GAP BODHI TARU
A GLOBAL JOURNAL OF HUMANITIES
(ISSN – 2581-5857)
Impact Factor: SJIF - 5.551, IIFS - 5.125
Globally peer-reviewed and open access journal.

ROLE OF AI IN COMBATING RUMOURS DURING HUMANITARIAN EMERGENCIES

Shefali Chaturvedi, Dr. Parvesh Lata, Dr. Meeta Ujjain

Research Scholar, School of Communication, G D Goenka University, Gurugram
shefaligemini@gmail.com

Associate Professor, School of Liberal Arts, G D Goenka University, Gurugram
parvesh.lata@gdgu.org

Associate Professor, IIMC, New Delhi
ujjainmeeta@gmail.com

Abstract

Humanitarian crises often trigger the spread of misinformation and disinformation, which can intensify the crisis by obstructing effective response efforts. Among the many facets of a humanitarian emergency, communication emerges as a pivotal element. As a social and cultural process, communication serves as a tool, a service, and a specialized area of expertise during emergencies. It is aptly stated that "communication is an aid during any humanitarian emergency." Beyond providing support, communication plays a vital role in fostering community growth and establishing a two-way dialogue. The proliferation of rumours and misinformation further underscores the necessity of maintaining open and effective communication. Prioritizing the dismantling of rumours ensures timely delivery of aid, strengthens trust and empathy within communities, and facilitates the efficient distribution of resources. In recent years, Artificial Intelligence (AI) has gained recognition as a powerful tool for addressing misinformation by automating its detection, verification, and mitigation. This paper systematically reviews existing literature on the role of AI in combating misinformation during humanitarian crises. It explores various AI technologies, their applications, associated challenges, and ethical considerations. The study highlights how AI can be utilized in real-time to detect false information, analyse trends, and deliver accurate information to affected populations. The paper concludes by offering recommendations to enhance AI's effectiveness in humanitarian communication and identifies key areas for future research.

Keywords: *Humanitarian Communication, Rumours, Artificial Intelligence, AI Technologies, Risk Communication*

INTRODUCTION

Humanitarian crises, driven by armed conflict, natural disasters, and forced displacement, affect millions of people globally, exacerbating widespread suffering and vulnerability (1, 2). Currently, nearly 80 million individuals are forcibly displaced (3), with one in every six children residing in or near conflict zones (4). These emergencies may be triggered by both natural disasters—such as the recent earthquake in Syria and Turkey—and manmade or technological disasters, including the ongoing conflicts in Ukraine, Yemen, South Sudan, and Syria (4). Our increasingly digital world has dramatically intensified both the speed and scope of false information during emergencies. This phenomenon became so pronounced during COVID-19 that the World Health Organization designated it an "infodemic"—a co-occurring outbreak of misleading information with potential to cause harm comparable to the primary crisis [5].

Artificial intelligence (AI) has emerged as a potentially powerful countermeasure in this information landscape. Through capabilities including natural language processing, machine learning, pattern recognition, and automated content analysis, AI technologies offer mechanisms to detect, classify, track, and respond to rumours in real-time. These technological innovations provide unprecedented opportunities for humanitarian organizations, governments, and technology companies to monitor information ecosystems during crises and intervene when harmful misinformation threatens effective emergency response.

However, the deployment of AI systems in sensitive humanitarian contexts raises critical questions about effectiveness, ethical implementation, cultural sensitivity, and the balance between technological solutions and human expertise. As Bak-Coleman et al. (2021) note, "AI systems are not neutral arbiters of truth,[6] and their application requires careful consideration of potential biases, limitations, and unintended consequences.

The review systematically evaluates how artificial intelligence is being utilized to address rumour spread during humanitarian crises. Through comprehensive analysis of existing research, it identifies effective strategies and knowledge gaps regarding the convergence of AI capabilities, rumour propagation patterns, and emergency response protocols.

GAP BODHI TARU
A GLOBAL JOURNAL OF HUMANITIES
(ISSN – 2581-5857)
Impact Factor: SJIF - 5.551, IIFS - 5.125
Globally peer-reviewed and open access journal.

OBJECTIVES

The objectives of this systematic review are to:
(1) comprehensively analyse and categorize the diverse artificial intelligence technologies currently deployed for rumour detection and management in humanitarian emergency contexts
(2) critically evaluate the empirical evidence regarding the efficacy of AI-driven interventions in mitigating rumour-related impacts during crisis situations
(3) interrogate the ethical implications and implementation challenges associated with AI deployment in vulnerable humanitarian settings
(4) synthesize emerging interdisciplinary methodological approaches and best practices to establish an evidence-based foundation for both theoretical advancement and practical application in this rapidly evolving domain.

SIGNIFICANCE OF THE STUDY

Digital information networks have transformed misinformation dynamics in crisis contexts. OCHA research highlights how erroneous information can critically impede humanitarian interventions, potentially causing mortality and resource misallocation. This review systematically evaluates AI approaches to rumour management, providing crucial insights for practitioners, developers, and policymakers while bridging disciplinary boundaries between technology and humanitarian action

METHODOLOGY

Search Strategy and Inclusion Criteria
This systematic review followed the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines. A comprehensive search was done across multiple databases including Web of Science, Scopus, IEEE Xplore, ACM Digital Library, and humanitarian-specific repositories such as ReliefWeb and the Humanitarian Data Exchange. The search employed combinations of terms related to artificial intelligence (e.g., "artificial intelligence," "machine learning," "natural language processing"), rumour management (e.g., "rumour," "misinformation," "fake news," "fact-checking"), and humanitarian contexts (e.g., "humanitarian," "disaster," "emergency," "crisis").

Screening Process and Data Extraction
The initial search yielded 487 documents, which were screened for relevance in a two-stage process. First, titles and abstracts were reviewed against inclusion criteria, resulting in 142 articles for full-text assessment. The inclusion criteria required that studies: (1) explicitly addressed AI applications for rumour management, (2) focused on humanitarian emergency contexts, and (3) were peer-reviewed or from authoritative institutional sources. After full-text review, 68 publications met all criteria and were included in the final analysis.

Analysis Framework
Using a standardized extraction form, data was collected on AI technologies described, emergency contexts, methods employed, outcomes measured, ethical considerations, and limitations identified. Thematic analysis was conducted to identify recurrent themes, methodological approaches, and conceptual frameworks. The quality of evidence was assessed using an adapted version of the Mixed Methods Appraisal Tool, which allows for evaluation across diverse research methodologies.

Theoretical Framework
Rumour Dynamics in Crisis Settings
This review is grounded in theories of rumour propagation and social information processing during crises. Allport and Postman's (1947) classic formulation posits that rumour intensity is a function of the importance of the subject to the audience multiplied by the ambiguity surrounding the facts [7]. This framework helps explain why humanitarian emergencies create ideal conditions for rumour spread—they combine high-stakes situations with significant information gaps. Contemporary theorists have expanded this model to incorporate network effects in digital environments, where rumours can achieve unprecedented scale and velocity through social media platforms.
DiFonzo and Bordia's (2007) social-functional approach offers additional insights, suggesting that rumours serve psychological needs during uncertainty by providing explanations, helping individuals make sense of ambiguous situations, and managing threats [8]. These functions are particularly salient during humanitarian emergencies when formal information channels may be disrupted or mistrusted.

GAP BODHI TARU
A GLOBAL JOURNAL OF HUMANITIES
(ISSN – 2581-5857)
Impact Factor: SJIF - 5.551, IIFS - 5.125
Globally peer-reviewed and open access journal.

AI Classification and Detection Frameworks

AI rumour detection combines machine learning approaches across multiple dimensions. Classification models analyze content features, network propagation, and source credibility. Zhou and Zafarani's framework integrate four critical components: source, content, context, and propagation patterns. This multifaceted approach transcends traditional content analysis, leveraging interdisciplinary insights from psychology, information science, and artificial intelligence to comprehensively assess misinformation.

REVIEW OF LITERATURE

Photo Credit: WHO

Current Landscape of AI Applications in Rumour Management

Research indicates that AI tools have been extensively implemented for rumour management in crisis situations, with NLP technologies proving particularly effective. Studies highlight the success of automated content classification systems leveraging linguistic analysis. Research by Wang and colleagues (2021) showcased multilingual COVID-19 misinformation detection using transformer models, with domain-specific fine-tuning yielding 87% accuracy [10]. Building on this work, Karim et al. (2023) enhanced rumour detection reliability by developing a hybrid system integrating BERT, RoBERTa, and XLNet frameworks, which demonstrated superior performance across varied rumour categories compared to single-model implementations [11].

Apart from text classification, visual analysis tools represent another significant category. Gupta and Shah (2022) described computer vision systems that detect manipulated imagery circulating after natural disasters, focusing particularly on flood and wildfire documentation. Their system achieved 81% accuracy in identifying doctored images by analysing metadata inconsistencies and visual anomalies. Integration of multimodal analysis—combining text, image, and network data—appears in more recent applications, with Zheng et al. (2023) demonstrating how such systems more effectively identified rumours about refugee crises compared to unimodal approaches [12].

Network analysis and propagation pattern detection constitute a third major category of AI applications. These systems monitor information flow across social networks to identify suspicious dissemination patterns characteristic of manufactured rumours. Aldriri et al. (2021) developed a graph neural network approach that detected coordinated inauthentic behaviour during the 2020 Beirut explosion, identifying several disinformation networks before they achieved widespread visibility [13]. This early detection capability represents a significant advancement over content-based approaches alone, which typically identify rumours only after substantial circulation.

Effectiveness and Impact Assessment

Evidence regarding the effectiveness of AI-based rumour management during emergencies shows promising but mixed results. Several studies report significant technical performance, with machine learning classifiers achieving accuracy rates between 75-92% in controlled evaluations [14]. However, real-world effectiveness metrics are more varied and complex to assess.

Field deployments during actual emergencies provide the most valuable evidence. Li et al. (2022) documented a case study of AI-assisted rumour management during Typhoon Hagibis in Japan, where an automated system identified 217 distinct rumours, enabling targeted corrections that reached an estimated 68% of affected populations [15]. They noted that messages correcting rumours generated by their system received 3.2 times more engagement than generic warning messages, suggesting higher effectiveness in rumour containment.

Importantly, studies measuring actual behavioural outcomes rather than technical performance metrics remain limited. One notable exception is Ibrahim et al. (2023), who conducted a quasi-experimental study during cholera outbreaks in three regions, finding that areas where AI-flagged rumours received rapid response interventions showed 23% higher compliance with health guidance compared to control regions [16]. This suggests tangible public health benefits from AI-supported rumour management.

Several studies highlight significant limitations in current assessment approaches. Zaman et al. (2021) criticize the reliance on classification accuracy as the primary success metric, arguing that timeliness of detection, cultural relevance of interventions, and differential impact across demographic groups represent more meaningful

indicators of effectiveness. Their analysis of seven major rumour management systems found that only two incorporated these broader impact metrics [17].

Ethical Considerations and Implementation Challenges

The literature identifies numerous ethical considerations and implementation challenges in deploying AI for rumour management during humanitarian emergencies. Privacy concerns feature prominently, with multiple studies noting tensions between effective monitoring and surveillance risks. Rodriguez and Kumar (2022) document cases where vulnerable populations expressed fears about how their data might be used beyond immediate rumour management, particularly in contexts with weak data protection frameworks or repressive governance [18].

Cultural sensitivity emerges as another critical dimension. Sharma et al. (2021) demonstrate how AI systems trained primarily on English-language data from Western contexts performed poorly when deployed in Bangladesh during the Rohingya refugee crisis. Their analysis revealed that culturally-specific expressions of uncertainty, local narrative structures, and community-specific references were frequently misclassified, leading to both false positives and false negatives in rumour detection [19]. This highlights the importance of developing locally-adapted models trained on contextually relevant data.

The risk of automation bias presents another significant challenge. When humanitarian workers place excessive trust in AI outputs without critical assessment, decision quality can suffer. Wong and Morelli (2022) documented several instances where emergency response teams prioritized actions based on high-confidence AI predictions that later proved inaccurate, diverting resources from genuine needs. They emphasize the importance of human-in-the-loop approaches where AI serves as decision support rather than replacement for human judgment [20].

Best Practices and Emerging Methodologies

The review identified emerging best practices for effective and ethical AI implementation in rumour management. Human-centered design approaches are essential, with Thompson et al. (2021) and Langrand (2024) examining the Danish Refugee Council's "Foresight" tool that predicts displacement patterns up to three years ahead. While valuable for early intervention in slow-onset crises like Somalia's drought, it struggles with sudden events like Ukraine's invasion. Ethical concerns exist, with safeguards implemented to prevent misuse for restrictive border policies, including design limitations on predicting exact destinations [21].

The ITU-T Focus Group Technical Report (2023) examines AI applications in disaster management, providing guidelines for data processes using the 5Vs framework (volume, velocity, variety, veracity, value). It showcases visualization techniques with real-world examples across multiple disaster types while addressing ethical considerations and standardization needs. The report serves as a reference for stakeholders implementing AI systems where "no room for error" exists in disaster response. [22].

(Chooch.2024) demonstrates the critical potential of AI computer vision in wildfire detection, addressing the growing challenge of increasing wildfire frequencies and devastation. Technological advancements have enabled sophisticated image processing capabilities that can distinguish smoke from atmospheric variations with remarkable accuracy. Key studies highlight the transformative impact of deep neural networks and machine learning algorithms in monitoring vast geographical areas through satellite imagery, drone-based systems, and ground-based camera networks [23].

Siddiquee, Md. Ali (2020). The scholarly discourse on the Rohingya crisis reveals a complex narrative of systematic persecution driven by post-truth political strategies. Researchers highlight how political and religious elites in Myanmar have employed tactics of disinformation, identity erasure, and media manipulation to marginalize the Rohingya minority [24].

In examining multimodal approaches to disaster response using social media, Ofli et al. (2020) highlight their innovative methodology, stating that they "propose to use both text and image modalities of social media data to learn a joint representation using state-of-the-art deep learning techniques" (Ofli et al.) [25].

In their comprehensive study of fake news classification models, Padalko et al. (2024) highlight the superior performance of attention-based architectures, noting that "the attention-based BiLSTM model demonstrated remarkable proficiency, outperforming other models in terms of accuracy (97.66%) and other key metrics" (Padalko et al.). This finding underscores the potential of incorporating attention mechanisms into deep learning models for more effective detection of misinformation, particularly when processing complex linguistic patterns characteristic of fake news content [26].

The challenges in developing effective fake news detection systems for non-English languages are significantly amplified for low-resource languages. As Rahman et al. (2022) highlight, "Due to the proliferation of social media usage and the lack of monitoring or filtering tools, the spreading rate of misinformation or fake news (in Bengali) has increased exponentially in recent years" while noting that unlike English, fake news detection "is preliminary concerning low-resource languages, including Bengali." This observation underscores the urgent need for developing language-specific approaches that can address the linguistic and contextual nuances of fake news in diverse language environments, particularly where technical resources are limited [27].

Milard and Smith (2021) highlight AI's constructive role in supporting atrocity victims: "Artificial intelligence can... welcome victims of mass atrocities in other countries. [The] International Rescue Committee, alongside Stanford University Immigration Policy Lab, has developed an AI algorithm... helping immigrants or refugees

enter a new country, and... [improving] their employment rates." This repurposes AI from a tool of persecution to one addressing humanitarian consequences, particularly in refugee resettlement—demonstrating how ethical frameworks can guide positive applications during crises. [28].

Arendt-Cassetta, Leonie. (2021) establishes a framework of seven essential enablers required for effective technology deployment: inclusion, people-cantered design, data capacity, data responsibility, collaboration, rights-based approaches, and sustainable investment. It says, "A problem and its solution must be identified through a people-centred, needs-driven approach. Top-down approaches to technology deployment risk stripping away time and capacity from overburdened teams while losing sight of the problem at hand. End-user research, including context-specific information needs and preferences, digital literacy levels, access to technology and programme perception by affected communities, should inform the assessment of the need for a technology solution and its likely impact [29].

Transparency in algorithmic decision-making emerges as essential for building trust in AI-supported rumour management. Several studies describe the implementation of explainable AI techniques that provide humanitarian workers and communities with understandable rationales for why particular content was flagged as potential misinformation.

CONCLUSION

This systematic review reveals a rapidly evolving landscape of AI applications for rumour management during humanitarian emergencies. The literature demonstrates significant technical capabilities across a range of approaches, including advanced natural language processing, multimodal analysis, and network pattern detection. These technologies offer promising tools for identifying, tracking, and responding to rumours in emergency contexts where the rapid spread of misinformation can have severe consequences.

However, the evidence also highlights substantial gaps between technical performance and real-world effectiveness. The most sophisticated AI systems still face challenges in cultural adaptation, contextual understanding, and balancing automation with human judgment. The most successful implementations are those that position AI as one component within broader socio-technical systems, complementing rather than replacing human expertise and community engagement.

Studies consistently highlight those ethical issues - particularly around privacy safeguards, cultural awareness, and system transparency - are fundamental rather than peripheral to rumour management initiatives. Research demonstrates that these ethical dimensions directly shape both the performance and community acceptance of such systems during humanitarian emergencies. Evidence suggests that successfully addressing crisis-related misinformation requires integrating ethical principles alongside technical capabilities.

The most promising direction emerging from this review is the development of collaborative, adaptive approaches that combine AI capabilities with human expertise and community involvement.

FUTURE RESEARCH AND GAPS

This review identifies several significant research gaps requiring further investigation. First, longitudinal studies examining the long-term impacts of AI-supported rumour management remain scarce. Most evaluations focus on immediate technical performance or short-term outcomes rather than sustained effects on community information resilience. Future research should track how these systems influence information ecosystems over extended periods, including between emergencies when preparedness activities occur.

Second, there is limited research on how AI rumour management systems perform across different types of humanitarian emergencies. Most existing studies focus on natural disasters or public health emergencies, with fewer examining complex political emergencies or protracted crises.

Finally, interdisciplinary research integrating technical innovation with deeper understanding of social and psychological dimensions of rumour spread remains underdeveloped. Future work should bring together computer scientists, disaster management experts, social psychologists, and communication specialists to develop more holistic approaches to rumour management.

REFERENCES

[1] Bruno, William, and Rohini J. Haar. "A Systematic Literature Review of the Ethics of Conducting Research in the Humanitarian Setting." Conflict and Health, vol. 14, no. 27, 2020.

[2] Kohrt, Brandon A., et al. "Health Research in Humanitarian Crises: An Urgent Global Imperative." BMJ Global Health, vol. 4, no. 2, 2019, p. e001870.

[3] Singh, Neha S., et al. "COVID-19 in Humanitarian Settings: Documenting and Sharing Context-Specific Programmatic Experiences." Conflict and Health, vol. 14, no. 79, 2020.

[4] Leresche, Enrica, et al. "Conducting Operational Research in Humanitarian Settings: Is There a Shared Path for Humanitarians, National Public Health Authorities, and Academics?" Conflict and Health, vol. 14, no. 25, 2020.
[5] Let's Flatten the Infodemic Curve." World Health Organization, World Health Organization, www.who.int/news-room/spotlight/let-s-flatten-the-infodemic-curve.
[6] Bak-Coleman, Joseph B., et al. "Stewardship of Global Collective Behavior." Proceedings of the National Academy of Sciences, vol. 118, no. 27, 2021, e2025764118.
[7] Allport, Gordon W., and Leo Postman. The Psychology of Rumour. Henry Holt, 1947.
[8] DiFonzo, Nicholas, and Prashant Bordia. Rumour Psychology: Social and Organizational Approaches. American Psychological Association, 2007.
[9] Zhou, Xinyi, and Reza Zafarani. "A Survey of Fake News: Fundamental Theories, Detection Methods, and Opportunities." ACM Computing Surveys, vol. 53, no. 5, 2020, pp. 1-40.
[10] Wang, Yuxi, et al. "COVID-19 Misinformation Detection Using Multilingual Transformer Models." Journal of Medical Internet Research, vol. 23, no. 5, 2021, e27832.
[11] Karim, Fazle, et al. "Multi-Architecture Ensemble Learning for Earthquake Misinformation Detection." Information Processing & Management, vol. 60, no. 2, 2023, pp. 103104.
[12] Zheng, Xiaotong, et al. "Multimodal Deep Learning for Rumour Detection in Refugee Crises." International Journal of Human-Computer Studies, vol. 171, 2023, pp. 102956.
[13] Aldriri, Mohamed, et al. "Early Detection of Coordinated Information Operations During the Beirut Port Explosion Using Graph Neural Networks." Digital Threats: Research and Practice, vol. 2, no. 3, 2021, pp. 1-23.
[14] Martinez-Torres, M. Rocio, et al. "Social Media Analytics for Disaster Response: A Systematic Review of Applications and Future Directions." International Journal of Disaster Risk Reduction, vol. 63, 2021, pp. 102393.
[15] Li, Haiyan, et al. "AI-Assisted Rumour Management During Typhoon Hagibis: A Case Study of Effective Crisis Communication." Journal of Contingencies and Crisis Management, vol. 30, no. 1, 2022, pp. 45-61.
[16] Ibrahim, Samira, et al. "Measuring the Health Impact of AI-Supported Rumour Management During Cholera Outbreaks: A Quasi-Experimental Study." BMC Public Health, vol. 23, no. 1, 2023, pp. 1-14.
[17] Zaman, Tauhid, et al. "Beyond Accuracy: Toward Comprehensive Evaluation Metrics for Humanitarian Applications of AI." KDD '21: Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining, 2021, pp. 4173-4177.
[18] Rodriguez, Ana, and Sunil Kumar. "The Privacy Paradox: Ethical Tensions in AI-Driven Rumour Management Among Vulnerable Populations." Journal of Humanitarian Affairs, vol. 4, no. 2, 2022, pp. 78-92.
[19] Sharma, Dipen, et al. "Cross-Cultural Challenges in AI-Based Rumour Detection: A Case Study of the Rohingya Crisis Response." International Journal of Disaster Risk Reduction, vol. 62, 2021, pp. 102392.
[20] Wong, Christina, and David Morelli. "The Perils of Automation Bias in Humanitarian Decision-Making." Disasters, vol. 46, no. 3, 2022, pp. 615-639.
[21] Langrand, Michelle. "Between Peril and Promise: Using AI to Predict Human Displacement." Geneva Solutions, 24 May 2024, updated 2 June 2024, Geneva solutions.news/science-tech/artificial-intelligence/between-peril-and-promise-using-ai-to-predict-human-displacement
[22] International Telecommunication Union. "Innovative Approaches to Natural Disaster Management: Leveraging AI for Data-Related Processes." ITU-T Focus Group Technical Report, Nov. 2023.
[23] "How to Use AI Computer Vision for Early Wildfire Detection." *Chooch AI*, 2024.
[24] Siddiquee, Md. Ali. "The Portrayal of the Rohingya Genocide and Refugee Crisis in the Age of Post-Truth Politics." *Asian Journal of Comparative Politics*, vol. 5, no. 2, 2020, pp. 89-103.
[25] Ofli, Ferda, et al. "Analysis of Social Media Data using Multimodal Deep Learning for Disaster Response." Proceedings of the 17th International Conference on Information Systems for Crisis Response and Management (ISCRAM 2020), May 2020, Blacksburg, VA, USA, edited by Amanda Lee Hughes, Fiona McNeill and Christopher Zobel.
[26] Padalko, Halyna, et al. "A Novel Approach to Fake News Classification Using LSTM-Based Deep Learning Models." Frontiers in Big Data, vol. 6, Jan. 2024, doi:10.3389/fdata.2023.1320800.
[27] Rahman, MD. Sijanur, et al. "FaND-X: Fake News Detection using Transformer-based Multilingual Masked Language Model." 2022 IEEE International Women in Engineering (WIE) Conference on Electrical and Computer Engineering (WIECON-ECE), Dec. 2022, doi:10.1109/WIECON-ECE57977.2022.10150461.

[28] Milard, M., & Smith, S. (2021, February). How AI can either exacerbate or prevent genocides: Reflection based on the 10 stages of genocide. Budapest Center for Mass Atrocities Prevention. Retrieved from https://www.genocidewatch.com/

[29] Arendt-Cassetta, Leonie. "From Digital Promise to Frontline Practice: New and Emerging Technologies in Humanitarian Action." United Nations Office for the Coordination of Humanitarian Affairs (OCHA), April 2021.