KMANPUB | AI and Tech in Behavioral and Social Sciences Vol. 4 No. 4 (2026): 1-13

Article history:
Received 27 March 2026
Revised 24 June 2026
Accepted 02 July 2026
Initial Publication 08 July 2026
Final Publication 01 October 2026

An Applied Artificial Intelligence Model for Social Media-Based Natural Crisis Management in Iran

Parinaz. Mohajerani[1], Mahdi. Zare[2*], Akbar. Nasrollahi Kasmani[3]

[1] PhD Student in Media Management, CT.C., Islamic Azad University, Tehran, Iran
[2] Professor; International Research Institute of Seismology and Earthquake Engineering, Tehran, Iran
[3] Associate Professor, Department of Communication, Journalism and Media, Faculty of Social Sciences, Communication and Media, CT.C., Islamic Azad University, Tehran, Iran

*** Corresponding author email address**: mzare@iiees.ac.ir

Article Info

Article type:
Original Research

How to cite this article:
Mohajerani, P., Zare, M., & Nasrollahi Kasmani, A. (2026). An Applied Artificial Intelligence Model for Social Media-Based Natural Crisis Management in Iran. *AI and Tech in Behavioral and Social Sciences, 4*(4), 1-13.
https://doi.org/10.61838/kman.aitech.5739

ABSTRACT

Natural hazards in Iran frequently generate complex communication demands because earthquakes, floods, droughts, and other disasters unfold in environments where social media platforms rapidly become both information resources and rumor amplifiers. This study developed an applied model for the use of artificial intelligence (AI) in social media for natural crisis management in Iran, with particular attention to detecting, analyzing, and countering misinformation. The article is based on the findings of a mixed-method, exploratory-sequential study conducted in five consecutive phases. First, semi-structured interviews with 16 experts in crisis management, media, artificial intelligence, and relief operations were analyzed thematically, producing nine major themes and ten key factors; instrument validity was confirmed through CVR and CVI, and reliability was supported by Cohen's kappa of 0.847. Second, interpretive structural modeling (ISM) with an eight-member expert panel identified an integrated crisis data monitoring center and transparent privacy legislation as foundational drivers, while reducing response time emerged as the ultimate system outcome. Third, a quantitative analysis of 384 social media messages from the Varzaghan-Ahar earthquake, the Khoy earthquake, and the Lorestan flood showed that 16.9% of messages contained misinformation and more than 45% of rumors appeared during the first six hours of crisis. Fourth, six AI algorithms were evaluated; ParsBERT performed best for Persian text with 88.2% test-set accuracy and an F1 score of 85.7%, CNN achieved 90.0% accuracy in the reported image test subset, and RNN-LSTM reached 83.3% accuracy in the reported video test subset. Finally, the findings were integrated into a five-layer operational model consisting of data, preprocessing, intelligent analysis, human verification and feedback, and decision/action layers. The model specifies three major objectives: preventing rumors and misinformation, reducing response time, and strengthening community resilience. The findings indicate that AI-supported social media monitoring can improve crisis communication, but implementation requires technical, ethical, legal, and social safeguards.

Keywords: *artificial intelligence; crisis communication; misinformation detection; natural disasters; social media; community resilience; Iran*

1. Introduction

Natural hazards are no longer only physical events; they are also information events. Earthquakes, floods, droughts, storms, and landslides generate immediate needs for rescue, coordination, trust, and public guidance. At the same time, they create intense uncertainty, emotional pressure, and fragmented information flows. In this environment, social media platforms have become central spaces where citizens, journalists, volunteers, relief organizations, and public authorities produce and circulate crisis-related information. During the first hours of a disaster, posts, images, short videos, hashtags, and messages may identify affected areas, missing persons, urgent needs, road blockages, damaged infrastructure, and public emotions faster than many official channels can respond.

The value of social media in emergencies has been widely recognized in crisis informatics. Studies of disaster communication have shown that social media can support situational awareness, citizen reporting, resource mobilization, volunteer coordination, and public accountability (Alexander, 2014; Palen & Anderson, 2016; Reuter & Kaufhold, 2018). Crowdsourced geographic information and cross-crisis communication studies further show that citizen-generated data can improve situational mapping and emergency interpretation when verification mechanisms are available (Goodchild & Glennon, 2010; Olteanu et al., 2015).

Research on online misinformation has demonstrated that false information can spread faster and more widely than verified information, partly because novel and emotionally charged claims attract attention and sharing (Vosoughi et al., 2018). In disaster contexts, this dynamic becomes more dangerous. A rumor about a collapsed dam, an exaggerated death toll, an incorrect aftershock warning, or a false message about relief distribution can directly affect public behavior. The distinction between misinformation, disinformation, and malinformation is analytically useful: misinformation may be false but not necessarily intentional, disinformation is deliberately deceptive, and malinformation may use genuine information in harmful ways (Wardle & Derakhshan, 2017). Crisis management systems must be able to recognize all three types, while avoiding overcorrection that suppresses useful citizen-generated information.

Artificial intelligence provides new possibilities for managing these complex information environments. Natural language processing can classify messages, extract locations, detect requests for help, identify rumors, and recognize emotional signals. Computer vision can evaluate images and videos for damage assessment, duplicate detection, and content verification. Machine learning and deep learning models can identify patterns in large volumes of noisy data that would be impossible to process manually. AI systems such as AIDR have already demonstrated the feasibility of combining machine learning with human computation for humanitarian response (Imran et al., 2014). More recent advances in transformer-based language models, including BERT and Persian-specific models such as ParsBERT, have expanded the capacity to analyze language-sensitive crisis content (Devlin et al., 2019; Farahani et al., 2020).

Iran is a highly relevant context for developing an applied AI model for social media-based natural crisis management. The country is exposed to multiple hazards, including destructive earthquakes, seasonal floods, droughts, landslides, and climate-related extreme events. Events such as the Varzaghan-Ahar earthquake, the Khoy earthquake, and the Lorestan flood illustrate both the vulnerability of communities and the importance of timely, accurate, and trusted information. Persian-language social media ecosystems also have distinctive linguistic, cultural, and platform-specific characteristics. Informal Persian, code-switching, sarcasm, local dialects, Telegram channels, Instagram posts, and X/Twitter hashtags all create challenges for automated analysis. Therefore, models developed for English-language crisis communication cannot be transferred directly without local adaptation.

The central problem addressed in this article is the absence of a comprehensive, context-sensitive, and operational model for using AI in social media to support natural crisis management in Iran. Existing international research has produced valuable tools and concepts, but many studies remain either technically narrow or insufficiently adapted to local governance, language, ethics, and crisis communication systems. Conversely, domestic crisis communication studies often identify the importance of social media and misinformation but do not develop an integrated AI-based architecture. This study responds to that gap by presenting an applied model that integrates expert knowledge, structural modeling, social media data analysis, algorithmic testing, and operational design.

This study aims to develop and present a context-sensitive applied model for using artificial intelligence in social media-based natural crisis management in Iran. It

E-ISSN: 3041-9433

addresses three research questions: (1) What are the key organizational, legal, technical, and social factors affecting the use of AI in social media-based crisis management? (2) What patterns of misinformation are visible in Iranian natural crisis-related social media messages? (3) Which AI algorithms perform most effectively for text-, image-, and video-based misinformation detection, and how can these findings be integrated into an operational crisis-management model?

2. Literature Review

2.1. Social Media in Crisis Communication

Social media has changed the temporal structure of crisis communication. In earlier models of disaster communication, official authorities, broadcast media, and emergency organizations were the primary producers of public information. Today, crisis information is distributed across networks of citizens, local communities, journalists, volunteers, public agencies, and automated accounts. This distributed structure creates opportunities for faster information sharing but also makes verification more difficult. Reuter and Kaufhold (2018) described social media as a major component of crisis informatics, emphasizing that public communication during emergencies is no longer a one-directional process from authorities to citizens (Reuter & Kaufhold, 2018). Instead, crisis communication is interactive, participatory, and rapidly evolving.

Citizen-generated information is especially important during the early stages of disasters. Vieweg et al. (2010) showed that microblogged information during hazards can contain actionable data, including location-specific reports and situational updates (Vieweg et al., 2010). Gao et al. (2011) similarly emphasized the potential of crowdsourced social media data for disaster relief and situational mapping (Gao et al., 2011). Social media messages may reveal needs that are not yet visible through official channels; however, these messages vary substantially in accuracy, relevance, completeness, and credibility. For this reason, emergency organizations require systems that can separate actionable information from noise without excluding valuable local knowledge.

The integration of social media into crisis management also raises institutional challenges. Emergency organizations may lack staff, technical infrastructure, and protocols for monitoring multiple platforms in real time. They may also face legal and ethical constraints concerning privacy, data retention, and public communication. In many contexts, delayed or unclear official communication creates an information vacuum that rumors rapidly fill. Therefore, social media monitoring must be embedded in an institutional framework that combines rapid analysis, transparent communication, human oversight, and public trust.

2.2. Misinformation, Rumors, and Trust During Disasters

Misinformation in disasters differs from misinformation in ordinary political or commercial communication because its consequences are immediate and operational. A false statement about the location of aid, a fabricated image of destruction, or an incorrect safety warning may directly influence movement, resource allocation, and emotional responses. Review studies indicate that fake news, disinformation, and misinformation on social media require both technical detection and social mitigation strategies (Aïmeur et al., 2023; Kumar & Shah, 2018).

Credibility assessment is a long-standing challenge in social media research. Castillo et al. (2011) showed that information credibility on Twitter can be estimated using features related to users, messages, topics, and propagation patterns (Castillo et al., 2011). However, crisis misinformation is dynamic: the truth value of a claim may change as new evidence emerges, and early reports may be incomplete rather than intentionally false. This means that an AI system should not only label information as true or false, but should also support verification workflows, uncertainty labels, confidence scores, and human review.

From a public communication perspective, trust is central. If citizens perceive official channels as slow, opaque, or politically motivated, they may rely more heavily on informal networks. Conversely, if official agencies provide rapid, specific, and verifiable updates, misinformation may lose influence. AI can assist by detecting misinformation early, but it cannot replace the legitimacy of transparent communication. Therefore, technological systems must be connected to trusted institutions and to clear public information strategies.

2.3. Artificial Intelligence for Crisis Data Analysis

AI methods have been widely applied to crisis-related data processing. Natural language processing can classify crisis messages into categories such as requests for help, infrastructure damage, eyewitness reports, emotional

AITBSS AI and Tech in Behavioral and Social Sciences
E-ISSN: 3041-9433

reactions, and misinformation. Machine learning models can detect anomalies, cluster topics, rank urgent messages, and support triage. Sentiment-oriented big-data analytics has also been used to support disaster response and recovery by identifying public concerns and emotional patterns (Ragini et al., 2018). Imran et al. (2014) reviewed the computational processing of social media messages in mass emergencies and identified several core tasks, including filtering, classification, extraction, summarization, and credibility assessment (Imran et al., 2014). These tasks remain central to modern AI-supported crisis systems.

Deep learning has expanded this capacity by allowing systems to learn complex representations from text, image, and video data. Transformer-based models such as BERT improved performance across many language understanding tasks by using contextualized word representations (Devlin et al., 2019). For Persian-language content, ParsBERT provides a more culturally and linguistically appropriate foundation than general multilingual models because it is trained on Persian corpora and better captures Persian morphology, syntax, and vocabulary (Farahani et al., 2020). For visual crisis data, convolutional neural networks can classify images and detect damage patterns, while recurrent neural networks and LSTM architectures can model sequential information in video and temporal data.

Nevertheless, AI systems in disasters face serious limitations. Crisis data are noisy, multilingual, emotionally charged, incomplete, and often geographically ambiguous. Labelled datasets are limited, especially for Persian-language crisis misinformation. Algorithms may reproduce bias, misinterpret sarcasm, over-rely on superficial cues, or fail when new types of misinformation emerge. For these reasons, human verification and feedback are not optional additions but central components of any responsible AI-based crisis management model.

2.4. Research Gap and Conceptual Rationale

The literature indicates that social media can improve situational awareness and coordination, but it also exposes crisis management systems to misinformation. AI can process large volumes of data, but algorithmic tools need institutional integration, ethical safeguards, and local linguistic adaptation. Existing studies often address these issues separately. Technical studies may evaluate classification models without considering the governance system in which they will operate. Communication studies may describe rumor patterns without testing AI-based detection models. Policy studies may recommend digital transformation without specifying operational layers or validation criteria.

The present study fills this gap by integrating five evidence sources: qualitative expert interviews, interpretive structural modeling, quantitative analysis of crisis-related social media messages, algorithmic performance testing, and model synthesis. This design allows the final model to be simultaneously empirical, technical, and operational.

3. Methods and Materials

The study employed an applied-developmental mixed-method design using an exploratory sequential logic. The design was organized into five consecutive phases: qualitative exploration, interpretive structural modeling, quantitative analysis of social media messages, algorithmic testing, and final model integration. This design was selected because the research problem required both conceptual exploration and empirical evaluation. The qualitative phase clarified the dimensions of the problem, the structural modeling phase identified relationships among key factors, the quantitative phase described patterns in crisis-related social media data, and the algorithmic phase tested the feasibility of AI-supported misinformation detection.

In the first phase, semi-structured interviews were conducted with 16 experts from four specialized areas: crisis management, media and communication, artificial intelligence, and relief operations. Participants were selected purposively on the basis of professional experience, direct relevance to crisis communication or AI, and willingness to participate. The interview guide covered misinformation challenges, social-media monitoring, AI applications, privacy and ethics, and operational requirements. Interviews were transcribed, coded by two researchers, and analyzed through thematic analysis following Braun and Clarke's six-phase logic (Braun & Clarke, 2006). Instrument validity was assessed using CVR and CVI, and inter-coder reliability was confirmed using Cohen's kappa coefficient of 0.847. Disagreements were resolved through discussion and supervisory review, and qualitative trustworthiness was considered through credibility, transferability, dependability, and confirmability criteria (Lincoln & Guba, 1985).

E-ISSN: 3041-9433

In the second phase, interpretive structural modeling was used to identify the structural relationships among key factors extracted from the qualitative phase. An eight-member expert panel participated in the development of the structural self-interaction matrix, reachability matrix, level partitioning, and MICMAC analysis. ISM was appropriate because the study sought not only to list factors but also to understand which factors drive the system and which represent dependent outcomes.

In the third phase, 384 public social-media items were sampled from three major natural-crisis cases: the Varzaghan-Ahar earthquake, the Khoy earthquake, and the Lorestan flood. The sample size was determined using Cochran's formula. Duplicate, irrelevant, and non-crisis-related items were excluded. Messages were coded using a codebook derived from the qualitative phase and were analyzed for misinformation prevalence, rumor timing, hashtag patterns, and spatial distribution. Publicly available metadata were anonymized before analysis. TF-IDF analysis of hashtags was conducted in Python, and heat-map analysis was used to examine spatial patterns.

In the fourth phase, six algorithms were evaluated for misinformation detection: Random Forest, Naive Bayes, linear SVM, ParsBERT, CNN/ResNet-50, and RNN-LSTM. The 384-message sample was used for descriptive content analysis, whereas algorithmic testing used modality-specific labelled subsets for text, image, and video data. Within each subset, data were divided into training and testing partitions, and the confusion matrices reported in the Results section refer only to the held-out test cases. Accuracy, precision, recall, F1 score, and processing speed were used as performance metrics. In the fifth phase, results from all previous phases were integrated into a comprehensive applied model consisting of five operational layers and four implementation-consideration domains.

Table 1

Overview of the Five-Phase Research Design

Phase	Purpose	Participants/Data	Analytical Procedure
1. Qualitative exploration	Identify themes and key factors	16 experts in crisis management, media, AI, and relief	Semi-structured interviews; thematic analysis; CVR/CVI; Cohen's kappa
2. ISM modeling	Structure relationships among factors	8-member expert panel	SSIM, reachability matrix, level partitioning, MICMAC
3. Social media analysis	Measure misinformation and rumor patterns	384 messages from three Iranian natural crises	Cochran sampling; TF-IDF; heat-map analysis
4. Algorithmic testing	Evaluate AI detection performance	Text, image, and video crisis data	Comparative testing of six algorithms
5. Model integration	Develop final applied model	Integrated findings from phases 1–4	Synthesis into layered operational model

4. Findings and Results

The preliminary qualitative content analysis of crisis-related social media data showed that Iranian crisis messages follow relatively identifiable linguistic and content patterns. Help-seeking messages were usually short, unstructured, and emotionally marked. Rumors and fake news showed recurring patterns, including false certainty, unidentified sources, numerical exaggeration, and the use of old or out-of-context images.

In the first phase, semi-structured interviews with 16 experts and thematic analysis produced nine main themes and ten key factors. In the second phase, Interpretive Structural Modeling and MICMAC analysis showed that establishing an integrated crisis data monitoring center and addressing the absence of transparent privacy legislation were foundational drivers, while reducing response time emerged as the final system outcome. In the third phase, 384 messages from the Varzaghan-Ahar earthquake, the Khoy earthquake, and the Lorestan flood were analyzed. Overall, 16.9% of the messages contained misinformation, and more than 45% of rumors were disseminated during the first six hours of crisis. In the fourth phase, algorithmic testing was interpreted separately for each data type because the text, image, and video subsets had different test-set sizes.

E-ISSN: 3041-9433

Table 2

Performance of classical machine-learning algorithms in rumor detection

Algorithm	Accuracy (%)	Precision (%)	Recall (%)	F1 score (%)	Processing speed (ms/message)
Random Forest (RF)	81.2	78.5	76.3	77.4	8
Naive Bayes (NB)	72.8	68.4	70.1	69.2	4
Linear SVM	83.5	80.2	79.8	80.0	10

As shown in Table 2, SVM achieved the best performance among the classical algorithms, with 83.5% accuracy. Random Forest ranked second with 81.2% accuracy, and Naive Bayes ranked third with 72.8% accuracy. Although these classical algorithms were faster than ParsBERT, their lower accuracy restricts their use in applications where maximum reliability is required.

Table 3

Confusion matrices of the best-performing algorithms in the test sets

Algorithm / data type	True negative (TN)	False positive (FP)	False negative (FN)	True positive (TP)
ParsBERT / text (76 messages)	41	5	4	26
CNN / images (20 images)	11	1	1	7
RNN-LSTM / video (12 videos)	6	1	1	4
SVM / text (44 messages)	23	3	5	13

Table 3 presents the confusion matrices used to clarify the classification errors of the algorithms. Based on the reported test-set counts, ParsBERT achieved 88.2% accuracy for text (67/76), CNN achieved 90.0% accuracy for images (18/20), RNN-LSTM achieved 83.3% accuracy for video (10/12), and SVM achieved 81.8% accuracy for the reported text test subset (36/44).

Table 4

Selected algorithms for each data type in the proposed model

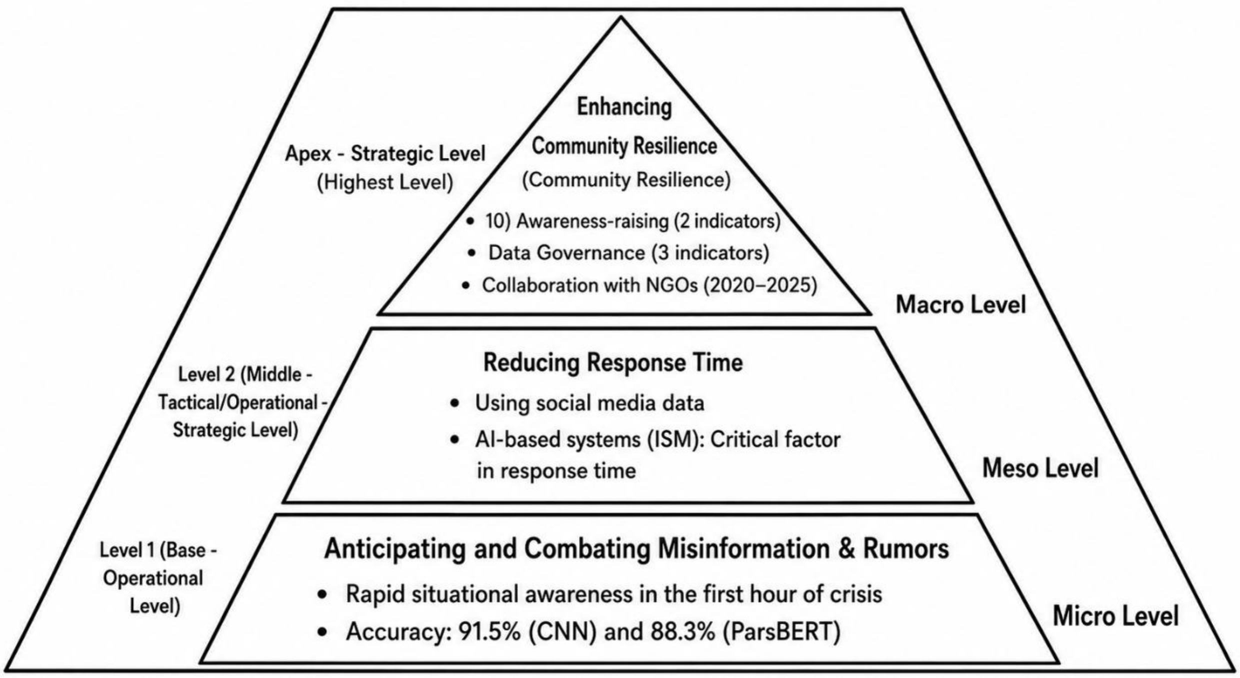
Data type	Selected algorithm	Key selection criterion	Operational implication
Text messages	ParsBERT	Highest text test-set accuracy (88.2%) and F1 score (85.7%)	Main model for textual rumor detection when accuracy is the priority
Text messages / high-speed mode	SVM	Acceptable test-set accuracy (81.8%) and faster speed (10 ms/message)	Alternative model when computational resources or time are limited
Images	CNN (ResNet-50)	Highest image test-set accuracy (90.0%)	Detection of old, fake, or misleading crisis images
Video	RNN-LSTM	Sequential frame analysis with 83.3% test-set accuracy	Detection of manipulated or misleading video content

As summarized in Table 4, the fourth phase therefore indicated that ParsBERT is the most appropriate algorithm for Persian textual data when accuracy is the priority, while SVM can be used as a faster text-processing alternative when lower accuracy is acceptable. For image data, CNN/ResNet-50 achieved the strongest reported result, and for video data, RNN-LSTM was selected because it can analyze sequential frame-level information, although its reported test-set accuracy should be interpreted cautiously because of the small video test set.

E-ISSN: 3041-9433

Figure 1

Hierarchy of the proposed model objectives (objective pyramid)



As shown in Figure 1, based on the integration of the second and third phases, the final model has three hierarchical objectives. The operational objective is prevention and countering of rumors and misinformation. The operational-strategic objective is reduction of crisis response time. The strategic objective is enhancement of community resilience, including informational, psychological, and structural resilience, which is consistent with the resilience-oriented logic of the Sendai Framework for Disaster Risk Reduction (United Nations Office for Disaster Risk, 2015).

Table 5

Acceptable performance thresholds for selected algorithms

Selected algorithm	Data type	Minimum acceptable accuracy	Minimum acceptable F1 score
ParsBERT	Text / accuracy priority	85%	82%
SVM	Text / speed priority	80%	78%
CNN (ResNet-50)	Image	88%	85%
RNN-LSTM	Video	80%	78%

Table 6

Acceptable processing-speed thresholds

Data type	Selected algorithm	Maximum acceptable processing time
Text	ParsBERT	50 ms/message
Text	SVM	15 ms/message
Image	CNN (ResNet-50)	100 ms/image
Video	RNN-LSTM	3 seconds/video

As shown in Tables 5 and 6, the proposed AI system should be evaluated using two main criteria: prediction accuracy and data-processing speed. Prediction accuracy was assessed through accuracy, precision, recall, and F1

AITBSS
E-ISSN: 3041-9433

score, while processing speed represents the time required to convert raw social-media data into an analytical output such as a warning, classification, or situational report.

Figure 2

Operational cycle of the proposed AI system

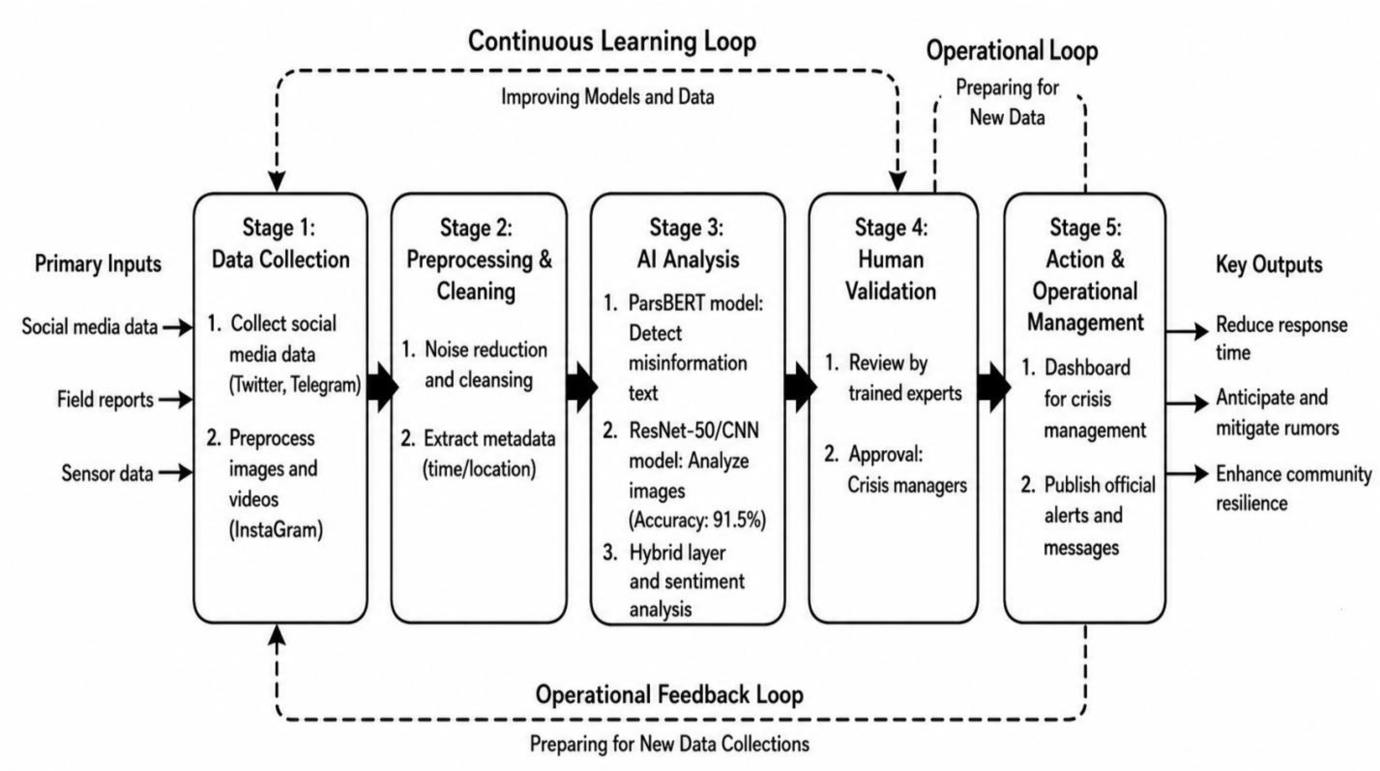


Table 7

Components and processes of the proposed model

Process	Input	Output	Responsible unit
Real-time monitoring	Raw social-media data	Initially filtered data	Multi-source data collector
Filtering and standardization	Raw and labelled data	Cleaned and standardized data	Data preprocessing engine
Intelligent analysis and classification	Preprocessed data	Classification output: real/false information	Misinformation detection engine
Human validation and feedback	Low-confidence cases	Verified data and AI feedback	Crisis expert and human-in-the-loop system
Alert and operational action	Final classification output	Alerts, public information, and operational recommendations	Decision dashboard and crisis managers

Figure 2 shows the internal mechanism through which crisis data are transformed into decisions. Table 7 summarizes the main components and responsible units. Before a crisis, historical data are monitored, labelled datasets are developed, algorithms are trained, and technical and human infrastructure is prepared. During a crisis, misinformation, fake news, old images, and manipulated videos are detected; sentiment is analyzed; content is classified; and rapid warnings are sent to crisis managers. After a crisis, user feedback and system performance are evaluated to improve future performance.

AITBSS
AI and Tech in Behavioral and Social Sciences
E-ISSN: 3041-9433

Figure 3

Comprehensive applied AI model in social media for natural crisis management

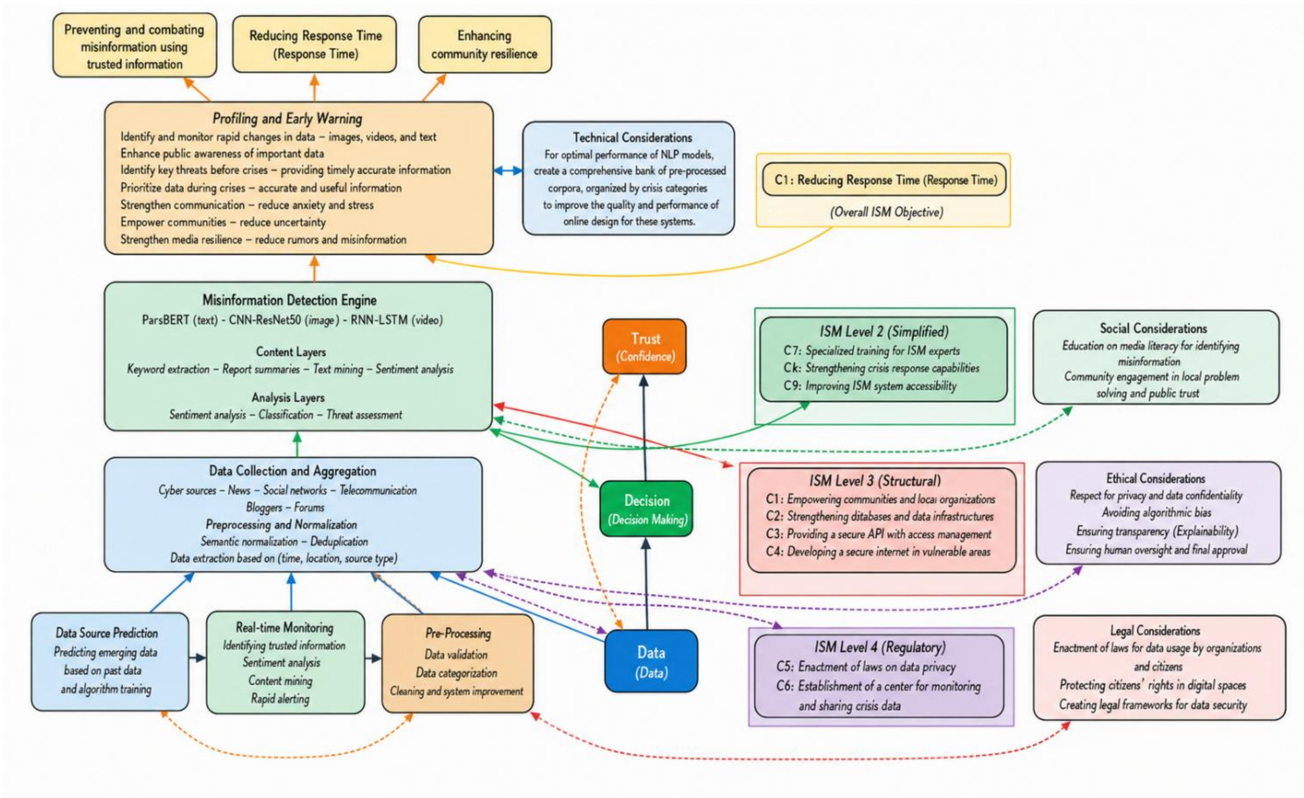


Table 8

Key considerations for implementing the proposed model

Consideration domain	Related issue	Operational recommendation
Technical	Language diversity, lack of labelled datasets, API restrictions, weak infrastructure, and cloud-processing requirements	Localize Persian NLP models for dialects; build labelled datasets from past crises; use domestic cloud infrastructure; design offline capacity for internet disruption
Social	Training needs, media literacy, public trust, and community resilience	Train citizens in media literacy; empower local journalists; increase public trust through transparent system outputs
Ethical	Privacy, anonymization, algorithmic bias, explainability, and human oversight	Use public and anonymized data only; avoid bias toward regions or groups; make AI decisions explainable; maintain human final review
Legal	Absence of transparent privacy rules, need for inter-agency protocols, and liability for system errors	Develop rules for emergency use of user data; create protocols among the Red Crescent, crisis-management bodies, broadcasters, and platforms; define liability for operational errors

The final model is presented at two complementary levels. At the operational level, it consists of five layers: data collection, preprocessing, intelligent analysis, human validation and feedback, and decision/action. At the conceptual level, Figure 3 summarizes the model through a three-layer DDT logic: Data, Decision, and Trust. The data layer includes multi-platform collection, preprocessing, noise reduction, and metadata extraction. The decision layer includes the misinformation detection engine, content classification, sentiment analysis, and human validation. The trust layer includes the expected outcomes of reduced rumor diffusion, increased public trust in official institutions, and strengthened social resilience. Table 8 summarizes the technical, social, ethical, and legal considerations required for implementation.

AITBSS
AI and Tech in Behavioral and Social Sciences
E-ISSN: 3041-9433

5. Discussion and Conclusion

The findings support the argument that AI-supported social media monitoring can strengthen natural crisis management, but only when implemented as part of a broader communication and governance system. The prevalence of misinformation in 16.9% of analyzed crisis messages confirms that misinformation is not a marginal issue. Its concentration in the first six hours is even more important because early rumor circulation overlaps with the most time-sensitive phase of emergency response. This finding is consistent with broader misinformation research showing that false information can diffuse rapidly in online networks (Vosoughi et al., 2018).

The superiority of ParsBERT for Persian text analysis is theoretically and practically important. Persian crisis communication contains informal expressions, local place names, spelling variation, and culturally specific rumor forms. General multilingual models may not adequately capture these features. The result therefore supports the use of language-specific transformer models for high-stakes crisis communication. However, a reported test-set accuracy of 88.2% also means that errors remain. In crisis management, even a small percentage of false positives or false negatives may have serious consequences. For this reason, the proposed model includes a human verification and feedback layer rather than a fully automated decision pipeline.

The high performance of CNN for image analysis suggests that visual data can play an important role in crisis verification. Images and videos often circulate faster than official damage assessments. They may reveal collapsed buildings, blocked roads, flooding, injured people, or aid distribution problems. However, visual misinformation is increasingly difficult to detect because old images may be recirculated, locations may be mislabelled, and AI-generated images may appear realistic. Therefore, computer vision tools should be combined with metadata analysis, reverse-image checks, geolocation, temporal verification, and human review.

The ISM finding that an integrated crisis data monitoring center is a root factor is one of the most important policy implications of the study. Fragmented monitoring by separate agencies may produce duplication, inconsistent messaging, and delays. A national or inter-agency monitoring center could integrate data from social media platforms, emergency hotlines, official reports, meteorological and seismological systems, and field teams. Such a center should not function as a censorship unit; rather, it should operate as a transparent public-interest infrastructure for verification, analysis, and communication.

Transparent privacy legislation emerged as another foundational factor. This is crucial because AI-supported monitoring can easily create public concern about surveillance, misuse of personal data, and political filtering. Disaster contexts may justify rapid analysis of public information, but they do not remove ethical obligations. The model therefore requires clear rules on data minimization, anonymization, purpose limitation, access control, auditability, and public accountability. Without these safeguards, technological systems may reduce rather than increase trust.

The model also emphasizes community resilience. Resilience is not only a technical property of infrastructure; it is also a communicative and social property. Communities are more resilient when they can access reliable information, correct rumors, coordinate local help, and maintain trust in response systems. AI can support resilience by identifying needs, detecting misinformation, and accelerating communication, but it cannot generate trust by itself. Trust depends on institutional transparency, consistent public messaging, and meaningful engagement with communities.

Compared with earlier crisis informatics frameworks, the proposed model contributes by integrating algorithmic performance results with structural, ethical, and operational considerations. Many AI studies focus on model accuracy, while many communication studies focus on platform behavior. The present model combines these perspectives and positions AI as one component within a five-layer crisis management architecture. This is especially important for developing contexts, where limited resources, fragmented governance, and data quality challenges make purely technical solutions insufficient.

6. Practical Implications

First, crisis management authorities should establish an integrated AI-supported social media monitoring center with clear legal authority, transparent governance, and inter-agency participation. The center should include experts in crisis management, data science, communication, ethics, law, and field relief operations. Its primary functions should be early rumor detection, situation monitoring, public communication support, and operational decision support.

Second, Persian-language crisis datasets should be developed and continuously updated. The lack of labelled Persian crisis data remains a major barrier to model improvement. Datasets should include text, images, videos, geolocation, rumor labels, credibility labels, and crisis categories. Data annotation should be performed by trained teams familiar with disaster terminology, Persian informal language, and local contexts.

Third, official crisis communication protocols should be connected to AI outputs. Detecting a rumor is not enough. Authorities must have procedures for verifying the claim, preparing a correction, publishing it through trusted channels, and monitoring whether the correction reduces misinformation. Public messaging should be rapid, specific, and evidence-based.

Fourth, the model should be implemented gradually through pilot projects. For example, a pilot system could first monitor text-based misinformation during flood alerts in one province before expanding to image and video analysis at the national level. Pilot evaluation should measure not only algorithmic accuracy but also response time, user trust, institutional usability, ethical compliance, and operational impact.

Fifth, public education and media literacy should be treated as part of crisis management. AI can identify misinformation, but citizens also need to know how to evaluate sources, avoid forwarding unverified claims, and rely on official updates during emergencies. Education campaigns should be designed before disasters occur, not only during crisis response.

7. Limitations and Future Research

This study has several limitations. First, the social media dataset consisted of 384 messages from three natural crises. Although the sample was adequate for the research design and provided meaningful insights, larger datasets would enable stronger statistical analysis and more robust algorithmic evaluation. Second, the study focused on Iranian crisis cases and Persian-language communication. The findings are contextually valuable but should be tested in other linguistic and cultural settings before broader generalization.

Third, the algorithmic evaluation used standard performance metrics such as accuracy and F1 score, but future studies should also assess latency, explainability, fairness, robustness to adversarial misinformation, and performance under real-time operational conditions. Fourth, ethical and legal considerations were identified conceptually, but future research should develop detailed governance protocols, data protection standards, and accountability mechanisms for AI-supported crisis monitoring. Finally, the rapid emergence of generative AI and synthetic media requires further investigation because future misinformation may increasingly include realistic fake images, videos, and audio.

8. Conclusion

This article presented an applied AI model for social media-based natural crisis management in Iran, focusing on the detection, analysis, and mitigation of misinformation. Based on a five-phase mixed-method study, the findings show that misinformation is a significant feature of crisis-related social media communication, with 16.9% of analyzed messages containing false information and more than 45% of rumors appearing within the first six hours of crisis. Algorithmic testing showed that ParsBERT performed best for Persian text, CNN for image data, and RNN-LSTM for video data. These results support the need for a multimodal AI system adapted to Persian-language and Iranian crisis contexts.

The final model consists of five operational layers: data, preprocessing, intelligent analysis, human verification and feedback, and decision/action. It also includes three objectives—preventing misinformation, reducing response time, and increasing community resilience—and two performance criteria: prediction accuracy and processing speed. The model demonstrates that AI can improve crisis management only when combined with integrated data governance, human oversight, ethical safeguards, legal clarity, and trusted public communication.

For Iran, the proposed model can inform the design of national systems for monitoring and analyzing social media data during natural crises. Organizations such as the Red Crescent, national crisis management bodies, emergency services, and public broadcasters could use the model as a framework for developing interoperable crisis information systems. More broadly, the study contributes to crisis informatics by showing how AI, social media analysis, and disaster governance can be integrated into a practical and locally adapted model.

Authors' Contributions

Conceptualization was performed by Parinaz Mohajerani and Mahdi Zare. Methodology was developed by Parinaz Mohajerani, Mahdi Zare, and Akbar Nasrollahi Kasmani. Data collection, data curation, formal analysis, visualization, and writing—original draft were conducted by Parinaz Mohajerani. Validation, writing—review and editing, and supervision were carried out by Mahdi Zare and Akbar Nasrollahi Kasmani. Project administration was led by Mahdi Zare. All authors contributed to the interpretation of the findings, critically reviewed the manuscript for important intellectual content, approved the final version, and agreed to be accountable for all aspects of the work.

Declaration

Artificial intelligence (AI)-assisted tools were used to improve the linguistic quality, readability, and grammatical accuracy of the manuscript. The authors retained full responsibility for the study design, data collection, data analysis, interpretation of the findings, and final content. All AI-assisted outputs were reviewed, verified, and edited by the authors before submission. No AI tool was used as an author of the manuscript.

Transparency Statement

Data are available for research purposes upon reasonable request to the corresponding author.

Acknowledgments

We would like to express our gratitude to all individuals helped us to do the project.

Declaration of Interest

The authors report no conflict of interest.

Funding

According to the authors, this article has no financial support.

Ethics Considerations

The study was conducted according to ethical principles for research involving human participants. Participants were informed about the purpose of the research, participation was voluntary, and informed consent was obtained before interviews. Confidentiality and anonymity were preserved during transcription, coding, and reporting. The research involved analysis of social media messages in aggregated form and did not report identifiable user information.

References

Aïmeur, E., Amri, S., & Brassard, G. (2023). Fake news, disinformation and misinformation in social media: A review. *Applied Sciences*, *13*(1), Article 30. https://doi.org/10.3390/app13010030

Alexander, D. E. (2014). Social media in disaster risk reduction and crisis management. *Science and Engineering Ethics*, *20*, 717-733. https://doi.org/10.1007/s11948-013-9502-z

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, *3*(2), 77-101. https://doi.org/10.1191/1478088706qp063oa

Castillo, C., Mendoza, M., & Poblete, B. (2011). Information credibility on Twitter. Proceedings of the 20th International Conference on World Wide Web,

Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. Proceedings of NAACL-HLT 2019,

Farahani, M., Gharachorloo, M., Farahani, M., & Manthouri, M. (2020). *ParsBERT: Transformer-based model for Persian language understanding* (arXiv, Issue. https://arxiv.org/abs/2005.12515

Gao, H., Barbier, G., & Goolsby, R. (2011). Harnessing the crowdsourcing power of social media for disaster relief. *IEEE Intelligent Systems*, *26*(3), 10-14. https://doi.org/10.1109/MIS.2011.52

Goodchild, M. F., & Glennon, J. A. (2010). Crowdsourcing geographic information for disaster response: A research frontier. *International Journal of Digital Earth*, *3*(3), 231-241. https://doi.org/10.1080/17538941003759255

Imran, M., Castillo, C., Diaz, F., & Vieweg, S. (2014). Processing social media messages in mass emergency: A survey. *ACM Computing Surveys*, *47*(4), Article 67. https://doi.org/10.1145/2771588

Kumar, S., & Shah, N. (2018). *False information on web and social media: A survey* (arXiv, Issue. https://arxiv.org/abs/1804.08559

Lincoln, Y. S., & Guba, E. G. (1985). *Naturalistic inquiry*. SAGE. https://doi.org/10.1016/0147-1767(85)90062-8

Olteanu, A., Vieweg, S., & Castillo, C. (2015). What to expect when the unexpected happens: Social media communications across crises. Proceedings of the ACM Conference on Computer Supported Cooperative Work & Social Computing,

Palen, L., & Anderson, K. M. (2016). Crisis informatics: New data for extraordinary times. *Science*, *353*(6296), 224-225. https://doi.org/10.1126/science.aag2579

Ragini, J. R., Anand, P. M. R., & Bhaskar, V. (2018). Big data analytics for disaster response and recovery through sentiment analysis. *International Journal of Information Management*, *42*, 13-24. https://doi.org/10.1016/j.ijinfomgt.2018.05.004

Reuter, C., & Kaufhold, M. A. (2018). Fifteen years of social media in emergencies: A retrospective review and future directions for crisis informatics. *Journal of Contingencies and Crisis Management*, *26*(1), 41-57. https://doi.org/10.1111/1468-5973.12196

United Nations Office for Disaster Risk, R. (2015). *Sendai Framework for Disaster Risk Reduction 2015-2030*.

https://www.undrr.org/publication/sendai-framework-disaster-risk-reduction-2015-2030

Vieweg, S., Hughes, A. L., Starbird, K., & Palen, L. (2010). Microblogging during two natural hazards events: What Twitter may contribute to situational awareness. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems,

Vosoughi, S., Roy, D., & Aral, S. (2018). The spread of true and false news online. *Science*, *359*(6380), 1146-1151. https://doi.org/10.1126/science.aap9559

Wardle, C., & Derakhshan, H. (2017). *Information Disorder: Toward an Interdisciplinary Framework for Research and Policy Making*. http://tverezo.info/wp-content/uploads/2017/11/PREMS-162317-GBR-2018-Report-desinformation-A4-BAT.pdf

AITBSS
AI and Tech in Behavioral and Social Sciences
E-ISSN: 3041-9433