International Journal of Disaster Risk Reduction 127 (2025) 105642

Contents lists available at ScienceDirect

International Journal of Disaster Risk Reduction

journal homepage: www.elsevier.com/locate/ijdrr

Large language model applications in disaster management: An interdisciplinary review

Fengyi Xu [a], Jun Ma [a,b,*], Nan Li [c,d], Jack C.P. Cheng [e]

[a] *Department of Urban Planning and Design, The University of Hong Kong, Hong Kong Special Administrative Region of China*
[b] *Urban Systems Institute, The University of Hong Kong, Hong Kong Special Administrative Region of China*
[c] *Department of Construction Management, Tsinghua University, Beijing 100084, China*
[d] *Hang Lung Center for Real Estate, Tsinghua University, Beijing 100084, China*
[e] *Department of Civil and Environmental Engineering, The Hong Kong University of Science and Technology, Hong Kong Special Administrative Region of China*

ARTICLE INFO

Keywords:
Disaster management
Large language models
Information processing
Multi-modal data fusion
Emergency response system

ABSTRACT

Disasters increasingly challenge urban resilience, demanding advanced computational approaches for effective information management and response coordination. This interdisciplinary review systematically assesses Large Language Model (LLM) applications in disaster management, analyzing 70 LLM-focused studies within the broader landscape of AI-driven disaster management. Our analysis establishes a phase-based framework spanning detection, tracking, analysis, and action, and reveals three critical gaps in current disaster management solutions: limited advancement beyond disaster response to include preparedness, recovery, and mitigation phases; insufficient integration across diverse stakeholder groups and available resources; and inadequate transformation of situation awareness data into actionable insights. Leveraging cross-modal semantic reasoning, knowledge graph-constrained entity extraction, and advanced code generation, LLMs are well positioned to overcome information ambiguity and verification challenges often encountered in rapidly evolving disaster contexts. These capabilities also enable automation in disaster investigation and communication, effectively orchestrating diverse analytical tools and resources. To harness these advantages and promote further progress, we introduce the “3M” framework for intelligent disaster information management: multi-modal data fusion for integrated assessment, multi-source information validation for robust truth-finding, and multi-agent collaboration in physical–virtual disaster systems. This framework provides a systematic foundation for advancing next-generation LLM-driven disaster management research and practice in increasingly complex contexts

Contents

* Correspondence to: Department of Urban Planning and Design, The University of Hong Kong, Room816, 8/F, Knowles Building, Pokfulam Road, Hong Kong Special Administrative Region of China.
E-mail address: junma@hku.hk (J. Ma).

https://doi.org/10.1016/j.ijdrr.2025.105642
Received 10 March 2025; Received in revised form 9 June 2025; Accepted 9 June 2025
Available online 26 June 2025
2212-4209/

1. Introduction

Natural hazards and extreme events, characterized by their complex spatiotemporal dynamics and cascading effects, pose unprecedented challenges to emergency management systems worldwide [1,2]. These events, ranging from climate-related disasters (e.g., hurricanes, floods, droughts) to geological phenomena (e.g., earthquakes, collapse, landslides), generate multifaceted impacts that threaten both human populations and critical infrastructure systems [3]. With global economic losses from natural disasters reaching $380 billion in 2023 [4], the imperative for effective disaster management has never been more critical. While rapid response remains crucial, as emphasized by Agency [5] that "a quick and overwhelming response is better than a well-planned and thought out response", contemporary disaster management faces fundamental challenges in implementing timely actions, particularly during the critical preparedness phase where complex scenario prediction is essential. Emergency evacuations exemplify this complexity, requiring simultaneous consideration of multiple variables including evacuation route characteristics, dynamic driving conditions, specialized needs of vulnerable populations, and complex social interactions among evacuees [6–8]. These intertwined factors, compounded by changing climate patterns, distinctive local environmental conditions, and varying vulnerability profiles across populations, introduce significant uncertainties into disaster modeling and response planning.

To address these multidimensional uncertainties in disaster preparedness and response, researchers have increasingly integrated artificial intelligence (AI) techniques – particularly machine learning and deep learning – to model non-linear relationships and ambiguous scenarios. Recent applications demonstrate this trend across diverse hazard types: physical information-fused deep learning for flood dynamics prediction [9], comparative analyses of neural network architectures for wildfire susceptibility modeling [10], and vision transformer applications for earthquake detection and magnitude estimation [11]. Despite their analytical capabilities, these traditional AI approaches face two fundamental limitations in operational disaster contexts [12]. First, they require substantial high-quality training data, yet disaster events are relatively rare and unevenly documented compared to continuous environmental parameters like temperature or moisture—creating inherent constraints in both predictive accuracy and spatial–temporal resolution. Second, conventional machine learning models excel at predefined tasks with structured data formats but demonstrate limited flexibility in generating contextualized, personalized outputs – such as localized wildfire response recommendations that integrate real-time event analysis – in adaptable formats that permit specific refinements. These limitations significantly constrain the effectiveness of traditional AI models in facilitating real-time, intelligent decision-making during disaster emergencies.

To elevate the effectiveness of disaster management systems, it is essential to pursue advancements in disaster information systems that enable the acquisition of fine-grained, real-time situational awareness during unfolding disaster events. This situational intelligence not only enhances the quality of training data necessary for sophisticated AI prediction models but also facilitates timely decision-making. A critical component of this approach is the integration of digital twins technology into disaster management, as articulated by Ford and Wolf [13] and Fan et al. [14], plays a critical role, as it leverages real-time data and simulations to into iterative feedback loops and leads to more resilient response to disasters [15]. However, achieving effective disaster situation awareness is complicated by the inherent limitations of current monitoring approaches and the challenges of integrating diverse information sources. Contemporary disaster monitoring employs various technological solutions, including specialized environmental sensors such as street inundation sensor [16] and global remote sensing systems [17]. However, these technologies face significant operational constraints: sensor networks often have limited coverage and reliability in extreme conditions, while remote sensing systems, despite their broad adoption, encounter challenges in adverse weather conditions and struggle to detect localized incidents such as small-scale landslides. Consequently, comprehensive situation awareness continues to heavily depend on human-driven investigation processes, though this approach presents challenges in information collection, analysis, and dissemination across diverse stakeholders [18].

In response to disaster coordination challenges, governments worldwide have developed increasingly sophisticated assessment frameworks and communication protocols, evolving from hierarchical processes—such as the four-step method of measuring, reporting, verifying, and evaluating disasters [19]—to integrated systems like the European Union's Common Emergency Communication and Information System (CECIS) that facilitates cross-border information sharing [20]. Despite these institutional advancements, significant challenges persist in the timely extraction and effective synthesis of critical information across jurisdictional and organizational boundaries. These integration difficulties are particularly pronounced when managing multiple complementary information streams, which are essential for comprehensive disaster situation awareness [21]: institutional channels comprising government surveys and emergency response teams' field reports, intermediary sources including authorized technical evaluations and media coverage, and public-sourced information through both structured reporting systems and spontaneous communications [22].

Four critical challenges persist in processing these unstructured data from diverse information sources. First, social media data, despite its immediacy, broad user and content coverage for event detection for achieving effective situation awareness and coordinating response actions across diverse stakeholder networks [23], presents significant processing challenges, including informal and varied language expressions in disaster awareness [24–26]. Second, although agency news coverage provides extensive event tracking and survey reports offer detailed descriptions, the diverse reporting formats across multiple agencies complicate integrated event tracking and verification processes. Third, the heterogeneous nature of event types, stakeholders, and response activities in both survey reports and during-disaster communications impedes systematic analysis. Fourth, the coordination of response actions faces operational complexities where standard protocols are disrupted by irregular disaster events [18]. The involvement of diverse stakeholders beyond first responders necessitates sophisticated knowledge integration across multiple phases - from pre-disaster community education to post-event learning and resource mobilization. These four fundamental challenges underscore the critical need for advanced natural language understanding capabilities that cannot only process diverse information streams effectively, but also provide intelligent operational support for diverse stakeholders through automated knowledge retrieval and action suggestion in integrated disaster management. The evolution of AI techniques, particularly in the domain of natural language processing, offers promising approaches to address these multifaceted challenges.

Natural language processing (NLP) has emerged as a critical technological foundation for disaster situation awareness and response coordination. Traditional rule-based NLP approaches, however, demonstrate significant limitations in managing the complexity and diversity of disaster-related information [25]. These constraints extend beyond simple pattern matching difficulties to fundamental challenges in contextual understanding and information extraction from informal communications [27]. The recent advancement of large language models (LLMs) presents transformative opportunities, offering enhanced capabilities in natural language understanding, multi-source information fusion, and context-aware content generation. This potential has gained institutional recognition globally, evidenced by FEMA's strategic integration of AI technologies into disaster management frameworks [28], national policy initiatives such as China's "14th Five-Year Plan for National Informatization'', which emphasizes artificial intelligence integration in emergency management systems [29], and focused attention on LLM integration in crisis communication by the European Crisis Management Laboratory [30]. While existing literature addresses AI applications in disaster management – including traditional NLP reviews [31], specialized implementations of neural networks for disaster image analysis [17], and general AI approaches for extreme weather modeling [12] – these works do not adequately address the distinctive capabilities of LLMs. Unlike previous deep learning approaches that require extensive domain-specific training and feature engineering, LLMs represent a fundamental shift in AI application through their contextual reasoning abilities and zero/few-shot capabilities. This technological evolution enables novel applications in disaster contexts, from real-time information synthesis across heterogeneous sources to adaptive response recommendation systems. Despite growing institutional interest and initial experimental applications, a critical knowledge gap exists regarding the integration frameworks and implementation pathways for LLM capabilities within established disaster management practices.

This review addresses this gap by synthesizing emerging research, conceptualizing application domains, and developing a structured framework for effective LLM deployment across the disaster management cycle. Our analysis focuses primarily on language model applications in disaster management from 2020–2024. Given the emergent nature of this field, we adopt an inclusive approach that considers both LLMs and relevant medium-sized language models (e.g., BERT) with contextual understanding capabilities applicable to disaster management information processing. Based on our review, we identified that current LLM applications naturally cluster around four key operational activities: detection (identifying potential disaster events), tracking (monitoring event development), analysis (assessing impacts and response performance), and action (supporting response coordination). This classification draws from our detailed examination of 70 language model applications, supplemented by comparative analysis with relevant AI approaches to contextualize LLM capabilities and limitations.

Based on this framework, our investigation addresses three research questions: (1) How do LLM capabilities enhance disaster management effectiveness across these four operational activities? (2) What technical approaches and implementation strategies enable effective LLM deployment in disaster management applications? (3) Drawing from experiences across these activities, how can LLM be systematically integrated into existing disaster management frameworks to address the persistent challenge in integrated disaster management? The interdisciplinary analysis identifies both current capabilities and limitations of LLM applications, particularly in connecting response-focused insights to preparedness and mitigation phases. Building on these findings, we propose strategic directions for LLM integration through a "3M" framework that aims to bridge gaps in the disaster management cycle. The resulting theoretical insights and practical guidelines contribute to advancing disaster information management systems toward more integrated approaches, ultimately enhancing communities' capacity to anticipate, respond to, and recover from disasters in urban environments.

2. Technical foundation: LLM in disaster information management

2.1. Challenges in managing disaster information for conventional NLP approaches

The diverse nature of disaster-related information streams presents fundamental challenges for traditional Natural Language Processing (NLP) approaches. These challenges manifest across three primary dimensions: data characteristics, processing requirements, and deployment demands, significantly impacting the effectiveness of conventional information management approaches in disaster scenarios.

First, disaster-related communications exhibit complex data characteristics that challenge conventional NLP methods. The informal expression style in social media posts requires robust processing of irregular punctuation, abbreviations, and ambiguous descriptions [32]. Cross-regional disaster impacts necessitate multi-linguistic processing capabilities, particularly crucial in large-scale disasters such as wildfires, floods, and earthquakes [33–35]. Text length variations between concise social media posts and detailed news reports further complicate processing approaches [36]. Additionally, the increasing integration of multimedia content in disaster reporting demands sophisticated multi-modal processing capabilities [37], as visual information often provides critical context for understanding disaster impacts and response needs.

Second, effective disaster information management requires advanced NLP processing capabilities across multiple tasks. Early event detection demands sophisticated classification of multilingual public communications with robust contextual understanding [38]. Information extraction requires advanced Named Entity Recognition (NER) for identifying locations, organizations, and temporal references, alongside relationship extraction for capturing causal, spatial, and temporal relationships between entities [39]. The complexity of these requirements is exemplified in recent research; Ma et al. [40] developed specialized models to extract diverse entity types (including geohazard types, geographic objects, emergency management objects) and multiple relationship categories (cause–effect, distance, spatial, and attribute relationships) to construct disaster ontology-based knowledge graphs. These sophisticated extraction tasks must operate reliably across varying communication styles and information formats while maintaining high accuracy under time-critical conditions.

Third, deployment demands in disaster management introduce distinct technical requirements for NLP systems. These systems must process and synthesize information across emergency response agencies, humanitarian organizations, and affected communities, while operating under severe time and resource constraints. The challenge extends beyond mere information processing to generating actionable insights that align with different stakeholders' operational protocols and decision-making needs. While traditional NLP approaches have achieved success in structured environments, their limitations in handling contextual complexity and generating adaptive responses have constrained their effectiveness in dynamic disaster scenarios. LLMs present a promising direction through their advanced capabilities in contextual understanding and adaptive information generation.

2.2. Evolution of backbone language models

The evolution of language models in disaster information processing has progressed from traditional feature-based approaches to increasingly sophisticated neural architectures. Early disaster information systems relied on manual feature engineering and statistical models, demonstrating limited effectiveness in processing the contextual nuances of disaster communications [25]. The emergence of transformer-based architectures [41] has fundamentally transformed natural language processing capabilities through several key innovations: self-attention mechanisms enabling long-range dependency modeling crucial for disaster narrative understanding, parallel processing architectures supporting rapid response requirements, and context-aware representations critical for disaster-specific terminology interpretation.

The effectiveness of modern language models stems from their pre-training process on massive text corpora, where models learn statistical patterns and linguistic representations through self-supervised learning [42]. Pre-trained parameters represent the learned weights within the neural network that encode this linguistic knowledge, with larger parameter counts generally enabling more sophisticated language understanding and generation capabilities. These parameters capture hierarchical linguistic patterns ranging from basic syntax to complex semantic relationships, making them particularly valuable for processing disaster-related communications despite linguistic variations, contextual ambiguities, and the time-critical nature of disaster reporting.

Backbone language models, such as GPT and BERT, serve as foundational architectures that establish core language processing capabilities through extensive pre-training on general-domain text. The choice of backbone architecture significantly influences both

Table 1
Representative language model architectures and their characteristics.

Introduced	Architecture	Capability	Parameters	Representative
2018	Decoder-only	Content generation and reasoning	>1T*	o3, GPT-4, Claude 3.5
			100B-1T	o3-mini, DeepSeek-R1
			1B-100B	Llama 3-8B, Llama 3.2-90B, DeepSeek-7B, DeepSeek R1-70B
2018	Encoder-only	Contextual understanding	100M-500M	BERT, RoBERTa
			<100M	DistilBERT, ALBERT, ELMo
2014	Encoder–decoder	Text transformation	100M-1B	T5, BART, Seq2Seq

Notes: Parameter counts marked with * are based on industry estimates, as official specifications have not been disclosed by the respective companies. o3 is estimated to exceed 100T parameters, while GPT-4 and Claude 3.5 are estimated to be in the range of 1.7T-2T parameters.

the model's capabilities and its operational characteristics in disaster management systems. Contemporary LLM architectures have evolved along three primary developmental paths, each offering distinct capabilities for disaster information processing (Table 1).

Decoder-only architectures represent the current frontier of language model development, with state-of-the-art models like OpenAI o3 [43], Claude 3.5 [44], and DeepSeek [45] achieving significant advancements through parameter scaling. Their autoregressive decoding architecture enables superior content generation capabilities by modeling sequential dependencies in language, crucial for generating detailed situation reports and response recommendations. These models demonstrate advanced reasoning capabilities through chain-of-thought processes, essential for complex disaster scenario analysis. Earlier implementations required integrating LLMs with separate computer vision models like YOLO for object detection and ResNet or VGG for damage classification in disaster imagery. Contemporary decoder-only architectural advances, exemplified by models such as Gemini Ultra [46] from early 2023, now incorporate native multi-modal processing capabilities, enabling simultaneous analysis of text, imagery, and structured data sources within a unified architecture. This architectural evolution has particular significance for disaster management, where multi-source information integration often leads to substantially improved accuracy in nuanced situation awareness such as damage assessment. These architectures have further evolved to support embodied AI applications, where models interact with physical environments through robotic systems, and demonstrate the ability to autonomously perceive, reason, and make decisions within their operational context. More deployment-oriented implementations, such as OpenAI o3-mini [47], DeepSeek-7B [48] and Llama-3-8B [49], maintain significant generation capabilities while offering practical deployment flexibility for resource-constrained systems.

Encoder-only architectures, which dominated the field's early development through models like BERT [50] and RoBERTa [51], excel in bidirectional contextual understanding and maintain significant value in disaster management applications despite the emergence of decoder-only LLMs. Their enduring value are demonstrated in domain-specific tasks requiring stable performance, low-latency inference, and cost-effective computation, where practical considerations outweigh the general-purpose capabilities of larger decoder models in operational disaster contexts. As a result, these architectures continue to function effectively in specialized tasks such as classification and named entity recognition, particularly in resource-limited deployment environments for disaster management. These models have evolved to address operational constraints through efficient implementations like DistilBERT [52] and ALBERT [53], offering practical deployment options where task-specific performance is prioritized over general-purpose capabilities. Similarly, The encoder–decoder architecture, exemplified by T5 [54] and BART [55], provides specialized capabilities in text transformation tasks, though their application in disaster management contexts remains more limited.

2.3. Model adaptation for disaster applications

The deployment of LLM in disaster contexts requires systematic adaptation approaches based on model accessibility and operational requirements [56]. While traditional NLP models often rely on neural network ensembling – combining RNNs for sequential patterns, GNNs for graph structures, and CNNs for local pattern recognition – modern LLMs like Llama 3 and GPT4 possess inherent language understanding capabilities through extensive pre-training. However, these general-purpose models require specialized adaptation for disaster management through two primary mechanisms: fine-tuning and prompt engineering (Table 2).

Fine-tuning approaches modify pre-trained model parameters to optimize performance for domain-specific tasks while preserving general language understanding capabilities [57]. This adaptation is particularly valuable in disaster management scenarios where models must recognize domain-specific terminology, understand emergency protocols, and generate contextually appropriate responses. For instance, fine-tuning enables models to accurately classify specific disaster types, extract critical information from emergency reports, and generate situation-specific rescue recommendations. Model accessibility determines the applicable adaptation methods: open-source models (e.g., DeepSeek-7B) enable comprehensive parameter modification, while partially open models (e.g., Llama 3) commonly utilize parameter-efficient methods such as Low-Rank Adaptation (LoRA) [58].

For closed-source models accessible via APIs (e.g., GPT-4), adaptation relies on prompt engineering techniques, a method of structuring input queries to elicit desired model behaviors without modifying model parameters. This approach encompasses structured instruction design, contextual few-shot learning, and output refinement using Chain-of-Thought (CoT) reasoning [59]. In disaster management contexts, well-crafted prompts enable models to analyze emergency situations, prioritize critical information, and generate protocol-aligned response recommendations while maintaining consistency with established emergency protocols. For example, prompts can be designed to extract key details from disaster reports, analyze potential risks, and suggest response strategies based on historical disaster management practices.

Table 2
Adaptation methods for language models.

Methods	Components	Example
Fine-tuning	Training dataset	Data source, content coverage, scale, distribution
	Configuration	Adaptation layer, rank–size, loss function
Prompt engineering	Instruction design	Role-play, meta-prompts, etc.
	Contextualization	Few-shot learning
	Output refinement	Chain-of-Thought (COT), self-evaluation, etc.

3. Methodology and analytical overview

3.1. Review methodology

This systematic review follows the PRISMA guidelines [60], employing a structured approach to literature identification, screening, analysis, and synthesis (Fig. 1). The search encompassed academic databases (Web of Science, IEEE Xplore, ScienceDirect, and Scopus) and preprint repositories (arXiv), combining disaster-related terms (flood, storm, hurricane, earthquake, drought, wildfire, forest fire, landslide, extreme heat) with LLM-related keywords (large language model, LLM, natural language processing, GPT, BERT, etc.). We employed Boolean operators to refine results (e.g., "disaster AND (LLM OR 'large language model')") and supplemented database searches with forward and backward citation tracking of key papers. The scope was confined to English-language publications from 2020 to 2024, including journal articles and conference proceedings.

We applied a two-stage screening process. Initial screening excluded following papers: (1) studies using language models only for literature review purposes without empirical disaster applications; (2) papers mentioning language models without specific implementation details; (3) studies focusing exclusively on computer vision approaches without language processing components; and (4) general AI disaster applications without employing pre-trained language models. After identifying 90 potentially relevant papers through database searches and citation tracking, we conducted a full-text review, resulting in 70 papers that met our inclusion criteria of employing language models (both large-scale models like GPT and medium-sized models like BERT) for disaster management applications.

The classification framework examines the included studies through four operational activities – detection, tracking, analysis, and action – to systematically identify implementation patterns and technical evolution in disaster management applications. Our analysis proceeds in three stages: (1) examining the temporal distribution of backbone LLM architectures, (2) conducting cross-sectional analysis between operational activities and disaster types (Section 3.2), and (3) summarizing application characteristics in disaster contexts (Section 3.3). The subsequent review results (Section 4) are structured according to these four operational activities, enabling systematic examination of LLM capabilities, technical approaches, and implementation strategies within existing disaster management frameworks.

Last, the discussion part (Section 5) identifies patterns in current LLM applications and evaluates their alignment with integrated disaster management theory. We identify research gaps and unexplored opportunities in the field. Building upon these insights, we propose future directions for addressing the challenges in integrated disaster management through LLM integration, incorporating dual considerations: the technological trajectory of language models and the operational requirements of disaster information systems. Our analytical framework is anchored in the DIKW (Data-Information-Knowledge-Wisdom) hierarchy [61], which provides a structured approach to examining how LLMs can optimize information processing, knowledge generation and application across the disaster information management cycle.

3.2. Evolution and distribution of LLM applications in disaster management

The systematic analysis of 70 identified studies reveals distinct patterns in both the architectural evolution and operational deployment of LLM. This analysis demonstrates a clear progression in model architectures and their application across different disaster types and management phases. The temporal distribution of research publications shows consistent growth, from 5 papers in 2020 to 9 in 2021, 10 in 2022, 19 in 2023, and accelerating to 27 papers in 2024. This trajectory reflects rapidly increasing interest in language model applications for disaster management, with publication volume quadrupling over the five-year period.

The architectural trajectory shows a significant transition from encoder-only to decoder-only models. BERT-based architectures initially dominated disaster-related research (2020–2022), primarily serving detection and documentation functions through their robust bidirectional encoding capabilities. However, recent years (2023–2024) have witnessed an accelerated adoption of decoder-only architectures, particularly GPT-4 and open-source alternatives such as Llama-2, which excel in generating contextually appropriate responses and complex reasoning tasks. This architectural shift has been accompanied by reduced implementation latency; while GPT-3's integration into disaster management frameworks showed considerable delay, subsequent models like GPT-4 and Gimini 1.5 Pro have experienced notably faster adoption cycles (Fig. 2).

The operational distribution analysis reveals a distinct emphasis on early-stage disaster management, with event detection applications (n=37), followed by analysis (n=17), tracking (n=8) and action-coordination applications (n=8). This imbalance suggests a critical gap in research addressing continuous monitoring and response coordination. Contemporary architectural advances, particularly in multi-modal processing and embodied agency (eg. house construction robotics) across physical and virtual

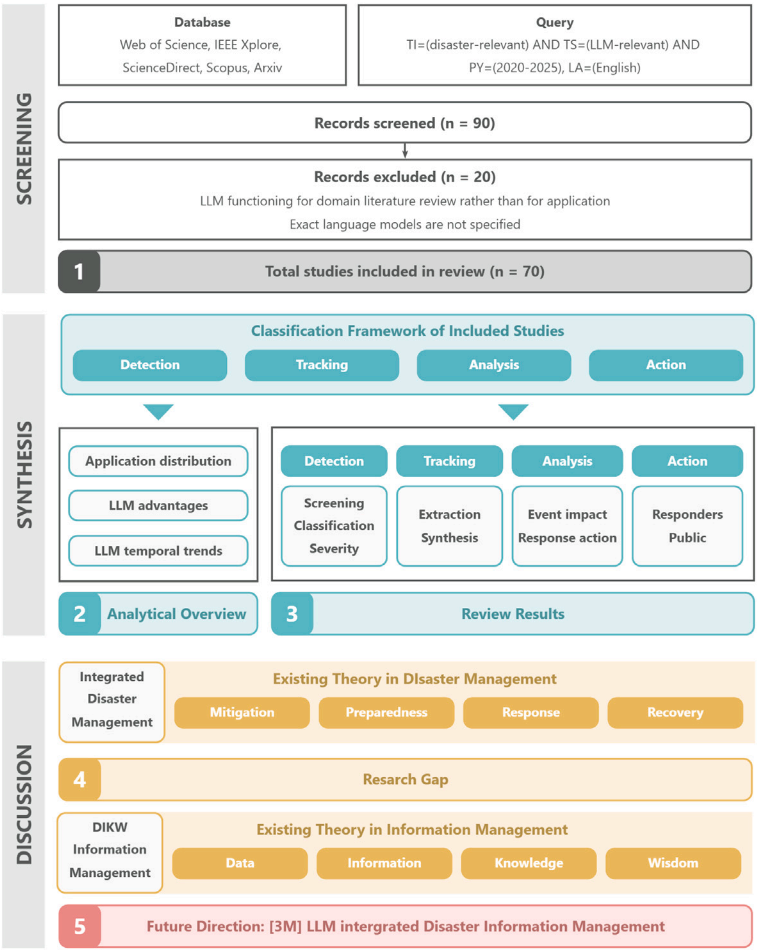

Fig. 1. Flow diagram of the systematic review process.

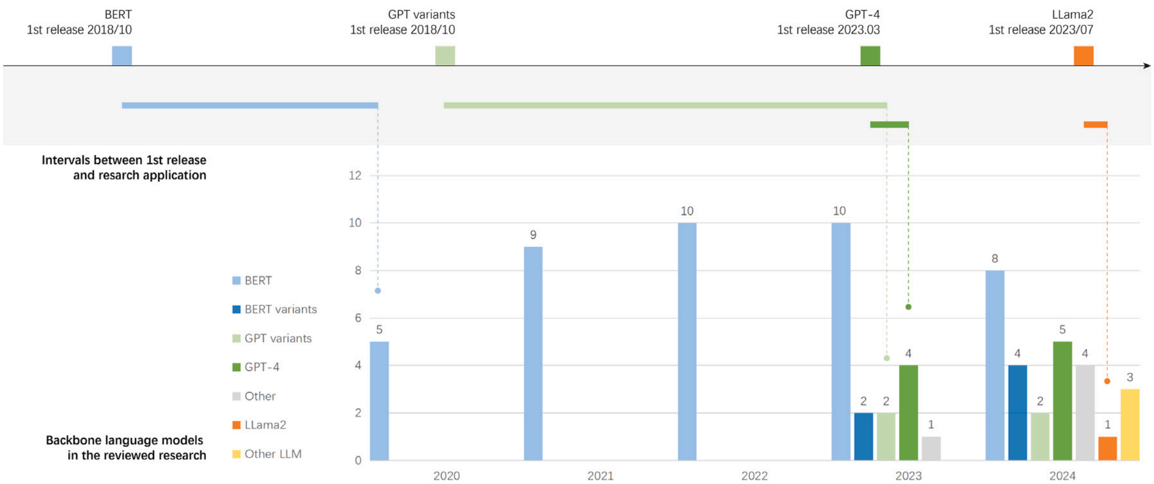

Fig. 2. Summary of base language model selected in current disaster management application.

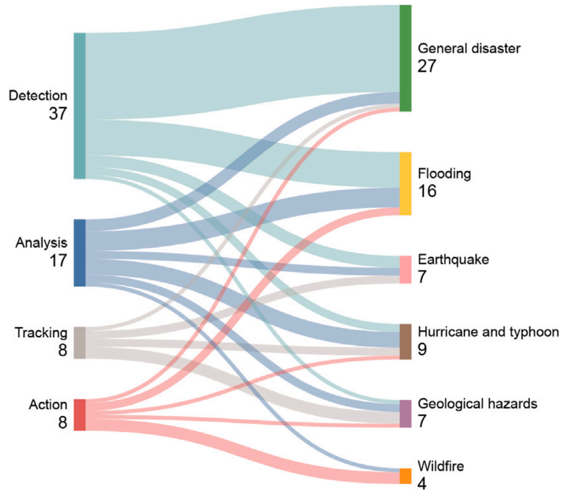

Fig. 3. The share and relationship of disaster type and LLM application.

domains [62,63], offer promising directions for addressing these gaps, though their implementation in disaster contexts remains nascent.

Cross-sectional analysis of disaster types reveals notable variations in LLM application focus (Fig. 3). General disaster scenarios constitute the primary category (n=27), reflecting the requirements of detecting universal disaster events and followed by specific disaster types:

- Flooding events studies (n=16) emerge as the most studied specific disaster type, with research predominantly focusing on flash flood detection and large-scale impact assessment.
- Hurricane and typhoon studies (n=9) emphasize cascading impact analysis, particularly in relation to subsequent flooding events.
- Geological hazard studies (n=7) prioritizes event documentation and inter-hazard relationship analysis
- Earthquake studies (n=7) span the full operational spectrum from detection to social impact assessment, especially the fatality number tracking.
- Wildfire research (n=4), while numerically smaller, exhibits a particular emphasis on integrating multiple data sources, including environmental sensor data, and historical cases for response action suggestion.

3.3. Summary of LLM advantages in disaster management operations

The analysis reveals LLMs distinct values throughout the operational phases in disaster management (Fig. 4).

For detection, LLMs demonstrate significant utility in disaster event identification through the systematic processing of online communications and severity assessment. The primary capability centers on unstructured data classification, where LLMs excel in three aspects: (1) processing informal and colloquial expressions and multi-lingual contents in disaster-related communications, (2) integrating multi-modal information streams for comprehensive severity assessment, particularly in enhancing granular measurements through the fusion of visual object recognition, and (3) guiding dataset synthesis for enhancing classifier training.

For tracking, LLMs enhance disaster event monitoring in entity extraction and synthesis through three aspects: (1) extracting disambiguating temporal and spatial references while establishing robust entity-relationship mapping, (2) facilitating multi-scale spatial integration for diverse location representations from precise coordinates to descriptive references, addressing varied geometric features and hierarchical granularity, and (3) resolve conflicting arguments in conjunction with complementary statistical models.

For analysis, LLMs contribute substantively to disaster impact assessment and response effectiveness evaluation through multidimensional information classification. The prevalent classification metrics encompass: (1) information source attribution, including stakeholder role classification in digital communications; (2) content characterization, from thematic clustering to entity-specific analysis; and (3) communication approach assessment, utilizing established frameworks such as Competing Values Framework (CVF) dimensions.

For action, LLMs serve as sophisticated intermediaries between diverse stakeholders and disaster information systems through three fundamental mechanisms: (1) intelligent task formulation derived from personalized queries, accommodating varied disaster management scenarios; (2) contextual knowledge retrieval, facilitating the identification and application of relevant historical precedents in complex disaster scenarios; and (3) adaptive response generation through iterative self-evaluation and human-machine collaborative protocols.

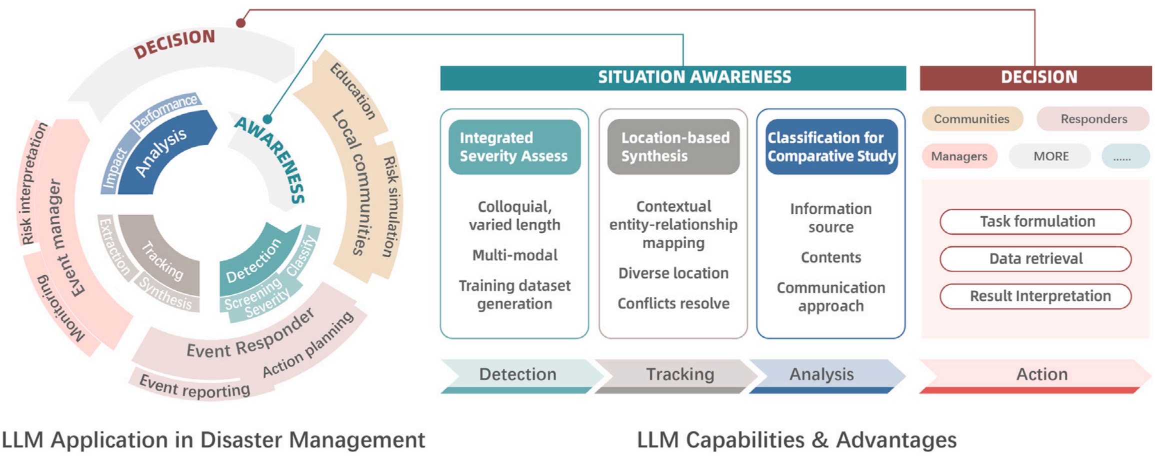

Fig. 4. Analytical framework of LLM capabilities in disaster management operations.

4. Review results

4.1. Detection

Event detection serves as the foundational first step in this analytical framework of disaster information management (detection, tracking, analysis, and synthesis). The fundamental challenges in disaster-related social media communications—informal expressions, multi-linguistic content, multi-length texts, and multi-modal data—require sophisticated language processing capabilities. This section examines how LLMs address these challenges through three sequential aspects: message screening, classification and severity assessment.

4.1.1. Message screening

The identification and classification of event-relevant messages from social media has emerged as a crucial crowdsourcing mechanism for disaster event detection [64–67]. Of particular significance are specialized message categories, such as behavioral facilitation (BF) messages, especially those with regional specificity, which have demonstrated substantial influence on public response during disaster events [68,69]. Following the architectural evolution discussed in Section 3.2, various BERT-based specialized binary classifiers have been utilized for screening, as summarized in Table 3.

The evolution of disaster-related message screening has progressed from early transformer-based approaches to more sophisticated integrated frameworks. Initial applications leveraging BERT's contextual understanding capabilities in rumor and misinformation detection [70]. Later approaches have expanded beyond pure content analysis to incorporate behavioral-statistical modeling for enhanced credibility assessment. Hembree et al. [71] developed a comprehensive framework for evaluating linguistic patterns, incorporating multiple dimensions including contrastive word connotations, temporal mood variations, readability metrics, and sentiment analysis. Complementing this linguistic approach, Fan et al. [72] proposed a cluster-based statistical methodology, establishing a correlation between information cluster size and event credibility assessment.

Despite these integrated methodologies in screening credible disaster events, significant opportunities remain for developing more specialized models to address the challenges of real-time information verification and context-specific reliability metrics. The emergence of generative AI has introduced new challenges in disaster information management, particularly with the proliferation of synthetic images and videos. This evolution in misinformation formats necessitates the development and implementation of multimodal LLM approaches for comprehensive verification protocols, setting the stage for advanced tracking mechanisms discussed in Section 4.2.

4.1.2. Message classification

The evolution toward multi-label classification systems enables rapid disaster response and management by facilitating stakeholder-specific information identification (Table 3). In humanitarian aid classification, systems have evolved to address complex rescue needs identification [73,74], incorporating hierarchical classifications from broad categories (affected populations, infrastructure damage) to specific subcategories (displaced individuals, evacuation requirements, trapped victims) [75]. For rescue operations, classification frameworks address three sequential stages: rescue requests identification (location containing, resource needs, etc.), issue reporting (hazard situations, victim information, resource availability, etc.) [76,77], and rescue tracking (response status and outcome, etc.) [78]. In the management domain, classification systems have demonstrated particular effectiveness in enhancing established repositories, such as NOAA's Storm Event Database, through systematic documentation of flash flood events [79].

Table 3
Various classifiers for disaster communication.

Classifier	Metric	Category
Binary	Informative	Certain disaster-related and contain information
	Credibility	Rumor, misinformation
	Attention	Behavioral facilitation (BF), Regional BF
Multi-classify	Humanitarian aid	Affected individuals, infrastructure and utilities, donation and volunteering, caution and advice, sympathy and support, and other useful information
	Rescue operation	Rescue requests, issue reporting, rescue tracking
	Manage	Damage and economic impact, fatalities, injuries and rescue, warning and emergency

Table 4
Challenges of disaster severity estimation.

Object	Severity	Source	Challenges
Environment	Earthquake: shaking intensity	Crowdsourced text	Ambiguous, unprofessional expressions
	Flooding: water level	Crowdsourced image	Complex damage situation classification
	Collapse: range and depth	Remote sensing	Limited post-disaster training data
People	Injury severity	Crowdsourced text	Lack of labeled training data
		Medical logs	Professional diagnosis requirement

For social media posts classification, traditional transformer-based language models, particularly BERT and its variants such as RoBERTa and DistilBERT, initially demonstrated significant improvements over conventional NLP models like Word2Vec and ELMo [32,67,76]. These early transformer implementations addressed challenges through fine-tuning techniques, focusing on dataset characteristics, distribution, and scale [80,81]. Researchers developed sophisticated approaches to handle dataset imbalance, such as Class-Balanced Focal-Loss [79] and instruction-based fine-tuning [82].

To enhance classification capabilities, these early-stage models were integrated with specialized neural architectures in three key areas: (1) contextual understanding through RNN, particularly Bi-LSTM [83], (2) network structure mapping using GNN, specifically GCN for global information extraction and GAT for local semantic preservation [77], and (3) efficient processing of short communications via CNN [67,84]. For multimedia content, researchers developed fusion approaches combining text features with image classifiers such as ResNet, DenseNet, and ConvNeXt [85,86].

The emergence of advanced LLMs has fundamentally transformed this multi-component approach to disaster information processing. Modern models demonstrate superior capability in directly extracting actionable information through few-shot and zero-shot learning, as demonstrated in ChatGPT for diverse flooding relevant linguistic understanding [34]. Contemporary multi-modal LLMs (e.g., GPT-4) further streamline information processing by simultaneously analyzing text and visual content. While traditional classification metrics remain valuable for structuring information extraction, their integration into prompt engineering now enables more nuanced stakeholder and specific information delivery, from humanitarian aid coordination to rescue operation monitoring, through a unified, efficient interface.

4.1.3. Severity assessment

Disaster severity assessment has integrated both instrumental measurements and systematic human observations, with many established severity scales fundamentally rooted in observed effects. A prime example is earthquake assessment, which distinguishes between instrumentally measured magnitude and intensity scales based on observed impacts, including human experiences and structural effects [87]. This established practice of incorporating human observations into formal severity metrics provides a theoretical foundation for leveraging modern disaster-relevant communications in severity assessment. While traditional severity documentation relied on structured field surveys and expert observations, the ubiquity of social media communications now offers real-time, widespread observational data, though requiring sophisticated interpretation methods to maintain assessment reliability.

LLM-based disaster severity assessments primarily address two critical domains: affected population assessment and environmental impact evaluation, detailed in Table 4. In human impact assessment, emerging research demonstrates LLMs' capability in medical triage from textual descriptions [88] and rescue prioritization [89], though comprehensive frameworks for emergency severity assessment remain limited. Environmental impact assessment faces distinct challenges across disaster types, including interpretation of non-technical descriptions, complex visual environments, and limited training data availability.

Disaster damage severity assessment methodologies have evolved across multiple research paradigms. Within text-focused approaches, researchers leverage language models to extract and quantify disaster severity from witness testimonies. This is mainly achieved through expression transformation, which bridges the gap between informal communications and formal severity metrics, exemplified by QuakeBERT's enhanced keyword mapping for earthquake intensity assessment [90]. Multi-modal processing approaches have evolved from basic ensemble methods [91] to sophisticated location-aware systems that integrate text analysis, geographical context, and visual assessment [74]. Beyond severity-level classification, the integration of fine-grained quantitative measurements has demonstrated significant potential for enhancing predictive model performance. For instance, flooding depth data extracted from crowdsourced imagery provides valuable parameters for improved flood prediction accuracy [92], despite the inherent challenges in depth estimation for conventional computer vision systems. Recent advances in multi-modal LLMs have

Table 5
LLM adaptation in disaster severity estimation.

Task	Adaptation	Example
Text transformation	Keyword-based quantification	BERT + Modified severity table based on colloquial keywords
Multi-modal processing	Embedding merge Trigger vision model via location Multi-modal LLM	BERT + VGG BERT + location trigger + YOLO GPT-4 via in-context learning, chain of thought
Synthetic data generation	Text synthesis via prompt engineering Image synthesis via multi agents	Use template to prompt LLM generate corpus with required labels LLM optimize prompts for guiding VM to generate images with damage description

transformed visual analysis capabilities through enhanced reasoning mechanisms, as evidenced in Gemini 1.5 Pro's few-shot visual interpretation capabilities [93] and GPT-4's object-referenced depth estimation technique. The later employs prompt engineering to guide GPT-4 to use familiar object as dimension reference, which effectively emulate human annotation processes for flood imagery, and ends up improving flooding depth estimation accuracy [94]. These advancements in quantitative disaster assessment present considerable potential for developing granular disaster damage datasets, thereby contributing to more precise disaster prediction frameworks.

In parallel to direct application of language models for severity classification and quantitative data extraction, researchers have investigated LLMs' capacity for synthetic data generation to address the persistent challenge of training data scarcity in specialized vision models. LLMs have demonstrated effectiveness in generating domain-specific training data, as exemplified by LLaMA-2's application in creating realistic damage descriptions for critical infrastructure facilities [95]. For visual data synthesis, GPT-4 has shown sophisticated capabilities in guiding image generation through both targeted modification (inpainting) and complete scene generation approaches [96], particularly valuable for expanding limited post-disaster satellite imagery datasets. The disaster severity assessment approaches from above three perspectives are summarized in Table 5.

These LLM-based approaches have significantly advanced disaster severity assessment through improved interpretation of textual and static visual content. However, despite the increasing prevalence of disaster-related videos [97], their integration into severity assessment remains underexplored. The emergence of advanced multi-modal LLMs presents promising opportunities for comprehensive severity assessment, particularly in processing temporal patterns and dynamic visual information, suggesting a crucial direction for future research in disaster monitoring systems.

4.2. Tracking

Once a disaster event is detected, maintaining its accurate temporal evolution becomes crucial for response coordination and impact assessment. The tracking of severe disaster events traditionally relies on authoritative sources – emergency response agencies and professional field investigators – who provide systematic damage assessments, victim reports, and infrastructure status updates. While these formal channels ensure information reliability, significant technical challenges in managing heterogeneous data structures and verifying rapidly proliferating content emerge [98]. Traditional NLP approaches face fundamental limitations in this integration, particularly in constructing uniform event documentation and establishing causal relationships when processing informal social media content alongside structured official reports. This section examines how LLMs address these challenges in event information extraction and automated synthesis from heterogeneous event reporting sources.

4.2.1. Event information extraction

Event information extraction from disaster reports encompasses two fundamental components: entity recognition and argument extraction. Entity recognition focuses on identifying core elements (time, location, event type), while argument extraction captures event-specific attributes including casualty figures, property damage assessments, and response actions [98–100]. The heterogeneity of source materials – ranging from concise social media posts to comprehensive survey documents – necessitates distinct approaches to model training and adaptation (Table 6).

Traditional small-scale language models, particularly BERT-based implementations, rely on structured annotation frameworks incorporating Part-of-Speech (POS) tagging and Beginning-Inside-Outside (BIO) sequence labeling [101]. These approaches can be enhanced through sophisticated neural architectures, with LSTM/Bi-LSTM combinations augmented by Conditional Random Fields (CRF) and Self-Attention mechanisms emerging as predominant solutions [98,99,101,102]. While these architectures demonstrate robust performance in structured extraction tasks, they require extensive annotated training data and domain-specific adaptations.

In contrast, large-scale language models such as ChatGLM-3b present a paradigm shift through zero-shot learning capabilities, requiring only prompt templates with example entities rather than extensive annotated datasets [100]. This approach proves particularly effective for handling ambiguous cases like relative temporal references and hierarchical location recognition [103]. However, LLMs face limitations in fine-grained entity disambiguation tasks, particularly when distinguishing between similar event types, and the closed-source nature of some LLMs often precludes domain-specific fine-tuning [100].

Small-scale language models and LLMs present distinct advantages in disaster information extraction. The former excel in structured extraction tasks and require minimal computational resources, while the latter demonstrate superior capability in handling

Table 6
Entity extraction methods for various language models.

Base language model	Training dataset	Model adaptation
BERT (medium LM)	Annotation via POS tagging, BIO sequence labeling	Neural architecture: LSTM/Bi-LSTM/Bi-GRU+CRF+Self-Attention
ChatGLM-3b (LLM)	Prompt template showing entity examples	Zero-shot learning

Table 7
Methods of event location synthesis.

Perspective	Strategy	Methods
Scale/granularity	Merge	Geocoding API + location fusion based on statistical rules
Category	Distinguish	Multiple specialized classifiers
Contextual description	Disambiguation	Embed POI with spatial contextual relationship

ambiguous cases. The potential of hybrid architectures combining these strengths remains unexplored in disaster information extraction. Despite these complementary strengths, the development of hybrid architectures remains an unexplored yet promising direction for optimizing both extraction performance and computational efficiency in integrated disaster tracking systems, particularly when processing diverse information sources.

4.2.2. Event synthesis

Building upon extracted event information, the synthesis of disaster events from multi-source reporting represents a critical challenge in disaster information processing, particularly when integrating reports that span varying perspectives and spatial scales. The synthesis process follows a two-step methodology: location-based integration establishes fundamental grouping units through spatial correlation, followed by argument synthesis that resolves and validates event details within each event group.

Location-based integration is fundamental to disaster response, enabling coordinated rescue operations [80], aid distribution [72], and area analysis [101]. This integration encompasses three forms of crowd-sourced data: geotagged coordinates, user registration locations, and content-embedded mentions, with the latter requiring sophisticated processing approaches. The strategies can be systematically categorized into three dimensions (Table 7): multi-scale integration, where Fan et al. [72] developed coordinate-based fusion rules through Google Map Geocoding API; category differentiation, which evolved from Fan et al. [72]'s three-tier system (county, neighborhood, POI) to Hu et al. [103]'s eleven-category classification; and contextual disambiguation, addressed by Zhang et al. [101] through BERT with Bi-LSTM-CRF architecture for context-aware spatial integration.

Following spatial grouping, argument synthesis within each location group presents additional challenges in argument updates and verification. Wang et al. [35] established a framework for real-time earthquake fatality tracking by incorporating statistical constraints, temporal sequence analysis, and source reliability metrics. This approach, though still remains unexplored in application, demonstrates the potential for systematic argument synthesis in event tracking.

Current synthesis approaches face fundamental limitations in both spatial grouping and argument verification. The prevalent coordinate-based merging methods often sacrifice urban geo-characteristics during statistical estimation, while argument synthesis lacks comprehensive verification frameworks beyond specific cases like fatality number synthesis. Recent developments in information fusion theory, particularly Dempster-Shafer Theory (DST), which is systematically demonstrated in disaster information fusion [104], offer promising theoretical foundations for addressing these challenges for handling uncertainty and conflicting information. The integration of information fusion theories like DST with advanced language models could enable evolution from single-perspective, coordinate-based event recording to multi-perspective, context-aware truth-finding in disaster event tracking systems.

4.3. Analysis

Comprehensive disaster event tracking enables systematic analysis across multiple dimensions, with classification serving as a fundamental prerequisite for comparative studies. The analytical framework encompasses three primary classification dimensions: information source (eyewitness accounts versus secondary reporting), content type (topic clusters and entity-specific information), and communication patterns (internal/external stakeholder interactions). These classifications facilitate measurements through four main approaches (Table 8): sentiment analysis for public reaction assessment, temporal analysis for event progression tracking, information dissemination network analysis for stakeholder response evaluation, and semantic network analysis for impact mechanism exploration.

Early transformer-based models, particularly BERT and its variants, established important foundations in contextual understanding for disaster analysis. These models demonstrated effectiveness in specific tasks such as topic modeling [105] and emotion identification [106], laying groundwork for structured disaster analysis including disaster impact analysis and response action performance evaluation, as examined in the following sections.

Table 8
Disaster event analysis approaches.

Measure	Techniques	Compare group	Analysis example
Sentiment	Sentiment category: Plutchik's wheel	Source	Affected public's emotion type
	Sentiment score: positive, negative	Source	Affected public's emotion tendency
		Content: response action	Disaster response action performance
Temporal evolution	Space–time scan statistics (STSS)	Content: hot topic	Hot topic temporal evolution pattern
		Content: severity topic	Emergency severity topic indicator
	Spatio-temporal point process (STPP)	Content: hot topic	Topic diffusion magnitude and velocity
Dissemination network	Node importance	Source:	Disaster information propagation patterns
	Public engagement	Communication approach: CVF dimension	Optimal mobilizing strategy
Semantic network	Knowledge graph	Content: event entity	Disaster impact chain visualization
	Topic modeling + WCN	Content: event type	Hazard interconnections

4.3.1. *Disaster impact analysis*

Sentiment analysis serves as a fundamental analytical tool for evaluating social impacts during and after disasters, implemented through sentiment categorization and quantitative scoring approaches. The Plutchik's wheel of emotions provides the theoretical framework for sentiment categorization, adopted by Ahmad et al. [107], which Oikawa et al. [108] extended using BERT models to classify communications into primary, secondary, and tertiary (interpretative) information layers. This framework enables quantitative sentiment scoring across different topics, as demonstrated in Praveen and Deepika [105]'s analysis of survivors' guilt emotions and Chen [109]'s post-typhoon impact assessments.

Temporal analysis is incorporated for event evolution primarily through Space-Time Scan Statistics (STSS) and Spatio-Temporal Point Process Models (STPP). STSS applications manifest in two approaches: topic-temporal and severity-focused assessment. In topic-temporal analysis, Fan et al. [72] mapped topics along timelines for situation monitoring, while Du et al. [36] identified consistent "situation-tips-rescue" patterns during disasters. For severity-focused assessment, employing entropy methods for severity measuring, Zhang et al. [110] combined BERT and VGG-16 to process multi-modal posts, identifying casualties and loss-related content as key severity indicators. Ma et al. [111] extended this framework using the Susceptible-Infected-Recovered (SIR) model with BERTopic for cross-regional wildfire analysis, examining correlations between topic diffusion patterns and actual disaster propagation.

Semantic network analysis, particularly through ontology-based knowledge graphs, enables impact chain exploration. BERT-BiGRU has emerged as an efficient model for entity and relationship extraction, optimized for shorter text processing. Ma et al. [40] employed this architecture to extract comprehensive disaster ontologies, while Liu et al. [112] developed a quintuple structure (Concept, Property, Relationship, Rule, Instance) for typhoon event representation. These Neo4j format knowledge graphs support SPARQL-based retrieval [113] and integrate with topic mining for damage pattern analysis across different geological hazards through Word Co-occurrence Networks (WCNs) [114].

4.3.2. *Response action performance evaluation*

Response effectiveness evaluation provides crucial insights for disaster response operations through two complementary measurement approaches: granular sentiment analysis of public reactions and information dissemination network analysis. Temporal sentiment analysis examines public responses to specific institutional actions throughout the disaster management cycle. Zhang et al. [115] developed a framework for evaluating sentiment progression across critical response phases, from investigation team establishment to accountability measures, enabling quantitative assessment of institutional response effectiveness. This analytical framework has been technically advanced through Target Sentiment Classification (TSC), demonstrated by Rachel and Bhuvaneswari [116] in assessing sentiment toward specific disaster-related needs (e.g., medical supplies, food distribution, emergency services).

Information dissemination network provides another lens for response effectiveness evaluation by examining static node importance metrics and dynamic engagement patterns. Static node analysis, primarily utilizing PageRank algorithms, reveals systematic disparities in information propagation effectiveness across stakeholder categories. Babvey et al. [73]'s implementation of fine-tuned BERT models for user classification demonstrated a hierarchical diffusion pattern: government and media posts achieved significantly higher propagation rates compared to eyewitness accounts, highlighting the critical role of institutional authority in disaster information dissemination. Dynamic engagement analysis has evolved through the integration of theoretical frameworks with communication approaches. Yan et al. [117]'s application of Sentence-BERT to analyze three million posts through the Competing Values Framework (CVF) revealed disaster-specific communication patterns: biological disasters showed higher engagement rates with internal and stable-oriented messages, while weather-related disasters demonstrated superior response rates to external and flexible-oriented communications, providing empirical support for disaster-specific communication strategy optimization.

While categorization-based approaches have enabled systematic analysis of disaster impacts and response performance, modern LLMs offer unprecedented capabilities in flexible and dynamic analysis without rigid classification constraints. Traditional research, focused on predefined stakeholder categories and disaster types, often overlooked emergent patterns and complex interactions. Advanced LLMs demonstrate superior capability in identifying these nuanced relationships, potentially benefiting cross-disaster pattern recognition and stakeholder interaction dynamics. These models enable expert-collaborative real-time analysis and suggestion generation, supporting the evolution of disaster management toward more integrated, context-aware approaches that can rapidly adapt to emerging disaster situations while maintaining analytical rigor.

4.4. Action

LLMs' semantic understanding and reasoning capabilities, which facilitate disaster detection, tracking and analysis, discussed in the previous section, have begun to transform disaster management by bridging the traditional gap between situation awareness and actionable response. This integration addresses critical knowledge barriers for key stakeholders during emergencies. For instance, emergency responders must analyze rapidly evolving scenarios and extract actionable insights from overwhelming data streams, while simultaneously formulating effective intervention strategies under extreme time pressure. At the same time, affected communities may struggle to contextualize their localized risks or communicate ground-level insights that could refine response efforts. LLMs can support both groups, enhancing coordination and information flow during crises. This section examines the evolving role of LLMs in two complementary domains: responder coordination and public engagement. A synthesis of key integration pathways tailored for the disaster emergency context concludes the section.

4.4.1. Responders coordination

Recent literature demonstrates that LLMs can support the development of emergency action plans by enabling rapid synthesis of evolving disaster information and facilitating interactive dialogue with emergency management professionals. For example, experimental applications have shown LLMs generating comprehensive survey reports from on-site photos [118] and enhancing web-based flood monitoring with advanced summary queries [119]. Unlike conventional deep learning models that rely on static datasets, for example, hydrodynamic-integrated machine learning models for flood risk assessment [120], LLMs provide dynamic, natural-language interpretations, helping bridge the gap between raw data and actionable insights.

Action plan formulation in disaster management typically follows standardized operational protocols, with most current LLM integration approaches built around modular frameworks and systematic questioning sequences. In-context learning underpins these methods, utilizing prompt engineering to embed professional disaster terminology and knowledge structures into the LLM's reasoning process. For instance, Goecks and Waytowich [121] integrated FEMA guideline terminology and exemplar outputs to enable geographically specific, protocol-compliant response plans. However, such prompt-based strategies often lack adaptability to complex scenarios requiring nuanced, domain-specific knowledge scattered across databases and literature. Systematic extraction of specialized knowledge, such as academic flooding mitigation strategies, can help inform localized responses [122], yet integrating this into action planning remains less explored. Compounding this are cognitive load issues during emergencies, with decision-makers frequently overwhelmed by data volume and time pressure, highlighting limitations of traditional information systems [123].

To mitigate these issues, advanced retrieval-augmented generation (RAG) frameworks have been developed. For example, Xie et al. [124] employed LLMs in wildfire risk analysis through structured profiling – collecting user expertise, risk priorities, and area interests – and iteratively refining action plans with human-expert feedback. Enhanced dialogue memory capacities in modern LLMs further support sustained human-AI collaboration. Despite these advances, key limitations remain. Integrating real-time and complex environmental data is still underdeveloped, as shown by attempts to apply LLMs to large-scale forest fire databases [125] and real-time sensor data streams [126]. Automated pipelines for real-time condition analysis and resource identification are not yet robust, and establishing reliable performance standards and constraints for LLM-generated recommendations remains a fundamental challenge given the high-stakes nature of emergency management.

4.4.2. Public engagement

Disaster risk reduction relies on collaboration between authorities and the general public. LLMs are increasingly used to enhance public education by providing accessible, natural language interfaces. Compared to professional systems, public-facing platforms typically address a narrower range of disaster scenarios but must flexibly interpret diverse user queries and translate technical information into understandable, actionable advice—especially for individuals with limited disaster knowledge and management experience.

A key challenge is accurately representing the specialized terminology common in disaster domains. General pre-trained models often lack this precision. Research has shown that fine-tuning, such as medium-sized T5 models trained on curated disaster corpora, can significantly improve typhoon-related question answering and provide credible, comprehensible advice to the public [127]. However, fine-tuning alone is insufficient for complex or location-specific queries—for instance, determining flood risk at a specific address during a storm. Thus, disambiguation and multi-database reasoning become crucial. To address this, knowledge graphs have proven effective for structuring the relationships between disaster knowledge entities. This enables LLMs to standardize and resolve

Table 9
Knowledge integration strategy for different task phases.

Phase	Knowledge strategy	Integration techniques
Task formulation	User initialize formulation	Continue training with corpus
		Replace query through semantic similarity
		Disambiguate user query through knowledge graph
	LLM initialize formulation	Pose the question and shape the task in defined framework
Data retrieval	Unstructured data embedding	RAG: Corpus embedding + semantic similarity search
	Structured data query	Code generation to operate external tool
Output synthesis	In-context learning	Prompt engineering with relevant corpus as context
	Interactive planning	Dialogue-based refinement with human feedback
	Self-improvement	Self-ask and self-evaluation
	External statistical constraint	Statistical model assessing information time and source

user questions into precise information needs. Such approaches have been successfully deployed in flood education systems that return individualized risk profiles, tailored to different inundation scenarios for users' specific locations [128].

Another limitation arises when users require information that is not pre-calculated—such as the impact of live or hypothetical environmental changes. Conventional retrieval systems like BERT+FAISS work only with pre-defined results. In contrast, advanced LLMs can generate code to operate analytical software tools. For example, in flood emergency contexts, ChatGPT-4 guided by software knowledge graphs can automate ArcGIS (via ArcPy scripting) to generate real-time, location-specific hazard assessments in response to user queries about rainfall or river levels [128]. This integration bridges the semantic gap between user intent and complex spatial analysis, enabling richer, tailored disaster insights for non-technical users.

Although recent advances have substantially expanded the public's ability to access and comprehend disaster information, the scalable and reliable adoption of LLM-enabled solutions remains an open challenge. Continued research and systematic evaluation will be essential to realize their full potential in improving public engagement during disaster scenarios.

4.4.3. Summary of LLM agency for disaster management

As highlighted above, while knowledge-constrained and tool-integrated LLM approaches have demonstrated value in disaster education and risk communication for specific hazards, scaling these solutions to accommodate diverse and evolving disaster scenarios remains less explored. Recent advances in tool-augmented LLMs, which enable language models to interact directly with external tools and data systems, present new opportunities for more flexible and responsive public engagement. For instance, in disaster contexts, such systems could automatically generate personalized evacuation routes in response to real-time flood or wildfire data, or synthesize up-to-date hazard advisories by querying multiple official databases simultaneously. In order to achieve these promising developments, a systematic framework to integrate LLMs as intelligent agent is needed.

The literature demonstrates that advanced LLMs enable more integrated disaster action workflows by bridging analysis and response planning, which have traditionally functioned as discrete processes. Across reviewed works, a common architecture emerges: (1) intelligent task formulation, facilitated by either user-driven query refinement or structured LLM-initiated profiling; (2) multi-modal data retrieval – drawing on both unstructured text and structured sensor or spatial data – using mechanisms such as retrieval-augmented generation (RAG), similarity search, and on-demand code generation; and (3) contextual interpretation of results, achieved through domain-specific prompt engineering, iterative refinement with expert input, and constraint verification against external models (Table 9).

These findings indicate a shift toward AI-mediated knowledge integration that enhances real-time situational awareness and response plan formulation. Nonetheless, the literature reveals persisting limitations, including incomplete frameworks for continuous integration of real-time data, insufficient evaluation of decision transparency and reliability, and a lack of large-scale, real-world validation in operational disaster settings. Addressing these limitations will be critical for future practical deployment of LLMs in disaster management.

5. Discussion

5.1. Broadening the roles of LLMs across the disaster management cycle

Despite recent progress, the integration of state-of-the-art LLMs such as GPT-4 and Gemini 2.5 into disaster management research remains in its early stages. Our review highlights a disconnect between the rapidly evolving capabilities of LLMs and their practical adoption across the full spectrum of disaster management activities. While LLMs demonstrate sophisticated abilities in disaster response, from situation awareness to action coordination, their implementation within comprehensive disaster management frameworks remains fragmented, particularly across the four established phases of mitigation, preparedness, response, and recovery [1]. Current research has focused mainly on event data collection and the generation of emergency action plans, addressing primarily the response phase, while LLM applications in preparedness, mitigation, and recovery are largely unexplored.

However, LLMs present significant opportunities throughout the entire disaster management cycle by bridging communication gaps and synthesizing complex information to support decision-making:

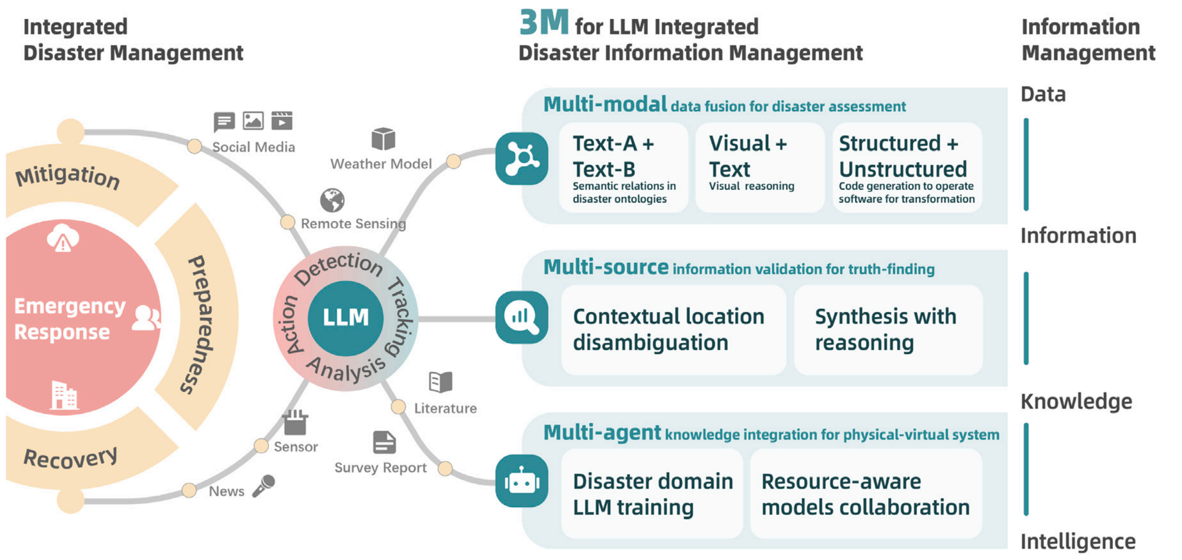

Fig. 5. Framework of incorporate LLM for integrated disaster management.

- Preparedness: LLMs can assist disaster monitoring agencies by analyzing local environmental factors, selecting suitable prediction models, and generating clear explanations of these choices. For first responders and community leaders, LLMs can summarize large amounts of historical disaster data, produce risk assessment reports tailored to specific communities, and even help design and simulate training exercises, such as role-playing coordinated action during earthquake drills.
- Recovery: After a disaster, LLMs can support managers by evaluating the severity of damage using data from public social media discussions and official channels. In addition, when connected with official disaster recovery platforms, LLMs can provide mental health guidance, self-help instructions, and connect affected individuals with medical resources—particularly valuable in disaster-prone or remote areas.
- Mitigation: LLMs can automate the generation of actionable recommendations by analyzing local hazard reports, updated regulatory documents, and lessons from similar cities worldwide. For example, LLMs can assist planners in drafting targeted mitigation plans, such as optimal evacuation routes or building retrofitting priorities, and efficiently publish relevant sections to appropriate agencies or initiatives, streamlining communication and accelerating the implementation of mitigation strategies.

These examples demonstrate how LLMs can transform real-time situation awareness into actionable, stakeholder-specific insights for researchers, monitoring agencies, responders, and the affected public. However, most existing applications of LLMs in disaster management still operate primarily at the data and information levels of the DIKW (Data-Information-Knowledge-Wisdom) hierarchy [61], focusing largely on extracting and summarizing data rather than enabling higher-level decision-making. Traditional NLP and machine learning approaches share this limitation, as their outputs are typically restricted to pre-defined tasks or categories. Although these models are useful for prediction and explanation, they often fall short in facilitating cross-domain communication and supporting urgent, complex decisions under uncertainty. In contrast, modern LLMs have the potential to fundamentally advance disaster management by offering two key capabilities: first, the integration of extensive general and disaster-specific knowledge through RAG technique; and second, the automation of analytical processes by generating tailored code and orchestrating specialized tools. Building on these strengths, the following section introduces a systematic “3M” framework to guide the comprehensive integration of LLMs into disaster information management.

5.2. “3M” framework for LLM integrated disaster information management

Building upon the above analysis, we propose a systematic framework for advancing LLM integration in disaster management. This “3M” framework addresses three critical dimensions that correspond to progressive levels in the DIKW hierarchy: multi-modal data fusion (data to information transformation), multi-source information validation (information to knowledge transformation), and multi-agent knowledge integration (knowledge to wisdom transformation) (Fig. 5).

5.2.1. Multi-modal data fusion for disaster assessment

The complexity of disaster scenarios demands seamless integration of diverse information sources to support comprehensive disaster assessment that spans multiple phases, stakeholders, scales, and disciplines. Current disaster information systems, while increasingly sophisticated in processing individual data types, face fundamental challenges in creating intelligent automation pipelines for multi-modal data fusion. Large language models offer promising capabilities for addressing these integration challenges:

Textual data synthesis across systems: Disaster information systems frequently encounter heterogeneous data streams, from real-time social media and sensor alerts to formal assessment reports, all in different technical formats. Traditional NLP models struggle to synthesize these inputs rapidly and flexibly for actionable decision-making. In contrast, LLMs not only directly interpret complex semantic meanings across diverse text types, but can also generate code to dynamically orchestrate specialized NLP tools – such as named entity recognition, sentiment analysis, or machine translation – for processing each type of text simultaneously. For example, during a hurricane, an LLM-powered platform can automate the extraction of crisis locations from emergency calls, summarize trends from social media, and integrate findings from expert field reports, generating tailored situational briefings to meet the different needs of emergency agencies. By leveraging structured emergency knowledge graphs [129] and automated tool use, LLMs enable efficient, adaptable, and fully automated information synthesis, significantly enhancing disaster decision-making processes.

Reasoning-based visual and textual data fusion: Multi-modal LLMs address the longstanding challenge of modal segregation in disaster assessment by enabling sophisticated, context-aware integration of both visual and textual information. The textual-visual-geo framework is invaluable in social sensing for built environment disruptions during disasters [130], as it significantly enhances the accuracy and relevance of information gathered from various sources. However, traditional frameworks that process satellite imagery, aerial photographs, and ground-level documentation separately from textual reports still lack robust data fusion mechanisms. These methods often result in ineffective and noisy feature combinations [131]. Recent incorporation of large language models (LLMs) allows for the extension of current statistical methods, such as re-ranking techniques, to achieve a more contextual integration of both textual and visual information, thereby facilitating better insight generation. LLMs like GPT-4 and Gemini 2.5 Pro can analyze and reason over the relationships between these modalities. For example, in post-flood damage assessment, a multi-modal LLM can seamlessly correlate satellite images with field investigation notes and news articles, detecting discrepancies such as misreported flood extents or unverified visual evidence. This "discordance-aware" processing is especially critical for validating information and detecting misinformation during crises, as demonstrated in recent work on multi-modal event classification [132], and plays an increasingly vital role in resource allocation and coordinated emergency responses. Furthermore, this capability directly addresses persistent challenges such as error-prone manual inspections, as well as the difficulties associated with handling occlusions and dynamic elements that are prevalent in traditional vision-only disaster prediction models [133].

Integration of unstructured and structured data: Advanced LLMs overcome major integration barriers between unstructured communications and specialized analytical systems by automating the translation of natural language inputs into structured analytical operations. Traditional disaster management workflows burden emergency managers with manually synthesizing outputs from diverse sources—such as meteorological models like GenCast [134], GraphCast [135], and vulnerability or cascade effect simulators [136]—often leading to bottlenecks under time pressure. In contrast, LLMs can dynamically generate code or queries to interface directly with these systems. For example, during wildfire response, an LLM can accept a manager's natural language request to "identify at-risk zones based on predicted wind changes and local fire history", then automatically orchestrate relevant data retrieval, analytics, and reporting across multiple platforms. This automation enables more efficient, adaptive, and coordinated decision-making, marking a transformative step in integrating complex analytical workflows for disaster prediction, damage assessment, and resource allocation.

5.2.2. Multi-source information validation for truth-finding

The transition from multi-modal fusion to actionable insights necessitates robust validation mechanisms in disaster management contexts. Effective disaster decision-making requires a comprehensive understanding of urban dynamics and the interplay of fast and slow change processes, as highlighted in [137], where the integration of varied spatial and temporal scales, alongside mixed-frequency data, is crucial. Traditional machine learning or deep learning models face significant obstacles in this process, due to their limited ability to reconcile data from distinct sources that employ diverse formats, terminology, and narrative styles. For instance, damage reports from governmental agencies often use formal, standard taxonomies, while social media users may post urgent, location-specific updates in informal language—sometimes with ambiguous references or missing context. Remote sensing imagery, emergency service logs, and citizen-generated content rarely align neatly, making it difficult for single-model approaches to validate and synthesize information reliably across all modalities.

In contrast, LLMs offer transformative advantages: their deep contextual understanding enables them to interpret and cross-reference differences in language and format between multiple information streams, such as reconciling a drone image's geospatial metadata with textual ground-truth from local authorities and real-time weather feeds. More importantly, LLMs can actively search for supplementary information—for example, automatically querying damage assessment databases, municipal infrastructure records, and relevant news feeds when inconsistencies are detected between initial field reports and sensor data. They can also orchestrate the use of specialized models and analytical tools, such as cascading impact prediction models on demand, thereby enabling dynamic, real-time validation and rapid escalation of urgent discrepancies. In a practical disaster scenario like an earthquake, LLMs can continuously aggregate building collapse reports from emergency services, cross-validate them with seismic sensor alerts and volunteer-submitted images, and trigger immediate follow-up investigations if validation thresholds are not met—all within a unified, adaptive workflow that addresses the urgent, diversified demands of crisis response from different disaster agencies.

Accurate event geolocation is foundational for synthesizing disaster information from multiple sources, yet varying and ambiguous references to locations – such as "just south of the old mill" or coordinates embedded in image metadata – have long challenged traditional validation pipelines. LLMs show significant advances in geolocation inference, as exemplified by recent work from Yin et al. [138], who demonstrated LLM-enhanced extraction of implicit geoinformation from noisy multimodal data (e.g., social media images, text reports) during Hurricane Harvey, addressing longstanding ambiguities in disaster reporting. Moreover, state-of-the-art models like OpenAI's O3 have outperformed expert human analysts in location inference and geographic reasoning tasks, such as the GeoGuessr challenge, by intelligently integrating visual cues and textual context. While LLM-based geolocalization still faces limitations when compared with traditional search engines [139] and leading visual-place recognition models like NetVLAD [140] or CosPlace [141], LLMs excel in orchestrating multiple tools and analytical pipelines, supporting iterative boundary determination and validation of disaster impact zones over time.

LLMs further elevate information triangulation beyond simple multi-source comparison by enabling intelligent synthesis and advanced reasoning, extending established multimodal fusion principles [142]. For example, ontology-enhanced frameworks, such as the approach by Elmhadhbi et al. [143], have improved disaster information validation through systematic relationship analysis. However, consistent and automated ontology alignment across varied disaster reports remains challenging. LLMs' nuanced understanding of disaster terminology enables sophisticated entity extraction and disambiguation, for instance, reconciling references to infrastructure damage ("bridge collapse", "riverbank washout") from distinctive sources and mapping them to a unified knowledge structure. Their ability to reason over relationships in time, space, and causality, including assessing the credibility of claims based on supporting evidence from different domains, supports more comprehensive and reliable disaster validation.

Nevertheless, challenges persist in operationalizing these capabilities within disaster management frameworks. Real-time processing optimization remains a critical constraint, with current LLM-based validation systems requiring substantial computational resources that may not be readily available in emergency contexts. Additionally, the integration of domain-specific validation rules and the maintenance of validation accuracy under extreme conditions require careful consideration. This underscores the ongoing need for human-in-the-loop verification and suggests that hybrid frameworks, combining the strengths of both LLMs and expert-driven processes, remain essential to ensuring robust and reliable disaster validation [144].

5.2.3. Multi-agent collaboration for physical-virtual system

With abundant and real-time situation awareness, the disaster decision system still faces challenges in optimal schema proposal regarding complex and evolving local scenarios. For instance, post the earthquake, decision-makers needed to synthesize rapidly incoming damage reports, medical needs, and on-the-ground rescue status across urban and rural regions. Even with access to extensive archives of global emergency response protocols and local historical reports, extracting the most relevant, situation-specific adaptation measures and align the plan with best practices remains difficult. The path forward requires robust methods to retrieve, integrate, and adapt research endeavors from both global disaster risk reduction experiences and emergent real-time sources, which is a long-lasting challenge in science for adaptation [145]. Traditional retrieval systems typically surface disconnected documents or isolated facts, often leaving practitioners with the burden of contextual adaptation. In contrast, multi-agent LLM-based systems can enable collaborative, context-aware knowledge synthesis, leveraging multiple agent perspectives and real-time negotiation to formulate actionable recommendations tailored to the unique challenges of each disaster context [146]. Such systems move beyond static retrieval, facilitating dynamic collaboration among human experts, local authorities, and AI agents to co-create robust decision support for the nuances of local disasters.

Effective LLM agency in disaster management involves addressing three principal barriers. First, task formulation and intelligent action planning remain challenging: LLMs must reliably interpret and translate broad emergency objectives into precise operational tasks, such as decomposing "optimize evacuation routes" into sequential drone-assisted damage assessment, hazard map refinement, and communication plan adjustment. Second, resource constraints in disaster emergencies hinder practical deployment, especially in physical-virtual systems. For example, autonomous search-and-rescue drones may have limited onboard hardware, making seamless real-time coordination between cloud-based language models and edge devices for victim localization or supply drops an engineering problem. Third, operational safety and data privacy are paramount in LLM-integrated autonomous systems. Systems used in disaster preparedness or recovery frequently manage sensitive data such as survivor lists, infrastructure vulnerability maps, or private communications between emergency institutions. LLM-driven recommendations influencing resource allocation or rescue priorities must be auditable, safe, and compliant with privacy requirements—particularly when agents autonomously trigger physical actions or inform critical decisions.

Advancing disaster-domain LLMs requires integrating specialized knowledge with continuous operational feedback to increase their precision and relevance in rapidly changing conditions. For instance, LLMs pre-trained on historical wildfire mitigation data and construction standards could intelligently select the most suitable prediction models for identifying at-risk zones, leveraging real-time environmental sensor input to recommend targeted mitigation strategies that consider specific terrain features and evolving weather patterns. During disasters, LLMs could dynamically facilitate interpreting live information streams, such as UAV surveillance video or field investigation reports, to recommend site-specific triage. In the recovery phase, LLMs trained on rehabilitation case studies may generate actionable, context-adapted guidance for post-disaster housing reconstruction or community relief prioritization, drawing on both local data and global knowledge repositories.

Multi-agent and multi-modal LLM systems are positioned to enhance robustness and adaptability in disaster management by dynamically assigning collaborative tasks to agents or models based on the complexity of the scenario and available computational resources. As outlined by situation awareness theory [147], such collaborations facilitate the progression from situation perception to comprehension, projection, and ultimately action execution. For example, in a major urban flood scenario, lightweight LLM agents could rapidly parse and categorize incoming social media feeds to detect flooding-affected regions. Simultaneously, advanced multi-modal assessment models – integrating satellite imagery, drone reconnaissance, and weather sensor data – could provide real-time evaluations of flood severity and infrastructure status. An LLM, operating as the intelligent assistant, pretrained on domain knowledge and equipped with a comprehensive flooding case library via RAG, can orchestrate these models, actively synthesize situational updates, and offer actionable rescue strategies—such as prioritizing medical supply routes in response to rising water levels and identifying safe zones for temporary shelters.

Human-expert integration in the decision process further enhances the reliability and transparency of these intelligent systems. Within this framework, domain-pretrained LLMs plan the action and operate specialized models, for example, statistical models for rainfall-runoff modeling, deep learning models for inundation mapping using drone imagery, and lightweight onboard models for autonomous drone delivery and search operations. LLMs also play a critical role in translating and communicating these complex, multi-agent arrangements to human decision-makers, ensuring transparency and fostering trust in the automated action plans. By mirroring established human team coordination patterns, these systems bridge the automation-practice divide—supporting both dynamic field operations and strategic decision-making in municipal recovery centers. To maintain public trust and uphold ethical standards, it is essential to incorporate robust safeguards such as privacy-preserving protocols, comprehensive decision audit trails, and continuous human-in-the-loop oversight, especially when sensitive outputs are deployed in critical disaster interventions.

In summary, the LLM-driven multi-agent frameworks and physical-virtual systems, which are built on multi-modal fusion and multi-source validation as conceptualized in the 3M framework (Fig. 5), holds transformative potential for disaster management. This roadmap enables context-adaptive, efficient, and transparent collaboration between human experts and language models. Addressing the challenges of intelligent task planning and resource optimization, while maintaining a consistent focus on reliability, safety, transparency, and societal trust, will be essential. By prioritizing these foundational principles, continued interdisciplinary research and practical deployment of LLM-integrated systems can not only enhance disaster preparedness, response, and recovery, but also set new standards for responsible and trustworthy innovation in critical emergency contexts.

6. Conclusion

As climate change intensifies the impacts of natural hazards globally, intelligent information processing has become increasingly crucial for effective disaster management. This systematic review advances the understanding of LLM applications in disaster management through three significant contributions. First, through a systematic examination of 70 LLM-relevant disaster articles within the broader landscape of AI-driven disaster management, we establish an operational phase-based review framework encompassing detection, tracking, analysis, and action phases. This framework synthesizes diverse disaster types and stakeholder needs, while revealing patterns in current LLM applications across different operational contexts. Second, by examining fundamental theories of integrated information management and the information hierarchy, we identify three critical gaps: limited progression beyond the immediate response to longer-term disaster management functions, insufficient integration across disaster types and stakeholder groups, and inadequate transformation of disaster situation awareness data into coordinated multi-stakeholder actions. These gaps reflect broader theoretical limitations in current LLM implementations, which primarily operate at the data-to-information level without fully achieving knowledge generation and intelligence emerging. Third, building upon these critical insights and emerging LLM capabilities, we propose the "3M" framework, extending existing information management theories through multi-modal data fusion, multi-source validation, and multi-agent collaboration, thereby enhancing the adaptation of LLM capabilities to evolving human-AI collaboration requirements.

Despite these contributions, several limitations should be noted. The field of LLM-integrated disaster management is rapidly evolving, with new models, benchmarks, and applications emerging at an unprecedented pace. As a result, this review may not capture the very latest advancements published after the study's search window, especially as high-impact LLM research has accelerated only recently. Additionally, while we prioritized comprehensiveness within our defined scope, necessary decisions regarding inclusion criteria, such as language and database selection, may have resulted in the omission of certain studies. Nonetheless, by systematically synthesizing the current body of research, this review aims to map the contours of this fast-emerging domain and to provide a conceptual framework that can guide and contextualize future work in intelligent disaster management.

These contributions advance disaster management theory and practice by establishing a systematic foundation for LLM integration, marking a crucial step toward proactive disaster resilience. As climate-related challenges intensify, this framework bridges the gap between technological capabilities and operational needs, offering a theoretically grounded approach for next-generation disaster information systems. The synthesis presented here not only guides future research directions but also provides practical insights for implementing intelligent disaster management solutions, ultimately supporting the global imperative of building more resilient communities in an increasingly uncertain climate future.

CRediT authorship contribution statement

Fengyi Xu: Writing – original draft, Visualization, Methodology, Conceptualization. **Jun Ma:** Writing – review & editing, Supervision, Methodology. **Nan Li:** Writing – review & editing. **Jack C.P. Cheng:** Writing – review & editing.

Declaration of Generative AI and AI-assisted technologies in the writing process

During the preparation of this work the author(s) used Claude 3.5 (Anthropic) in order to improve language clarity and readability through grammar refinement and word choice suggestions. After using this tool/service, the author(s) reviewed and edited the content as needed and take(s) full responsibility for the content of the published article.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Dr. Li, Nan is a co-author of this manuscript and serves as Associate Editor of the International Journal of Disaster Risk Reduction. Given his role as Associate Editor, Dr. Li, Nan had no involvement in the peer review of this article and had no access to information regarding its peer review. Full responsibility for the editorial process for this article was delegated to another journal editor. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The authors sincerely thank Professor Jiangping Zhou and Professor Shenjing He for their invaluable guidance and insightful suggestions, which significantly enriched this manuscript.

Data availability

No data was used for the research described in the article.

References

[1] C.f.P.R. National Governors' Association, Comprehensive Emergency Management: A Governor's Guide, Defense Civil Preparedness Agency, U.S. Department of Defense, Washington, D.C., 1979.
[2] P. Basabe, T. Beer, N. Catto, P.T. Bobrowsky, P.T. Bobrowsky, Encyclopedia of natural hazards, first ed., in: Encyclopedia of Earth Sciences Series, Springer Netherlands, Dordrecht, 2013.
[3] P.P.P.P. Karan, S.P. Subbiah, D. Gilbreath (Eds.), The Indian Ocean tsunami : The global response to a natural disaster, in: The Indian Ocean Tsunami : The Global Response To a Natural Disaster, University Press of Kentucky, 2011.
[4] S.R. Institute, Natural catastrophes in 2023: High earthquake losses, extreme weather events cause widespread damage, Sigma research, Swiss Re Institute, 2024.
[5] F.E.M. Agency, Hurricane Frances Morning Briefing, Emergency Response Briefing, Federal Emergency Management Agency, Florida, 2004, Briefing materials presented during Hurricane Frances emergency response.
[6] M.A. Dulebenets, O.F. Abioye, E.E. Ozguven, R. Moses, W.R. Boot, T. Sando, Development of statistical models for improving efficiency of emergency evacuation in areas with vulnerable population, Reliab. Eng. Syst. Saf. 182 (2019) 233–249, http://dx.doi.org/10.1016/j.ress.2018.09.021.
[7] M.A. Dulebenets, J. Pasha, O.F. Abioye, M. Kavoosi, E.E. Ozguven, R. Moses, W.R. Boot, T. Sando, Exact and heuristic solution algorithms for efficient emergency evacuation in areas with vulnerable populations, Int. J. Disaster Risk Reduct. 39 (2019-10-01) 101114.
[8] A. Fathianpour, B. Evans, M.B. Jelodar, S. Wilkinson, Tsunami evacuation modelling via micro-simulation model, Prog. Disaster Sci. 17 (2023-01-01) 100277, http://dx.doi.org/10.1016/j.pdisas.2023.100277.
[9] C. Li, Z. Han, Y. Li, M. Li, W. Wang, J. Dou, L. Xu, G. Chen, Physical information-fused deep learning model ensembled with a subregion-specific sampling method for predicting flood dynamics, J. Hydrol. 620 (2023-05) 129465, http://dx.doi.org/10.1016/j.jhydrol.2023.129465.
[10] G. Zhang, M. Wang, K. Liu, Deep neural networks for global wildfire susceptibility modelling, Ecol. Indic. 127 (2021) 107735, http://dx.doi.org/10.1016/j.ecolind.2021.107735.
[11] O.M. Saad, Y. Chen, A. Savvaidis, S. Fomel, Y. Chen, Real-time earthquake detection and magnitude estimation using vision transformer, J. Geophys. Res. Solid Earth 127 (5) (2022) e2021JB023657.
[12] G. Camps-Valls, M.-Á. Fernández-Torres, K.-H. Cohrs, A. Höhl, A. Castelletti, A. Pacal, C. Robin, F. Martinuzzi, I. Papoutsis, I. Prapas, J. Pérez-Aracil, K. Weigel, M. Gonzalez-Calabuig, M. Reichstein, M. Rabel, M. Giuliani, M.D. Mahecha, O.-I. Popescu, O.J. Pellicer-Valero, S. Ouala, S. Salcedo-Sanz, S. Sippel, S. Kondylatos, T. Happé, T. Williams, Artificial intelligence for modeling and understanding extreme weather and climate events, Nat. Commun. 16 (1) (2025-02-24) 1919, http://dx.doi.org/10.1038/s41467-025-56573-8.
[13] D.N. Ford, C.M. Wolf, Smart cities with digital twin systems for disaster management, J. Manag. Eng. 36 (4) (2020) 04020027, http://dx.doi.org/10.1061/(ASCE)ME.1943-5479.0000779.
[14] C. Fan, C. Zhang, A. Yahja, A. Mostafavi, Disaster city digital twin: A vision for integrating artificial and human intelligence for disaster management, Int. J. Inf. Manage. 56 (2021-02-01) 102049, http://dx.doi.org/10.1016/j.ijinfomgt.2019.102049.
[15] M.R.M.F. Ariyachandra, G. Wedawatta, Digital twin smart cities for disaster risk management: A review of evolving concepts, Sustainability 15 (15) (2023) 11910.
[16] C. Mydlarz, P. Sai Venkat Challagonda, B. Steers, J. Rucker, T. Brain, B. Branco, H.E. Burnett, A. Kaur, R. Fischman, K. Graziano, K. Krueger, E. Hénaff, V. Ignace, E. Jozwiak, J. Palchuri, P. Pierone, P. Rothman, R. Toledo-Crow, A.I. Silverman, FloodNet: Low-cost ultrasonic sensors for real-time measurement of hyperlocal, street-level floods in new york city, Water Resour. Res. 60 (5) (2024-05) e2023WR036806, http://dx.doi.org/10.1029/2023WR036806.
[17] A. Akhyar, M. Asyraf Zulkifley, J. Lee, T. Song, J. Han, C. Cho, S. Hyun, Y. Son, B.-W. Hong, Deep artificial intelligence applications for natural disaster management systems: A methodological review, Ecol. Indic. 163 (2024-06) 112067, http://dx.doi.org/10.1016/j.ecolind.2024.112067.
[18] E.L. Quarantelli, Ten Research Derived Principles of Disaster Planning, University Archives, 1982, A, Carton: 5, Folder: 52, Penelope Ploughman Love Canal collection, MS-0100.
[19] Q. Ge, Integrated assessment of natural disaster risks in China, in: Integrated Assessment of Natural Disaster Risks in China, first ed., Scientific Press, Beijing, 2008.
[20] E. Commission, Emergency Response Coordination Centre (ERCC), European Civil Protection and Humanitarian Aid Operations, 2024, Website. Accessed 09 February 2024.
[21] L. Palen, K.M. Anderson, Crisis informatics—New data for extraordinary times, Sci. (Am. Assoc. Adv. Sci.) 353 (6296) (2016) 224–225.
[22] C. Reuter, M.-A. Kaufhold, Fifteen years of social media in emergencies: A retrospective review and future directions for crisis informatics, J. Contingencies Crisis Manag. 26 (1) (2018) 41–57.
[23] L.G. Canton, Emergency Management: Concepts and Strategies for Effective Programs, John Wiley & Sons, 2019.
[24] T. Sakaki, M. Okazaki, Y. Matsuo, Earthquake shakes Twitter users: Real-time event detection by social sensors, in: Proceedings of the 19th International Conference on World Wide Web, ACM, 2010, pp. 851–860, http://dx.doi.org/10.1145/1772690.1772777.

[25] M. Imran, C. Castillo, F. Diaz, S. Vieweg, Processing social media messages in mass emergency: A survey, ACM Comput. Surv. 47 (4) (2015) 1–38, http://dx.doi.org/10.1145/2771588.

[26] C. Zhang, C. Fan, W. Yao, X. Hu, A. Mostafavi, Social media for intelligent public information and warning in disasters: An interdisciplinary review, Int. J. Inf. Manage. 49 (2019) 190–207, http://dx.doi.org/10.1016/j.ijinfomgt.2019.04.004.

[27] S. Vieweg, A.L. Hughes, K. Starbird, L. Palen, Microblogging during two natural hazards events: what twitter may contribute to situational awareness, in: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '10, Association for Computing Machinery, New York, NY, USA, 2010, pp. 1079–1088, http://dx.doi.org/10.1145/1753326.1753486.

[28] United States Department of Homeland Security, Federal Emergency Management Agency – AI Use Cases, U.S. Department of Homeland Security, 2024, DHS Artificial Intelligence Use Case Inventory, (Accessed 08 February 2025).

[29] C.C.A. Commission, The 14th Five-Year Plan for National Informatization, Technical Report, State Council of the People's Republic of China, 2021, Document No. 40 of 2021.

[30] E.C.J.R. Centre, Artificial Intelligence Applied to Disasters and Crises Management, Publications Office, 2024.

[31] A. Tounsi, M. Temimi, A systematic review of natural language processing applications for hydrometeorological hazards assessment, Nat. Hazards 116 (3) (2023-04) 2819–2870, http://dx.doi.org/10.1007/s11069-023-05842-0.

[32] T. Alcantara, O. Garcia-Vazquez, H. Calvo, J.A. Torres-Leon, Disaster tweets: Analysis from the metaphor perspective and classification using LLM's, in: H. Calvo, L. Martinez-Villasenor, H. Ponce (Eds.), in: Advances in Soft Computing, Micai 2023, Pt II, vol. 14392, Springer International Publishing Ag, Cham, 2024, pp. 106–117.

[33] J.R. Chowdhury, C. Caragea, D. Caragea, Cross-lingual disaster-related multi-label tweet classification with manifold mixup, in: 58th Annual Meeting of the Association for Computational Linguistics (ACL 2020): Student Research Workshop, Assoc Computational Linguistics-Acl, Stroudsburg, 2020, pp. 292–298.

[34] J. Qian, Y. Du, F. Liang, J. Yi, N. Wang, W. Tu, S. Huang, T. Pei, T. Ma, Quantifying urban linguistic diversity related to rainfall and flood across China with social media data, ISPRS Int. J. Geo Inf. 13 (3) (2024) 92, http://dx.doi.org/10.3390/ijgi13030092.

[35] C. Wang, D. Engler, X. Li, J. Hou, D.J. Wald, K. Jaiswal, S. Xu, Near-real-time earthquake-induced fatality estimation using crowdsourced data and large-language models, Int. J. Disaster Risk Reduct. 111 (2024) 104680, http://dx.doi.org/10.1016/j.ijdrr.2024.104680.

[36] W. Du, C. Ge, S. Yao, N. Chen, L. Xu, Applicability analysis and ensemble application of BERT with TF-IDF, TextRank, MMR, and LDA for topic classification based on flood-related VGI, ISPRS Int. J. Geo Inf. 12 (6) (2023) 240, http://dx.doi.org/10.3390/ijgi12060240.

[37] Y. Sun, C. Wang, Y. Peng, Reducing hallucinations: Enhancing VQA for flood disaster damage assessment with visual contexts, in: Proceedings of International Conference on Algorithms, Software Engineering, and Network Security, Asens 2024, Assoc Computing Machinery, New York, 2024, pp. 587–590, http://dx.doi.org/10.1145/3677182.3677287.

[38] M. Imran, C. Castillo, F. Diaz, P. Meier, Crisis computing: Leveraging social media for disaster management, IEEE Intell. Syst. 29 (6) (2014) 64–71.

[39] M. Imran, S. Elbassuoni, C. Castillo, F. Diaz, P. Meier, Practical extraction of disaster-relevant information from social media, in: Proceedings of the 22nd International Conference on World Wide Web, in: WWW '13 Companion, Association for Computing Machinery, New York, NY, USA, 2013, pp. 1021–1024, http://dx.doi.org/10.1145/2487788.2488109.

[40] K. Ma, M. Tian, Y. Tan, Q. Qiu, Z. Xie, R. Huang, Ontology-based BERT model for automated information extraction from geological hazard reports, J. Earth Sci. 34 (5) (2023) 1390–1405.

[41] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A.N. Gomez, Ł. Kaiser, I. Polosukhin, Attention is all you need, in: Advances in Neural Information Processing Systems, vol. 30, Curran Associates, Inc., 2017, pp. 5998–6008.

[42] X. Qiu, T. Sun, Y. Xu, Y. Shao, N. Dai, X. Huang, Pre-trained models for natural language processing: A survey, Sci. China Technol. Sci. 63 (2020) 1872–1897, http://dx.doi.org/10.1007/s11431-020-1647-3.

[43] OpenAI, Openai o3 and o3-mini: Next generation simulated reasoning models, Technical Report, OpenAI, 2024, Announced December 20, 2024.

[44] Anthropic, Introducing Claude 3.5 Sonnet, Technical Report, Anthropic, 2024, Claude 3.5 family includes Sonnet and Haiku models, with Sonnet setting new industry benchmarks for graduate-level reasoning (GPQA), undergraduate-level knowledge (MMLU), and coding proficiency.

[45] DeepSeek, DeepSeek: Architecture and Capabilities, Technical Report, DeepSeek, Shanghai, China, 2025.

[46] G. Team, G. Research, Gemini: A Family of Highly Capable Multimodal Models, Technical Report, Google, 2023.

[47] OpenAI, Launching O3-mini in the API, Technical Report, OpenAI, 2025, Released January 31, 2025.

[48] DeepSeek, DeepSeek-7B: A Practical Large Language Model for Resource-Constrained Systems, Technical Report, DeepSeek, Shanghai, China, 2024.

[49] M. AI, LLaMA-3-8B: A Scalable Language Model for Versatile Applications, Technical Report, Meta AI, Menlo Park, CA, 2024.

[50] J. Devlin, M.-W. Chang, K. Lee, K. Toutanova, BERT: Pre-training of deep bidirectional transformers for language understanding, in: Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Association for Computational Linguistics, 2019, pp. 4171–4186.

[51] Y. Liu, M. Ott, N. Goyal, J. Du, M. Joshi, D. Chen, O. Levy, M. Lewis, L. Zettlemoyer, V. Stoyanov, RoBERTa: A robustly optimized BERT pretraining approach, 2019.

[52] V. Sanh, L. Debut, J. Chaumond, T. Wolf, DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter, 2019.

[53] Z. Lan, M. Chen, S. Goodman, K. Gimpel, P. Sharma, R. Soricut, ALBERT: A lite BERT for self-supervised learning of language representations, in: International Conference on Learning Representations, 2020.

[54] C. Raffel, N. Shazeer, A. Roberts, K. Lee, S. Narang, M. Matena, Y. Zhou, W. Li, P.J. Liu, Exploring the limits of transfer learning with a unified text-to-text transformer, J. Mach. Learn. Res. 21 (1) (2020).

[55] M. Lewis, Y. Liu, N. Goyal, M. Ghazvininejad, A. Mohamed, O. Levy, V. Stoyanov, L. Zettlemoyer, BART: Denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension, in: D. Jurafsky, J. Chai, N. Schluter, J. Tetreault (Eds.), Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, Association for Computational Linguistics, Online, 2020, pp. 7871–7880.

[56] R. Bommasani, D.A. Hudson, E. Adeli, R. Altman, S. Arora, S. von Arx, M.S. Bernstein, J. Bohg, A. Bosselut, E. Brunskill, et al., On the opportunities and risks of foundation models, ACM Comput. Surv. 55 (5) (2022) 1–166, http://dx.doi.org/10.1145/3560324.

[57] J. He, C. Gao, S. Wang, Y. Yang, X. Xie, Understanding fine-tuning of pre-trained language models: A survey, IEEE Trans. Pattern Anal. Mach. Intell. 45 (12) (2023) 14086–14110, http://dx.doi.org/10.1109/TPAMI.2023.3236085.

[58] E.J. Hu, Y. Shen, P. Wallis, Z. Allen-Zhu, Y. Li, S. Wang, W. Chen, LoRA: Low-rank adaptation of large language models, in: International Conference on Learning Representations, 2021.

[59] J. Wei, X. Wang, D. Schuurmans, M. Bosma, B. Ichter, F. Xia, E. Chi, Q. Le, D. Zhou, Chain of thought prompting elicits reasoning in large language models, in: Advances in Neural Information Processing Systems, vol. 35, 2022, pp. 24824–24837.

[60] M.J. Page, J.E. McKenzie, P.M. Bossuyt, I. Boutron, T.C. Hoffmann, C.D. Mulrow, L. Shamseer, J.M. Tetzlaff, E.A. Akl, S.E. Brennan, R. Chou, J. Glanville, J.M. Grimshaw, A. Hróbjartsson, M.M. Lalu, T. Li, E.W. Loder, E. Mayo-Wilson, S. McDonald, L.A. McGuinness, L.A. Stewart, J. Thomas, A.C. Tricco, V.A. Welch, P. Whiting, D. Moher, The PRISMA 2020 statement: An updated guideline for reporting systematic reviews, BMJ (2021) n71.

[61] R.L. Ackoff, From data to wisdom, J. Appl. Syst. Anal. 16 (1) (1989) 3–9.

[62] D. Driess, F. Xia, M.S.M. Sajjadi, C. Lynch, A. Chowdhery, B. Ichter, A. Wahid, J. Tompson, Q. Vuong, T. Yu, W. Huang, Y. Chebotar, P. Sermanet, K. Duan, V. Vanhoucke, K. Hausman, S. Levine, K. Greff, I. Mordatch, M. Toussaint, P. Florence, PaLM-E: An embodied multimodal language model, in: International Conference on Machine Learning, PMLR, 2023, pp. 8789–8811.

[63] Z. Durante, Q. Huang, N. Wake, R. Gong, J.S. Park, B. Sarkar, R. Taori, Y. Noda, D. Terzopoulos, Y. Choi, K. Ikeuchi, H. Vo, L. Fei-Fei, J. Gao, Agent AI: Surveying the horizons of multimodal interaction, 2024, Retrieved from https://arxiv.org/abs/2401.03568.
[64] Z. Wang, T. Zhu, S. Mai, Disaster detector on twitter using bidirectional fincoder representation from transformers with keyword position information, in: H.B. Sun (Ed.), Proceedings of 2020 IEEE 2nd International Conference on Civil Aviation Safety and Information Technology, ICCASIT, IEEE, New York, 2020, pp. 474–477, http://dx.doi.org/10.1109/ICCASIT50869.2020.9368610.
[65] S. Deb, A.K. Chanda, Comparative analysis of contextual and context-free embeddings in disaster prediction from twitter data, Mach. Learn. Appl. 7 (2022) 100253, http://dx.doi.org/10.1016/j.mlwa.2022.100253.
[66] G. Indra, N. Duraipandian, Modeling of optimal deep learning based flood forecasting model using twitter data, Intell. Autom. Soft Comput. 35 (2) (2023) 1455–1470, http://dx.doi.org/10.32604/iasc.2023.027703.
[67] S. Saleem, M. Mehrotra, Context-aware transfer learning approach to detect informative social media content for disaster management, Int. J. Adv. Comput. Sci. Appl. 15 (1) (2024) 680–689.
[68] Y. Suzuki, Y. Yoneda, A. Nadamoto, Analysis of behavioral facilitation tweets for large-scale natural disasters dataset using machine learning, in: C. Strauss, G. Kotsis, A.M. Tjoa, I. Khalil (Eds.), in: Database and Expert Systems Applications, DEXA 2021, Pt II, vol. 12924, Springer International Publishing Ag, Cham, 2021, pp. 161–169, http://dx.doi.org/10.1007/978-3-030-86475-0_16.
[69] F. Yamamoto, Y. Suzuki, A. Nadamoto, Extraction and analysis of regionally specific behavioral facilitation information in the event of a large-scale disaster, in: 2021 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology, WI-IAT 2021, Assoc Computing Machinery, New York, 2021, pp. 538–543, http://dx.doi.org/10.1145/3486622.3493991.
[70] L. Diao, Z. Tang, X. Guo, Z. Bai, S. Lu, L. Li, Weibo disaster rumor recognition method based on adversarial training and stacked structure, KSII Trans. Internet Inf. Syst. 16 (10) (2022) http://dx.doi.org/10.3837/t?s.2022.10.001.
[71] A. Hembree, A. Beggs, T. Marshall, E.N. Ceesay, Decoding linguistic ambiguity in times of emergency based on twitter disaster datasets, in: R. Paul (Ed.), 2021 IEEE 11th Annual Computing and Communication Workshop and Conference, CCWC, IEEE, New York, 2021, pp. 1239–1244, http://dx.doi.org/10.1109/CCWC51732.2021.9375916.
[72] C. Fan, F. Wu, A. Mostafavi, A hybrid machine learning pipeline for automated mapping of events and locations from social media in disasters, IEEE Access 8 (2020) 10478–10490, http://dx.doi.org/10.1109/ACCESS.2020.2965550.
[73] P. Babvey, G. Gongora-Svartzman, C. Lipizzi, J.E. Ramirez-Marquez, Content-based user classifier to uncover information exchange in disaster-motivated networks, PLoS ONE 16 (11) (2021) e0259342, http://dx.doi.org/10.1371/journal.pone.0259342.
[74] S. Kamoji, M. Kalla, Effective flood prediction model based on twitter text and image analysis using BMLP and SDAE-HHNN, Eng. appl. Artif. Intell. 123 (2023) 106365, http://dx.doi.org/10.1016/j.engappai.2023.106365.
[75] A. Romascanu, H. Ker, R. Sieber, S. Greenidge, S. Lumley, D. Bush, S. Morgan, R. Zhao, M. Brunila, Using deep learning and social network analysis to understand and manage extreme flooding, J. Contingencies Crisis Manage. 28 (3) (2020) 251–261.
[76] B. Zhou, L. Zou, A. Mostafavi, B. Lin, M. Yang, N. Gharaibeh, H. Cai, J. Abedin, D. Mandal, VictimFinder: Harvesting rescue requests in disaster response from social media with BERT, Comput. Environ. Urban Syst. 95 (2022) 101824, http://dx.doi.org/10.1016/j.compenvurbsys.2022.101824.
[77] S. Ghosh, S. Maji, M.S. Desarkar, Unsupervised domain adaptation with global and local graph neural networks under limited supervision and its application to disaster response, IEEE Trans. Comput. Soci. Sys. 10 (2) (2023) 551–562.
[78] H.M. Zahera, R. Jalota, M.A. Sherif, A.-C.N. Ngomo, I-AID: Identifying actionable information from disaster-related tweets, IEEE Access 9 (2021) 118861–118870, http://dx.doi.org/10.1109/ACCESS.2021.3107812.
[79] R.S. Wilkho, S. Chang, N.G. Gharaibeh, FF-BERT: A BERT-based ensemble for automated classification of web-based text on flash flood events, Adv. Eng. Inform. 59 (2024) 102293, http://dx.doi.org/10.1016/j.aei.2023.102293.
[80] R. Suwaileh, T. Elsayed, M. Imran, H. Sajjad, When a disaster happens, we are ready: Location mention recognition from crisis tweets*, Int. J. Disaster Risk Reduct. 78 (2022) 103107, http://dx.doi.org/10.1016/j.ijdrr.2022.103107.
[81] S. Naaz, Z. Ul Abedin, D.R. Rizvi, Sequence classification of tweets with transfer learning via BERT in the field of disaster management, EAI Endorsed Trans. scalable Inf. Syst. 8 (31) (2021) 8.
[82] K. Yin, C. Liu, A. Mostafavi, X. Hu, CrisisSense-LLM: Instruction fine-tuned large language model for multi-label social media text classification in disaster informatics, 2024, URL https://arxiv.org/abs/2406.15477.
[83] V. Balakrishnan, Z. Shi, C.L. Law, R. Lim, L.L. Teh, Y. Fan, J. Periasamy, A comprehensive analysis of transformer-deep neural network models in twitter disaster detection, Mathematics 10 (24) (2022) 4664, http://dx.doi.org/10.3390/math10244664.
[84] R. Franceschini, A. Rosi, F. Catani, N. Casagli, Detecting information from twitter on landslide hazards in italy using deep learning models, GEOENVIRONMENTAL DISASTERS 11 (1) (2024) 22, http://dx.doi.org/10.1186/s40677-024-00279-4.
[85] S. Khanmohammadi, E. Golafshani, Y. Bai, H. Li, M. Bazli, M. Arashpour, Multi-modal mining of crowd-sourced data: Efficient provision of humanitarian aid to remote regions affected by natural disasters, Int. J. Disaster Risk Reduct. 96 (2023) 103972, http://dx.doi.org/10.1016/j.ijdrr.2023.103972.
[86] H. Krishnan, A. Roy, A.K. Menon, S. Devnandan, H.M. Babu, Natural disaster detection using social media, in: J.C. Prasad (Ed.), 2023 Advanced Computing and Communication Technologies for High Performance Applications, ACCTHPA, IEEE, New York, 2023, http://dx.doi.org/10.1109/ACCTHPA57160.2023.10083339.
[87] P.T. Bobrowsky (Ed.), Encyclopedia of natural hazards, in: Encyclopedia of Earth Sciences Series, Springer, 2013.
[88] J.M. Franc, A.J. Hertelendy, L. Cheng, R. Hata, M. Verde, Accuracy of a commercial large language model (ChatGPT) to perform disaster triage of simulated patients using the simple triage and rapid treatment (START) protocol: Gage repeatability and reproducibility study, J. Med. Internet Res. 26 (2024) e55648, http://dx.doi.org/10.2196/55648.
[89] S. Koju, K. Takeuchi, A. Watanabe, T. Hirayama, H. Nakao, Estimating task priority in japanese disaster chronology logs, in: Proceedings of 2023 7th International Conference on Natural Language Processing and Information Retrieval, NLPIR 2023, Assoc Computing Machinery, New York, 2023, pp. 304–309, http://dx.doi.org/10.1145/3639233.3639248.
[90] J. Han, Z. Zheng, X.-Z. Lu, K.-Y. Chen, J.-R. Lin, Enhanced earthquake impact analysis based on social media texts via large language model, Int. J. Disaster Risk Reduct. 109 (2024) 104574.
[91] A.K. Kanth, P. Chitra, G.G. Sowmya, Deep learning-based assessment of flood severity using social media streams, Stoc. Environ. Res. Risk Assess. 36 (2) (2022) 473–493, http://dx.doi.org/10.1007/s00477-021-02161-3.
[92] C. Songchon, G. Wright, L. Beevers, The use of crowdsourced social media data to improve flood forecasting, J. Hydrol. 622 (2023-07) 129703, http://dx.doi.org/10.1016/j.jhydrol.2023.129703.
[93] S.M. Mousavi, M. Stogaitis, T. Gadh, R.M. Allen, A. Barski, R. Bosch, P. Robertson, N. Thiruverahan, Y. Cho, A. Raj, Gemini & physical world: Large language models can estimate the intensity of earthquake shaking from multi-modal social media posts, 2024, Retrieved from http://arxiv.org/abs/2405.18732.
[94] T. Akinboyewa, H. Ning, M.N. Lessani, Z. Li, Automated floodwater depth estimation using large multimodal model for rapid flood mapping, Comput. Urban Sci. 4 (1) (2024) 12, http://dx.doi.org/10.1007/s43762-024-00123-3.
[95] A.W. Ziaullah, F. Ofli, M. Imran, Monitoring critical infrastructure facilities during disasters using large language models, 2024, Retrieved from http://arxiv.org/abs/2404.14432.

[96] R. Ou, H. Yan, M. Wu, C. Zhang, A method of efficient synthesizing post-disaster remote sensing image with diffusion model and LLM, in: 2023 Asia Pacific Signal and Information Processing Association Annual Summit and Conference, Apsipa Asc, IEEE, New York, 2023, pp. 1549–1555.
[97] W. Gao, X. Zhu, Y. Wang, L. Li, Detecting disaster-related tweets via multimodal adversarial neural network, IEEE Multimedia 27 (4) (2020) 28–37, http://dx.doi.org/10.1109/MMUL.2020.3012675.
[98] Z. Wang, M. Huang, C. Li, J. Feng, S. Liu, G. Yang, Intelligent recognition of key earthquake emergency Chinese information based on the optimized BERT-BiLSTM-CRF algorithm, Applied Sci. Basel 13 (5) (2023) 3024, http://dx.doi.org/10.3390/app13053024.
[99] Z. Li, X. Zhang, Research on named entity recognition methods for urban underground space disasters based on text information extraction, in: N. El-Sheimy, A.A. Abdelbary, N. El-Bendary, Y. Mohasseb (Eds.), Geospatial Week 2023, vol. 48-1, Copernicus Gesellschaft Mbh, Gottingen, 2023, pp. 547–552, http://dx.doi.org/10.5194/isprs-archives-XLVIII-1-W2-2023-547-2023.
[100] Y. Hao, J. Qi, X. Ma, S. Wu, R. Liu, X. Zhang, An LLM-based inventory construction framework of urban ground collapse events with spatiotemporal locations, ISPRS Int. J. Geo Inf. 13 (4) (2024) 133, http://dx.doi.org/10.3390/ijgi13040133.
[101] Y. Zhang, Z. Chen, X. Zheng, N. Chen, Y. Wang, Extracting the location of flooding events in urban systems and analyzing the semantic risk using social sensing data, J. Hydrol. 603 (2021) http://dx.doi.org/10.1016/j.jhydrol.2021.127053.
[102] X. Lei, W. Song, R. Fan, R. Feng, L. Wang, Semi-supervised geological disasters named entity recognition using few labeled data, Geoinformatica 27 (2) (2023) 263–288, http://dx.doi.org/10.1007/s10707-022-00474-1.
[103] Y. Hu, G. Mai, C. Cundy, K. Choi, N. Lao, W. Liu, G. Lakhanpal, R.Z. Zhou, K. Joseph, Geo-knowledge-guided GPT models improve the extraction of location descriptions from disaster-related social media messages, Int. J. Geogr. Inf. Sci. 37 (11) (2023) 2289–2318.
[104] L. Fei, T. Li, W. Ding, Dempster–Shafer theory-based information fusion for natural disaster emergency management: A systematic literature review, Inf. Fusion 112 (2024-12) 102585.
[105] S.V. Praveen, R. Deepika, Unveiling the perceptions of syrian and turkish citizens afflicted by survivor guilt in the aftermath of the 2023 earthquake: A study based on deep learning, Asian J. Psychiatry 86 (2023) 103672, http://dx.doi.org/10.1016/j.ajp.2023.103672.
[106] T. Bui, A. Hannah, S. Madria, R. Nabaweesi, E. Levin, M. Wilson, L. Nguyen, Emotional health and climate-change-related stressor extraction from social media: A case study using hurricane harvey, Mathematics 11 (24) (2023) 4910, http://dx.doi.org/10.3390/math11244910.
[107] Z. Ahmad, A. Kumar, A. Ekbal, P. Bhattacharyya, Emotion driven crisis response: A benchmark setup for multi-lingual emotion analysis in disaster situations, in: 2021 International Joint Conference on Neural Networks, IJCNN, IEEE, New York, 2021.
[108] Y. Oikawa, M. Ptaszynski, F. Masui, AI training for thunderstorm training: Better situational awareness for disaster tweets using context and emotions, in: 2022 10th International Conference on Affective Computing and Intelligent Interaction Workshops and Demos, ACIIW, IEEE, New York, 2022, pp. 1–10, http://dx.doi.org/10.1109/ACIIW57231.2022.10086020.
[109] X. Chen, Monitoring of public opinion on typhoon disaster using improved clustering model based on single-pass approach, Sage Open 13 (3) (2023) 21582440231200098, http://dx.doi.org/10.1177/21582440231200098.
[110] M. Zhang, Q. Huang, H. Liu, A multimodal data analysis approach to social media during natural disasters, Sustainability 14 (9) (2022) 5536, http://dx.doi.org/10.3390/su14095536.
[111] Z. Ma, L. Li, L. Hemphill, G.B. Baecher, Y. Yuan, Investigating disaster response for resilient communities through social media data and the susceptible-infected-recovered (SIR) model: A case study of 2020 western US wildfire season, Sustain. Cities Soc. 106 (2024) 105362.
[112] H. Liu, N. Luo, Q. Zhao, Research on the construction of typhoon disaster chain based on Chinese web corpus, J. Marine Sci. Eng. 10 (1) (2022) 44, http://dx.doi.org/10.3390/jmse10010044.
[113] Y.-C. Liu, C.-L. Kuo, Constructing spatio-temporal disaster knowledge graph from social media, in: A. Heppenstall, M. Wang, U. Demsar, R. Lemmens, J. Yao (Eds.), 27TH Agile Conference on Geographic Information Science Geographic Information Science for a Sustainable Future, 5, Copernicus Gesellschaft Mbh, Gottingen, 2024, p. 37, http://dx.doi.org/10.5194/agile-giss-5-37-2024.
[114] W. Li, L. Wu, X. Xu, Z. Xie, Q. Qiu, H. Liu, Z. Huang, J. Chen, Deep learning and network analysis: Classifying and visualizing geologic hazard reports, J. Earth Sci. 35 (4) (2024) 1289–1303.
[115] P. Zhang, H. Zhang, F. Kong, Research on online public opinion in the investigation of the "7-20" extraordinary rainstorm and flooding disaster in Zhengzhou, China, Int. J. Disaster Risk Reduct. 105 (2024) http://dx.doi.org/10.1016/j.ijdrr.2024.104422.
[116] J.J.L. Rachel, A. Bhuvaneswari, An xlnet based target dependent sentiment classification along with a smart webapp for post-disaster management using django, in: T. Senjyu, C. So-In, A. Joshi (Eds.), in: Smart Trends in Computing and Communications, VOL 4, SMARTCOM 2024, vol. 948, Springer-Verlag Singapore Pte Ltd, Singapore, 2024, pp. 419–431, http://dx.doi.org/10.1007/978-981-97-1329-5_34.
[117] B. Yan, F. Mai, C. Wu, R. Chen, X. Li, A computational framework for understanding firm communication during disasters, Inf. Syst. Res. 35 (2) (2024) 590–608, http://dx.doi.org/10.1287/isre.2022.0128.
[118] K. Wolf, D. Winecki, A. Nandi, Camera-first form filling: Reducing the friction in climate hazard reporting, in: Workshop on HUMAN-in-the-Loop Data Analytics, Hilda 2023, Assoc Computing Machinery, New York, 2023, pp. 1–7, http://dx.doi.org/10.1145/3597465.3605218.
[119] G. Colverd, P. Darm, L. Silverberg, N. Kasmanoff, FloodBrain: Flood disaster reporting by web-based retrieval augmented generation with an LLM, 2023, URL https://arxiv.org/abs/2311.02597.
[120] A. Khoshkonesh, R. Nazari, M.R. Nikoo, M. Karimi, Enhancing flood risk assessment in urban areas by integrating hydrodynamic models and machine learning techniques, Sci. Total Environ. 952 (2024-11) 175859.
[121] V.G. Goecks, N.R. Waytowich, DisasterResponseGPT: Large language models for accelerated plan of action development in disaster response scenarios, 2023, Retrieved from http://arxiv.org/abs/2306.17271.
[122] M. Zhang, J. Wang, Automatic extraction of flooding control knowledge from rich literature texts using deep learning, APPLIED SCIENCES- BASEL 13 (4) (2023) 2115, http://dx.doi.org/10.3390/app13042115.
[123] I. D'Alessio, A. Quaglieri, J. Burrai, A. Pizzo, E. Mari, U. Aitella, G. Lausi, G. Tagliaferri, P. Cordellieri, A.M. Giannini, C. Cricenti, "Leading through crisis": A systematic review of institutional decision-makers in emergency contexts, Behav. Sci. 14 (6) (2024-06-06) 481, http://dx.doi.org/10.3390/bs14060481.
[124] Y. Xie, B. Jiang, T. Mallick, J.D. Bergerson, J.K. Hutchison, D.R. Verner, J. Branham, M.R. Alexander, R.B. Ross, Y. Feng, L.-A. Levy, W. Su, C.J. Taylor, WildfireGPT: Tailored large language model for wildfire analysis, 2024, Retrieved from http://arxiv.org/abs/2402.07877.
[125] Y.J. Chew, S.Y. Ooi, Y.H. Pang, Z.Y. Lim, Framework to create inventory dataset for disaster behavior analysis using google earth engine: A case study in peninsular malaysia for historical forest fire behavior analysis, Forests 15 (6) (2024) 923, http://dx.doi.org/10.3390/f15060923.
[126] R. Chandra, S.S. Kumar, R. Patra, S. Agarwal, Decision support system for forest fire management using ontology with big data and LLMs, 2024, URL https://arxiv.org/abs/2405.11346.
[127] Y. Xia, Y. Huang, Q. Qiu, X. Zhang, L. Miao, Y. Chen, A question and answering service of typhoon disasters based on the T5 large language model, ISPRS Int. J. Geo Inf. 13 (5) (2024) 165, http://dx.doi.org/10.3390/ijgi13050165.
[128] J. Zhu, P. Dang, Y. Cao, J. Lai, Y. Guo, P. Wang, W. Li, A flood knowledge-constrained large language model interactable with GIS: Enhancing public risk perception of floods, Int. J. Geogr. Inf. Sci. 38 (4) (2024) 603–625, http://dx.doi.org/10.1080/13658816.2024.2306167.
[129] M. Chen, Z. Tao, W. Tang, T. Qin, R. Yang, C. Zhu, Enhancing emergency decision-making with knowledge graphs and large language models, Int. J. Disaster Risk Reduct. 113 (2024-10) 104804, http://dx.doi.org/10.1016/j.ijdrr.2024.104804.
[130] C. Fan, Y. Jiang, A. Mostafavi, Social sensing in disaster city digital twin: Integrated textual–Visual–geo framework for situational awareness during built environment disruptions, J. Manag. Eng. 36 (3) (2020) 04020002.

[131] S. Abdali, S. Shaham, B. Krishnamachari, Multi-modal misinformation detection: Approaches, challenges and opportunities, 2024-09-18, URL http://arxiv.org/abs/2203.13883.
[132] A. El-Niss, A. Alzu'Bi, A. Abuarqoub, Multimodal fusion for disaster event classification on social media: A deep federated learning approach, in: Proceedings of the 7th International Conference on Future Networks and Distributed Systems, ACM, 2023-12-21, pp. 758–763.
[133] M. Kamari, Y. Ham, AI-based risk assessment for construction site disaster preparedness through deep learning-based digital twinning, Autom. Constr. 134 (2022) 104091, http://dx.doi.org/10.1016/j.autcon.2021.104091.
[134] I. Price, A. Sanchez-Gonzalez, F. Alet, T.R. Andersson, A. El-Kadi, D. Masters, T. Ewalds, J. Stott, S. Mohamed, P. Battaglia, R. Lam, M. Willson, Probabilistic weather forecasting with machine learning, Nature 637 (8044) (2025-01-02) 84–90.
[135] R. Lam, A. Sanchez-Gonzalez, M. Willson, P. Wirnsberger, M. Fortunato, F. Alet, S. Ravuri, T. Ewalds, Z. Eaton-Rosen, W. Hu, A. Merose, S. Hoyer, G. Holland, O. Vinyals, J. Stott, A. Pritzel, S. Mohamed, P. Battaglia, Learning skillful medium-range global weather forecasting, Science 382 (6677) (2023) 1416–1421.
[136] J.A.C. Remes, A. Horowitz (Eds.), Critical Disaster Studies, University of Pennsylvania Press, 2021.
[137] N. Mohammadi, J.E. Taylor, Thinking fast and slow in disaster decision-making with smart city digital twins, Nat. Comput. Sci. 1 (12) (2021-12) 771–773, http://dx.doi.org/10.1038/s43588-021-00174-0.
[138] W. Yin, Y. Xue, Z. Liu, H. Li, M. Werner, LLM-enhanced disaster geolocalization using implicit geoinformation from multimodal data: A case study of hurricane harvey, Int. J. Appl. Earth Obs. Geoinf. 137 (2025-03) 104423, http://dx.doi.org/10.1016/j.jag.2025.104423.
[139] A. Wazzan, S. MacNeil, R. Souvenir, Comparing traditional and LLM-based search for image geolocation, in: Proceedings of the 2024 ACM SIGIR Conference on Human Information Interaction and Retrieval, 2024-03-10, pp. 291–302, http://dx.doi.org/10.1145/3627508.3638305.
[140] R. Arandjelović, P. Gronat, A. Torii, T. Pajdla, J. Sivic, NetVLAD: CNN architecture for weakly supervised place recognition, 2016-05-02, URL http://arxiv.org/abs/1511.07247.
[141] G. Berton, C. Masone, B. Caputo, Rethinking visual geo-localization for large-scale applications, in: 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR, IEEE, 2022-06, pp. 4868–4878.
[142] D. Lahat, T. Adali, C. Jutten, Multimodal data fusion: An overview of methods, challenges, and prospects, Proc. IEEE 103 (9) (2015) 1449–1477.
[143] L. Elmhadhbi, M.-H. Karray, B. Archimède, J.N. Otte, B. Smith, An ontological approach to enhancing information sharing in disaster response, Information 12 (10) (2021-10-19) 432, http://dx.doi.org/10.3390/info12100432.
[144] R. Chen, R. Sharman, H.R. Rao, S.J. Upadhyaya, Decision support for crisis management: Requirements and challenges, MIS Q. Exec. 15 (1) (2016) 13–33.
[145] J. Nalau, B. Verrall, Mapping the evolution and current trends in climate change adaptation science, Clim. Risk Manag. 32 (2021) 100290, http://dx.doi.org/10.1016/j.crm.2021.100290.
[146] K.-T. Tran, D. Dao, M.-D. Nguyen, Q.-V. Pham, B. O'Sullivan, H.D. Nguyen, Multi-agent collaboration mechanisms: A survey of LLMs, 2025-01-10, http://dx.doi.org/10.48550/arXiv.2501.06322.
[147] M.R. Endsley, Toward a theory of situation awareness in dynamic systems, Hum. Factors 37 (1) (1995) 32–64, http://dx.doi.org/10.1518/001872095779049543.