Harnessing Large Language Models for Disaster Management: A Survey

Anonymous ACL submission

Abstract

Large Language Models (LLMs) have demonstrated remarkable capabilities across various domains, including their emerging role in mitigating threats to human life, infrastructure, and the environment during natural disasters. Despite increasing research on disaster-focused LLMs, there remains a lack of systematic reviews and in-depth analyses of their applications in natural disaster management. To address this gap, this paper presents a comprehensive survey of LLMs in disaster response, introducing a taxonomy that categorizes existing works based on disaster phases and application scenarios. By compiling public datasets and identifying key challenges and opportunities, this study aims to provide valuable insights for the research community and practitioners in developing advanced LLM-driven solutions to enhance resilience against natural disasters.

1 Introduction

Natural disasters are becoming increasingly frequent and severe, posing unprecedented threats to human life, infrastructure, and the environment(Manyena, 2006; Yu et al., 2018; Chaudhary and Piracha, 2021). The 2010 Haiti earthquake, for instance, resulted in over 200,000 fatalities and widespread infrastructure devastation (DesRoches et al., 2011). Similarly, the 2020 Australian bushfires caused the deaths of at least 33 people and an estimated loss of one billion animals (Deb et al., 2020). The profound impact of such catastrophic events underscores the urgent need for effective disaster management strategies. Recently, large language models (LLMs) have transformed research and technological innovation with their exceptional capabilities in contextual understanding, logical reasoning, and complex problem-solving across multiple modalities (Zhang et al., 2024b,a). These capabilities position LLMs as powerful tools for natural disaster management, enabling them to analyze vast real-time disaster data, facilitate dynamic communication with affected communities, and support critical decision-making (Otal et al., 2024).

Despite their potential, a systematic review of LLMs in disaster management remains absent, limiting researchers and practitioners in identifying best practices, addressing research gaps, and optimizing LLM deployment for disaster-related challenges. To bridge this gap, this paper presents a comprehensive survey of LLM applications in disaster management, categorizing them across three model architectures and the four key disaster phases: mitigation, preparedness, response, and recovery. We introduce a novel taxonomy that integrates application scenarios, specific tasks, and model architectures tailored to disaster-related challenges. Additionally, we summarize publicly available datasets, identify key challenges, and explore avenues for enhancing the effectiveness, efficiency, and trustworthiness of LLMs in disaster response. This review aims to inspire and guide AI researchers, policymakers, and practitioners toward developing LLM-driven disaster management frameworks. Our key contributions are as follows:

- **Systematical Review:** We provide the first systematical review of explorations of LLMs applications in disaster management across four key disaster phases.
- **Novel Taxonomy:** We propose a taxonomy integrating application scenarios, specific tasks, and model architectures, providing both practical and technical insights into this survey.
- **Resource Compilation:** We compile essential resources (e.g., datasets), and highlight key challenges and future research directions to advance LLM-driven disaster management.

2 Background

Disaster management is a multidisciplinary field that integrates resources, expertise, and strategies to mitigate the impact of increasingly severe disasters. Its primary goal is to minimize immediate damage while fostering long-term resilience and adaptive recovery. Disaster management comprises four interconnected phases (Sun et al., 2020):

- **Mitigation** involves identifying risks and vulnerabilities while implementing proactive measures to prevent disasters.
- **Preparedness** includes developing comprehensive plans and public education initiatives to enhance readiness for potential disasters.
- **Response** identifies and addresses immediate needs during a disaster, including emergency rescue operations and resource distribution.
- **Recovery** involves rebuilding affected areas, addressing both physical and social impacts to facilitate a return to normalcy.

In general, LLMs have the potential to serve as general-purpose foundations for developing specialized AI tools that enhance various aspects of disaster management. Here, we categorize LLM architectures into three main types: (1) encoder-based LLM (e.g., BERT (Devlin, 2018)), which excel in contextual understanding; (2) decoder-based LLM (e.g., GPT (Brown, 2020)), which are optimized for sequential prediction; and (3) multimodal LLMs, which integrate multiple modalities to enhance information processing (Tiong et al., 2022; Madichetty et al., 2021) In disaster management, common downstream tasks include classification (e.g., damage classification), estimation (e.g., severity estimation), extraction (e.g., knowledge extraction), and generation (e.g., report generation). To tailor LLMs for these tasks, techniques such as fine-tuning and prompting are commonly employed.

3 LLM For Disaster Management

Foundation models can be utilized across the four disaster management phases: mitigation, preparedness, response, and recovery. Within each phase, existing works are categorized based on application scenarios, specific tasks, and model architectures. Figure 1 presents an overview of our taxonomy, with detailed summaries provided in Appendix A.

3.1 Disaster Mitigation

Assessing vulnerabilities is a crucial component of disaster mitigation, where LLMs have demonstrated promising potential. This process involves identifying and analyzing infrastructure and communities at risk, enabling proactive measures to reduce disaster impact.

***Vulnerability* Classification.** A system named *Infrastructure Ombudsman* has leveraged supervised learning with encoder-based LLMs and zero-shot prompt learning with decoder-based LLMs to detect and classify concerns about potential infrastructure failures from social media data (Chowdhury et al., 2024). This approach enables decision-makers to effectively prioritize resources and address critical issues in a timely manner.

***Answer* Generation.** Beyond infrastructure vulnerability assessment, decoder-based LLMs can assist in addressing community vulnerability-related queries by retrieving and leveraging the Social Vulnerability Index (SVI) (Martelo and Wang, 2024).

3.2 Disaster Preparedness

In the long term, LLMs can play a pivotal role in disaster preparedness through (1) enhancing public awareness by disseminating accurate and accessible information, and (2) supporting disaster forecasting with advanced data analysis. Building on these forecasts, LLMs can aid decision-makers in issuing (3) timely disaster warnings, improving short-term preparedness. Furthermore, LLMs can support well-structured (4) evacuation planning, ensuring the safe relocation of individuals and assets.

3.2.1 Public Awareness Enhancement

Enhancing public awareness of disasters is crucial, particularly by providing insights and knowledge derived from past disaster experiences.

***Knowledge* Extraction.** *Encoder-based LLMs* have been fine-tuned to extract disaster-related knowledge from news articles and social media (Fu et al., 2024), as well as from extensive disaster literature (Zhang and Wang, 2023), using Named Entity Recognition (NER). To improve the logical coherence of extracted entities, Ma et al. propose BERT-BiGRU-CRF for NER, enabling the construction of disaster knowledge graphs (Ma et al., 2023). In addition, *decoder-based LLMs* have been fine-tuned with instructional learning to extract knowledge triplets from documents for knowledge graph construction (Wu et al., 2024).

***Answer* Generation.** The extracted disaster knowledge could be incorporated in *decoder-based LLMs*' prompts, facilitating disaster-related question answering (Hostetter et al., 2024; Martelo and Wang,

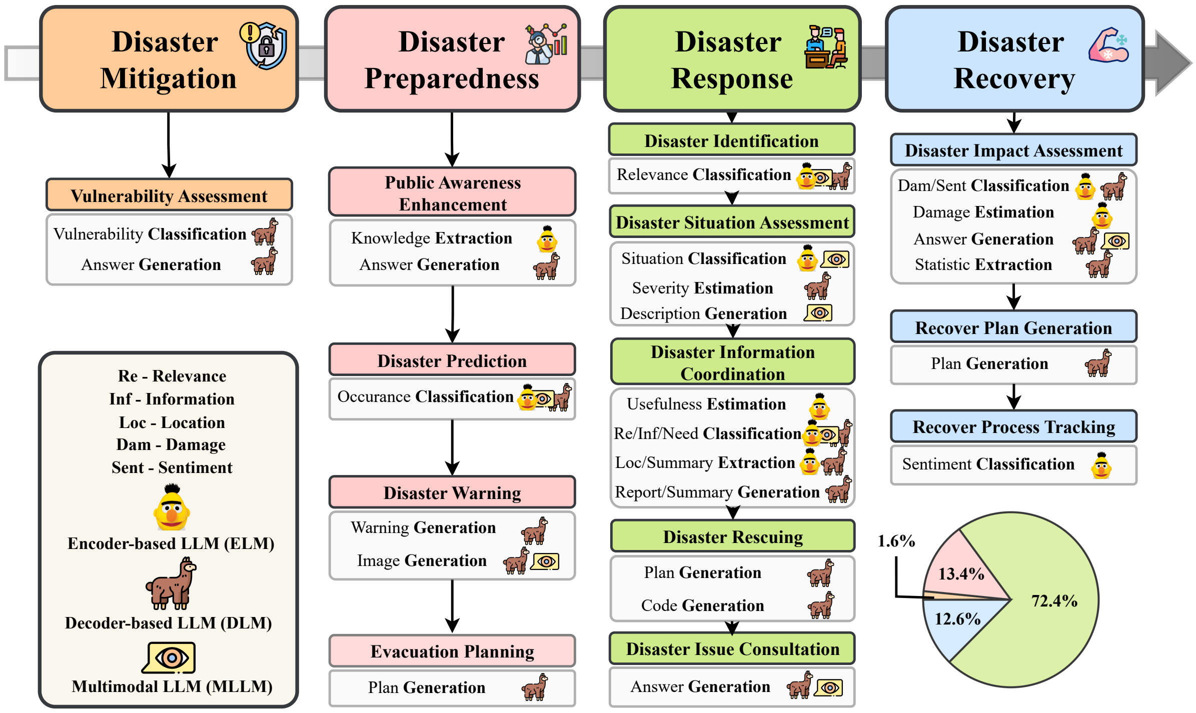

Figure 1: Taxonomy of applications of LLMs in disaster management. This survey categorizes the utilization of LLMs across four **disaster phases**, highlighting specific **applications** where **tasks** such as classification, estimation, extraction, and generation are performed by three **types of LLMs** (Encoder-based, Decoder-based, and Multimodal LLM). The chart in the bottom-right corner presents the distribution of surveyed papers across each phase.

2024; Li et al., 2023). Additionally, techniques such as retrieval-augmented generation (RAG) have been employed to further improve knowledge integration (Zhu et al., 2024).

3.2.2 Disaster Prediction

Effective disaster preparedness also relies on accurate and reliable disaster prediction.

***Occurrence* Classification.** *Encoder-based LLMs* have been widely employed for disaster prediction. For instance, BERT has been integrated with GRU and CNN to predict disaster (Indra and Duraipandian, 2023). However, textual data alone is often limited due to its subjective and imprecise nature, prompting the adoption of *multimodal LLMs* that incorporate multiple data modalities. For instance, Zeng et al. combine historical flood data with geographical descriptions of specific locations to assess disaster risk (Zeng and Bertsimas, 2023). Additionally, satellite imagery has been leveraged to provide visual context, enhancing predictive accuracy (Liu and Zhong, 2023). To further improve disaster prediction with explicit external knowledge, *decoder-based LLMs* have been integrated with retrieval-augmented generation (RAG) to retrieve historical flood data, aiding in risk assessment and action recommendation (Wang et al., 2024).

3.2.3 Disaster Warning

Once a disaster is anticipated, timely warnings are essential for ensuring public safety.

***Warning* Generation.** *Decoder-based LLMs* have proven valuable in generating warning messages based on rule-based alerts derived from streaming data (Chandra et al., 2024), significantly improving the responsiveness of warning systems. Additionally, RAG has enhanced LLMs by enabling the retrieval of disaster alerts from official APIs, providing real-time information on impending disasters (Martelo and Wang, 2024).

***Image* Generation.** In addition to textual warnings, visual warnings can provide more vivid and intuitive descriptions, effectively reaching a broader audience. To achieve this, *multimodal LLMs* enhanced by diffusion-based text-to-image generative models can generate detailed visual representations of impending disasters (Lubin et al., 2024), enhancing the clarity and impact of disaster alerts.

3.2.4 Evacuation Planning

***Plan* Generation.** To safeguard individuals and property from impending disasters, *decoder-based*

LLMs have been prompted to generate escape plans and provide evacuation recommendations (Hostetter et al., 2024).

3.3 Disaster Response

With accurate and real-time (1) disaster identification and (2) situation assessment, decision-makers can acquire critical insights to establish a solid foundation for response efforts. Additionally, LLMs can facilitate (3) disaster information coordination, enhancing collaboration among stakeholders for more effective disaster response. As a result, decision-makers can leverage LLMs to execute key actions, including (4) disaster rescue operations and (5) disaster-related consultations.

3.3.1 Disaster Identification

Effective disaster response begins with accurate and real-time identification, enabling efficient interventions (Said et al., 2019; Weber et al., 2020). Social media serves as a valuable resource in this process, offering real-time updates from affected individuals (Anderson, 2016; Trono et al., 2015).

***Relevance* Classification with Encoder-based LLMs.** Classifying social media posts to identify disaster-related content is a crucial step in disaster detection, where LLMs have proven to be highly effective. Encoder-based LLMs augmented with trainable adapters are commonly employed for this task through fine-tuning on annotated disaster corpora (Ningsih and Hadiana, 2021; Singh et al., 2022; Lamsal et al., 2024a). Recognizing the diverse sources of disaster data, ensemble methods combine predictions from multiple LLMs to leverage their complementary strengths in processing varied linguistic patterns (Mukhtiar et al., 2023). Pure LLM-based approaches may struggle to capture fine-grained structural features in disaster-related posts. To address this, hybrid architectures integrate CNNs to capture local n-gram patterns (Franceschini et al., 2024; Song and Huang, 2021; Meghatria et al., 2024), attention-based BiLSTMs to model sequential dependencies (Huang et al., 2022), and graph neural networks (GNNs) to represent semantic word relationships (Manthena, 2023; Ghosh et al., 2022). To tackle the challenge of limited labeled training data, active learning has been employed to automatically label informative samples (Paul et al., 2023).

***Relevance* Classification with Encoder-based LLMs.** Furthermore, decoder-based LLMs such as Llama-2 and GPT-4 have demonstrated strong performance in relevance classification using prompt learning techniques (Taghian Dinani et al., 2023).

***Relevance* Classification with Multimodal LLMs.** Image data also provide valuable insights for disaster analysis and can be integrated to enhance classification using multimodal LLMs. This integration can be achieved through simple aggregation (Kamoji et al., 2023; Madichetty et al., 2021) or attention-based mechanisms (Shetty et al., 2024). To address challenges arising from multimodal heterogeneity, Zhou et al. employ a CycleGAN combined with a mixed fusion strategy (Zhou et al., 2023b). Beyond multimodal heterogeneity, research also tackles other critical challenges in multimodal learning. These include addressing label scarcity through semi-supervised minimax entropy domain adaptation frameworks (Wang and Wang, 2022) and enhancing model performance by leveraging the complementary strengths of diverse LLMs and visual models using ensemble methods (Hanif et al., 2023). Beyond social media, data from sources such as satellite imagery and news articles can further enhance disaster analysis (Jang et al., 2024).

3.3.2 Disaster Situation Assessment

After disaster identification, assessing its severity and spread is essential for formulating effective response strategies.

***Situation* Classification.** *encoder-based LLMs* have been fine-tuned to for binary classification to identify situational posts (Madichetty and Sridevi, 2021). Raj et al. employ BERT and NER to extract disaster-related locations, using location counts as an indicator of disaster severity (Raj et al., 2023). Additionally, multimodal LLMs integrate visual data to further enhance disaster situational assessment (Kanth et al., 2022).

***Severity* Estimation.** While classification provides only a coarse understanding, severity estimation offers precise quantitative insights. *decoder-based LLMs* enhanced with chain-of-thought (CoT) reasoning have been used to estimate earthquake intensity, expressed as Modified Mercalli Intensity (MMI) (Mousavi et al., 2024). In addition, *multimodal LLMs* leverage rich image data for more accurate estimations. For example, FloodDepth-GPT employs prompt-based guidance with GPT-4 to estimate floodwater depth from flood images.

***Description* Generation.** Beyond categorical and statistical descriptions, multimodal LLMs can generate more comprehensible textual situational reports from disaster images (Hu and Rahnemoonfar, 2024; Wolf et al., 2023).

3.3.3 Disaster Information Coordination

Coordinating disaster-related information is crucial for ensuring an organized and collaborative response (Comfort et al., 2004; Bharosa et al., 2010). Social media plays a pivotal role in this process, as individuals actively share posts containing warnings, urgent needs, and other critical information (Lindsay, 2011; Imran et al., 2015).

***Usefulness* Estimation.** To improve the accessibility of valuable information, *encoder-base LLMs* are utilized to filter informative tweets by computing usefulness ratings (Yamamoto et al., 2022). However, this approach requires a predefined threshold to determine the relevance of a tweet.

***Relevance* Classification.** Several studies fine-tune *encoder-based LLMs* for binary relevance classification, as discussed in Section 3.3.1. Additionally, LLMs have been applied to multi-level relevance classification to further refine disaster-related information filtering (Blomeier et al., 2024).

***Information* Classification.** To facilitate information dissemination, several studies have fine-tuned *encoder-based LLMs* to classify posts based on different information types, including actionable types such as "important for managers" (Sharma et al., 2021); humanitarian types such as "Injured people" (Yuan et al., 2022); and disaster-specific types (Liu et al., 2021). When fine-tuning data is limited, augmentation strategies such as manual hashtag annotation (Boros et al., 2022) and self-training with soft labeling (Li et al., 2021) are employed to enhance classification performance.

Pure LLM-based methods may have limitations, as discussed in Section 3.3.1. In contrast, hybrid architectures enhance performance by integrating CNNs and BiLSTMs to improve local pattern comprehension (Zou et al., 2024) and employing Graph Attention Networks (GATs) to capture correlations between tweet embeddings and information types (Zahera et al., 2021). Additionally, FF-BERT leverages an ensemble of BERT and CNN to combine model strengths for improved classification (Wilkho et al., 2024). Other studies enhance the application of LLMs in disaster information classification by extracting rationales—evidence that supports classification decisions (Nguyen and Rudra, 2022b, 2023). RACLC (Nguyen and Rudra, 2022a) employs a two-stage framework, utilizing contrastive learning to refine rationale extraction and improve classification performance.

Decoder-based LLMs have also been employed for disaster type and humanitarian classification through instruction tuning (Otal and Canbaz, 2024; Yin et al., 2024), as well as zero-shot and few-shot prompting (Dinani et al., 2024).

Multimodal LLMs can integrate rich visual data from social media to enhance classification by leveraging multiple modalities. This integration can be achieved through simple feature aggregation (Zhang et al., 2022; Yu and Wang, 2024) or more advanced fusion techniques, such as cross-attention mechanisms (Abavisani et al., 2020) and dual transformer architectures (Zhou et al., 2023a). Additionally, Basit et al. classify posts into humanitarian or structural categories only when the text and image classification outputs align; otherwise, the posts are uninformative (Basit et al., 2023).

***Need* Classification.** Social media enables individuals to express urgent needs during disasters. *Encoder-based LLMs* have been employed to detect disaster-related needs (Yang et al., 2024; Vitiugin and Purohit, 2024) and rescue requests (Toraman et al., 2023). Responders also use social media to share available resources. Encoder-based LLMs have been employed to match needs with resources using cosine similarity-based retrieval methods, where both offer and request posts are embedded using XLM-RoBERTa (Conneau, 2019), optimizing resource allocation.

***Location* Extraction.** Additionally, various post-processing techniques enhance information dissemination, particularly through location extraction. Several studies fine-tune encoder-based LLMs for location reference recognition (LRR), classifying tokens into categories such as "Inside Locations" (ILOC) and "Other Tokens" (O) (Mehmood et al., 2024; Suwaileh et al., 2022; Koshy and Elango, 2024). LRR can be further improved by integrating a conditional random field (CRF) model, which enhances the logical consistency of extracted locations (Ma et al., 2022; Zhang et al., 2021). Furthermore, external knowledge corpora can support location extraction. For instance, Caillaut et al. use cosine similarity to match post entities with a knowledge base, ensuring the authenticity of extracted locations (Caillaut et al., 2024).

Decoder-based LLMs are widely used for extracting location-relevant information through prompt learning (Yu and Wang, 2024). To enhance accuracy, external knowledge has been incorporated into prompts, including geo-knowledge (Hu et al., 2023) and Object Character Recognition-based object descriptions (Firmansyah et al., 2024).

***Summary* Extraction.** Furthermore, summarizing disaster-related posts provides a macro-level understanding during crises. Several studies focus on identifying critical and informative posts for summarization by integrating advanced techniques into *encoder-based LLMs*, such as integer linear programming (ILP) (Nguyen and Rudra, 2022a; Nguyen et al., 2022) and Rapid Automatic Keyword Extraction (RAKE) (Garg et al., 2024).

***Summary* Generation.** *Decoder-based LLMs* extend summarization capabilities by generating summaries from retrieved text. For example, Vitiugin et al. rank key tweets using an LSTM model and then apply a T5 model to generate summaries based on the top-ranked tweets (Vitiugin and Castillo, 2022). Crisis2Sum performs query-focused summarization through a multi-step process, including query-informed document retrieval, reranking, fact extraction, clustering, fusion into event nuggets, and final selection for summarization (Seeberger and Riedhammer, 2024a). Additionally, agent-based approaches can enhance summary quality by leveraging multiple LLMs for document retrieval, reranking, and instruction-following summarization (Seeberger and Riedhammer, 2024b).

***Report* Generation.** *Decoder-based LLMs* have been employed for disaster report generation, utilizing techniques such as RAG to extract relevant web data (Colverd et al., 2023) and Chain-of-Thought reasoning to enhance the coherence and accuracy of generated reports.

3.3.4 Disaster Rescuing

Grounded in a comprehensive understanding of the disaster situation, disaster rescue focuses on saving lives and protecting property through timely and coordinated actions.

***Plan* Generation.** Effective rescue operations require well-structured rescue plans. *Decoder-based LLMs* have been prompted to generate actionable response plans, offering essential guidance for disaster response (Goecks and Waytowich, 2023).

***Code* Generation.** Once a plan is established, *decoder-based LLMs* can support its execution by assisting organizations and rescue teams. For instance, they can facilitate robotic system guidance during rescue operations by translating verbal inputs into actionable operational commands using RAG (Panagopoulos et al., 2024).

3.3.5 Disaster Issue Consultation

During disasters, affected individuals and organizations often seek reliable guidance. Disaster issue consultation provides advice, safety updates, and expert recommendations, helping them access resources, evaluate options, and make informed decisions (Jiang, 2024).

***Answer* Generation.** *Decoder-based LLMs* are employed to generate answers for frequently asked questions and provide disaster-related guidance (Rawat, 2024; Chen and Fang, 2024). To mitigate hallucination, RAG is integrated with verified disaster-related documents. For example, WildfireGPT retrieves wildfire-related literature and data to enhance prompts (Xie et al., 2024). Chen et al. introduce a prompt chain to guide LLM reasoning over a disaster knowledge graph, incorporating structured knowledge (Chen et al., 2024). Unlike traditional RAG approaches without training, Xia et al. combine fine-tuning for implicit knowledge updates with RAG for explicit knowledge, further improving response quality (Xia et al., 2024).

Additionally, *multi-modal LLMs* can integrate textual and visual data to enhance disaster response. For example, several visual question answering (VQA) models, such as Plug-and-Play VQA (Tiong et al., 2022), have been prompted for zero-shot VQA in disaster scenarios (Sun et al., 2023). To handle complex user queries, ADI introduces sequential modular tools, incorporating vision-language models (VLMs), object detection models, and semantic segmentation models (Liu et al., 2024). Furthermore, FloodLense combines ChatGPT with diffusion models to highlight disaster-affected areas in images, enhancing flood-related geographical question answering (Kumbam and Vejre, 2024).

3.4 Disaster Recovery

LLMs can play a crucial role in (1) disaster impact assessment, a vital step in the recovery process. By providing a comprehensive understanding of disaster impacts, LLMs can assist decision-makers in (2) generating recovery plans tailored to specific needs. Additionally, disaster responders have leveraged LLMs for (3) continuous recovery process tracking, ensuring effectiveness and progress throughout the recovery phase.

Get over it Choc you lost. Anyone with half a brain could see it. http://t.co/ythaSNX6	**off-topic**
Flooding hits eastern Australia: Hundreds of homes are inundated and and several people reported missing as flood waters rise in the ...	**on-topic**

Figure 2: A sample of dataset for disaster relevance classification from CrisisLexT6 (Olteanu et al., 2014).

3.4.1 Disaster Impact Assessment

Accurately assessing the extent of damage across both physical and social dimensions is essential for prioritizing recovery efforts effectively.

***Damage* Classification.** From the physical dimension, *encoder-based LLMs* have been employed to identify and categorize disaster-related damage (e.g., human/infrastructure damage (Malik et al., 2024), water/power supply damage (Chen and Lim, 2021)) Additionally, Zou et al. propose a BERT-BiLSTM-Sit-CNN framework, improving textual understanding for damage-related post identification and damage-type classification (Zou et al., 2024). Beyond type classification, LLMs have been utilized to assess damage severity. For instance, Jeba et al. employ BERT to classify damage impact severity in social media posts and news articles (Jeba et al., 2024).

***Damage* Estimation.** Damage severity can be more effectively quantified through fine-grained estimation. Chen et al. compute damage severity scores by measuring the similarity between post tokens and predefined seed words' embeddings, both of which are derived from encoder-based LLMs (Chen and Lim, 2021).

***Answer* Generation.** In addition, *decoder-based LLMs* can answer specific assessment questions. Ziaullah et al. employ RAG-enhanced LLMs to retrieve operational status updates of critical infrastructure facilities from social media data (Ziaullah et al., 2024). *Multimodal LLMs* further incorporate remote sensing data for enhanced assessment. Estevao et al. prompt GPT-4o to generate damage assessments based on building images (Estêvão, 2024). To improve modality alignment, SAM-VQA employs a supervised attention-based vision-language model (VLM) to integrate image and question features for visual question answering (VQA) tasks (Sarkar et al., 2023). Additionally, auxiliary tasks have been leveraged to enhance VQA performance. For instance, DATWEP dynamically balances the significance of segmentation and VQA tasks by adjusting class weights during training (Alsan and Arsan, 2023).

***Statistic* Extraction.** *Decoder-based LLMs* have also used few-shot learning to extract fatality information from social media (Hou and Xu, 2022), offering timely insights into human loss.

***Sentiment* Classification.** From the social dimension, disasters can influence public sentiment, where *encoder-based LLMs* (Han et al., 2024a; Berbère et al., 2023) have been fine-tuned to classify social media posts into positive and negative emotions. In addition, Li et al. employ *decoder-based LLM* (e.g., GPT 3.5) to classify posts into five emotional types, such as "panic" and "sadness", using zero-shot prompting (Li et al., 2025). This approach helps responders better understand and address the emotional impact of disasters.

3.4.2 Recovery Plan Generation

Based on impact assessment, a recovery plan is formulated to rebuild infrastructure, restore services, and strengthen resilience (Hallegatte et al., 2018).

***Plan* Generation.** *Decoder-based LLMs* have been applied in certain recovery scenarios to generate recovery and reconstruction plans. For example, ChatGPT has been prompted to develop disaster recovery strategies for business restoration (White and Liptak, 2024; Lakhera, 2024).

3.4.3 Recovery Process Tracking

Continuous tracking of the recovery process ensures that progress remains aligned with the planned timeline, allowing decision-makers to adapt recovery strategies to evolving needs.

***Sentiment* Classification.** *Encoder-based LLMs* (e.g., BERTweet) have been employed to assess public sentiment throughout the recovery period (CONTRERAS et al.), enabling responders to tailor recovery efforts to effectively address the emotional needs of affected populations.

4 Datasets

Multiple disaster-related datasets have been employed to evaluate LLMs in disaster management. A comprehensive list of publicly available datasets is provided in Appendix B.

Classification datasets primarily consist of textual inputs from platforms such as Twitter and news

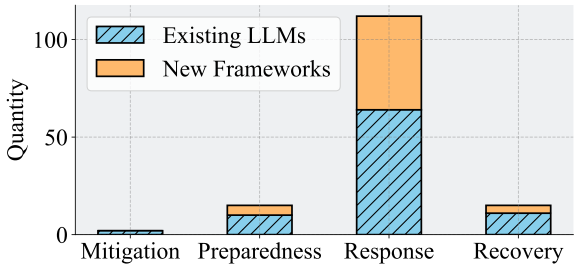

Figure 3: Publication counts utilizing existing LLMs and developing new models across the four phases of disaster management.

outlets, categorizing data based on informativeness (relevance) (Olteanu et al., 2014) (illustrated in Figure 2), humanitarian types (Imran et al., 2016), damage levels (Alam et al., 2021b), and other relevant attributes. Some datasets also incorporate visual data, including satellite imagery and social media images (Alam et al., 2018). Model performance is typically evaluated using metrics such as accuracy and F1 score.

Estimation datasets usually provide quantitative labels such as flood depths (Akinboyewa et al., 2024). Metrics like Mean Absolute Error (MAE) are used for evaluation.

Generation datasets are also extensively used and primarily fall into two categories: question answering and summarization. Question-answering datasets provide disaster-related questions paired with crowdsourced annotated answers (Rawat, 2024). Additionally, multimodal question-answering datasets, which incorporate disaster-related images as contextual information, are widely utilized (Sun et al., 2023). For the summarization task, large collections of documents serve as inputs, with reference summaries curated by domain experts (McCreadie and Buntain, 2023). Both question-answering and summarization tasks are evaluated using metrics such as BLEU.

Extraction datasets identify and label specific elements within a sentence, such as keywords (Nguyen and Rudra, 2022a) and locations (Suwaileh et al., 2022). Tokens are labeled as "outside," "start," or "end" to indicate their extraction status. These datasets are primarily used for token-level classification tasks and are evaluated using classification metrics.

5 Challenges and Opportunities

Large Language Models (LLMs) hold great promise for disaster management but face several key limitations. Most studies deploy generic LLMs as universal solutions, overlooking domain-specific challenges and the need for tailored frameworks, as shown in Figure 3. Additionally, current applications are heavily concentrated on disaster response, leaving other phases underexplored, as illustrated in Figure 1. To fully harness the potential of LLMs in disaster management, researchers must address the disaster-specific challenges outlined below.

Dataset Construction. Current datasets are heavily skewed toward classification tasks, leaving other areas underexplored. Additionally, raw disaster data often contains uncertainty and bias (Smith and Katz, 2013), posing challenges in constructing reliable datasets. Innovative approaches, such as synthetic data generation (Kalluri et al., 2024), offer a promising solution to enhance dataset coverage across diverse disaster-related scenarios.

Efficient Deployment. Large-scale LLMs face efficiency challenges (Ramesh Raja et al., 2024), limiting their viability for real-time decision-making in emergency disaster scenarios. While lightweight models offer a more efficient alternative (Saleem et al., 2024), they often compromise robustness in disaster-related tasks. Developing models that balance efficiency and reliability is essential for effective disaster management.

Robust Generation. Decoder-based LLMs are prone to hallucination, generating factually inaccurate outputs that pose serious risks in disaster contexts, such as false evacuation routes, resource misallocation, and potential loss of lives. To mitigate these risks, strategies such as integrating RAG with external knowledge bases (Colverd et al., 2023), domain-specific training (Lamsal et al., 2024a), and uncertainty estimation (Xu et al.) can help reduce hallucinated outputs and improve reliability.

6 Conclusion

This paper surveys the application of LLMs in disaster management across the four disaster phases, introducing a taxonomy that integrates application scenarios, specific tasks, and the architectures of models addressing these tasks. By presenting publicly available datasets and identifying key challenges, we aim to inspire collaborative efforts between AI researchers and decision-makers, ultimately enabling the full potential of LLMs to build more resilient communities and advance proactive disaster management practices.

Limitations

Survey Scope. This work focuses exclusively on disaster management applications only where existing LLMs have been utilized, leaving out other potential scenarios (e.g. repair cost evaluation during the recovery phase) that have yet to be explored in current LLM research. While these unexplored areas hold significant promise for future advancements, they fall beyond the scope of this study due to space constraints.

Datasets. Additionally, we include only a subset of datasets used in existing studies, prioritizing those that are easily accessible. Many datasets either are not open-sourced, have restrictive access policies, or lack assured quality, making them less suitable for reproducibility and further research.

References

Mahdi Abavisani, Liwei Wu, Shengli Hu, Joel Tetreault, and Alejandro Jaimes. 2020. Multimodal categorization of crisis events in social media. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 14679–14689.

Ademola Adesokan, Sanjay Madria, and Long Nguyen. 2023. Tweetace: A fine-grained classification of disaster tweets using transformer model. In *2023 IEEE Applied Imagery Pattern Recognition Workshop (AIPR)*, pages 1–9. IEEE.

Temitope Akinboyewa, Huan Ning, M Naser Lessani, and Zhenlong Li. 2024. Automated floodwater depth estimation using large multimodal model for rapid flood mapping. *Computational Urban Science*, 4(1):12.

Firoj Alam, Ferda Ofli, and Muhammad Imran. 2018. Crisismmd: Multimodal twitter datasets from natural disasters. In *Proceedings of the international AAAI conference on web and social media*, volume 12.

Firoj Alam, Umair Qazi, Muhammad Imran, and Ferda Ofli. 2021a. Humaid: Human-annotated disaster incidents data from twitter with deep learning benchmarks. In *Proceedings of the International AAAI Conference on Web and social media*, volume 15, pages 933–942.

Firoj Alam, Hassan Sajjad, Muhammad Imran, and Ferda Ofli. 2021b. Crisisbench: Benchmarking crisis-related social media datasets for humanitarian information processing. In *Proceedings of the International AAAI conference on web and social media*, volume 15, pages 923–932.

Aymen Omar Alharm and Syibrah Naim. Enhancing natural disaster response: A deep learning approach to disaster sentiment analysis using bert and lstm. *Available at SSRN 4755638*.

Huseyin Fuat Alsan and Taner Arsan. 2023. Dynamic task and weight prioritization curriculum learning for multimodal imagery. *arXiv preprint arXiv:2310.19109*.

Ben Anderson. 2016. Governing emergencies: The politics of delay and the logic of response. *Transactions of the institute of British geographers*, 41(1):14–26.

Stelios Andreadis, Aristeidis Bozas, Ilias Gialampoukidis, Anastasia Moumtzidou, Roberto Fiorin, Francesca Lombardo, Thanassis Mavropoulos, Daniele Norbiato, Stefanos Vrochidis, Michele Ferri, et al. 2022. Disastermm: Multimedia analysis of disaster-related social media data task at mediaeval 2022. In *MediaEval*.

Gail M Atkinson and David J Wald. 2007. "did you feel it?" intensity data: A surprisingly good measure of earthquake ground motion. *Seismological Research Letters*, 78(3):362–368.

Mohammad Basit, Bashir Alam, Zubaida Fatima, and Salman Shaikh. 2023. Natural disaster tweets classification using multimodal data. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, pages 7584–7594.

Riheme Berbère, Safa Elkefi, Safa Bhar Layeb, and Achraf Tounsi. 2023. Exploring cognitive sustainability concerns in public responses to extreme weather events: An nlp analysis of twitter data. *Cognitive Sustainability*, 2(4).

Nitesh Bharosa, JinKyu Lee, and Marijn Janssen. 2010. Challenges and obstacles in sharing and coordinating information during multi-agency disaster response: Propositions from field exercises. *Information systems frontiers*, 12:49–65.

Eike Blomeier, Sebastian Schmidt, and Bernd Resch. 2024. Drowning in the information flood: Machine-learning-based relevance classification of flood-related tweets for disaster management. *Information*, 15(3):149.

Emanuela Boros, Gaël Lejeune, Mickaël Coustaty, and Antoine Doucet. 2022. Adapting transformers for detecting emergency events on social media. In *14th International Conference on Knowledge Discovery and Information Retrieval*, pages 300–306. SCITEPRESS-Science and Technology Publications.

Tom B Brown. 2020. Language models are few-shot learners. *arXiv preprint arXiv:2005.14165*.

Gaëtan Caillaut, Samuel Auclair, Cécile Gracianne, Nathalie Abadie, and Guillaume Touya. 2024. Entity linking for real-time geolocation of natural disasters from social network posts. *PloS one*, 19(10):e0307254.

Ritesh Chandra, Shashi Shekhar Kumar, Rushil Patra, and Sonali Agarwal. 2024. Decision support system for forest fire management using ontology with big data and llms. *arXiv preprint arXiv:2405.11346*.

S Chandrakala and S Albert Antony Raj. 2022. Identifying the label of crisis related tweets using deep neural networks for aiding emergency planning. In *2022 International Conference on Innovative Computing, Intelligent Communication and Smart Electrical Systems (ICSES)*, pages 1–6. IEEE.

Muhammad T Chaudhary and Awais Piracha. 2021. Natural disasters—origins, impacts, management. *Encyclopedia*, 1(4):1101–1131.

Minze Chen, Zhenxiang Tao, Weitong Tang, Tingxin Qin, Rui Yang, and Chunli Zhu. 2024. Enhancing emergency decision-making with knowledge graphs and large language models. *International Journal of Disaster Risk Reduction*, 113:104804.

Wei Chen and Jiing Fang. 2024. Optimizing ai-driven disaster management through llms.

Zi Chen and Samsung Lim. 2021. Social media data-based typhoon disaster assessment. *International Journal of Disaster Risk Reduction*, 64:102482.

Md Towhidul Absar Chowdhury, Soumyajit Datta, Naveen Sharma, and Ashiqur R KhudaBukhsh. 2024. Infrastructure ombudsman: Mining future failure concerns from structural disaster response. In *Proceedings of the ACM on Web Conference 2024*, pages 4664–4673.

Grace Colverd, Paul Darm, Leonard Silverberg, and Noah Kasmanoff. 2023. Floodbrain: Flood disaster reporting by web-based retrieval augmented generation with an llm. *arXiv preprint arXiv:2311.02597*.

Louise K Comfort, Kilkon Ko, and Adam Zagorecki. 2004. Coordination in rapidly evolving disaster response systems: The role of information. *American behavioral scientist*, 48(3):295–313.

A Conneau. 2019. Unsupervised cross-lingual representation learning at scale. *arXiv preprint arXiv:1911.02116*.

Diana CONTRERAS, Dimosthenis ANTYPAS, Sean WILKINSON, Jose CAMACHO-COLLADOS, Philippe GARNIER, and Cécile CORNOU. Assessing post-disaster recovery using sentiment analysis: The case of haiti.

Anusha Danday and T Satyanarayana Murthy. 2022. Twitter data analysis using distill bert and graph based convolution neural network during disaster.

Jens A de Bruijn, Hans de Moel, Brenden Jongman, Marleen C de Ruiter, Jurjen Wagemaker, and Jeroen CJH Aerts. 2019. A global database of historic and real-time flood events based on social media. *Scientific data*, 6(1):311.

Proloy Deb, Hamid Moradkhani, Peyman Abbaszadeh, Anthony S Kiem, Johanna Engström, David Keellings, and Ashish Sharma. 2020. Causes of the widespread 2019–2020 australian bushfire season. *Earth's Future*, 8(11):e2020EF001671.

Reginald DesRoches, Mary Comerio, Marc Eberhard, Walter Mooney, and Glenn J Rix. 2011. Overview of the 2010 haiti earthquake. *Earthquake spectra*, 27(1_suppl1):1–21.

Jacob Devlin. 2018. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.

Soudabeh Taghian Dinani, Doina Caragea, and Nikesh Gyawali. 2024. Disaster tweet classification using fine-tuned deep learning models versus zero and few-shot large language models. In *Data Management Technologies and Applications: 12th International Conference, DATA 2023, Rome, Italy, July 11-13, 2023, Revised Selected Papers*, volume 2105, page 73. Springer Nature.

Wenying Du, Chang Ge, Shuang Yao, Nengcheng Chen, and Lei Xu. 2023. Applicability analysis and ensemble application of bert with tf-idf, textrank, mmr, and lda for topic classification based on flood-related vgi. *ISPRS International Journal of Geo-Information*, 12(6):240.

Premkumar Duraisamy and Yuvaraj Natarajan. 2024. Twitter disaster prediction using different deep learning models. *SN Computer Science*, 5(1):179.

João MC Estêvão. 2024. Effectiveness of generative ai for post-earthquake damage assessment. *Buildings*, 14(10):3255.

Hafiz Budi Firmansyah, Valerio Lorini, Mehmet Oguz Mulayim, Jorge Gomes, and Jose Luis Fernandez-Marquez. 2024. Improving social media geolocation for disaster response by using text from images and chatgpt. In *Proceedings of the 2024 11th Multidisciplinary International Social Networks Conference*, pages 67–72.

Spyros Fontalis, Alexandros Zamichos, Maria Tsourma, Anastasios Drosou, and Dimitrios Tzovaras. 2023. A comparative study of deep learning methods for the detection and classification of natural disasters from social media. In *ICPRAM*, pages 320–327.

Rachele Franceschini, Ascanio Rosi, Filippo Catani, and Nicola Casagli. 2024. Detecting information from twitter on landslide hazards in italy using deep learning models. *Geoenvironmental Disasters*, 11(1):22.

Shengnan Fu, David M Schultz, Heng Lyu, Zhonghua Zheng, and Chi Zhang. 2024. Extracting spatiotemporal flood information from news texts using machine learning for a national dataset in china. *Hydrology and Earth System Sciences Discussions*, 2024:1–32.

Piyush Kumar Garg, Roshni Chakraborty, Srishti Gupta, and Sourav Kumar Dandapat. 2024. Ikdsumm: Incorporating key-phrases into bert for extractive disaster tweet summarization. *Computer Speech & Language*, 87:101649.

Samujjwal Ghosh, Subhadeep Maji, and Maunendra Sankar Desarkar. 2022. Gnom: graph neural network enhanced language models for disaster related multilingual text classification. In *Proceedings of the 14th ACM Web Science Conference 2022*, pages 55–65.

Vinicius G Goecks and Nicholas R Waytowich. 2023. Disasterresponsegpt: Large language models for accelerated plan of action development in disaster response scenarios. *arXiv preprint arXiv:2306.17271*.

Aneela Habib, Yasir Saleem Afridi, Madiha Sher, and Tiham Khan. 2024. Relevance classification of flood-related tweets using xlnet deep learning model.

Stéphane Hallegatte, Jun Rentschler, and Brian Walsh. 2018. *Building back better: achieving resilience through stronger, faster, and more inclusive post-disaster reconstruction*. World Bank.

Jin Han, Zhe Zheng, Xin-Zheng Lu, Ke-Yin Chen, and Jia-Rui Lin. 2024a. Enhanced earthquake impact analysis based on social media texts via large language model. *International Journal of Disaster Risk Reduction*, page 104574.

Jin Han, Zhe Zheng, Xin-Zheng Lu, Ke-Yin Chen, and Jia-Rui Lin. 2024b. Quakebert: Accurate classification of social media texts for rapid earthquake impact assessment. *arXiv preprint arXiv:2405.06684*.

Muhammad Hanif, Muhammad Waqas, Amgad Muneer, Ayed Alwadain, Muhammad Atif Tahir, and Muhammad Rafi. 2023. Deepsdc: Deep ensemble learner for the classification of social-media flooding events. *Sustainability*, 15(7):6049.

Haley Hostetter, MZ Naser, Xinyan Huang, and John Gales. 2024. Large language models in fire engineering: An examination of technical questions against domain knowledge. *arXiv preprint arXiv:2403.04795*.

James Hou and Susu Xu. 2022. Near-real-time seismic human fatality information retrieval from social media with few-shot large-language models. In *Proceedings of the 20th ACM Conference on Embedded Networked Sensor Systems*, pages 1141–1147.

Xiyang Hu and Maryam Rahnemoonfar. 2024. Flood simulation: Integrating uas imagery and ai-generated data with diffusion model. In *IGARSS 2024-2024 IEEE International Geoscience and Remote Sensing Symposium*, pages 565–568. IEEE.

Yingjie Hu, Gengchen Mai, Chris Cundy, Kristy Choi, Ni Lao, Wei Liu, Gaurish Lakhanpal, Ryan Zhenqi Zhou, and Kenneth Joseph. 2023. Geo-knowledge-guided gpt models improve the extraction of location descriptions from disaster-related social media messages. *International Journal of Geographical Information Science*, 37(11):2289–2318.

Lida Huang, Panpan Shi, Haichao Zhu, and Tao Chen. 2022. Early detection of emergency events from social media: A new text clustering approach. *Natural Hazards*, 111(1):851–875.

Muhammad Imran, Carlos Castillo, Fernando Diaz, and Sarah Vieweg. 2015. Processing social media messages in mass emergency: A survey. *ACM Computing Surveys (CSUR)*, 47(4):1–38.

Muhammad Imran, Carlos Castillo, Ji Lucas, Patrick Meier, and Sarah Vieweg. 2014. Aidr: Artificial intelligence for disaster response. In *Proceedings of the 23rd international conference on world wide web*, pages 159–162.

Muhammad Imran, Shady Elbassuoni, Carlos Castillo, Fernando Diaz, and Patrick Meier. 2013. Practical extraction of disaster-relevant information from social media. In *Proceedings of the 22nd international conference on world wide web*, pages 1021–1024.

Muhammad Imran, Prasenjit Mitra, and Carlos Castillo. 2016. Twitter as a lifeline: Human-annotated Twitter corpora for NLP of crisis-related messages. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)*, pages 1638–1643, Portorož, Slovenia. European Language Resources Association (ELRA).

G Indra and N Duraipandian. 2023. Modeling of optimal deep learning based flood forecasting model using twitter data. *Intell. Autom. Soft Comput*, 35:1455–1470.

Tarun Jain, Dinesh Gopalani, and Yogesh Kumar Meena. 2024. Classification of humanitarian crisis response through unimodal multi-class textual classification. In *2024 International Conference on Emerging Systems and Intelligent Computing (ESIC)*, pages 151–156. IEEE.

Youngsun Jang, Maryam Moshrefizadeh, Abir Mohammad Hadi, Kwanghee Won, and John Kim. 2024. Multimodal fusion of heterogeneous representations for anomaly classification in satellite imagery. In *Proceedings of the 39th ACM/SIGAPP Symposium on Applied Computing*, pages 1056–1058.

Samiha Maisha Jeba, Tanjim Taharat Aurpa, and Md Rawnak Saif Adib. 2024. From facebook posts to news headlines: using transformer models to predict post-disaster impact on mass media content. *Social Network Analysis and Mining*, 14(1):200.

Yue Jiang. 2024. The applications of large language models in emergency management. In *2024 IEEE 6th Advanced Information Management, Communicates, Electronic and Automation Control Conference (IMCEC)*, volume 6, pages 199–202. IEEE.

Tarun Kalluri, Jihyeon Lee, Kihyuk Sohn, Sahil Singla, Manmohan Chandraker, Joseph Xu, and Jeremiah Liu. 2024. Robust disaster assessment from aerial imagery using text-to-image synthetic data. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 7449–7459.

Supriya Kamoji, Mukesh Kalla, and Chandani Joshi. 2023. Fusion of multimodal textual and visual descriptors for analyzing disaster response. In *2023 5th International Conference on Smart Systems and Inventive Technology (ICSSIT)*, pages 1614–1619. IEEE.

A Krishna Kanth, P Chitra, and G Gayathri Sowmya. 2022. Deep learning-based assessment of flood severity using social media streams. *Stochastic Environmental Research and Risk Assessment*, 36(2):473–493.

Rishav Karanjit, Vidya Samadi, Amanda Hughes, Pamela Murray-Tuite, and Keri Stephens. 2024. Converging human intelligence with ai systems to advance flood evacuation decision making. *Natural Hazards and Earth System Sciences Discussions*, 2024:1–29.

Anuradha Khattar and SMK Quadri. 2022. Camm: cross-attention multimodal classification of disaster-related tweets. *IEEE Access*, 10:92889–92902.

Rani Koshy and Sivasankar Elango. 2023. Multimodal tweet classification in disaster response systems using transformer-based bidirectional attention model. *Neural Computing and Applications*, 35(2):1607–1627.

Rani Koshy and Sivasankar Elango. 2024. Applying social media in emergency response: an attention-based bidirectional deep learning system for location reference recognition in disaster tweets. *Applied Intelligence*, 54(7):5768–5793.

Saideshwar Kota, Smitha Haridasan, Ajita Rattani, Aaron Bowen, Glyn Rimmington, and Atri Dutta. 2022. Multimodal combination of text and image tweets for disaster response assessment. In *D2R2*.

Pranath Reddy Kumbam and Kshitij Maruti Vejre. 2024. Floodlense: A framework for chatgpt-based real-time flood detection. *arXiv preprint arXiv:2401.15501*.

Prashant Lakhera. 2024. Leveraging large language models (llms) for enhanced disaster recovery in aws. *Authorea Preprints*.

Rabindra Lamsal, Maria Rodriguez Read, and Shanika Karunasekera. 2024a. Crisistransformers: Pre-trained language models and sentence encoders for crisis-related social media texts. *Knowledge-Based Systems*, 296:111916.

Rabindra Lamsal, MariaRodriguez Read, Shanika Karunasekera, and Muhammad Imran. 2024b. Crema: Crisis response through computational identification and matching of cross-lingual requests and offers shared on social media. *IEEE Transactions on Computational Social Systems*.

Ben-Xiang Li and Chih-Yuan Chen. 2024. Typhoondig: Distinguishing, identifying and geo-tagging typhoon-related social media posts in taiwan. In *2024 9th International Conference on Big Data Analytics (ICBDA)*, pages 149–156. IEEE.

Hongmin Li, Doina Caragea, and Cornelia Caragea. 2021. Combining self-training with deep learning for disaster tweet classification. In *The 18th international conference on information systems for crisis response and management (ISCRAM 2021)*.

Rong Li, Lei Zhao, ZhiQiang Xie, Chunhou Ji, Jiamin Mo, Zhibing Yang, and Yuyun Feng. 2025. Mining and analyzing the evolution of public opinion in extreme disaster events from social media: A case study of the 2022 yingde flood in china. *Natural Hazards Review*, 26(1):05024015.

Xiangpeng Li, Yuqin Jiang, and Ali Mostafavi. 2023. Ai-assisted protective action: Study of chatgpt as an information source for a population facing climate hazards. *arXiv preprint arXiv:2304.06124*.

Bruce R Lindsay. 2011. Social media and disasters: Current uses, future options, and policy considerations.

Gengyin Liu and Huaiyang Zhong. 2023. Harnessing diverse data for global disaster prediction: A multimodal framework. *arXiv preprint arXiv:2309.16747*.

Junhua Liu, Trisha Singhal, Lucienne TM Blessing, Kristin L Wood, and Kwan Hui Lim. 2021. Crisisbert: a robust transformer for crisis classification and contextual crisis embedding. In *Proceedings of the 32nd ACM conference on hypertext and social media*, pages 133–141.

Zhuoran Liu, Danpei Zhao, and Bo Yuan. 2024. Rescueadi: Adaptive disaster interpretation in remote sensing images with autonomous agents. *arXiv preprint arXiv:2410.13384*.

Won Lubin, Choi Min-ji, Choi Ji-hoon, and Bae Byeong-jun. 2024. Text-to-3d cinemagraphs for generation of visual content in disaster alerts: A generative ai framework with llms and diffusion models. *JOURNAL OF BROADCAST ENGINEERING*, 29(5):662–675.

Kai Ma, YongJian Tan, Zhong Xie, Qinjun Qiu, and Siqiong Chen. 2022. Chinese toponym recognition with variant neural structures from social media messages based on bert methods. *Journal of Geographical Systems*, 24(2):143–169.

Kai Ma, Miao Tian, Yongjian Tan, Qinjun Qiu, Zhong Xie, and Rong Huang. 2023. Ontology-based bert model for automated information extraction from geological hazard reports. *Journal of Earth Science*, 34(5):1390–1405.

Sreenivasulu Madichetty and Sreekanth Madisetty. 2023. A roberta based model for identifying the multi-modal informative tweets during disaster. *Multimedia Tools and Applications*, 82(24):37615–37633.

Sreenivasulu Madichetty, Sridevi Muthukumarasamy, and P Jayadev. 2021. Multi-modal classification of twitter data during disasters for humanitarian response. *Journal of ambient intelligence and humanized computing*, 12:10223–10237.

Sreenivasulu Madichetty and M Sridevi. 2021. A neural-based approach for detecting the situational information from twitter during disaster. *IEEE Transactions on Computational Social Systems*, 8(4):870–880.

Muhammad Shahid Iqbal Malik, Muhammad Zeeshan Younas, Mona Mamdouh Jamjoom, and Dmitry I Ignatov. 2024. Categorization of tweets for damages: infrastructure and human damage assessment using fine-tuned bert model. *PeerJ Computer Science*, 10:e1859.

Satya Pranavi Manthena. 2023. Leveraging tweets for rapid disaster response using bert-bilstm-cnn model.

Siambabala Bernard Manyena. 2006. The concept of resilience revisited. *Disasters*, 30(4):434–450.

Rafaela Martelo and Ruo-Qian Wang. 2024. Towards democratized flood risk management: An advanced ai assistant enabled by gpt-4 for enhanced interpretability and public engagement. *arXiv preprint arXiv:2403.03188*.

Richard McCreadie and Cody Buntain. 2023. Crisisfacts: building and evaluating crisis timelines.

Richard McCreadie, Cody Buntain, and Ian Soboroff. 2019. Trec incident streams: Finding actionable information on social media.

Emma McDaniel, Samuel Scheele, and Jeff Liu. 2024. Zero-shot classification of crisis tweets using instruction-finetuned large language models. *arXiv preprint arXiv:2410.00182*.

Riadh Meghatria, Safa Ferrah, and Hadjer Belhenniche. 2024. Harnessing social media for natural disaster detection. In *2024 8th International Conference on Image and Signal Processing and their Applications (ISPA)*, pages 1–7. IEEE.

Ayaz Mehmood, Muhammad Tayyab Zamir, Muhammad Asif Ayub, Nasir Ahmad, and Kashif Ahmad. 2024. A named entity recognition and topic modeling-based solution for locating and better assessment of natural disasters in social media. *arXiv preprint arXiv:2405.00903*.

S Mostafa Mousavi, Marc Stogaitis, Tajinder Gadh, Richard M Allen, Alexei Barski, Robert Bosch, Patrick Robertson, Nivetha Thiruverahan, and Youngmin Cho. 2024. Gemini & physical world: Large language models can estimate the intensity of earthquake shaking from multi-modal social media posts. *arXiv preprint arXiv:2405.18732*.

Wisal Mukhtiar, Waliiya Rizwan, Aneela Habib, Yasir Saleem Afridi, Laiq Hasan, and Kashif Ahmad. 2023. Relevance classification of flood-related twitter posts via multiple transformers. *arXiv preprint arXiv:2301.00320*.

Sumera Naaz, Zain Ul Abedin, and Danish Raza Rizvi. 2021. Sequence classification of tweets with transfer learning via bert in the field of disaster management. *EAI Endorsed Transactions on Scalable Information Systems*, 8(31):e8–e8.

Thi Huyen Nguyen and Koustav Rudra. 2022a. Rationale aware contrastive learning based approach to classify and summarize crisis-related microblogs. In *Proceedings of the 31st ACM International Conference on Information & Knowledge Management*, pages 1552–1562.

Thi Huyen Nguyen and Koustav Rudra. 2022b. Towards an interpretable approach to classify and summarize crisis events from microblogs. In *Proceedings of the ACM Web Conference 2022*, pages 3641–3650.

Thi Huyen Nguyen and Koustav Rudra. 2023. Learning faithful attention for interpretable classification of crisis-related microblogs under constrained human budget. In *Proceedings of the ACM Web Conference 2023*, pages 3959–3967.

Thi Huyen Nguyen, Miroslav Shaltev, and Koustav Rudra. 2022. Crisicsum: Interpretable classification and summarization platform for crisis events from microblogs. In *Proceedings of the 31st ACM International Conference on Information & Knowledge Management*, pages 4941–4945.

AK Ningsih and AI Hadiana. 2021. Disaster tweets classification in disaster response using bidirectional encoder representations from transformer (bert). In *IOP Conference Series: Materials Science and Engineering*, volume 1115, page 012032. IOP Publishing.

Alexandra Olteanu, Carlos Castillo, Fernando Diaz, and Sarah Vieweg. 2014. Crisislex: A lexicon for collecting and filtering microblogged communications in crises. In *Proceedings of the international AAAI conference on web and social media*, volume 8, pages 376–385.

Alexandra Olteanu, Sarah Vieweg, and Carlos Castillo. 2015. What to expect when the unexpected happens: Social media communications across crises. In *Proceedings of the 18th ACM conference on computer supported cooperative work & social computing*, pages 994–1009.

Hakan T Otal and M Abdullah Canbaz. 2024. Ai-powered crisis response: Streamlining emergency management with llms. In *2024 IEEE World Forum on Public Safety Technology (WFPST)*, pages 104–107. IEEE.

Hakan T Otal, Eric Stern, and M Abdullah Canbaz. 2024. Llm-assisted crisis management: Building advanced llm platforms for effective emergency response and public collaboration. In *2024 IEEE Conference on Artificial Intelligence (CAI)*, pages 851–859. IEEE.

Tejit Pabari, Beth Tellman, Giannis Karamanolakis, Mitchell Thomas, Max Mauerman, Eugene Wu, Upmanu Lall, Marco Tedesco, Michael S Steckler, Paolo Colosio, et al. 2023. Flood event extraction from news media to support satellite-based flood insurance. *arXiv preprint arXiv:2312.14943*.

Dimitrios Panagopoulos, Adolfo Perrusquia, and Weisi Guo. 2024. Selective exploration and information gathering in search and rescue using hierarchical learning guided by natural language input. *arXiv preprint arXiv:2409.13445*.

Nayan Ranjan Paul, Rakesh Chandra Balabantaray, and Deepak Sahoo. 2023. Fine-tuning transformer-based representations in active learning for labelling crisis dataset of tweets. *SN Computer Science*, 4(5):553.

Jayr Pereira, Robson Fidalgo, Roberto Lotufo, and Rodrigo Nogueira. 2023. Crisis event social media summarization with gpt-3 and neural reranking. In *Proceedings of the 20th International ISCRAM Conference*, pages 371–384.

Courtney J Powers, Ashwin Devaraj, Kaab Ashqeen, Aman Dontula, Amit Joshi, Jayanth Shenoy, and Dhiraj Murthy. 2023. Using artificial intelligence to identify emergency messages on social media during a natural disaster: A deep learning approach. *International Journal of Information Management Data Insights*, 3(1):100164.

Maryam Rahnemoonfar, Tashnim Chowdhury, Argho Sarkar, Debvrat Varshney, Masoud Yari, and Robin Roberson Murphy. 2021. Floodnet: A high resolution aerial imagery dataset for post flood scene understanding. *IEEE Access*, 9:89644–89654.

SN Gokul Raj, P Chitra, AK Silesh, and R Lingeshwaran. 2023. Flood severity assessment using distilbert and ner. In *Machine Learning and Computational Intelligence Techniques for Data Engineering: Proceedings of the 4th International Conference MISP 2022, Volume 2*, volume 998, page 391. Springer Nature.

Sabarish Raja Ramesh Raja, MS Antony Vigil, Muthukumar Pattaiah, and B Sudarson. 2024. Analyzing the computational efficiency of llm models for nlp tweet classification during emergency-crisis. In *International Conference on Computational Intelligence in Data Science*, pages 3–15. Springer.

Rajat Rawat. 2024. Disasterqa: A benchmark for assessing the performance of llms in disaster response. *arXiv preprint arXiv:2410.20707*.

Naina Said, Kashif Ahmad, Michael Riegler, Konstantin Pogorelov, Laiq Hassan, Nasir Ahmad, and Nicola Conci. 2019. Natural disasters detection in social media and satellite imagery: a survey. *Multimedia Tools and Applications*, 78:31267–31302.

Saima Saleem, Nabeela Hasan, Anuradha Khattar, Priti Rai Jain, Tarun Kumar Gupta, and Monica Mehrotra. 2024. Deltran15: A deep lightweight transformer-based framework for multiclass classification of disaster posts on x. *IEEE Access*.

Cinthia Sánchez, Hernan Sarmiento, Andres Abeliuk, Jorge Pérez, and Barbara Poblete. 2023. Cross-lingual and cross-domain crisis classification for low-resource scenarios. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 17, pages 754–765.

Argho Sarkar, Tashnim Chowdhury, Robin Roberson Murphy, Aryya Gangopadhyay, and Maryam Rahnemoonfar. 2023. Sam-vqa: Supervised attention-based visual question answering model for post-disaster damage assessment on remote sensing imagery. *IEEE Transactions on Geoscience and Remote Sensing*, 61:1–16.

Philipp Seeberger and Korbinian Riedhammer. 2024a. Crisis2sum: An exploratory study on disaster summarization from multiple streams. *ISCRAM Proceedings*, 21.

Philipp Seeberger and Korbinian Riedhammer. 2024b. Multi-query focused disaster summarization via instruction-based prompting. *arXiv preprint arXiv:2402.09008*.

Sandeep Sharma, Saurabh Basu, Niraj Kant Kushwaha, Anugandula Naveen Kumar, and Pankai Kumar Dalela. 2021. Categorizing disaster tweets into actionable classes for disaster managers: An empirical analysis on cyclone data. In *2021 International Conference on Electrical, Computer, Communications and Mechatronics Engineering (ICECCME)*, pages 1–5. IEEE.

Nisha P Shetty, Yash Bijalwan, Pranav Chaudhari, Jayashree Shetty, and Balachandra Muniyal. 2024. Disaster assessment from social media using multimodal deep learning. *Multimedia Tools and Applications*, pages 1–26.

Pardeep Singh, Monika, Bhawna Shishodia, and Satish Chand. 2022. Twitter-based disaster response framework using electra transformer. In *Workshop on Mining Data for Financial Applications*, pages 507–516. Springer.

Adam B Smith and Richard W Katz. 2013. Us billion-dollar weather and climate disasters: data sources, trends, accuracy and biases. *Natural hazards*, 67(2):387–410.

Guizhe Song and Degen Huang. 2021. A sentiment-aware contextual model for real-time disaster prediction using twitter data. *Future Internet*, 13(7):163.

Kevin Stowe, Martha Palmer, Jennings Anderson, Marina Kogan, Leysia Palen, Kenneth M Anderson, Rebecca Morss, Julie Demuth, and Heather Lazrus. 2018. Developing and evaluating annotation procedures for twitter data during hazard events. In *Proceedings of the Joint Workshop on Linguistic Annotation, Multiword Expressions and Constructions (LAW-MWE-CxG-2018)*, pages 133–143.

Wenjuan Sun, Paolo Bocchini, and Brian D Davison. 2020. Applications of artificial intelligence for disaster management. *Natural Hazards*, 103(3):2631–2689.

Yimin Sun, Chao Wang, and Yan Peng. 2023. Unleashing the potential of large language model: Zero-shot vqa for flood disaster scenario. In *Proceedings of the 4th International Conference on Artificial Intelligence and Computer Engineering*, pages 368–373.

Reem Suwaileh, Tamer Elsayed, Muhammad Imran, and Hassan Sajjad. 2022. When a disaster happens, we are ready: Location mention recognition from crisis tweets. *International Journal of Disaster Risk Reduction*, 78:103107.

Soudabeh Taghian Dinani, Doina Caragea, and Nikesh Gyawali. 2023. Disaster tweet classification using fine-tuned deep learning models versus zero and few-shot large language models. In *International Conference on Data Management Technologies and Applications*, pages 73–94. Springer.

Anthony Meng Huat Tiong, Junnan Li, Boyang Li, Silvio Savarese, and Steven CH Hoi. 2022. Plug-and-play vqa: Zero-shot vqa by conjoining large pretrained models with zero training. *arXiv preprint arXiv:2210.08773*.

Cagri Toraman, Izzet Emre Kucukkaya, Oguzhan Ozcelik, and Umitcan Sahin. 2023. Tweets under the rubble: Detection of messages calling for help in earthquake disaster. *arXiv preprint arXiv:2302.13403*.

Edgar Marko Trono, Yutaka Arakawa, Morihiko Tamai, and Keiichi Yasumoto. 2015. Dtn mapex: Disaster area mapping through distributed computing over a delay tolerant network. In *2015 Eighth International Conference on Mobile Computing and Ubiquitous Networking (ICMU)*, pages 179–184. IEEE.

Ihsan Ullah, Anum Jamil, Imtiaz Ul Hassan, and Byung-Seo Kim. 2023. Unveiling the power of deep learning: A comparative study of lstm, bert, and gru for disaster tweet classification. *IEIE Transactions on Smart Processing & Computing*, 12(6):526–534.

Sherin R Varghese, Sujitha Juliet, and NS Athish. 2024. Social media text analysis for disaster management using distilbert model. In *2024 international conference on science technology engineering and management (ICSTEM)*, pages 1–7. IEEE.

Fedor Vitiugin and Carlos Castillo. 2022. Cross-lingual query-based summarization of crisis-related social media: An abstractive approach using transformers. In *Proceedings of the 33rd ACM Conference on Hypertext and Social Media*, pages 21–31.

Fedor Vitiugin and Hemant Purohit. 2024. Multilingual serviceability model for detecting and ranking help requests on social media during disasters. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 18, pages 1571–1584.

Junaid Abdul Wahid, Lei Shi, Yufei Gao, Bei Yang, Lin Wei, Yongcai Tao, Shabir Hussain, Muhammad Ayoub, and Imam Yagoub. 2022. Topic2labels: A framework to annotate and classify the social media data through lda topics and deep learning models for crisis response. *Expert Systems with Applications*, 195:116562.

Congcong Wang, Paul Nulty, and David Lillis. 2021. Transformer-based multi-task learning for disaster tweet categorisation. *arXiv preprint arXiv:2110.08010*.

Gelan Wang, Yu Liu, Shukai Liu, Ling Zhang, and Liqun Yang. 2024. Remflow: Rag-enhanced multi-factor rainfall flooding warning in sponge airports via large language model.

Jing Wang and Kexin Wang. 2022. Bert-based semi-supervised domain adaptation for disastrous classification. *Multimedia Systems*, 28(6):2237–2246.

Ethan Weber, Nuria Marzo, Dim P Papadopoulos, Aritro Biswas, Agata Lapedriza, Ferda Ofli, Muhammad Imran, and Antonio Torralba. 2020. Detecting natural disasters, damage, and incidents in the wild. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XIX 16*, pages 331–350. Springer.

Gwendolyn White and Sadie Liptak. 2024. Small business continuity and disaster recovery plans using ai and chatgpt. *Issues in Information Systems*, 25(4):118–126.

Rohan Singh Wilkho, Shi Chang, and Nasir G Gharaibeh. 2024. Ff-bert: A bert-based ensemble for automated classification of web-based text on flash flood events. *Advanced Engineering Informatics*, 59:102293.

Kristina Wolf, Dominik Winecki, and Arnab Nandi. 2023. Camera-first form filling: Reducing the friction in climate hazard reporting. In *Proceedings of the Workshop on Human-In-the-Loop Data Analytics*, pages 1–7.

Zhengrong Wu, Haibo Yang, Yingchun Cai, Bo Yu, Chuangheng Liang, Zheng Duan, and Qiuhua Liang. 2024. Intelligent monitoring applications of landslide disaster knowledge graphs based on chatglm2. *Remote Sensing*, 16(21):4056.

Yongqi Xia, Yi Huang, Qianqian Qiu, Xueying Zhang, Lizhi Miao, and Yixiang Chen. 2024. A question and answering service of typhoon disasters based on the t5 large language model. *ISPRS International Journal of Geo-Information*, 13(5):165.

Yangxinyu Xie, Bowen Jiang, Tanwi Mallick, Joshua David Bergerson, John K Hutchison, Duane R Verner, Jordan Branham, M Ross Alexander, Robert B Ross, Yan Feng, et al. 2024. Wildfiregpt: Tailored large language model for wildfire analysis. *arXiv preprint arXiv:2402.07877*.

Wei Xu, Xuanhua Xu, and Weiwei Zhang. Enhancing emergency decision making through risk quantification and action adjustment of human-llm dual agents. *Available at SSRN 5037356*.

Futo Yamamoto, Tadahiko Kumamoto, Yu Suzuki, and Akiyo Nadamoto. 2022. Methods of calculating usefulness ratings of behavioral facilitation tweets in disaster situations. In *Proceedings of the 11th International Symposium on Information and Communication Technology*, pages 88–95.

Pingjing Yang, Ly Dinh, Alex Stratton, and Jana Diesner. 2024. Detection and categorization of needs during crises based on twitter data. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 18, pages 1713–1726.

Kai Yin, Chengkai Liu, Ali Mostafavi, and Xia Hu. 2024. Crisissense-llm: Instruction fine-tuned large language model for multi-label social media text classification in disaster informatics. *arXiv preprint arXiv:2406.15477*.

Chen Yu and Zhiguo Wang. 2024. Multimodal social sensing for the spatio-temporal evolution and assessment of nature disasters. *Sensors*, 24(18):5889.

Manzhu Yu, Chaowei Yang, and Yun Li. 2018. Big data in natural disaster management: a review. *Geosciences*, 8(5):165.

Faxi Yuan, Chao Fan, Hamed Farahmand, Natalie Coleman, Amir Esmalian, Cheng-Chun Lee, Flavia I Patrascu, Cheng Zhang, Shangjia Dong, and Ali Mostafavi. 2022. Smart flood resilience: harnessing community-scale big data for predictive flood risk monitoring, rapid impact assessment, and situational awareness. *Environmental Research: Infrastructure and Sustainability*, 2(2):025006.

Hamada M Zahera, Rricha Jalota, Mohamed Ahmed Sherif, and Axel-Cyrille Ngonga Ngomo. 2021. I-aid: identifying actionable information from disaster-related tweets. *IEEE Access*, 9:118861–118870.

Kiran Zahra, Muhammad Imran, and Frank O Ostermann. 2020. Automatic identification of eyewitness messages on twitter during disasters. *Information processing & management*, 57(1):102107.

Cynthia Zeng and Dimitris Bertsimas. 2023. Global flood prediction: a multimodal machine learning approach. *arXiv preprint arXiv:2301.12548*.

Mengna Zhang, Qisong Huang, and Hua Liu. 2022. A multimodal data analysis approach to social media during natural disasters. *Sustainability*, 14(9):5536.

Min Zhang and Juanle Wang. 2023. Automatic extraction of flooding control knowledge from rich literature texts using deep learning. *Applied Sciences*, 13(4):2115.

Qiang Zhang, Keyang Ding, Tianwen Lyv, Xinda Wang, Qingyu Yin, Yiwen Zhang, Jing Yu, Yuhao Wang, Xiaotong Li, Zhuoyi Xiang, et al. 2024a. Scientific large language models: A survey on biological & chemical domains. *arXiv preprint arXiv:2401.14656*.

Xinsheng Zhang and Yulong Ma. 2023. An albert-based textcnn-hatt hybrid model enhanced with topic knowledge for sentiment analysis of sudden-onset disasters. *Engineering Applications of Artificial Intelligence*, 123:106136.

Yan Zhang, Zeqiang Chen, Xiang Zheng, Nengcheng Chen, and Yongqiang Wang. 2021. Extracting the location of flooding events in urban systems and analyzing the semantic risk using social sensing data. *Journal of Hydrology*, 603:127053.

Yu Zhang, Xiusi Chen, Bowen Jin, Sheng Wang, Shuiwang Ji, Wei Wang, and Jiawei Han. 2024b. A comprehensive survey of scientific large language models and their applications in scientific discovery. *arXiv preprint arXiv:2406.10833*.

Tingting Zhao, Shubo Tian, Jordan Daly, Melissa Geiger, Minna Jia, and Jinfeng Zhang. 2024. Information retrieval and classification of real-time multi-source hurricane evacuation notices. *arXiv preprint arXiv:2401.06789*.

Bing Zhou, Lei Zou, Ali Mostafavi, Binbin Lin, Mingzheng Yang, Nasir Gharaibeh, Heng Cai, Joynal Abedin, and Debayan Mandal. 2022. Victimfinder: Harvesting rescue requests in disaster response from social media with bert. *Computers, Environment and Urban Systems*, 95:101824.

Jinyan Zhou, Xingang Wang, Ning Liu, Xiaoyu Liu, Jiandong Lv, Xiaomin Li, Hong Zhang, and Rui Cao. 2023a. Visual and linguistic double transformer fusion model for multimodal tweet classification. In *2023 International Joint Conference on Neural Networks (IJCNN)*, pages 1–8. IEEE.

Jinyan Zhou, Xingang Wang, Jiandong Lv, Ning Liu, Hong Zhang, Rui Cao, Xiaoyu Liu, and Xiaomin Li. 2023b. Public crisis events tweet classification based on multimodal cycle-gan. In *2023 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, pages 2251–2257. IEEE.

Jun Zhu, Pei Dang, Yungang Cao, Jianbo Lai, Yukun Guo, Ping Wang, and Weilian Li. 2024. A flood knowledge-constrained large language model interactable with gis: enhancing public risk perception of floods. *International Journal of Geographical Information Science*, 38(4):603–625.

Abdul Wahab Ziaullah, Ferda Ofli, and Muhammad Imran. 2024. Monitoring critical infrastructure facilities during disasters using large language models. *arXiv preprint arXiv:2404.14432*.

Liwei Zou, Zhi He, Chengle Zhou, and Wenbing Zhu. 2024. Multi-class multi-label classification of social media texts for typhoon damage assessment: a two-stage model fully integrating the outputs of the hidden layers of bert. *International Journal of Digital Earth*, 17(1):2348668.

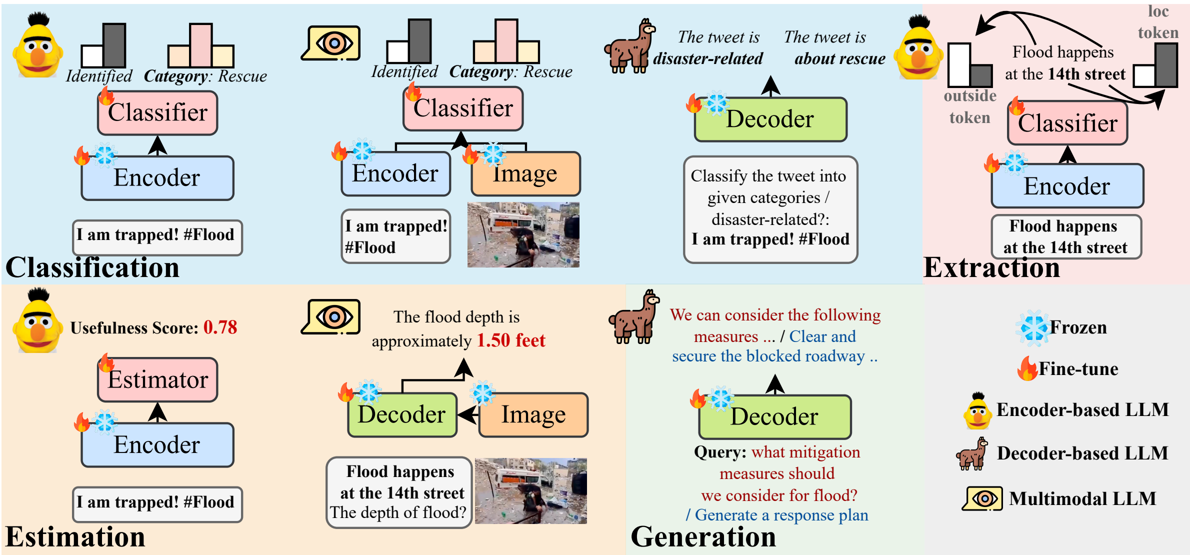

Figure 4: Pipeline of major tasks performed by different types of LLMs in disaster management.

A Summary of Papers

A.1 Summary Table

Table 1 summarizes the surveyed papers, detailing their disaster phases, application scenarios, specific tasks, and architecture types.

A.2 Pipeline Illustration

In this section, we present Figure 4, which illustrates the role of LLMs in disaster management. The figure outlines the major pipelines of three LLM architectures—encoder-based, decoder-based, and multimodal—applied across the four task types covered in this survey: classification, extraction, estimation, and generation. This visualization provides key insights into their mechanisms and applications in disaster management.

A.3 Statistics

To provide a comprehensive overview of the current state of LLMs in disaster management, we present statistics from the surveyed papers, highlighting a significant gap between the NLP and disaster management communities. This gap underscores the urgent need for stronger interdisciplinary collaboration to bridge these fields and fully harness the potential of LLMs in addressing disaster-related challenges.

Figure 3 illustrates the number of publications leveraging existing LLMs versus those developing new frameworks, revealing that most studies are heavily application-focused. The majority rely on fine-tuning or prompting existing LLMs for disaster management tasks, rather than designing novel architectures. While some efforts have provided valuable insights, most research remains concentrated on the response phase, with limited exploration across other critical disaster management scenarios. Figure 5 illustrates the distribution of publications across academic venues, revealing that relatively few disaster management papers appear in NLP- or AI-specific conferences and journals. This trend reflects limited engagement from the LLM research community in this domain, underscoring the need to increase awareness and foster greater collaboration within the field.

B Datasets

Table 2 summarizes existing publicly available datasets. For classification tasks, we exclude datasets that focus on a single disaster type if they are already incorporated into comprehensive benchmarks such as CrisisBench (Alam et al., 2021b).

B.1 Classification Datasets

- **CrisisLexT6** (Olteanu et al., 2014): This dataset is designed for relevance classification. It contains data from six crisis events between October 2012 and July 2013.

- **CrisisLexT26** (Olteanu et al., 2015): This dataset is an updated version of CrisisLexT6, which contains public data from 26 crisis events in 2012 and 2013 with relevance information and six humanitarian categories.

- **CrisisNLP** (Imran et al., 2016): This dataset is a large-scale dataset that includes classes from humanitarian disaster responses and classes related to health emergencies. It is collected from 19 different disaster events that happened between 2013 and 2015.

- **SWDM2013** (Imran et al., 2013): This dataset is utilized for relevance classification that consists of tweets from two events: (i) the Joplin collection contains tweets from the tornado that struck Joplin, Missouri on May 22, 2011; (ii) The Sandy collection contains tweets collected from Hurricane Sandy that struck the Northeastern US on Oct 29, 2012.

- **ISCRAM2013** (Imran et al., 2013): This dataset consists of tweets collected from the same events as in SWDM2013, containing both relevance and humanitarian categories.

- **Disaster Response Data (DRD)** (Alam et al., 2021b): This dataset consists of tweets collected during various crisis events that took place in 2010 and 2012. This dataset is annotated using 36 classes that include relevance as well as humanitarian categories.

- **Disasters on Social Media (DSM)** (Alam et al., 2021b): This dataset comprises 10K tweets annotated with relevance labels.

- **AIDR** (Imran et al., 2014): This dataset contains data obtained from the AIDR system on September 25, 2013, collecting tweets using hashtags such as "#earthquake". It is utilized for relevance and humanitarian classification.

- **CrisisMMD** (Alam et al., 2018): This dataset is a multimodal and multitask dataset comprising $16k$ labeled tweets and corresponding images. Tweets have been sourced from seven natural disaster events that took place in 2017. Each sample is annotated with relevance, humanitarian (eight classes), and damage severity categories (mild, severe, and none).

- **Multi-Crisis** (Sánchez et al., 2023): This dataset was proposed to evaluate transfer learning scenarios where data from high-resource languages (e.g., English) is used to classify messages in low-resource languages (e.g., Spanish, Italian) and unseen crisis domains, with relevance and humanitarian categories. It is collected from 7 existing datasets, 53 crisis events, and contains 9 domains.

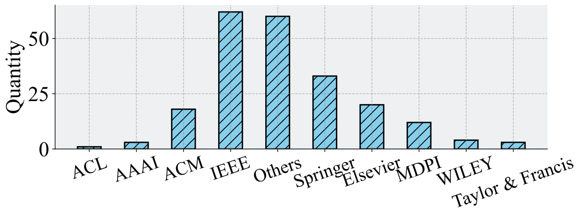

Figure 5: Publication distribution across various academic venues, with a focus on model design on the left and application-based research on the right.

- **CrisisBench** (Alam et al., 2021b): This dataset is a comprehensive benchmark consolidated from 9 existing datasets, utilized for relevance and humanitarian classification.

- **Eyewitness Messages** (Zahra et al., 2020): This dataset is designed to identify disaster eyewitness-related tweets and classify them into three categories: direct eyewitnesses, indirect eyewitnesses, and vulnerable eyewitnesses—individuals who anticipate a disaster and are present in regions where disaster warnings have been issued. It comprises 14,000 tweets collected from earthquakes, hurricanes, and wildfires.

- **TREC Incident Streams** (McCreadie et al., 2019): This dataset has been developed as part of the TREC-IS 2018 evaluation challenge and consists of $20k$ tweets labeled for actionable information identification and information criticality assessment.

- **HumAID** (Alam et al., 2021a): This dataset contains $77k$ labeled tweets, which are sampled from 24 million tweets collected during 19 disasters between 2016 and 2019, including hurricanes, earthquakes, wildfires, and floods. It is balanced in terms of disaster types and contains 7 humanitarian categories.

- **EPIC**: This dataset contains data primarily collected from Hurricane Sandy, including tweets from 93 users across four annotation schemes, with data spanning three weeks around the hurricane's landfall. It is used for relevance and humanitarian classification.

B.2 Estimation Datasets

- **Did You Feel It (DYFI)** (Atkinson and Wald, 2007): This dataset includes ground shaking intensity and geographic distribution information, collected from post-earthquake reports through 750,000 online questionnaire responses from individuals who experienced the event.

- **FloodDepth** (Akinboyewa et al., 2024): This dataset consists of 150 flood photos collected online, used to estimate floodwater depth based on various reference objects, including stop signs, vehicles, and humans.

- **Behavioral Facilitation (BF)** (Yamamoto et al., 2022): This dataset, collected after the 2018 Hokkaido earthquake, includes data labeled with usefulness ratings based on behavioral facilitation information.

B.3 Extraction Datasets

- **(Fu et al., 2024)**: This dataset contains county-level data from news media collected during urban flood events from 2000 to 2022. It is utilized to extract information such as the time and location of disasters.

- **(Ma et al., 2023)**: This dataset is designed for entity and relation extraction, comprising 5,560 annotated instances, 12,980 entities, and 6,895 relations derived from reports on geological hazards.

- **DisasterMM** (Andreadis et al., 2022): This dataset was collected from Twitter by searching for flood-related keywords. It consists of two subsets: RCTP, which includes 6,672 tweets for relevance classification, and LETT, which contains 4,992 tweets used for location extraction. In the LETT subset, words are annotated with "B-LOC" for the first word in a sequence referring to a location, "I-LOC" for subsequent words within the same location sequence, and "O" for words that do not correspond to a location.

- **(Suwaileh et al., 2022)**: This dataset contains 22,000 crisis-related tweets from various disasters, including floods, earthquakes, and hurricanes. It is annotated with location-related tags such as "inLOC" and "outLOC."

- **Re´SoCIO** (Caillaut et al., 2024): This dataset is constructed by merging Wikipedia datasets and multiple disaster-related datasets, annotated with a set of 9 NER labels with different types of information.

- **(Nguyen and Rudra, 2022a)**: This dataset contains tweet data with annotated rationales from 4 subsets of CrisisNLP. It is used for rationale extraction, and the extracted rationales can assist in disaster classification.

B.4 Generation Datasets

- **(Vitiugin and Castillo, 2022)**: This dataset is used to generate summaries of various disaster events, with the official report of each event serving as the ground truth.

- **CrisisFACTS** (McCreadie and Buntain, 2023): This dataset is a multi-stream collection comprising data from eight crisis events gathered across various platforms. It is designed to process daily multi-platform streams and generate summaries based on specific information needs, such as "Have airports closed?"

- **DisasterQA** (Rawat, 2024): This dataset includes disaster-related multiple choice questions from 7 different sources, examples could be "What causes a tsunami?".

- **FFD-IQA** (Sun et al., 2023): This dataset comprises 2,058 images and 22,422 question-meta ground truth pairs related to the safety of individuals trapped in disaster sites and the availability of emergency services. It includes three types of questions: free-form, multiple-choice, and yes-no questions.

- **FloodNet** (Rahnemoonfar et al., 2021): This dataset consists of 4,500 question-image pairs collected after Hurricane Harvey. The questions pertain to buildings, roads, and entire scenes, categorized into four groups: "Simple Counting," "Complex Counting," "Yes/No," and "Condition Recognition."

Table 1: Summary of LLMs in disaster management with their disaster phases, application scenarios, specific tasks, and architecture types. "Arch": Type of LLM architectures used; "NM": Whether the paper presents novel methods.

Paper	**Phase**	**Application**	**Task**	**Arch**	**NM**
(Chowdhury et al., 2024)	Mitigation	Vulnerability Assessment	Vulnerability Classification	Decoder	No
(Martelo and Wang, 2024)	Mitigation	Vulnerability Assessment	Answer Generation	Decoder	Yes
(Fu et al., 2024)	Preparedness	Public Awareness Enhancement	Knowledge Extraction	Encoder	No
(Zhang and Wang, 2023)	Preparedness	Public Awareness Enhancement	Knowledge Extraction	Encoder	No
(Ma et al., 2023)	Preparedness	Public Awareness Enhancement	Knowledge Extraction	Encoder	Yes
(Wu et al., 2024)	Preparedness	Public Awareness Enhancement	Knowledge Extraction	Decoder	No
(Hostetter et al., 2024)	Preparedness	Public Awareness Enhancement	Answer Generation	Decoder	No
(Martelo and Wang, 2024)	Preparedness	Public Awareness Enhancement	Answer Generation	Decoder	No
(Li et al., 2023)	Preparedness	Public Awareness Enhancement	Answer Generation	Decoder	No
(Indra and Duraipandian, 2023)	Preparedness	Disaster Forecast	Occurrence Classification	Encoder	Yes
(Zeng and Bertsimas, 2023)	Preparedness	Disaster Forecast	Occurrence Classification	Multimodal	Yes
(Liu and Zhong, 2023)	Preparedness	Disaster Forecast	Occurrence Classification	Multimodal	Yes
(Wang et al., 2024)	Preparedness	Disaster Forecast	Occurrence Classification	Decoder	No
(Chandra et al., 2024)	Preparedness	Disaster Warning	Warning Generation	Decoder	No
(Martelo and Wang, 2024)	Preparedness	Disaster Warning	Warning Generation	Decoder	No
(Lubin et al., 2024)	Preparedness	Disaster Warning	Image Generation	Multimodal	Yes
(Hostetter et al., 2024)	Preparedness	Evacuation Planning	Plan Generation	Decoder	No
(Ningsih and Hadiana, 2021)	Response	Disaster Identification	Relevance Classification	Encoder	No
(Madichetty and Madisetty, 2023)	Response	Disaster Identification	Relevance Classification	Encoder	No
(Singh et al., 2022)	Response	Disaster Identification	Relevance Classification	Encoder	No
(Powers et al., 2023)	Response	Disaster Identification	Relevance Classification	Encoder	No
(Duraisamy and Natarajan, 2024)	Response	Disaster Identification	Relevance Classification	Encoder	No
(Ullah et al., 2023)	Response	Disaster Identification	Relevance Classification	Encoder	No
(Li and Chen, 2024)	Response	Disaster Identification	Relevance Classification	Encoder	No
(Zhao et al., 2024)	Response	Disaster Identification	Relevance Classification	Encoder	No
(Karanjit et al., 2024)	Response	Disaster Identification	Relevance Classification	Encoder	No
(Pabari et al., 2023)	Response	Disaster Identification	Relevance Classification	Encoder	No
(de Bruijn et al., 2019)	Response	Disaster Identification	Relevance Classification	Encoder	No
(Zhao et al., 2024)	Response	Disaster Identification	Relevance Classification	Encoder	No
(Wang et al., 2021)	Response	Disaster Identification	Relevance Classification	Encoder	No
(Habib et al., 2024)	Response	Disaster Identification	Relevance Classification	Encoder	No
(Liu et al., 2021)	Response	Disaster Identification	Relevance Classification	Encoder	No
(Fontalis et al., 2023)	Response	Disaster Identification	Relevance Classification	Encoder	No
(Mehmood et al., 2024)	Response	Disaster Identification	Relevance Classification	Encoder	No
(Paul et al., 2023)	Response	Disaster Identification	Relevance Classification	Encoder	Yes
(Lamsal et al., 2024a)	Response	Disaster Identification	Relevance Classification	Encoder	Yes
(Manthena, 2023)	Response	Disaster Identification	Relevance Classification	Encoder	Yes
(Danday and Murthy, 2022)	Response	Disaster Identification	Relevance Classification	Encoder	Yes
(Ghosh et al., 2022)	Response	Disaster Identification	Relevance Classification	Encoder	Yes
(Taghian Dinani et al., 2023)	Response	Disaster Identification	Relevance Classification	Decoder	No
(Kamoji et al., 2023)	Response	Disaster Identification	Relevance Classification	Multimodal	Yes
(Madichetty et al., 2021)	Response	Disaster Identification	Relevance Classification	Multimodal	Yes
(Koshy and Elango, 2023)	Response	Disaster Identification	Relevance Classification	Multimodal	Yes
(Shetty et al., 2024)	Response	Disaster Identification	Relevance Classification	Multimodal	Yes
(Zhou et al., 2023b)	Response	Disaster Identification	Relevance Classification	Multimodal	Yes
(Yu and Wang, 2024)	Response	Disaster Identification	Relevance Classification	Multimodal	Yes
(Zhang et al., 2022)	Response	Disaster Identification	Relevance Classification	Multimodal	Yes
(Kota et al., 2022)	Response	Disaster Identification	Relevance Classification	Multimodal	Yes
(Wang and Wang, 2022)	Response	Disaster Identification	Relevance Classification	Multimodal	Yes
(Hanif et al., 2023)	Response	Disaster Identification	Relevance Classification	Multimodal	Yes

Paper	Phase	Application	Task	Arch	NM
(Jang et al., 2024)	Response	Disaster Identification	Relevance Classification	Multimodal	Yes
(Madichetty and Sridevi, 2021)	Response	Disaster Situation Assessment	Situation Classification	Encoder	Yes
(Raj et al., 2023)	Response	Disaster Situation Assessment	Situation Classification	Encoder	Yes
(Kanth et al., 2022)	Response	Disaster Situation Assessment	Situation Classification	Multimodal	Yes
(Mousavi et al., 2024)	Response	Disaster Situation Assessment	Severity Estimation	Decoder	No
(Akinboyewa et al., 2024)	Response	Disaster Situation Assessment	Severity Estimation	Multimodal	No
(Hu and Rahnemoonfar, 2024)	Response	Disaster Situation Assessment	Description Generation	Multimodal	No
(Wolf et al., 2023)	Response	Disaster Situation Assessment	Description Generation	Multimodal	No
(Yamamoto et al., 2022)	Response	Disaster Information Coordination	Usefulness Estimation	Encoder	No
(Blomeier et al., 2024)	Response	Disaster Information Coordination	Relevance Classification	Encoder	No
(Adesokan et al., 2023)	Response	Disaster Information Coordination	Information Classification	Encoder	No
(Wahid et al., 2022)	Response	Disaster Information Coordination	Information Classification	Encoder	No
(Chandrakala and Raj, 2022)	Response	Disaster Information Coordination	Information Classification	Encoder	No
(Naaz et al., 2021)	Response	Disaster Information Coordination	Information Classification	Encoder	No
(Du et al., 2023)	Response	Disaster Information Coordination	Information Classification	Encoder	No
(Adesokan et al., 2023)	Response	Disaster Information Coordination	Information Classification	Encoder	No
(Han et al., 2024b)	Response	Disaster Information Coordination	Information Classification	Encoder	No
(Sharma et al., 2021)	Response	Disaster Information Coordination	Information Classification	Encoder	No
(Yuan et al., 2022)	Response	Disaster Information Coordination	Information Classification	Encoder	No
(Liu et al., 2021)	Response	Disaster Information Coordination	Information Classification	Encoder	No
(Boros et al., 2022)	Response	Disaster Information Coordination	Information Classification	Encoder	Yes
(Li et al., 2021)	Response	Disaster Information Coordination	Information Classification	Encoder	Yes
(Zou et al., 2024)	Response	Disaster Information Coordination	Information Classification	Encoder	Yes
(Zahera et al., 2021)	Response	Disaster Information Coordination	Information Classification	Encoder	Yes
(Wilkho et al., 2024)	Response	Disaster Information Coordination	Information Classification	Encoder	Yes
(Nguyen and Rudra, 2022b)	Response	Disaster Information Coordination	Information Classification	Encoder	Yes
(Nguyen and Rudra, 2023)	Response	Disaster Information Coordination	Information Classification	Encoder	Yes
(Nguyen and Rudra, 2022a)	Response	Disaster Information Coordination	Information Classification	Encoder	Yes
(Otal and Canbaz, 2024)	Response	Disaster Information Coordination	Information Classification	Decoder	No
(Yin et al., 2024)	Response	Disaster Information Coordination	Information Classification	Decoder	No
(Dinani et al., 2024)	Response	Disaster Information Coordination	Information Classification	Decoder	No
(Zhang et al., 2022)	Response	Disaster Information Coordination	Information Classification	Multimodal	Yes
(Yu and Wang, 2024)	Response	Disaster Information Coordination	Information Classification	Multimodal	Yes
(Shetty et al., 2024)	Response	Disaster Information Coordination	Information Classification	Multimodal	Yes
(Abavisani et al., 2020)	Response	Disaster Information Coordination	Information Classification	Multimodal	Yes
(Zhou et al., 2023a)	Response	Disaster Information Coordination	Information Classification	Multimodal	Yes
(Basit et al., 2023)	Response	Disaster Information Coordination	Information Classification	Multimodal	Yes
(Yang et al., 2024)	Response	Disaster Information Coordination	Need Classification	Encoder	No
(Toraman et al., 2023)	Response	Disaster Information Coordination	Need Classification	Encoder	No
(Zhou et al., 2022)	Response	Disaster Information Coordination	Need Classification	Encoder	No
(Vitiugin and Purohit, 2024)	Response	Disaster Information Coordination	Need Classification	Encoder	Yes
(Conneau, 2019)	Response	Disaster Information Coordination	Need Classification	Encoder	Yes
(Lamsal et al., 2024b)	Response	Disaster Information Coordination	Need Classification	Encoder	Yes
(Mehmood et al., 2024)	Response	Disaster Information Coordination	Location Extraction	Encoder	No
(Suwaileh et al., 2022)	Response	Disaster Information Coordination	Location Extraction	Encoder	No
(Koshy and Elango, 2024)	Response	Disaster Information Coordination	Location Extraction	Encoder	Yes
(Ma et al., 2022)	Response	Disaster Information Coordination	Location Extraction	Encoder	Yes
(Zhang et al., 2021)	Response	Disaster Information Coordination	Location Extraction	Encoder	Yes
(Caillaut et al., 2024)	Response	Disaster Information Coordination	Location Extraction	Encoder	Yes
(Yu and Wang, 2024)	Response	Disaster Information Coordination	Location Extraction	Decoder	No
(Hu et al., 2023)	Response	Disaster Information Coordination	Location Extraction	Decoder	No
(Firmansyah et al., 2024)	Response	Disaster Information Coordination	Location Extraction	Decoder	No
(Nguyen and Rudra, 2022a)	Response	Disaster Information Coordination	Summary Extraction	Encoder	Yes
(Nguyen et al., 2022)	Response	Disaster Information Coordination	Summary Extraction	Encoder	Yes
(Garg et al., 2024)	Response	Disaster Information Coordination	Summary Extraction	Encoder	Yes

Paper	Phase	Application	Task	Arch	NM
(Vitiugin and Castillo, 2022)	Response	Disaster Information Coordination	Summary Extraction	Decoder	Yes
(Colverd et al., 2023)	Response	Disaster Information Coordination	Report Generation	Decoder	No
(Pereira et al., 2023)	Response	Disaster Information Coordination	Report Generation	Decoder	No
(Seeberger and Riedhammer, 2024b)	Response	Disaster Information Coordination	Report Generation	Decoder	Yes
(Seeberger and Riedhammer, 2024a)	Response	Disaster Information Coordination	Report Generation	Decoder	Yes
(Goecks and Waytowich, 2023)	Response	Disaster Rescuing	Plan Generation	Decoder	No
(Panagopoulos et al., 2024)	Response	Disaster Rescuing	Code Generation	Decoder	No
(Rawat, 2024)	Response	Disaster Issue Consultation	Answer Generation	Decoder	No
(Chen and Fang, 2024)	Response	Disaster Issue Consultation	Answer Generation	Decoder	No
(Xie et al., 2024)	Response	Disaster Issue Consultation	Answer Generation	Decoder	No
(Chen et al., 2024)	Response	Disaster Issue Consultation	Answer Generation	Decoder	Yes
(Xia et al., 2024)	Response	Disaster Issue Consultation	Answer Generation	Decoder	Yes
(Sun et al., 2023)	Response	Disaster Issue Consultation	Answer Generation	Multimodal	No
(Liu et al., 2024)	Response	Disaster Issue Consultation	Answer Generation	Multimodal	Yes
(Kumbam and Vejre, 2024)	Response	Disaster Issue Consultation	Answer Generation	Multimodal	Yes
(Malik et al., 2024)	Recovery	Disaster Impact Assessment	Damage Classification	Encoder	No
(Chen and Lim, 2021)	Recovery	Disaster Impact Assessment	Damage Classification	Encoder	No
(Jeba et al., 2024)	Recovery	Disaster Impact Assessment	Damage Classification	Encoder	No
(Zou et al., 2024)	Recovery	Disaster Impact Assessment	Damage Classification	Encoder	Yes
(Chen and Lim, 2021)	Recovery	Disaster Impact Assessment	Damage Estimation	Encoder	Yes
(Ziaullah et al., 2024)	Recovery	Disaster Impact Assessment	Answer Generation	Decoder	No
(Estêvão, 2024)	Recovery	Disaster Impact Assessment	Answer Generation	Multimodal	No
(Sarkar et al., 2023)	Recovery	Disaster Impact Assessment	Answer Generation	Multimodal	No
(Alsan and Arsan, 2023)	Recovery	Disaster Impact Assessment	Answer Generation	Multimodal	No
(Hou and Xu, 2022)	Recovery	Disaster Impact Assessment	Statistic Extraction	Decoder	No
(Han et al., 2024a)	Recovery	Disaster Impact Assessment	Sentiment Classification	Encoder	No
(Alharm and Naim)	Recovery	Disaster Impact Assessment	Sentiment Classification	Encoder	No
(Zhang and Ma, 2023)	Recovery	Disaster Impact Assessment	Sentiment Classification	Encoder	No
(Varghese et al., 2024)	Recovery	Disaster Impact Assessment	Sentiment Classification	Encoder	No
(Berbère et al., 2023)	Recovery	Disaster Impact Assessment	Sentiment Classification	Encoder	No
(Li et al., 2025)	Recovery	Disaster Impact Assessment	Sentiment Classification	Decoder	No
(White and Liptak, 2024)	Recovery	Recovery Plan Generation	Plan Generation	Decoder	No
(Lakhera, 2024)	Recovery	Recovery Plan Generation	Plan Generation	Decoder	No
(CONTRERAS et al.)	Recovery	Recovery Process Tracking	Sentiment Classification	Encoder	No

Table 2: Summary of publicly available datasets utilized in disaster management. For **Application**, "DI": Disaster Identification; "DInf": Disaster Information Coordination; "DIC": Disaster Issue Consultation; "DSA": Disaster Situation Assessment; "PAE": Public Awareness Enhancement; "DIA": Disaster Impact Assessment. For **Disaster Type**, "Mix" denotes the datasets contain various types of disasters.

Dataset	Phase	Application	Task	Disaster Type	Modality	Used in	#Sample
CrisisLexT6 (Olteanu et al., 2014)	Response	DI	Classification	Mix	Text	(McDaniel et al., 2024)	60,082
CrisisLexT26 (Olteanu et al., 2015)	Response	DI, DInf	Classification	Mix	Text	(McDaniel et al., 2024)	27,933
CrisisNLP (Imran et al., 2016)	Response	DI, DInf	Classification	Mix	Text	(Taghian Dinani et al., 2023)	52,656
SWDM13 (Imran et al., 2013)	Response	DI, DInf	Classification	Mix	Text	(McDaniel et al., 2024)	1,543
ISCRAM2013 (Imran et al., 2013)	Response	DI, DInf	Classification	Mix	Text	(McDaniel et al., 2024)	3,617
DRD (Alam et al., 2021b)	Response	DI, DInf	Classification	Mix	Text	(McDaniel et al., 2024)	26,235
DSM (Alam et al., 2021b)	Response	DI	Classification	Mix	Text	(McDaniel et al., 2024)	10,876

Dataset	Phase	Application	Task	Disaster Type	Modality	Used in	#Sample
AIDR (Imran et al., 2014)	Response	DI, DInf	Classification	Mix	Text	(McDaniel et al., 2024)	7,411
CrisisMMD (Alam et al., 2018)	Response	DI, DInf	Classification	Mix	Text, Image	(Jain et al., 2024)	16,058
Multi-Crisis (Sánchez et al., 2023)	Response	DI, DInf	Classification	Mix	Text	(Sánchez et al., 2023)	164,625
CrisisBench (Alam et al., 2021b)	Response	DI, DInf	Classification	Mix	Text	(McDaniel et al., 2024)	109,796
Eyewitness Messages (Zahra et al., 2020)	Response	DInf	Classification	Mix	Text	(Zahra et al., 2020)	14,000
TREC Incident Streams (McCreadie et al., 2019)	Response	DI, DInf	Classification	Mix	Text	(Khattar and Quadri, 2022)	19,784
HumAID (Alam et al., 2021a)	Response	DInf	Classification	Mix	Text	(Basit et al., 2023)	77,000
EPIC (Stowe et al., 2018)	Response	DI, DInf	Classification	Mix	Text	(Adesokan et al., 2023)	3469
Did You Feel It (DYFI) (Mousavi et al., 2024)	Response	DSA	Estimation	Earthquake	Text	(Mousavi et al., 2024)	750,000
FloodDepth (Akinboyewa et al., 2024)	Response	DSA	Estimation	Flood	Text, Image	(Akinboyewa et al., 2024)	150
Behavioral Facilitation (BF) (Yamamoto et al., 2022)	Response	DInf	Estimation	Earthquake	Text	(Yamamoto et al., 2022)	1,400
(Fu et al., 2024)	Preparedness	PAE	Extraction	Flood	Text	(Fu et al., 2024)	633
(Ma et al., 2023)	Preparedness	PAE	Extraction	Landslide	Text	(Ma et al., 2023)	5,560
DisasterMM (Andreadis et al., 2022)	Response	DI, DInf	Classification, Extraction	Flood	Text	(Mehmood et al., 2024)	6,672, 4,992
(Suwaileh et al., 2022)	Response	DInf	Extraction	Mix	Text	(Suwaileh et al., 2022)	22,137
Re´SoCIO (Caillaut et al., 2024)	Response	DInf	Extraction	Flood	Text	(Caillaut et al., 2024)	4,617
(Nguyen and Rudra, 2022a)	Response	DInf	Extraction	Mix	Text	(Nguyen and Rudra, 2022a)	32
(Vitiugin and Castillo, 2022)	Response	DInf	Generation	Mix	Text	(Vitiugin and Castillo, 2022)	5,791
CrisisFACTS (McCreadie and Buntain, 2023)	Response	DIC	Generation	Mix	Text	(Pereira et al., 2023)	748,466
DisasterQA (Rawat, 2024)	Response	PAE, DIC	Generation	Mix	Text	(Rawat, 2024)	707
FFD-IQA (Sun et al., 2023)	Response	DIC	Generation	Flood	Text, Image	(Sun et al., 2023)	22,422
FloodNet (Rahnemoonfar et al., 2021)	Recovery	DIA	Generation	Flood	Text, Image	(Sarkar et al., 2023)	4,500