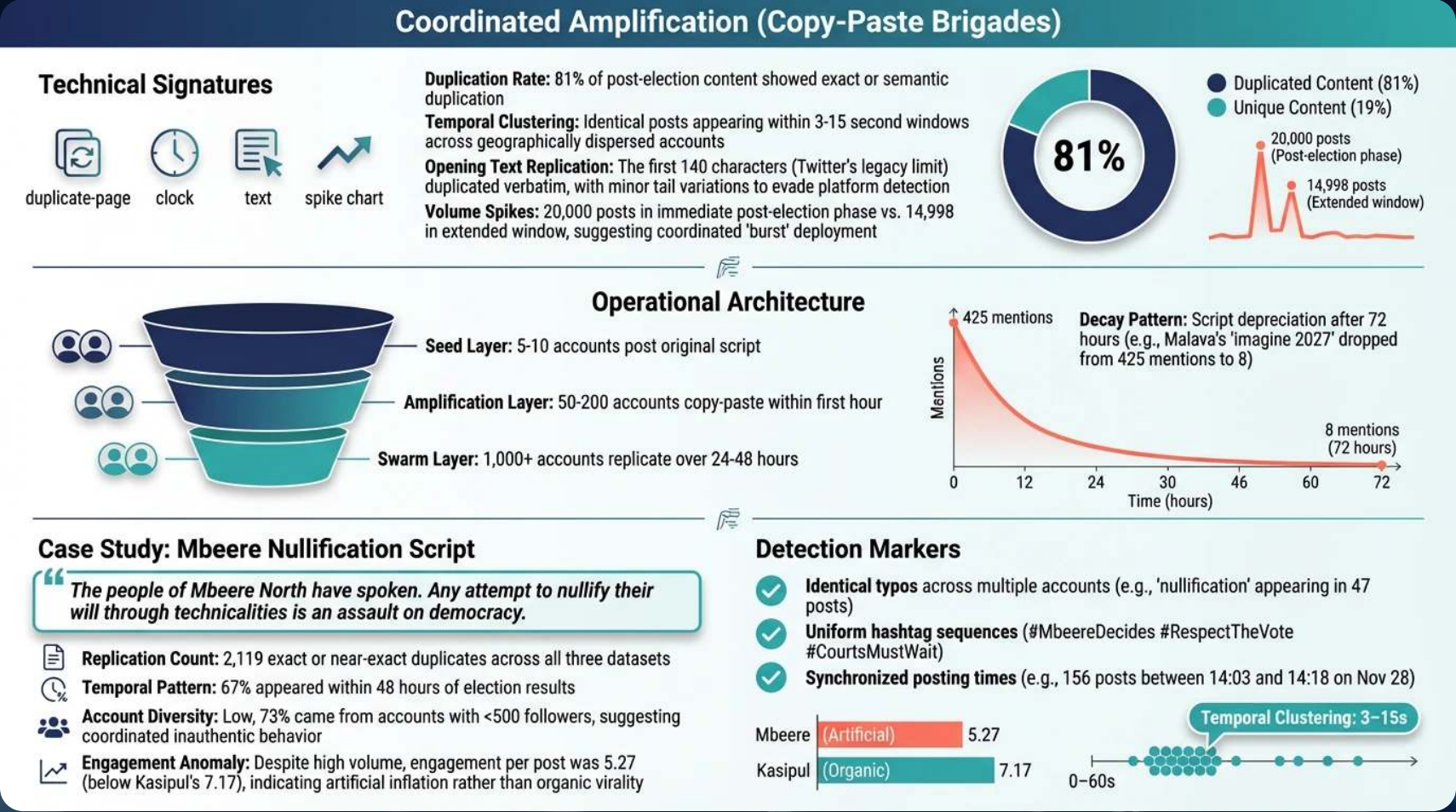
Tactical Analysis: The DISARM Playbook in Action

Viewed through a DISARM lens, the 2025 operations show consistent use of identifiable tactics. These six mechanisms are not isolated tactics; they are integrated components of an industrial production system:



Coordinated Amplification (Copy-Paste Brigades)

Definition: Industrial-scale duplication of identical or near-identical content across multiple accounts to simulate organic consensus.



Lawfare Framing (Weaponized Legal Jargon)

Definition: The systematic transformation of electoral contests into pre-litigated legal battles in the public sphere, conditioning audiences to expect and accept judicial intervention as the "real" outcome.

Strategic Function:

1

Delegitimization

Frames electoral victories as provisional, pending legal "correction."

2

Expectation Management

Primes supporters for court challenges, reducing shock if results are overturned.

3

Institutional Capture

Applies pressure on judiciary by creating public "consensus" that courts "must" act.

Linguistic Markers:

1

Legal Terminology Saturation

"nullification," "technicalities," "petition," "irregularities," "evidence," "affidavit."

2

Certainty Language

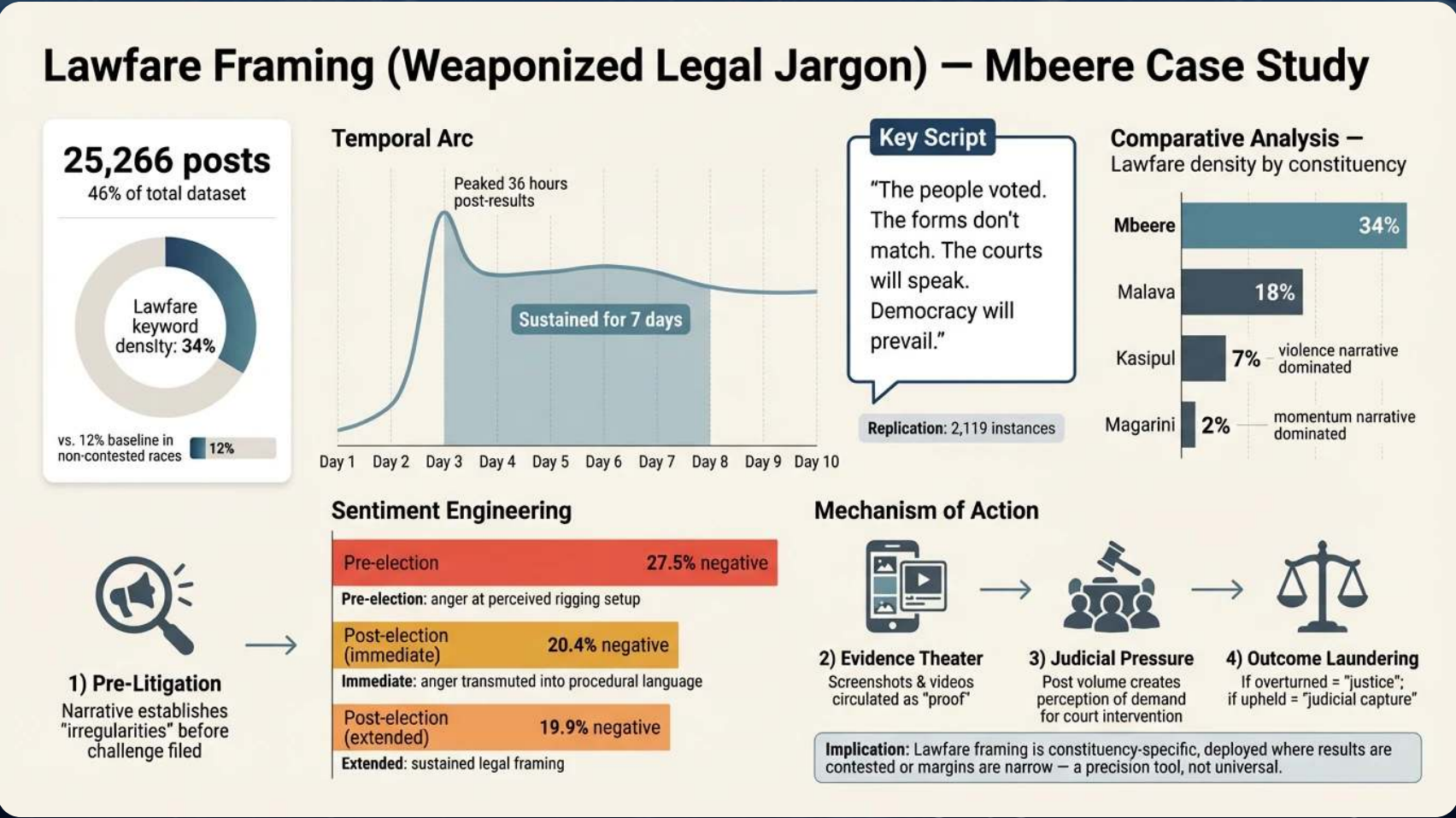
"will be overturned," "must be challenged," "cannot stand," "courts will decide."

3

Procedural Fetishism

Detailed references to Forms 35A, 35B, RTS failures, technical jargon as credibility signal.

Mbeere Case Study:



Temporal Projection ("Rehearsal" Scripts)

Definition: The use of current events as narrative vehicles to pre-condition audiences for future scenarios, normalizing crisis and conflict as inevitable horizons.

Strategic Function:

1

Desensitization:
Repeated exposure to "crisis" language reduces shock when actual crises occur

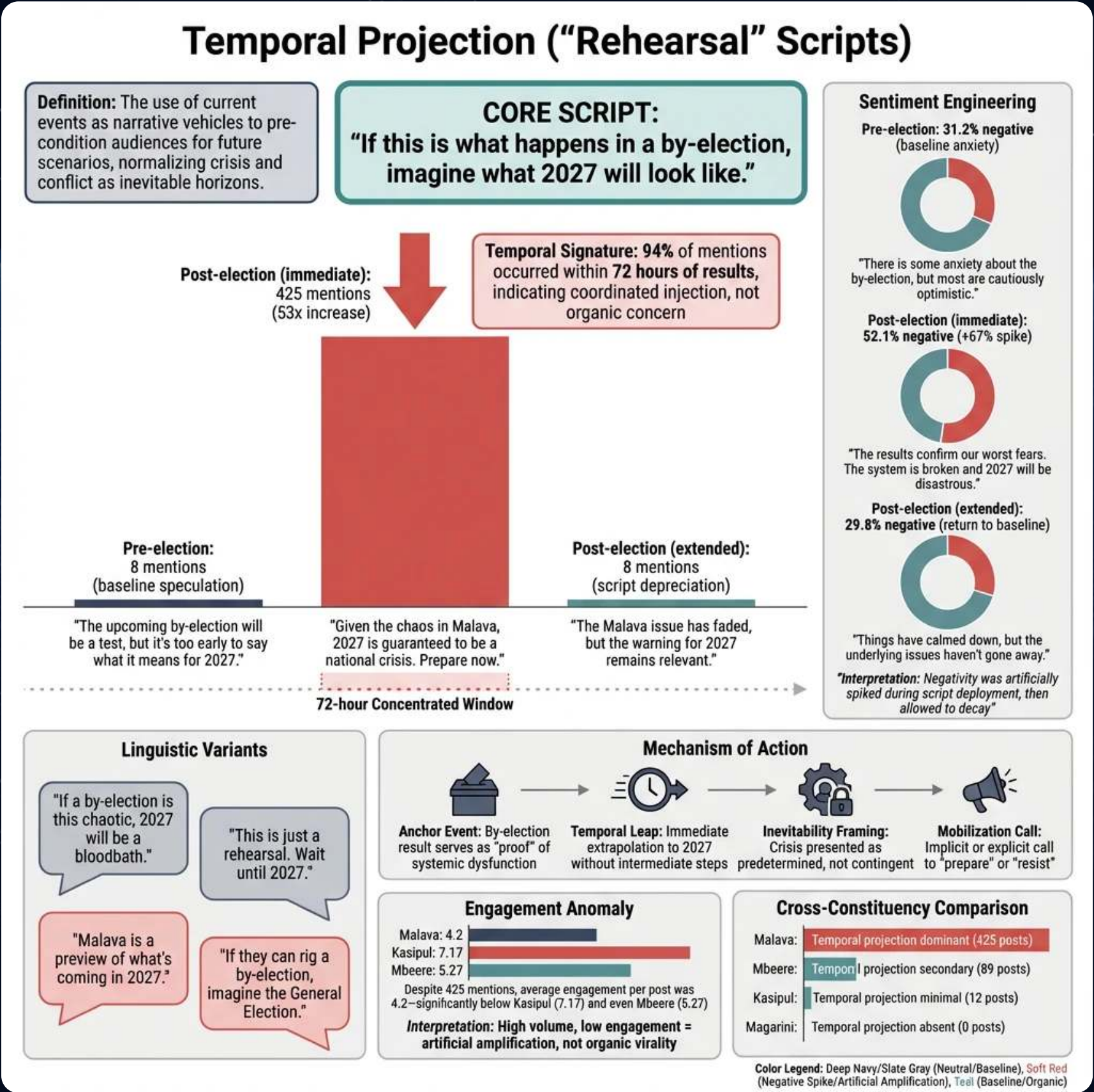
2

Expectation Setting
Frames 2027 General Election as predetermined catastrophe

3

Mobilization Priming
Creates sense of urgency and existential stakes for future contests.

Malava Case Study:



Legacy Media Hijacking

Definition: The strategic appropriation of mainstream media content (headlines, graphics, screenshots) and its recirculation with partisan captions, using institutional credibility as a "carrier wave" for manipulated narratives.

Strategic Function:

1

Credibility Transfer:

Mainstream outlet's brand authority lends legitimacy to partisan interpretation

2

Decontextualization

Headlines or images severed from original context, reframed to fit narrative

3

Institutional Capture

Media outlets inadvertently become distribution channels for manipulated narratives

Mbeere "Irregularities" Headline Case Study:

Case Study: Mbeere “Irregularities”

Original Headline (Nation):

Screenshot Caption (recirculated):

Replication: 347 instances of this screenshot across datasets

Engagement Comparison

Engagement: 6.8 per post (higher than Mbeere's 5.27 average), credibility transfer effect

Category	Engagement
Average	5.27
Hijacked screenshot	6.8

~500 ~2,300

Original Article (on outlet's platform): ~500 engagements

Hijacked Screenshot (recirculated): ~2,300 engagements (4.6x multiplier)

Interpretation: Partisan framing drives higher engagement than original journalism

Visual Signatures

Logo Prominence:
Mainstream outlet's logo prominently displayed to signal authority

Selective Cropping:
Only headline and key image shown; nuance in article body omitted

Timestamp Manipulation:
Some screenshots show altered timestamps to suggest 'breaking' news

Linguistic Patterns in Captions:

Forced Admission: "Even [outlet] admits..." / "Mainstream media can't hide..."

Selective Emphasis: "Notice they said 'concerns', that's code for rigging"

Call to Action: "Share this before they delete it" / "Mainstream media finally telling the truth"

Forced Admission: "Even [outlet] admits..." / "Mainstream media can't hide..."

Cross-Outlet Analysis

Outlet	Hijacked Screenshots
Nation Media	412
Standard Media	267
Citizen TV	198 (video stills)
KBC	76 (lower credibility, less useful in the campaign)

Mechanism of Institutional Capture:

Outlet Publishes: Attempting balanced, factual reporting

Content Hijacked: Screenshot extracted, reframed

Outlet Blamed: If they correct record, accused of "backtracking" or "bias"

Chilling Effect: Outlets may self-censor to avoid being weaponized

Replication: 347 instances of this screenshot across datasets

Interpretation: Partisan framing drives higher engagement than original journalism

Implication: Legacy media's credibility is now a liability, their content is systematically hijacked and weaponized. Traditional fact-checking (correcting individual posts) is futile; the production chain (screenshot → caption → recirculation) must be disrupted.

Violence and Crime Scene Narratives

Definition: The elevation of localized incidents into systemic proof of brutality, using emotional intensity and visual evidence to bypass rational scrutiny.

Strategic Function:

1

Emotional Hijacking

Violence imagery triggers visceral responses, short-circuiting analytical processing

2

Binary Construction

Forces audiences into "victim" or "perpetrator" camps, eliminating nuance

3

Institutional Delegitimization

Frames/captures state actors (police, IEBC) as complicit/perpetrators in violence.

Kasipul Case Study:

Strategic Function

- Emotional Hijacking:** Violence imagery triggers visceral responses, short-circuiting analytical processing
- Binary Construction:** Forces audiences into "victim" or "perpetrator" camps, eliminating nuance
- Institutional Delegitimization:** Frames state actors (police, IEBC) as complicit in violence

Kasipul Case Study - Agoro Sare Incident

Narrative Volume: 741 posts explicitly mentioning 'Agoro Sare,' "violence," or "brutality"

Peak volume: 9:00 PM - 11:00 PM (312 posts in 2 hours)

First mention: 6:47 PM on election day

Sustained mentions: 7 days post-election

Visual Saturation: 68% of posts included images or videos (vs. 23% baseline across other constituencies)

Sentiment Engineering

Pre-election: 22.4% negative (baseline anxiety)

Post-election (immediate): 77.5% negative (+246% spike)

Post-election (extended): 31.2% negative (partial decay, but sustained above baseline)

Interpretation: Violence narrative created sustained emotional intensity, unlike Malava's temporary spike

Engagement Dominance

Kasipul engagement per post: 7.17 (highest across all constituencies)

Mbeere engagement per post: 5.27

Malava engagement per post: 5.55

Magarini engagement per post: 6.23

Interpretation: Violence narratives generate highest organic engagement, emotion drives virality

Linguistic & Visual Patterns

"brutal attack"

"state-sponsored violence"

"innocent victims"

Heroism Pairing: "brave voters" resisting "thugs/goons" creates emotional binary

Decontextualization: Incident presented without proportional context

Visual Evidence Theater

- Image Replication:** Same 4-5 images circulated across 400+ posts
- Captioning Variance:** Identical images paired with different captions to fit multiple narratives
- Timestamp Manipulation:** Some images dated to different times/locations, suggesting coordinated visual library

Mechanism of Action

Incident Occurs: Real or alleged violence at polling station

Immediate Amplification: Seed accounts post images/videos within minutes

Emotional Saturation: Swann accounts replicate with outrage language

Sustained Fixation: Incident referenced for days, detached from resolution or context

Institutional Blame: Violence attributed to state actors, IEBC, or "system"

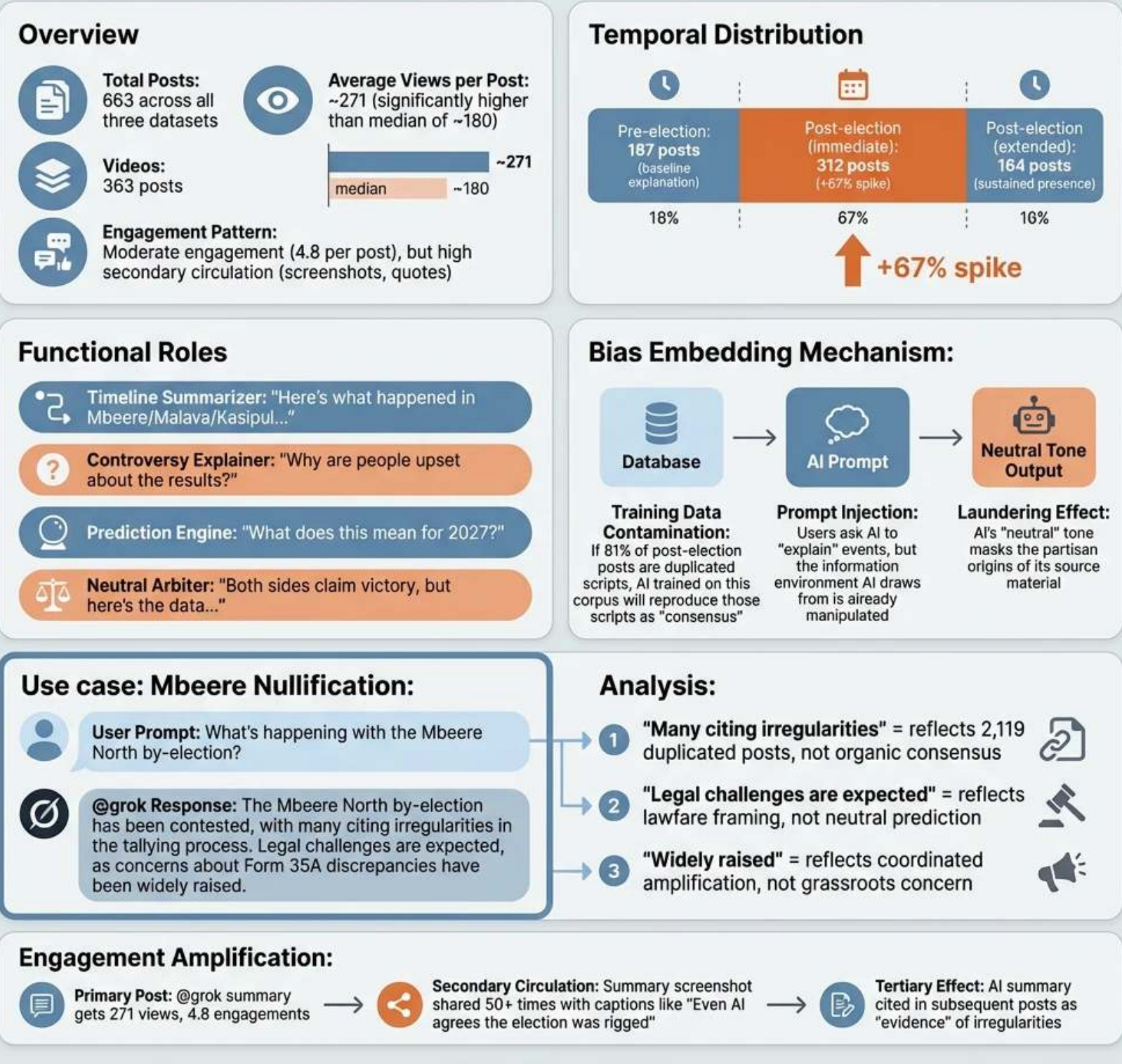
Cross-Constituency Comparison

Kasipul:	Violence narrative dominant (741 posts, 77.5% negative sentiment)
Malava:	Violence narrative secondary (127 posts, 82.1% negative sentiment)
Mbeere:	Violence narrative minimal (34 posts, 20.4% negative sentiment)
Magarini:	Violence narrative absent (3 posts, 5.2% negative sentiment)

Implication: Violence narratives are constituency-specific, deployed where incidents occur or can be manufactured. They generate highest engagement but require visual evidence (real or fabricated) to sustain intensity.

Force Multiplier: AI as Mediator (@grok)

@grok Case Study:





09

Legal and Regulatory Grey Zones: The Enforcement Gap When Law Meets Industrial Manipulation

Legal and Regulatory Grey Zones: The Enforcement Gap When Law Meets Industrial Manipulation

Core Problem: The observed operations during the 2025 by-elections exploit structural ambiguities in legal definitions, platform policies, and enforcement mechanisms. They operate in the space between "clearly illegal" and "clearly legitimate," a grey zone where coordination masquerades as consensus, manipulation presents as participation, and industrial production mimics organic conversation.

Current legal and regulatory frameworks are designed to address:

1

Individual bad actors (not coordinated networks).

2

False content (not true-but-misleading framing).

3

Automated bots (not human-operated astroturfing).

4

Direct incitement (not ambient narrative conditioning).

The 2025 by-elections revealed that none of these frameworks are adequate for the industrialized information environment we now inhabit.



Human-Operated Astroturfing: The "Real Person" Loophole

Definition: Astroturfing is the practice of creating artificial grassroots support through coordinated inauthentic behavior. Traditional definitions focus on automated bots, accounts operated by scripts or AI that post at inhuman speeds.

The Grey Zone: The 2025 by-elections demonstrated a shift to human-operated astroturfing: thousands of real people, operating real accounts, posting coordinated content on a production schedule.

Why Current Law Fails:

1 Kenyan Computer Misuse and Cybercrimes Act (2018)

Criminalizes "false publication" and "identity theft," but does not address coordinated authentic behavior.

2 Platform Policies (Twitter/X, Facebook, WhatsApp)

Define "inauthentic behavior" as bot-driven or fake accounts, not coordinated real accounts.

3 Electoral Code of Conduct

Prohibits "false statements" but not coordinated amplification of true-but-misleading statements.



10

Recommendations

Evolving Journalism Roles

1 Report the Operation, Not Just the Content

Trace how viral claims spread and who initiates them.

Newsrooms must investigate how narratives are manufactured and amplified, rather than simply debunking isolated falsehoods.

Actionable steps:

- When a claim goes viral, assign reporters to trace its origin: Who posted it first? How many accounts repeated it verbatim? What was the temporal pattern of spread?
- Publish investigative pieces that illuminate the industrial process: "How a Single Legal Tweet Became 911 Copies: Inside the Mbeere Nullification Script."
- Create visual timelines showing the choreography of narrative pivots (e.g., security fears → logistical chaos → legal inevitability).
- Interview data scientists, OSINT researchers, and platform experts to explain coordination indicators to the public.

3 Protect and Watermark Media Assets

Official election graphics, results sheets, and key explainers should be clearly watermarked and archived to prevent manipulation.

Actionable steps:

- Watermark all original graphics, infographics, and results tables with newsroom branding and timestamps.
- Maintain an easily accessible online repository of verified graphics, results tables, and key explainers (e.g., "How Petitions Work," "What Nullification Requires").
- When newsroom assets are misused (e.g., screenshots with partisan captions), explicitly call out the manipulation in coverage: "A graphic originally published by [Newsroom] has been recirculated with false claims. Here is what the data actually shows."
- Collaborate across newsrooms to create a shared "verified assets library" that all outlets can reference.

2 Build Internal Early-Warning Systems

Newsrooms should develop "volume vs. engagement" radar as an internal alert system for astroturfing.

Action Steps:

- Create simple dashboards to track:
 1. Text clusters with sensitive terms (e.g., "nullification," "military").
 2. Sudden post spikes with low engagement.
 3. Accounts posting identical content in short time frames.
- Establish a "CIB desk" to monitor dashboards during elections.
- Set alerts: 50+ identical posts in an hour triggers review.
- Collaborate with monitoring labs for real-time script alerts.

4 Invest in AI and CIB Literacy for Editorial Staff

All journalists, editors, and producers need training on how coordinated inauthentic behavior works and how AI mediators shape narratives.

Actionable steps:

- Conduct quarterly training sessions covering:
- How to recognize copy-paste brigades and paraphrase swarms.
 - How AI systems like @grok summarize timelines and the risks of laundering bias as neutrality.
 - How to interpret alerts from civil society CIB labs.
 - When not to quote obviously industrialized content (e.g., verbatim scripts from swarm accounts).
 - Develop internal editorial guidelines: "Do not quote identical text from multiple low-follower accounts as evidence of 'public opinion.'"
 - "When describing AI-generated summaries, clarify that they reflect the input data, which may itself be manipulated."
 - Invite external experts (data scientists, disinformation researchers) for newsroom "lunch and learns."

Reporter Checklist

- 1** Check for Duplication
Is the text duplicated across multiple accounts?
- 2** Look for Coordination Signs
Does the account show signs of coordination?
- 3** Identify Temporal Clustering
Is there evidence of posts clustered in short time windows?
- 4** Frame Industrialized Content
Frame as a coordinated campaign, not public opinion.
- 5** Use Screenshots Wisely
Incorporate screenshots with context about coordination.

Civic Engagement Strategies

1 Real-Time Monitoring

Utilize open-source tools for detecting manipulation.

2 Legal Myths Awareness

Create clear legal explainers to combat lawfare.

3 Civic Education Programs

Empower citizens to recognize coordinated narratives.



Proposed Solutions

Amend computer misuse laws or election laws to criminalize coordinated inauthentic amplification, defined to include:

1 Number of accounts

Three or more accounts posting identical or near-identical content.

2 Duration

A defined measurement window (e.g., content posted within 24 hours).

3 Disclosure

Coordination occurring without clear disclosure.

4 Intent

An intention to deceive the public about the existence of genuine grassroots support.

5 Penalty

Fines for organizers or account suspension for participants.

6 Defenses

including,

- Disclosure: If coordination is disclosed (e.g., "This is an organized campaign"), no offense.
- Organic: If coordination is coincidental (e.g., people independently using the same hashtag), no offense.
- Scale: If coordination involves <3 accounts or reaches <10,000 people, no offense.

Enforcement:

Regulators

Primary enforcement during election periods.

Prosecution of organizers

Platforms

Required to detect and report coordinated networks to IEBC/regulators.





Key Conclusions

- 1** **Factory of Ideas**
2025 by-elections reveal Kenya's digital sphere as a programmable factory floor.
- 2** **Need for Real-Time Monitoring**
We must monitor scripts and not just viral posts to understand the manipulations.
- 3** **Systematic Response Required**
Actors must be mapped; AI mediators need auditing to ensure accountability.
- 4** **Urgency of Recommendations**
Building resilience is critical before the 2027 election cycle.
- 5** **Democracy's Defense**
Capacity to see and resist manipulation is vital for Kenya's democratic future.

Thanks a Lot!

© AI Salon Kenya & Election Ledger 2025

To chat more about methods or results, contact

mwende@webworks.co.ke

lilian@webworks.co.ke