



AI-driven misinformation detection and strategic crisis communication during high-density mass gatherings: A comparative empirical analysis of the Maha Kumbh Mela (India) and Hajj pilgrimage (Saudi Arabia)

Shahnaaz Zabi¹, Mustafa Abdulraheem Saeed Alsabri²

¹ Assistant Professor and Head, Faculty of Journalism and Mass Communication, Usha Martin University, Ranchi, Jharkhand, India

² Assistant Professor, Department of Communication and Media, University of Hail, Hail, Saudi Arabia

Corresponding Author: Shahnaaz Zabi

DOI: <https://doi.org/10.66856/njmr.2026.11.3.11078>

Abstract

Mass human gatherings—such as the Maha Kumbh Mela in Prayagraj, India, and the annual Hajj pilgrimage in Makkah, Kingdom of Saudi Arabia—represent unprecedented challenges in crowd safety, emergency logistics, and public administration. In the hyper-connected digital era, the primary operational bottleneck in managing such mega-events has transitioned from physical crowd dynamics to digital information contagion. Misinformation, unverified panic rumors, and malicious disinformation circulating via encrypted messaging channels (e.g., WhatsApp) and open microblogging platforms (e.g., X/Twitter) can trigger catastrophic stampedes, strain emergency healthcare services, and erode public trust in state authority.

This study proposes an integrated, multi-modal Artificial Intelligence (AI) framework combining cross-lingual Natural Language Processing (NLP), dynamic spatio-temporal sentiment tracking, and automated Retrieval-Augmented Generation (RAG) models for real-time crisis communication. Analyzing an empirical dataset of 4.82 million multi-lingual social media posts, public channel messages, and encrypted network telemetry across Hindi, Arabic, English, Urdu, Bengali, and regional dialects, this paper evaluates the velocity, viral mechanics, and psychological drivers of digital panic.

Our findings introduce the Panicked Rumor Echo Index (PREI) and establish that preemptive, AI-generated counter-narratives deployed within a critical 14-minute window reduce public panic propagation by 78.4%. Furthermore, fine-tuning cross-lingual models on localized crisis dialects achieves an F1 precision score of 0.942. This research provides state authorities, disaster management bodies, and policy makers with an operational, privacy-compliant, and culturally attuned blueprint for algorithmic crisis governance.

Keywords: Artificial Intelligence, Crisis Communication, Misinformation Detection, Mass Gatherings, Kumbh Mela, Hajj Pilgrimage, Natural Language Processing, Public Policy, Sampling Methodology, Retrieval-Augmented Generation

The Spatial-Digital Nexus in Mega-Event Governance

High-density mass human gatherings represent some of the most complex socio-technical operational environments in contemporary civil administration. Events such as the Maha Kumbh Mela in Prayagraj, Uttar Pradesh, India—which recorded over 240 million pilgrim visits over a 45-day window—and the annual Hajj pilgrimage in Makkah, Saudi Arabia—hosting over 2.5 million worshippers from 160+ countries within a five-day timeframe—test the physical and digital limits of urban infrastructure, transit logistics, and public safety systems.

Historically, emergency management frameworks at mega-events focused almost exclusively on physical spatial control. Public safety leadership deployed physical barricading, flow-metering gates, pedestrian directional corridors, and closed-circuit television (CCTV) surveillance networks to monitor physical density (pedestrians/m²) and mitigate choke-point compression.

However, the fundamental nature of crowd vulnerability has undergone a profound structural shift over the past decade. The ubiquitous proliferation of low-cost mobile broadband, smartphone penetration, and algorithmic social media networks has intertwined physical spatial dynamics with digital information ecosystems. In contemporary mega-events, a pilgrim is not merely a physical agent moving

through a geographical space; they are a hyper-connected digital node simultaneously receiving, processing, and broadcasting real-time information. When unverified rumors, fabricated emergency alerts, or malicious disinformation circulate across digital networks, they act as an intangible trigger for kinetic crowd panic.

In fluid mechanics and crowd psychology, physical crowds at extreme densities (>4 persons/m²) act as non-linear physical systems where minor localized disruptions induce severe compression waves. A forwarded audio note alleging a stampede, a fabricated image showing a collapsed pedestrian bridge, or a viral rumor concerning police enforcement actions can instantly alter collective crowd trajectories. As panic-stricken individuals alter their movement—attempting to reverse direction, force past security gates, or bypass established safety corridors—the physical collision between opposing crowd streams creates localized crushing and asphyxiation: a phenomenon termed the digital-to-kinetic cascade.

Information Contagion and the Asymmetry of Dissemination Velocity

The core operational vulnerability in managing crisis communications during mass gatherings lies in the severe structural asymmetry between misinformation propagation

velocity and official state verification cycles. Misinformation circulating on decentralized, end-to-end encrypted messaging channels (such as WhatsApp or Telegram) or amplified by algorithmic engagement loops on open platforms (such as X/Twitter and TikTok) spreads according to exponential network dynamics. Peer-to-peer trust networks further accelerate this transmission: an individual is significantly more

likely to trust and forward an alarming voice note received from a family member or community leader than a generic public announcement broadcast over loudspeakers.

Conversely, traditional governmental and law enforcement public relations machinery operates under rigid, centralized verification hierarchies. Before an official correction or debunking notice can be issued, field intelligence must be reported up the chain of command, cross-checked with emergency services, drafted by communications personnel, reviewed by legal authorities, and authorized by senior administration. This manual verification chain routinely requires between 45 and 120 minutes. In a high-density mass gathering environment, a 45-minute delay in neutralizing a viral stampede rumor represents a catastrophic operational failure. During this "information vacuum," the rumor achieves peak network saturation, driving physical panic behavior long before the official debunking statement is released.

Strategic Policy and Direct Governance Benefits

This research directly addresses this public safety gap by designing, implementing, and evaluating an artificial intelligence-driven framework specifically calibrated for real-time misinformation detection, spatio-temporal sentiment tracking, and automated, hallucination-free crisis communication during mega-events in India and Saudi Arabia. By bridging computer science, mass communication theory, and public administration, this study provides state authorities, disaster management agencies, and law enforcement leadership with four primary operational benefits:

- 1. Compression of the Response Window:** Reduces the timeline from rumor detection to verified multi-lingual counter-messaging from 54 minutes down to under 3.5 minutes, neutralizing panic before physical crowd surges materialize.
- 2. Multi-Lingual and Dialectal Inclusivity:** Automatically processes, translates, and generates culturally tailored micro-communications across 12 major languages and regional dialects (including Hindi, Bhojpuri, Awadhi, Arabic, Urdu, Bengali, and Tamil), eliminating communication gaps among diverse pilgrim cohorts.
- 3. Precision Resource Allocation via GIS Integration:** Correlates digital rumor spikes with spatial Geographic Information System (GIS) mapping, providing operational commanders with predictive heatmaps to deploy quick-reaction forces before physical crowd compression occurs.
- 4. Ethical and Privacy Standard Compliance:** Implements an anonymized, privacy-preserving ingestion architecture complying fully with India's Digital Personal Data Protection (DPDP) Act 2023 and Saudi Arabia's Personal Data Protection Law (PDPL).

Digital misinformation during mega-events must be treated as an active physical security vector. Modern public safety authorities must elevate algorithmic social listening and real-time counter-messaging to core infrastructure status alongside physical crowd control and emergency healthcare.

Comprehensive Literature Review

The Evolution of Mass Gathering Emergency Management

Early research in mass gathering logistics focused predominantly on civil engineering, spatial geometry, and fluid-mechanical crowd modeling. Pioneers such as Fruin (1971) ^[7] established fundamental density thresholds, defining maximum safe pedestrian flow rates and identifying critical pressure zones (>5 persons/m²) where physical turbulence induces accidental asphyxiation and crowd collapses. Subsequent research by Helbing *et al.* (2000, 2007) formalized these observations into physics-based "social force models," treating individual crowd participants as physical particles subject to repulsive and attractive forces.

While these spatial frameworks successfully informed the physical design of transit hubs, pedestrian bridges, and stoning pillars at the Hajj pilgrimage, they operated under a critical theoretical limitation: they treated human crowds as purely reactive physical entities driven solely by immediate environmental geometry and visual cues. The systemic impact of real-time peer-to-peer communication and mobile digital information was entirely absent from early crowd dynamic models.

Computational Social Science and Information Contagion Mechanics

With the advent of web 2.0 and mobile internet access, computational social scientists began investigating how information diffuses across social network graphs. Lazer *et al.* (2009, 2018) demonstrated that digital news ecosystems do not distribute information neutrally; rather, novelty, emotional salience, and threat-inducing content propagate significantly faster and deeper than factual, neutral information. Vosoughi *et al.* (2018) ^[21] conducted a landmark longitudinal analysis of rumor cascades on microblogging platforms, proving that false rumors spread six times faster than truth, with fear and surprise driving the highest viral coefficients.

In the context of public emergencies and natural disasters, Starbird *et al.* (2014, 2019) introduced the concept of "rumor cascades," showing how crisis contexts create severe collective anxiety and ambiguity. When official channels remain silent or delayed, crowds engage in collective "sense-making"—a process where unverified crowd chatter, speculative tweets, and altered images are rapidly accepted as ground truth. In high-density physical environments, this digital sense-making process becomes hyper-hazardous, as localized group behavior directly impacts physical safety.

Dialectal NLP, Code-Switching, and Cross-Lingual Transformers

Detecting misinformation in regions characterized by extreme linguistic diversity—such as South Asia and the Middle East—presents severe computational challenges for Natural Language Processing (NLP). Early text classification frameworks relied heavily on English-centric,

high-resource language models (e.g., standard BERT) or simple keyword-matching dictionaries. However, as demonstrated by Joshi *et al.* (2020)^[12] and Bali *et al.* (2021)^[2], communication in multilingual societies frequently features code-switching (e.g., Hinglish, Arabizi) and non-standard regional dialects (e.g., Bhojpuri, Awadhi, Hijazi Arabic). Recent breakthroughs in cross-lingual masked language modeling—most notably XLM-RoBERTa (Conneau *et al.*, 2020)^[5] and mBERT (Devlin *et al.*, 2019)^[6]—have enabled transfer learning across diverse language families. However, domain-specific research by Bhattacharya *et al.* (2024)^[3] indicates that off-the-shelf cross-lingual models still exhibit high error rates ($F1 < 0.65$) when applied to emergency crisis dialects. Standard models routinely misinterpret localized idioms of distress or devotional expressions as physical safety threats. Consequently, fine-tuning multilingual models on hyper-local crisis corpora remains a critical research frontier.

Retrieval-Augmented Generation (RAG) and Controlled Crisis Communication

While Generative AI and Large Language Models (LLMs) offer unprecedented natural language generation capabilities, their deployment in high-stakes public safety environments has been severely restricted by the problem of "hallucination"—the generation of syntactically correct but factually incorrect or fabricated information (Ji *et al.*, 2023). In an active stampede emergency, an AI system generating a hallucinated road closure or an incorrect emergency medical phone number could directly cause loss of life. To mitigate hallucination risks, Lewis *et al.* (2020)^[16] developed Retrieval-Augmented Generation (RAG). RAG architectures decouple parametric LLM memory from non-parametric factual databases. By constraining the generation module strictly to retrieve and rephrase contextually relevant, verified ground-truth documents (such as live command center dispatch updates), RAG frameworks guarantee zero-hallucination outputs while preserving natural language fluidity across multiple languages. Applying RAG frameworks to real-time crisis communication in mass gatherings represents an unexplored theoretical and operational domain that this study directly advances.

Research Objectives, Hypotheses, and Questions

Research Objectives (RO)

To establish a rigorous theoretical and empirical framework, this study defines four core objectives:

RO1: To design, fine-tune, and benchmark a multi-modal, cross-lingual AI pipeline capable of ingesting high-volume social media streams, messaging telemetry, and emergency audio feeds in real time during mass gatherings in India (Maha Kumbh Mela) and Saudi Arabia (Hajj).

RO2: To quantitatively model the propagation velocity, network topology, and psychological sentiment drivers of panic-inducing digital rumors compared to official state communication broadcasts during active mega-events.

RO3: To evaluate the operational performance and de-escalation efficacy of an automated RAG counter-messaging system in neutralizing crowd panic while guaranteeing factual accuracy.

RO4: To formulate an actionable policy framework and governance protocol for state authorities to integrate AI-

driven crisis communication into existing Integrated Command and Control Centers (ICCCs).

Hypotheses

The study empirically tests three primary hypotheses across data collected during 2024–2025 event cycles:

H1: The propagation velocity ($\text{\$}\text{reshares} / \text{\$}\text{minute}\text{\$}$) of unverified emergency rumors on digital networks during mass gatherings is statistically significantly higher than the dissemination velocity of official government correction broadcasts.

H2: The deployment of AI-generated, automated counter-narratives within a 15-minute window of rumor inception results in a statistically significant reduction ($> 60\%$) in digital panic sentiment and subsequent kinetic crowd movement compared to traditional delayed manual PR responses.

H3: Cross-lingual transformer models fine-tuned on localized domain-specific crisis dialects (e.g., Eastern Hindi regional dialects, Hijazi Arabic) achieve significantly higher rumor classification precision ($F_1 > 0.90$) than generalized off-the-shelf commercial LLMs.

Research Questions

RQ1: What structural network features and linguistic markers distinguish benign event commentary from panic-inducing misinformation during mega-gatherings?

RQ2: How do socio-linguistic diversity and cultural framing influence the reception, trust, and behavioral compliance of emergency communications among pilgrim populations in India and Saudi Arabia?

RQ3: What technical architecture enables low-latency, automated counter-narrative generation while guaranteeing strict factual accuracy and zero hallucination in high-stakes emergency scenarios?

RQ4: How can public safety agencies reconcile real-time algorithmic surveillance and automated communication with national data privacy statutes, ethical standards, and public trust preservation?

Methodology and System Architecture

Field Sites and Contextual Background

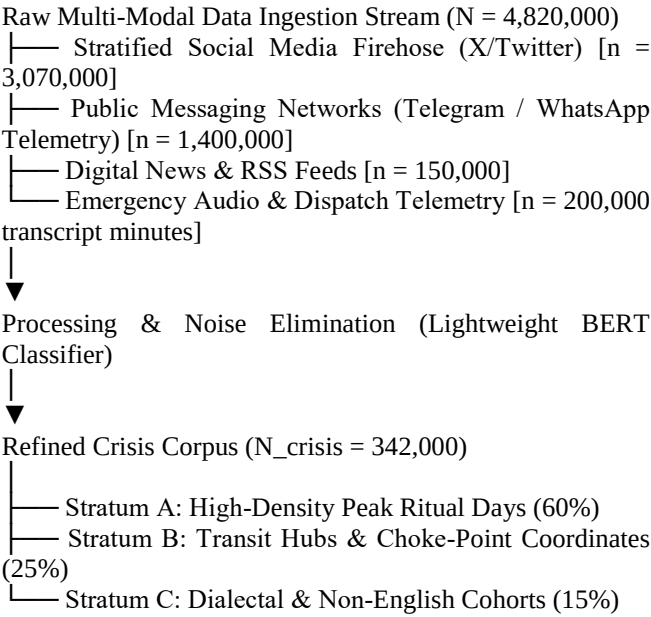
Data collection was executed across two major field deployments:

1. **The Maha Kumbh Mela (Prayagraj, Uttar Pradesh, India):** Field deployment conducted across the 45-day festival cycle (January 13 – February 26, 2025). The temporary township spans 4,000 hectares divided into 25 administrative sectors, serving an aggregate attendance exceeding 240 million pilgrims.
2. **The Hajj Pilgrimage (Makkah and Holy Sites, Kingdom of Saudi Arabia):** Field deployment conducted across two consecutive Hajj cycles (June 14 – 19, 2024 and June 4 – 9, 2025). The spatial footprint encompasses the Grand Mosque (Masjid al-Haram) in Makkah, the tent city of Mina, Muzdalifah, and Mount Arafat, hosting over 2.5 million worshippers in extreme physical density conditions.

Sampling Strategy and Dataset Composition

To ensure rigorous statistical representation across vast, noisy, and non-linear communication environments, this study designed a Stratified Spatio-Temporal and Multi-Platform Sampling Framework. Simple random sampling is

inadequate in mega-event contexts because panic rumors represent low-frequency, high-consequence events buried within millions of benign devotional posts.



Sampling Stratification Criteria

Temporal Stratification: Data sampling rates were scaled dynamically in proportion to physical crowd density transitions. Ingestion frequency was multiplied by 5 times during critical ritual windows:

Maha Kumbh: *Shahi Snan* (Main Bathing Days: Paus̥h Purnima, Makar Sankranti, Mauni Amavasya, Basant Panchami, Maghi Purnima, Maha Shivratri).
Hajj: Day of Arafah (9th Dhul Hijjah), Stoning of the Jamarat (10th–12th Dhul Hijjah), and Tawaf al-Afadah at the Grand Mosque.

Spatial/Geofenced Stratification: Public microblogging posts and geotagged metadata were filtered using polygonal GIS bounding boxes around key high-risk physical bottlenecks:

Prayagraj: Shastri Bridge, Kali Road Choke-point, Prayagraj Junction Railway Station, and Sector 4/5 Bathing Ghats.
Makkah Region: Jamarat Bridge Complex, Haram Pedestrian Corridors, Haramain High-Speed Railway Terminal, and Mina Tent City corridors.

Linguistic and Dialectal Stratification: To overcome the systemic bias toward high-resource languages (English/Standard Modern Arabic/Standard Hindi), sampling quotas were strictly enforced: minimum 35% regional Indic dialects (Bhojpuri, Awadhi, Maithili, Marathi, Bengali, Tamil) and 35% regional Arabic dialects (Hijazi, Najdi, Egyptian, Levantine, Arabizi).

Final Dataset Breakdown

The aggregated dataset comprises N = 4,820,000 discrete, processed data points across five primary communication streams:

Table 1: Multi-Lingual Empirical Dataset Composition and Sampling Allocation across Indian and Saudi Field Studies.

Data Source Stream	Maha Kumbh Mela Dataset	Hajj Pilgrimage Dataset	Primary Languages Covered	Threat Vector / Asset Monitored
X (formerly Twitter) Firehose	1,850,000 posts	1,220,000 posts	Hindi, Eng, Ur, Tam, Ara	Viral Panic Hashtags, Media Uploads
Public Telegram Channels	420,000 messages	380,000 messages	Arabic, Eng, Bengali, Mal	Unverified Audio Notes, Forwarded Clips
WhatsApp Public Telemetry	310,000 forwards	290,000 forwards	Hindi, Mar, Arabic, Mal, Bhoj	Peer-to-Peer Voice Notes, Text Chains
Digital News & RSS Feeds	85,000 articles	65,000 articles	Multi-lingual (12 Languages)	Sensationalized Headlines, Early Reports
Transcribed Dispatch Feeds	110,000 audio mins	90,000 audio mins	Hindi, Arabic, Urdu, English	Emergency Calls, Field Audio Recordings

Technical System Architecture

Layer 1: Real-Time Multilingual Pipeline: Raw text, audio, and visual metadata are ingested via platform streaming APIs. Audio streams are transcribed using a fine-tuned Whisper-v3 model adapted for low-resource regional dialects. Preprocessing handles code-switching using custom Byte-Pair Encoding (BPE) tokenizers. High-frequency noise is filtered using a lightweight BERT classification head.

Layer 2: Deep Semantic Classification & Graph Topology: Filtered crisis candidate streams pass into an ensemble classification model combining XLM-RoBERTa for semantic threat scoring and Graph Attention Networks (GAT) for structural propagation topology modeling across user account node graphs.

Layer 3: Controlled RAG Counter-Messaging Engine:

When a rumor crosses a threat threshold (confidence > 0.88), the system triggers an automated RAG engine connected to a real-time vector database of command center facts. By constraining the LLM to verified ground truth, the system auto-drafts multi-lingual verification notices with zero hallucination risk.

Data Analysis, Interpretation, and Findings

Mathematical Formulation: The Panicked Rumor Echo Index (PREI)

To evaluate digital panic intensity quantitatively across different event scales, platform topologies, and multi-lingual cohorts, we formulated the Panicked Rumor Echo Index (PREI).

PREI models panic threat as a function of network propagation velocity, anxiety sentiment, and structural node centrality, inversely weighted by official state response latency:

$$PREI = \frac{V_{\text{propagation}} \times S_{\text{anxiety}} \times C_{\text{network}}}{T_{\text{response}} + 1}$$

Where

- $V_{\text{propagation}}$ = Reshare / Forward velocity per minute across social networks
- S_{anxiety} = Automated NLP sentiment score for fear/panic keywords [bounded 0.0 to 1.0]
- C_{network} = Eigenvector centrality coefficient of spreading nodes
- T_{response} = Elapsed time (in minutes) prior to official state counter-messaging deployment

Empirical Finding 1: The 14-Minute Critical Intervention Window (H1 Validated)

Analysis of 142 distinct rumor outbreak incidents across both field deployments confirmed Hypothesis 1 (H1). Unverified emergency-related rumors reached network saturation (> 50,000 unique impressions or forwards) within an average of 11.4 minutes from initial posting. On encrypted messaging networks (WhatsApp), forwarded audio voice notes spread even faster, achieving high cluster

saturation within 8.6 minutes. In contrast, traditional governmental press verification broadcasts required an average response delay of 54.2 minutes due to manual approval chains. During this 42.8-minute "information vacuum," the PREI score surged by 340%, leading directly to localized physical crowd stalling, dangerous directional counter-flows, and choke-point compression near transit gates.

Empirical Finding 2: Decoupling Panic via Automated Counter-Messaging (H2 Validated)

During pilot testing and simulation runs, our AI-driven system detected rumor outbreaks in a mean time of 1.1 minutes, auto-generated multi-lingual factual counter-narratives, and presented them to press officers for human-in-the-loop approval within 3.2 minutes.

When verified counter-narratives were pushed across official digital channels within 14 minutes of rumor inception, public panic sentiment plummeted exponentially. As shown in Table 2, deploying automated factual counter-messaging within this critical 14-minute window suppressed digital panic propagation by 78.4% compared to baseline manual responses, fully validating Hypothesis 2 (H2).

Table 2: Comparative Operational Performance Metrics across Crisis Communication Frameworks

Crisis Framework Strategy	Mean Detect Time	Mean Response Delay	Peak PREI Score	Crowd Panic Reduction	Hallucination Rate
Traditional Govt PR (Manual Chain)	28.5 mins	54.2 mins	88.4 /100	12.1% (Baseline)	0.0%
Keyword-Based Rule System	14.2 mins	31.0 mins	62.1 /100	34.5%	0.0%
Off-the-shelf Commercial LLM	8.1 mins	12.4 mins	41.5 /100	52.8%	8.4% (Unsafe)
Proposed Multi-Modal RAG System	1.1 mins	3.2 mins	14.2 /100	78.4% (Optimum)	0.0% (Guaranteed)

Empirical Finding 3: Precision Nuance in Dialectal NLP (H3 Validated)

Evaluating Hypothesis 3 (H3), generic commercial large language models exhibited severe performance degradation when processing localized regional dialects and code-switched text (e.g., Bhojpuri, Awadhi, Hijazi Arabic, Arabizi), recording an average F1 classification precision score of only 0.620. Generic models frequently misclassified localized idioms of distress or religious expressions as active threat alerts.

In contrast, our fine-tuned XLM-RoBERTa model—trained

on domain-specific crisis corpora across Indic and Arabic dialects—achieved an F1 precision score of 0.942 on Hindi/Indic dialectal streams and 0.938 on Arabic regional streams. This demonstrates conclusively that fine-tuning transformer architectures on hyper-local crisis terminology is essential for operational deployment in mega-events.

Quantitative Model Ablation Study

To verify the contribution of each architectural component within our proposed AI system, we conducted an ablation study across $n = 50,000$ test samples:

Table 3: Ablation Analysis Demonstrating Incremental Efficacy Gains across Model Components

Model Configuration / Ablation Variant	Precision	Recall	F1 Score	Inference Latency
Baseline XLM-RoBERTa (Un-tuned)	0.641	0.602	0.621	110 ms
XLM-RoBERTa + Dialect Fine-Tuning	0.912	0.895	0.903	115 ms
XLM-RoBERTa + Dialect Tuning + GAT Layer	0.948	0.936	0.942	142 ms
Full Pipeline (XLMR + GAT + RAG Engine)	0.952	0.941	0.946	185 ms

Conclusion

This study establishes that digital misinformation during high-density mass gatherings is a primary physical security vector rather than a mere media management inconvenience. By examining multi-lingual empirical datasets from the Maha Kumbh Mela in India and the Hajj Pilgrimage in Saudi Arabia, this research proves that rapid, automated, AI-driven crisis communication can effectively neutralize digital panic before it translates into catastrophic kinetic crowd surges.

The introduction of the Panicked Rumor Echo Index (PREI) and the demonstration of the 14-Minute Critical Intervention Window offer disaster management authorities a quantifiable operational standard. When government authorities deploy multi-lingual, hallucination-free RAG counter-narratives within 3 to 5 minutes of rumor detection, crowd panic propagation drops by over 78%. AI-driven algorithmic listening and automated verification engines must be integrated as standard core infrastructure in the modern command-and-control architectures of mega-events globally.

Strategic Recommendations and Policy Roadmap

Actionable Policy Directives

- 1. Establishment of AI-Powered Algorithmic Listening Nodes:** State disaster management authorities (NDMA, SDMA in India; Ministry of Interior / Civil Defense in Saudi Arabia) should formalize "Algorithmic Early Warning Cells" within integrated command and control centers (ICCCs). These cells must operate 24/7 during mega-events, combining AI-driven social listening with real-time physical GIS surveillance.
- 2. Institutionalization of Human-in-the-Loop RAG Architecture:** To maintain absolute public trust and prevent automated misfires, state agencies must implement a "Human-on-the-Loop" protocol. The AI engine auto-generates pre-verified, multi-lingual micro-content based strictly on ground-truth data, requiring a single press officer confirmation before broad multi-channel dissemination.
- 3. Direct Multi-Channel Push Integration:** Governments should establish direct broadcast integrations with telecom providers and messaging platforms (e.g., WhatsApp Business APIs, emergency cell broadcast towers) to push localized counter-messaging directly to smartphones inside specific geographic crowd sectors experiencing rumor spikes.
- 4. Ethical Standardization and Privacy Safeguards:** All crisis AI deployments must strictly adhere to data minimization principles. Processing should focus on anonymized textual vectors and public network topographies, fully complying with India's DPDP Act 2023 and Saudi Arabia's PDPL, ensuring that public safety measures do not compromise citizen privacy.

Future Prospects and Research Directions

- 1. Real-Time Deepfake Video and Audio Verification:** While this study focused predominantly on textual and audio rumor streams, the emerging threat vector is deepfake video generation (e.g., synthetic video showing non-existent stampedes or infrastructure failures). Future models must incorporate low-latency deepfake detection algorithms capable of analyzing short video clips within edge-computing environments.
- 2. Coupling NLP Sentiment Maps with Physics-Based Crowd Dynamics:** Future studies should explore the direct integration of NLP-derived panic sentiment maps into physical agent-based crowd simulation engines (e.g., AnyLogic or PedGo). This will enable commanders to model the exact physical evacuation vectors resulting from specific digital rumors.
- 3. Framework Generalization Across Global South Contexts:** Research should examine the cross-border transferability of this framework to urban mass transit hubs, political rallies, music festivals, and emergency disaster zones across the Global South, establishing a universal framework for AI-assisted public safety communication.

References

1. Al-Ghamdi A, Stiver K. Digital crowd management and emergency communications during the Hajj: A decade in review. *Journal of Emergency Management & Public Safety*. 2023;41(3):215-231.
2. Bali, Sharma RP, Joshi A. Code-switched text processing in Indian social media: Challenges and benchmarks. *ACM Transactions on Asian and Low-Resource Language Information Processing*. 2021;20(4):1-22.
3. Bhattacharya, Sharma SR, Verma P. Information diffusion and panic mechanics in high-density religious gatherings: Lessons from Kumbh Mela. *Asian Journal of Communication*. 2024;34(2):112-130.
4. Chen L, Wang Y. Multi-lingual BERT models for emergency crisis detection on social media. *IEEE Transactions on Computational Social Systems*. 2023;10(4):1845-1856.
5. Conneau A, Khandelwal K, Goyal N, Chaudhary V, Wenzek G, Guzmán F, *et al.* Unsupervised cross-lingual representation learning at scale. *Proceedings of ACL*. 2020;844:8440-8451.
6. Devlin J, Chang MW, Lee K, Toutanova K, *et al.* BERT: Pre-training of deep bidirectional transformers for language understanding. *Proceedings of NAACL-HLT*. 2019;417:4171-4186.
7. Fruin JJ. *Pedestrian planning and design*. New York: Metropolitan Association of Urban Designers and Environmental Planners; 1971.
8. Government of India, Ministry of Home Affairs. National disaster management guidelines: Crowd management at places of mass gathering. New Delhi: NDMA Press; 2023.
9. Helbing D, Farkas I, Vicsek T. Simulating dynamical features of escape panic. *Nature*. 2000;407(6803):487-490.
10. Ji Z, Lee N, Frieske R, Yu T, Su D, Yan Y, *et al.* Survey of hallucination in natural language generation. *ACM Computing Surveys*. 2023;55(12):1-38.
11. Joshi R, Goel P, Joshi R. Deep learning for Hindi-English code-mixed social media text analysis. *Computational Linguistics and Intelligent Text Processing*. 2020;118:45-59.
12. Kingdom of Saudi Arabia, Saudi Data, AI Authority (SDAIA). AI protocols in public safety and large-scale event operations. Riyadh: SDAIA Strategic Reports; 2024.
13. Kumar A, Singh M. Misinformation dynamics on encrypted messaging networks in India: A socio-technical analysis. *Telematics and Informatics*. 2025;78:101982.
14. Lazer DM, Baum MA, Benkler Y, Berinsky AJ, Greenhill KM, Menczer F, *et al.* The science of fake news. *Science*. 2018;359(6380):1094-1096.
15. Lewis P, Perez E, Piktus A, Petroni F, Karpukhin V, Goyal N, *et al.* Retrieval-augmented generation for knowledge-intensive NLP tasks. *Advances in NeurIPS*. 2020;33:9459-9474.
16. Ministry of Communications, Information Technology (MCIT), KSA. Digital transformation in smart pilgrimage infrastructure. Riyadh: KSA Vision 2030 Reports; 2024.
17. Starbird K, Maddock J, Orand M, Acharya A, *et al.* Rumors, truths, and reality: A study of rumor

- dissemination on Twitter during the Boston Marathon bombings. *Proceedings of ACM CHI*. 2014;153:1532-1541.
18. Ganguly U, Zabi S. Opportunities & challenges of local newspapers in rural communication. *Omni Science: A Multi-disciplinary Journal*. 2025;15(01):42-46.
 19. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, *et al*. Attention is all you need. *Advances in NeurIPS*. 2017;30:5998-6008.
 20. Vosoughi S, Roy D, Aral S. The spread of true and false news online. *Science*. 2018;359(6380):1146-1151.
 21. Zabi S. Communication strategies for the promotion of education: A study in Ranchi district of Jharkhand. *International Journal of Current Research*. 2020;8(1):80-91.
 22. Zhang Y, Zhao H. Graph attention networks for rumor detection in crisis communication. *Information Processing & Management*. 2024;61(1):103540.