Towards Crisis-Aware Controllable Generation for IPAWS-Style Emergency Dissemination through Context Retrieval Inference

Anonymous ACL submission

Abstract

This study introduces a weakly supervised and controllable generation framework for transforming noisy citizen-generated disaster communication into official-style emergency dissemination messages. The proposed framework bridges the gap between informal public crisis discourse and structured public-warning systems by semantically aligning citizen-generated disaster reports with official IPAWS dissemination archives. Structured crisis-context extraction, context-control conditioning, retrieval-augmented generation, and LoRA-based fine-tuning are combined to generate calm, actionable, and operationally grounded dissemination outputs. Experiments conducted using Mistral 7B and Qwen2.5 demonstrate substantial improvements in dissemination quality, controllability, factual consistency, and hallucination reduction compared to the corresponding base models. Results further show that retrieval augmentation and operational context conditioning improve dissemination grounding and structural consistency. The study establishes a new citizen-to-alert generation benchmark that extends disaster Natural Language Processing (NLP) research beyond classification and situational analytics toward operational emergency communication generation.

1 Introduction

Social media platforms have become one of the earliest and largest sources of disaster-related information during emergencies, where citizens frequently publish real-time reports about floods, storms, earthquakes, fires, damaged infrastructure, and rescue needs before official alerts are released. These citizen-generated signals play an important role in disaster response coordination, situational awareness, and humanitarian assessment (Imran et al., 2015; Kriz and Hribljan, 2018; Yin et al., 2024; Lamsal et al., 2024). Consequently, disaster informatics and crisis NLP research have

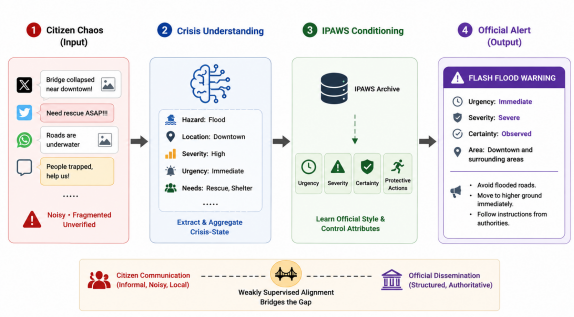

Figure 1: Overview of the proposed weakly supervised citizen-to-alert dissemination framework.

achieved substantial progress in disaster classification, humanitarian-category detection, event extraction, multimodal disaster understanding, and crisis summarization (Alam et al., 2021; Shaik and Oussalah, 2024; Shaik et al., 2025). However, most existing studies focus primarily on information identification or retrieval rather than generating structured operational dissemination outputs required by emergency management agencies. This creates a significant gap between noisy citizen-generated disaster discourse and official public-warning systems.

Bridging this gap is challenging because citizen-generated communication and official dissemination alerts belong to fundamentally different communication domains. Citizen reports are often fragmented, emotional, duplicated, geographically ambiguous, and temporally inconsistent (Palen, 2008; Vieweg et al., 2010), whereas official dissemination systems such as IPAWS require highly standardized communication structures containing hazard type, urgency, severity, certainty, affected regions, and protective actions. Existing NLP approaches mainly formulate disaster communication as classification, retrieval, or summarization tasks (Olteanu et al., 2014; Alam et al., 2021), while the downstream problem of transforming citizen-generated crisis signals into authoritative dissemination lan-

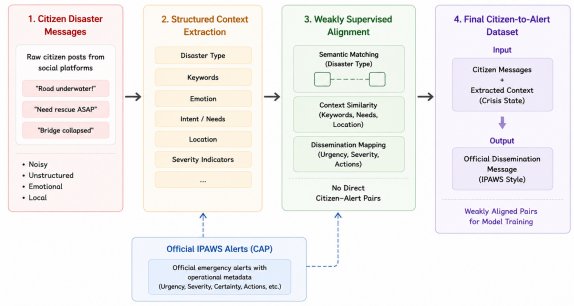

Figure 2: Overview of the weakly supervised dataset construction pipeline for citizen-to-alert dissemination generation.

guage remains largely unexplored. Although recent multimodal disaster studies such as CrisisMMD-R (Zhang et al., 2026) improve contextual understanding and explainability, they do not address operational emergency dissemination generation. Recent advances in transformer architectures, instruction tuning, and controllable generation (Vaswani et al., 2017; Raffel et al., 2020; Sanh et al., 2022; Ouyang et al., 2022) provide new opportunities for this task, but official emergency communication still requires factual consistency, low hallucination risk, operational clarity, and adherence to dissemination protocols.

To address these challenges, this study introduces a crisis-aware controllable dissemination framework for transforming noisy citizen-generated disaster communication into official-style emergency dissemination messages. Figure 1 presents an overview of the proposed framework. We formulate this problem as a weakly supervised citizen-to-alert generation task by semantically aligning citizen-generated disaster messages with official IPAWS dissemination archives to construct a weakly aligned dissemination corpus. The framework combines structured crisis-context extraction, semantic alignment, context-control conditioning, retrieval-augmented generation, and LoRA-based fine-tuning to generate calm, actionable, and operationally grounded dissemination outputs. The main contributions of this study are fourfold: (1) the introduction of a weakly aligned citizen-to-alert dissemination dataset, (2) a controllable crisis-aware generation framework that learns the transition from citizen discourse to official dissemination language, (3) dissemination control-conditioning mechanisms using operational attributes such as disaster type, urgency, severity, and protective-action cues, and (4) the establishment of a new benchmark task that extends disaster NLP research beyond classification and situational analytics toward operational public-warning communication generation.

2 Related Work

2.1 Crisis Informatics and Social Media Understanding

Recent crisis informatics research has shown that social media provides valuable real-time signals during disasters, but most NLP work has focused on identifying, filtering, or classifying crisis-related messages rather than generating operational dissemination outputs. Ponce-López and Spataru (Ponce-López and Spataru, 2022) proposed a social media data analysis framework for disaster response using machine learning and transformer-based classifiers across multiple disaster datasets, showing that social media messages can be filtered into disaster-relevant categories for response support. However, their framework remains classification-oriented and does not address how extracted signals can be converted into official public-warning language. Similarly, Dharrao et al. (Dharrao et al., 2024) proposed a BERT-embedding-based CNN model for disaster tweet classification using the Kaggle Disaster Tweets dataset, improving the detection of real disaster-related tweets but stopping at binary classification rather than alert generation. Hanny et al. (Hanny et al., 2024) explored active learning for identifying disaster-related tweets with limited annotation effort, comparing keyword filtering, RoBERTa fine-tuning, and active learning. Their work reduces labeling cost for disaster tweet detection, but it does not model the downstream transformation from detected citizen reports into structured official messages. These studies provide important foundations for crisis message understanding, but they do not address the citizen-to-alert generation problem studied in this work (Ponce-López and Spataru, 2022; Dharrao et al., 2024; Hanny et al., 2024).

2.2 Domain-Specific Crisis Language Models

A second line of work focuses on adapting language models to the crisis domain. Lamsal et al. introduced CrisisTransformers, a family of pre-trained language models and sentence encoders trained on more than 15 billion tokens from tweets across over 30 crisis events. Their models improve classification, semantic search, and sentence repre-

sentation for crisis-related social media texts. Yin et al. proposed CrisisSense-LLM, an instruction-tuned large language model for multi-label classification of disaster-related tweets, covering aspects such as event type, informativeness, and humanitarian relevance. These works show that domain adaptation and instruction tuning are useful for crisis-related NLP. However, they mainly produce labels or representations for situational awareness. They do not generate official-style dissemination messages, and they do not use an official alert archive such as IPAWS as a stylistic and operational supervision source. In contrast, our work uses citizen disaster messages to learn crisis-state representations and uses IPAWS alerts to learn official dissemination style, control attributes, and public-warning structure (Lamsal et al., 2024; Yin et al., 2024).

2.3 LLMs for Disaster Monitoring and Automated Reporting

Recent studies have started using large language models for disaster monitoring and social media analysis. Cantini et al. used prompt-based LLMs to analyze user-generated content and identify citizen-reported disaster issues such as damaged buildings, broken gas pipelines, and flooding. Linardos et al. combined LLMs and clustering to structure and label disaster-related social media content at scale. Lei et al. reviewed LLM applications in disaster management and showed that LLMs are increasingly used for disaster identification, situation assessment, response support, and information extraction. These works are closely related because they move beyond traditional classifiers and use LLMs for richer crisis understanding. However, they still focus mostly on structuring, labeling, or summarizing crisis information. They do not explicitly train models to transform noisy citizen messages into official dissemination messages with required alert fields such as urgency, severity, certainty, headline, and protective guidance. Our work fills this gap by treating the task as controlled generation rather than only disaster monitoring or automated reporting (Cantini et al., 2025; Linardos et al., 2025; Lei et al., 2025).

2.4 Urgency, Context, and Multimodal Crisis Signals

Other recent studies have focused on extracting urgency, context, and multimodal evidence from crisis data. Meunier et al. proposed CrisisTS, a multimodal and multilingual dataset for urgency classification that links crisis-related social media texts with geocoded meteorological time series. This is important because emergency communication requires more than topic detection; it also requires understanding urgency and operational relevance. Multimodal crisis studies such as Teng et al. also show that social media images and text can provide complementary signals for informativeness, humanitarian category detection, and damage severity assessment. These works contribute useful ideas for crisis-state modeling, especially the importance of urgency, location, damage, and external context. However, they do not solve the dissemination problem. They classify urgency or informativeness but do not generate public-ready alert messages. Our work builds on this direction by using extracted crisis context as a conditioning layer for official-style generation (Meunier et al., 2025; Teng et al., 2025).

2.5 Warning Message Generation and Public Alert Systems

The closest line of work to the present study is warning message generation. Gonella et al. introduced CrisiText, a large-scale dataset of warning messages for LLM training across multiple crisis scenarios. This work is important because it directly recognizes that NLP research on crisis communication has been dominated by classification tasks and that warning-message generation remains underexplored. However, CrisiText focuses on warning-message generation from crises and expert-written guidelines, whereas our study focuses on a different and more difficult weak-supervision setting: transforming noisy citizen-generated messages into official-style dissemination messages. In the public warning literature, Sadiq et al. reviewed public alert and warning systems and emphasized that such systems are designed to detect hazards, alert at-risk populations, and support protective decision-making. FEMA also describes IPAWS as the U.S. national system for authenticated emergency and life-saving information. These works motivate why generated messages must be structured, actionable, and operationally constrained. Our work connects these two areas by using the IPAWS archive as a source of official dissemination style and metadata while using citizen messages as noisy crisis evidence (Gonella et al., 2026; Sadiq et al., 2023).

Overall, existing research has made strong progress in disaster tweet classification, crisis repre-

sentation learning, urgency detection, multimodal crisis understanding, and LLM-based disaster monitoring. However, the main limitation is that these approaches usually stop at detecting, classifying, or structuring citizen-generated crisis information. They do not address the operational communication gap between citizen chaos and official dissemination. The current study compensates for this limitation in three ways. First, it constructs a weakly aligned citizen-to-alert dataset by combining Kaggle disaster messages with IPAWS official alerts. Second, it introduces structured context-control tokens to represent disaster type, emotion, intent, official style, factual constraints, and actionability. Third, it fine-tunes open-source LLMs to generate official-style dissemination messages with controlled fields such as event, urgency, severity, headline, and description. This shifts disaster NLP from situational information extraction toward operational emergency communication generation.

3 Methodology

Figure 2 represents the dataset transformation process, and Figure 3 illustrates the training phase, inference phase, and evaluation criteria followed in this study.

3.1 Dataset Construction

The pipeline transforms noisy citizen-generated disaster communication into structured crisis-state representations and subsequently performs weakly supervised semantic alignment with official IPAWS dissemination alerts to construct operational citizen-to-alert training pairs. The procedure to extract citizen context from the disaster messages and official IPAWS alerts processing is given in Appendix A.1.

3.1.1 Citizen Disaster Messages

Citizen-generated disaster communication data were constructed using the Disaster Response Messages dataset [1], containing 26,248 disaster-related social media and public communication records. Each instance consists of a message identifier, original message text, translated content, and communication metadata.

3.1.2 Official IPAWS Alerts

Official emergency dissemination alerts were constructed using the FEMA Integrated Public Alert and Warning System (IPAWS) Archived Alerts dataset [2], which contains approximately 12 GB of Common Alerting Protocol (CAP) 1.2 emergency alerts encoded in JSON and embedded XML format. To efficiently process the large-scale archive, a streaming-based reduction pipeline was implemented using the `ijson` framework, enabling incremental parsing without loading the entire dataset into memory.

3.1.3 Structured Context Extraction

To transform the aligned citizen-IPAWS data into a trainable instruction-format corpus, each citizen message was first converted into a structured input context containing the standardized disaster type, active sub-categories, extracted keywords, emotion label, emotion score, and original citizen signal. The official alert corpus was represented using cleaned IPAWS-style fields, including event, urgency, severity, headline, and description. For semantic matching, only the citizen message text was encoded on the citizen side, while the IPAWS headline and description were encoded on the official-alert side. Both representations were embedded using the SentenceTransformer model `all-MiniLM-L6-v2`, and cosine similarity was used to identify the most relevant official alert within the same disaster category.

Let the structured citizen context be denoted as:

$$x_i = \{d_i, s_i, k_i, e_i, q_i, t_i\} \tag{1}$$

where d_i is the disaster type, s_i is the active sub-category set, k_i represents extracted keywords, e_i is the emotion label, q_i is the emotion score, and t_i is the citizen signal. For each citizen message, the corresponding candidate IPAWS alerts were restricted to the same disaster category d_i. The matching representation is defined as:

$$\mathbf{u}_i = f_\theta(t_i), \qquad \mathbf{v}_j = f_\theta(h_j \oplus desc_j) \tag{2}$$

where h_j and $desc_j$ denote the IPAWS headline and description, $\oplus$ denotes textual concatenation, and f_θ is the sentence embedding model. The best official alert was selected using:

$$a_i^* = \arg \max_{a_j \in D_{IPAWS}^{d_i}} \frac{\mathbf{u}_i \cdot \mathbf{v}_j}{\|\mathbf{u}_i\| \|\mathbf{v}_j\|} \tag{3}$$

[1] https://www.kaggle.com/datasets/sidharth178/disaster-response-messages

[2] https://www.fema.gov/openfema-data-page/ipaws-archived-alerts-v1

Only pairs satisfying the semantic relevance threshold were retained:

$$Sim(c_i, a_i^*) \geq 0.35 \quad (4)$$

The target output y_i was then generated from the matched IPAWS alert using a cleaned official-alert template. For situational reports, the target preserved the IPAWS-style structure and added citizen-context notes when keywords or operational sub-categories indicated relevant concerns. For uncertainty-oriented citizen questions, a safer response template was used to avoid unsupported claims about event status. The final training corpus was saved in JSONL format as:

$$D_{train} = \{(I, x_i, y_i, \alpha_i, e_i^{IPAWS})\}_{i=1}^{N} \quad (5)$$

where I is the fixed instruction, x_i is the structured citizen context, y_i is the official dissemination target, α_i is the semantic match score, and e_i^{IPAWS} is the matched IPAWS event type. This process produced a supervised instruction-tuning dataset in which each training instance maps citizen-generated disaster context to an official, calm, structured, and actionable dissemination message.

The aligned instruction-format records were subsequently converted into a chat-style supervised fine-tuning format. Each example was reformulated as a three-turn conversation consisting of a fixed system instruction, a user message containing a structured citizen disaster context with control tokens, and an assistant response containing the official IPAWS-style dissemination output. Disaster type and emotion intensity were encoded using explicit control tokens such as `<DISASTER_STORM>` and `<EMOTION_MEDIUM>`, followed by style constraints including `<OFFICIAL_STYLE>`, `<NO_UNSUPPORTED_FACTS>`, and `<ACTIONABLE_MESSAGE>`. This design guides the model toward controlled, factual, and operationally safe emergency-message generation.

The chat-formatted corpus was then filtered using a quality-control script to remove examples with metadata contamination or incomplete outputs. Specifically, outputs containing unsupported dates, timestamps, years, time zones, coordinates, or latitude-longitude expressions were removed. In addition, examples were discarded if the generated response lacked a valid `[DESCRIPTION]` field or if the description contained fewer than ten words. This filtering step reduced the risk of training the model to reproduce unsupported operational details and ensured that each target message contained a complete dissemination description.

The final cleaned training corpus is represented as:

$$D_{chat} = \{(m_i^{sys}, m_i^{user}, m_i^{asst})\}_{i=1}^{3560} \quad (6)$$

Where m_i^{sys} is the fixed emergency-communication system prompt that m_i^{user} contains the controlled structured citizen context, and m_i^{asst} contains the cleaned official dissemination target. The resulting dataset contains 3,560 high-quality chat-style examples for supervised fine-tuning.

3.2 Fine-tuning of a Pre-trained Model based on Low-Rank Adaptation (LoRA)

To adapt the pre-trained large language model (LLM) for controlled emergency dissemination generation, this study employs Low-Rank Adaptation (LoRA), a parameter-efficient fine-tuning strategy proposed by Hu et al. (Hu et al., 2021). Instead of updating the entire parameter space of the pre-trained transformer model, LoRA introduces trainable low-rank decomposition matrices into selected transformer layers while keeping the original model weights frozen. This significantly reduces computational cost and memory requirements while maintaining competitive downstream performance.

Let the original weight matrix of a transformer layer be represented as:

$$W_0 \in \mathbb{R}^{d \times k} \quad (7)$$

where d and k denote the input and output dimensionalities of the layer, respectively. During conventional full fine-tuning, all parameters in W_0 are updated. In contrast, LoRA assumes that the task-specific parameter update can be approximated using a low-rank decomposition:

$$\Delta W = BA \quad (8)$$

where:

$$B \in \mathbb{R}^{d \times r}, \quad A \in \mathbb{R}^{r \times k}, \quad r \ll \min(d, k) \quad (9)$$

and r represents the low-rank adaptation dimension. The adapted weight matrix becomes:

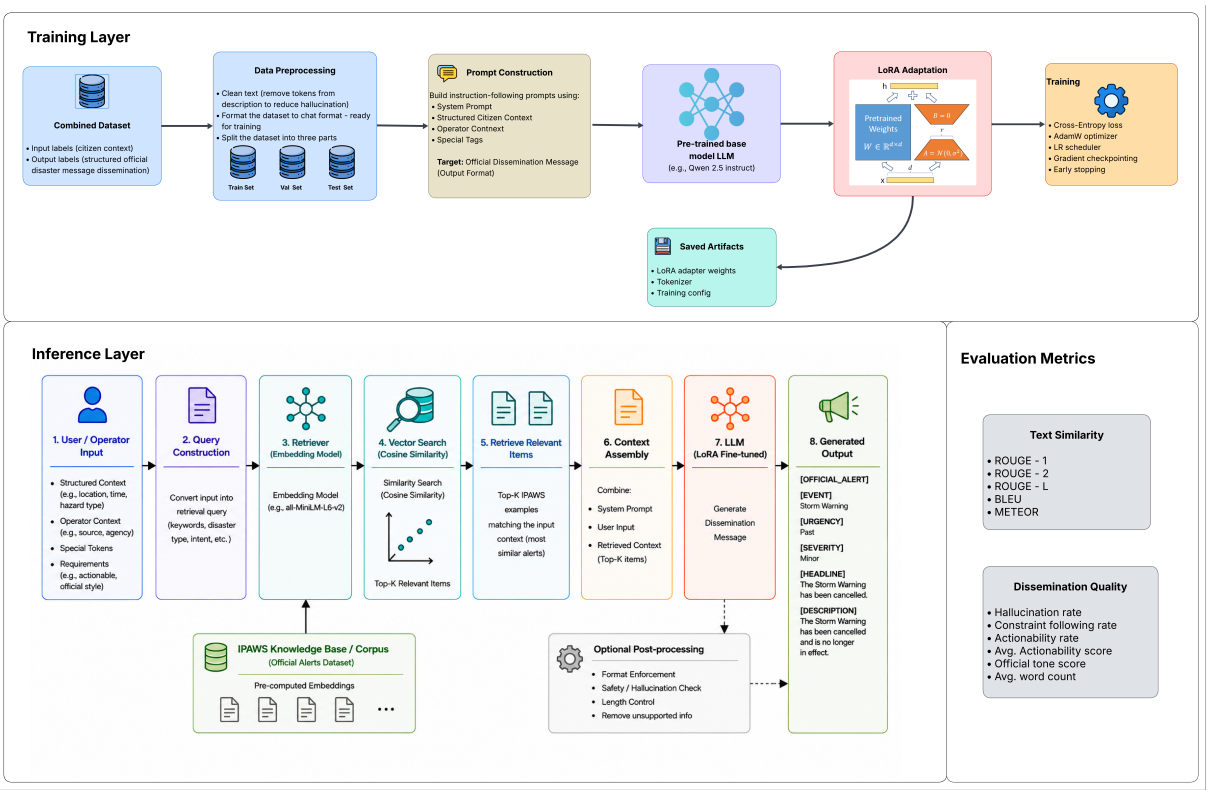

Figure 3: Overview of the proposed retrieval-augmented LoRA fine-tuned framework for structured emergency dissemination message generation and evaluation.

$$W = W_0 + \Delta W \tag{10}$$

Thus, the forward propagation of a transformer layer can be expressed as:

$$h = W_0 x + BAx \tag{11}$$

where x denotes the input hidden representation and h represents the transformed output embedding.

In the proposed framework, LoRA adapters are injected into the attention projection layers of the pre-trained instruction-tuned LLM. During training, the base model parameters remain frozen while only the low-rank matrices (A, B) are optimized. This substantially reduces the number of trainable parameters and enables efficient fine-tuning on disaster dissemination data with limited computational resources.

The training process utilizes structured dissemination prompts containing disaster-specific control tokens, structured incident context, and operator-provided grounding information. The objective is to adapt the model toward generating standardized official dissemination messages while preserving the general linguistic and reasoning capabilities of the original foundation model.

Formally, given an input prompt x_i and target dissemination message y_i, the model parameters are optimized using the autoregressive language modeling objective:

$$\mathcal{L}_{LM} = -\sum_{t=1}^{T} \log P_\theta(y_t \mid y_{<t}, x_i) \tag{12}$$

where y_t denotes the target token at timestep t, and θ represents the trainable LoRA parameters.

The use of LoRA is particularly suitable for disaster dissemination generation due to three reasons. First, it enables parameter-efficient adaptation of large instruction-tuned models without requiring full retraining. Second, it preserves the general language understanding capability of the base model while specializing the model toward emergency communication tasks. Third, the lightweight adapter-based design facilitates rapid experimentation and deployment across different dissemination scenarios and disaster categories.

Additionally, the LoRA-based fine-tuning strategy is combined with retrieval-augmented inference, where semantically similar official alerts are

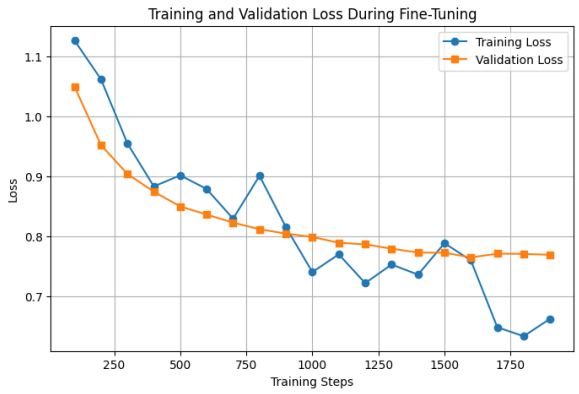

Figure 4: Training and validation loss curve for Qwen 2.5

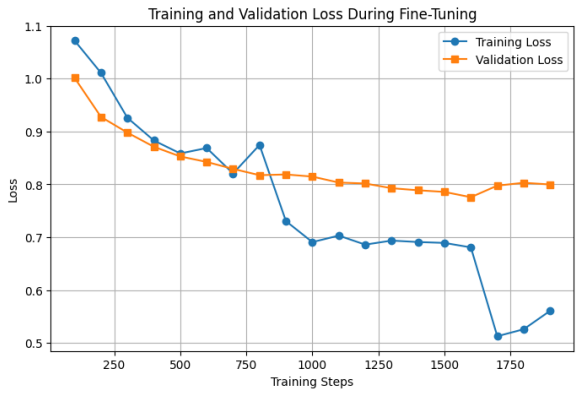

Figure 5: Training and validation loss curve for Mistral 7B

retrieved during generation. This hybrid design improves factual grounding, dissemination consistency, and structural controllability of generated emergency alerts. The inference time-controlled dissemination approach is explained in Appendix A.3.

4 Baselines and Evaluation

To evaluate the proposed crisis-aware controllable dissemination framework, experiments were conducted using Mistral 7B and Qwen2.5 under four configurations: base model, retrieval-augmented base model, LoRA fine-tuned model, and fine-tuned model with retrieval augmentation. Figures 4 and 5 show stable convergence during training, with validation losses stabilizing after approximately 1500 steps, indicating effective learning without severe overfitting. Table 1 presents the text generation evaluation results using ROUGE, METEOR, and BLEU metrics. Fine-tuned models substantially outperformed their corresponding base models, with the fine-tuned Qwen2.5 configuration achieving the best overall performance, obtaining ROUGE-1, ROUGE-2, and ROUGE-L scores of 0.5190, 0.4066, and 0.4798, respectively, together with a METEOR score of 0.4338 and BLEU score of 35.41. Similar improvements were observed for Mistral 7B, demonstrating the effectiveness of the proposed weakly supervised context-control framework for structured disaster dissemination generation. Retrieval augmentation additionally improved dissemination grounding and structural consistency, although slight reductions in BLEU and METEOR scores were observed in some cases due to stylistic variations introduced by retrieved dissemination examples.

Table 2 presents the safety, controllability, and dissemination-quality evaluation results. Fine-tuned models substantially reduced hallucination rates while improving constraint-following behavior and official dissemination tone compared to the corresponding base models. In particular, the retrieval-augmented fine-tuned Mistral 7B configuration achieved a hallucination rate of 0.0 and a perfect constraint-following score of 1.0, demonstrating strong dissemination controllability and factual consistency. Overall, the results indicate that the proposed citizen-to-alert generation framework successfully bridges noisy citizen-generated disaster discourse and structured IPAWS-style dissemination language despite the absence of directly aligned citizen-to-alert training pairs.

5 Discussions

The findings of this study demonstrate that controllable generation combined with structured contextual conditioning can substantially improve the quality, consistency, and operational reliability of disaster dissemination messages generated from noisy citizen reports. Existing disaster NLP research has primarily focused on disaster detection, humanitarian-category classification, urgency prediction, sentiment analysis, and situational awareness extraction (Yin et al., 2024; Lamsal et al., 2024; Dharrao et al., 2024), while recent work, such as CrisiText (Gonella et al., 2026) explored warning-message generation using structured crisis descriptions rather than noisy citizen discourse. In contrast, the proposed framework addresses a more challenging weak-supervision setting in which citizen reports are transformed into structured crisis-state representations for operational dissemination generation. The experimental re-

Table 1: Performance comparison of fine-tuned models on disaster dissemination generation.

Model	R-1	R-2	R-L	METEOR	BLEU
Mistral 7B Base Model	0.2234	0.0577	0.1736	0.1849	8.0960
Mistral 7B Base Model + Retrieval	0.2252	0.0651	0.1789	0.1783	7.8872
Mistral 7B Fine-tuned	0.5025	0.3583	0.4557	0.4036	26.8135
Mistral 7B Fine-tuned + Retrieval	0.4965	0.3730	0.4675	0.3902	24.9474
Qwen2.5 Base Model	0.2238	0.0524	0.1717	0.1930	7.9095
Qwen2.5 Base Model + Retrieval	0.2297	0.0607	0.1766	0.2043	8.3020
Qwen2.5 Fine-tuned	0.5190	0.4066	0.4798	0.4338	35.41
Qwen2.5 Fine-tuned + Retrieval	0.5124	0.4128	0.4834	0.4072	31.17

Table 2: Safety, controllability, and dissemination-quality evaluation of fine-tuned models.

Model	Hallucination Rate	Constraint Following Rate	Actionability Rate	Official Tone Score	Avg. Word Count
Mistral 7B Base Model	0.08	0.92	0.79	0.761	87.79
Mistral 7B Base Model + Retrieval	0.03	0.97	0.67	0.80	76.58
Mistral 7B Fine-tuned	0.03	0.97	0.02	0.8283	64.51
Mistral 7B Fine-tuned + Retrieval	0.0	1.0	0.02	0.875	56.44
Qwen2.5 Base Model	0.09	0.91	0.78	0.857	78.29
Qwen2.5 Base Model + Retrieval	0.05	0.95	0.87	0.917	70.62
Qwen2.5 Fine-tuned	0.03	0.97	0.02	0.8383	69.90
Qwen2.5 Fine-tuned + Retrieval	0.03	0.97	0.11	0.9133	57.64

sults further show that structured contextual signals, dissemination-specific control tokens, and retrieval augmentation significantly improve dissemination controllability and factual consistency. Prior controllable generation studies demonstrated that control codes and prompt-conditioning strategies can guide stylistic and semantic generation behavior (Keskar et al., 2019; Zhang et al., 2022), but such methods have rarely been explored for operational emergency communication. The proposed framework extends these ideas by incorporating hazard type, urgency, severity, emotional intensity, and official-style constraints into the generation process. Hallucination analysis revealed that base language models frequently generated unsupported operational details, consistent with concerns raised in recent studies on generative AI for emergency communication (Xu et al., 2025; Faiaz et al., 2024). The proposed context-control framework substantially reduced hallucination rates while improving constraint-following behavior, while retrieval augmentation improved dissemination tone, structural consistency, and protective-action alignment through semantically related IPAWS examples, supporting prior findings on message clarity and consistency in public-warning systems (Sadiq et al., 2023; Luhtakallio et al., 2023).

6 Conclusion

Although the proposed framework demonstrates promising results for controllable disaster dissemination generation, several limitations remain. The dataset construction process relies on weakly supervised semantic alignment between citizen-generated disaster messages and official IPAWS alerts, which may introduce imperfect citizen–official dissemination mappings due to the absence of manually annotated citizen-to-alert datasets. In addition, the current framework focuses primarily on textual disaster communication and does not incorporate multimodal crisis evidence such as images, videos, geospatial streams, or sensor observations that could further improve dissemination grounding and contextual understanding. While the proposed context-control framework substantially reduces hallucination and improves dissemination controllability, generated outputs should still be considered decision-support drafts rather than fully autonomous operational alerts in highly ambiguous crisis scenarios. Furthermore, although automatic and rule-based evaluation metrics provide useful quantitative insights, they do not fully capture real-world dissemination effectiveness or operational usability. Broader human evaluation involving emergency managers and public-warning experts, together with multilingual and culturally adaptive dissemination modeling, remains an important direction for future work.

References

Firoj Alam and 1 others. 2021. Crisisbench: Benchmarking crisis-related social media datasets for humanitarian nlp. In *Proceedings of ACL Findings*.

Satanjeev Banerjee and Alon Lavie. 2005. Meteor: An automatic metric for mt evaluation with improved correlation with human judgments. In *Proceedings of the ACL Workshop on Intrinsic and Extrinsic Evaluation Measures for Machine Translation and/or Summarization*, pages 65–72.

Ricardo Campos, Vítor Mangaravite, Arian Pasquali, Alípio Mário Jorge, Célia Nunes, and Adam Jatowt. 2018. Yake! collection-independent automatic keyword extractor. In *European conference on information retrieval*, pages 806–810. Springer.

Riccardo Cantini and 1 others. 2025. Harnessing prompt-based large language models for disaster management and social media analysis. *International Journal of Disaster Risk Reduction*.

Deepak Dharrao, M. R. Aadithyanarayanan, Rewaa Mital, and 1 others. 2024. An efficient method for disaster tweets classification using gradient-based optimized convolutional neural networks with bert embeddings. *MethodsX*, 13:102843.

Rahman Faiaz and 1 others. 2024. Ai-driven emergency communication systems: Opportunities and risks of generative language models. *IEEE Access*, 12:88214–88236.

Daniel Gillick, Alessandro Presta, and Gaurav Singh Tomar. 2018. End-to-end retrieval in continuous space. *arXiv preprint arXiv:1811.08008*.

Giulia Gonella and 1 others. 2026. Crisitext: A dataset of warning messages for llm training in crisis situations. *Preprint*, arXiv:2510.09243.

David Hanny, Sebastian Schmidt, and Bernd Resch. 2024. Active learning for identifying disaster-related tweets: A comparison with keyword filtering and generic fine-tuning. In *Intelligent Systems Conference*, pages 126–142. Springer.

Jochen Hartmann. 2023. Emotion english distilroberta-base; 2022. *URL: https://huggingface. co/j-hartmann/emotion-english-distilroberta-base [accessed 2024-06-22]*.

Edward J. Hu, Yelong Shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Lu Wang, and Weizhu Chen. 2021. Lora: Low-rank adaptation of large language models. *arXiv preprint arXiv:2106.09685*.

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.

Nitish Shirish Keskar, Bryan McCann, Lav R. Varshney, Caiming Xiong, and Richard Socher. 2019. Ctrl: A conditional transformer language model for controllable generation. *Preprint*, arXiv:1909.05858.

Stjepan Kriz and Vlatka Hribljan. 2018. Social media in crisis management: An evaluation and analysis of crisis informatics research. *International Journal of Disaster Risk Reduction*, 27:644–656.

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

Yu Lei and 1 others. 2025. Harnessing large language models for disaster management: Opportunities and challenges. *International Journal of Disaster Risk Reduction*.

Chin-Yew Lin. 2004. Rouge: A package for automatic evaluation of summaries. In *Text Summarization Branches Out*, pages 74–81.

Vasileios Linardos and 1 others. 2025. Utilizing llms and ml algorithms in disaster-related social media analytics. *AI*, 6(3):33.

Eeva Luhtakallio and 1 others. 2023. Effectiveness of sms-based emergency warning messages in public crisis communication. *Safety Science*, 163:106117.

Jean-Baptiste Meunier and 1 others. 2025. Crisists: A multimodal and multilingual dataset for crisis severity and urgency understanding. In *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics*.

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.

Long Ouyang, Jeffrey Wu, Xu Jiang, Diogo Almeida, Carroll Wainwright, Pamela Mishkin, Chong Zhang, Sandhini Agarwal, Katarina Slama, Alex Ray, and 1 others. 2022. Training language models to follow instructions with human feedback. In *Advances in Neural Information Processing Systems*, volume 35, pages 27730–27744.

Leysia Palen. 2008. Online social media in crisis events. *Educause quarterly*, 31(3):76–78.

Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu. 2002. Bleu: a method for automatic evaluation of machine translation. In *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics*, pages 311–318.

Víctor Ponce-López and Catalina Spataru. 2022. Social media data analysis framework for disaster response. *Discover Artificial Intelligence*, 2(10).

Colin Raffel, Noam Shazeer, Adam Roberts, Katherine Lee, Sharan Narang, Michael Matena, Yanqi Zhou, Wei Li, and Peter J. Liu. 2020. Exploring the limits of transfer learning with a unified text-to-text

transformer. *Journal of Machine Learning Research*, 21(140):1–67.

Nils Reimers and Iryna Gurevych. 2019. Sentence-bert: Sentence embeddings using siamese bert-networks. In *Proceedings of the 2019 conference on empirical methods in natural language processing and the 9th international joint conference on natural language processing (EMNLP-IJCNLP)*, pages 3982–3992.

Abdul-Akeem Sadiq and 1 others. 2023. Public alert and warning systems: A review of social and technical challenges. *International Journal of Disaster Risk Reduction*, 87:103567.

Victor Sanh, Albert Webson, Colin Raffel, Stephen H. Bach, Lintang Sutawika, Zaid Alyafeai, Antoine Chaffin, Arnaud Stiegler, Teven Le Scao, Arun Raja, Manan Dey, M Saiful Bari, Canwen Xu, Urmish Thakker, Shanya Sharma, and 1 others. 2022. Multitask prompted training enables zero-shot task generalization. In *International Conference on Learning Representations*.

Fuzel Shaik, Getnet Demil, and Mourad Oussalah. 2025. Ai-based approach in early warning systems: Focus on emergency communication ecosystem and citizen participation in nordic countries. *arXiv preprint arXiv:2506.18926*.

Fuzel Ahamed Shaik and Mourad Oussalah. 2024. On mining mobile emergency communication applications in nordic countries. *International Journal of Disaster Risk Reduction*, 108:104566.

Xiaoyu Teng and 1 others. 2025. Multimodal informative classification of disaster-related social media posts. *Information Processing & Management*, 62(1):104021.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. *Advances in neural information processing systems*, 30.

Sarah Vieweg, Amanda L Hughes, Kate Starbird, and Leysia Palen. 2010. Microblogging during two natural hazards events: what twitter may contribute to situational awareness. In *Proceedings of the SIGCHI conference on human factors in computing systems*, pages 1079–1088.

Fan Xu and 1 others. 2025. Large language model applications in disaster management. *International Journal of Disaster Risk Reduction*.

Kai Yin, Bo Li, 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*.

Han Zhang and 1 others. 2022. A survey of controllable text generation using transformer-based language models. *ACM Computing Surveys*, 55(12):1–38.

Yuanjun Zhang, Fuzel Ahamed Shaik, Suvojit Acharjee, Fahad Khalid, and Mourad Oussalah. 2026. Towards reliable multimodal disaster severity assessment through preference optimization and explainable vision-language reasoning. *Reliability Engineering & System Safety*, page 112674.

A Appendix

A.1 Dataset Construction

A.1.1 Citizen Context extracted from Disaster Messages

Let the raw citizen message dataset be represented as:

$$D_{raw} = \{m_i, t_i, g_i, c_i\}_{i=1}^{N} \tag{13}$$

where m_i denotes the message identifier, t_i the message text, g_i the communication genre, and c_i the associated categorical annotations.

The original annotations were represented as a sparse multi-label categorical structure:

$$c_i = \{k_1 - v_1, k_2 - v_2, \ldots, k_n - v_n\} \tag{14}$$

where k_n denotes the category label and $v_n \in \{0, 1\}$ represents binary activation for category k_n. The category sequence was transformed into a binary multi-hot encoding representation:

$$\mathbf{y}_i = [y_{i1}, y_{i2}, \ldots, y_{i35}] \tag{15}$$

where $y_{ij} = 1$ indicates the presence of category j in message i. Records containing anomalous category assignments (`related=2`) were removed to preserve binary label consistency. The `child_alone` category was additionally discarded due to zero discriminative variance. After preprocessing, the resulting citizen communication corpus contained 26,055 valid disaster-related instances.

To reduce semantic sparsity and improve contextual generalization, the original categories were hierarchically grouped into higher-level disaster-event representations. The disaster taxonomy was defined as:

$$\mathcal{D} = \{flood,\ storm,\ fire,\ earthquake,\ cold,\ weather_related,\ public_safety,\ infrastructure\} \tag{16}$$

where semantically related categories were merged using logical aggregation functions. For example:

$$weather_related = weather \vee other_weather \tag{17}$$

Similarly, message-level operational subcategories were constructed to capture communication semantics associated with citizen reports:

$$\mathcal{S} = \{basic_needs,\ rescue,\ aid,\ safety,\ infrastructure,\ direct_report,\ request_offer,\ medical\} \tag{18}$$

This hierarchical transformation substantially reduced categorical sparsity while preserving operational disaster semantics relevant to emergency communication.

To enrich contextual representation, emotion-aware semantic analysis was performed using the pretrained transformer model *Emotion English DistilRoBERTa-base* (Hartmann, 2023). Given a disaster message t_i, the transformer predicts an emotion probability distribution:

$$P(e_k|t_i) = \text{softmax}(z_k) \tag{19}$$

where z_k represents the transformer output logits for emotion class e_k. The dominant emotional state associated with message i is computed as:

$$\hat{e}_i = \arg\max_k P(e_k|t_i) \tag{20}$$

while the emotional confidence score is defined as:

$$s_i = \max_k P(e_k|t_i) \tag{21}$$

where $s_i \in [0, 1]$ represents the confidence associated with the predicted emotional state.

Semantic keyword extraction was performed using YAKE (Yet Another Keyword Extractor) (Campos et al., 2018), an unsupervised statistical keyword extraction framework. For a candidate keyword w, the importance score is computed using local textual statistics:

$$K(w) = \prod_{f \in F} score_f(w) \tag{22}$$

where F denotes the set of local textual features including frequency, position, casing, contextual diversity, and sentence dispersion. Lower values of $K(w)$ correspond to higher semantic relevance. The top-k ranked keywords ($k = 5$) were retained for each disaster message.

Named Entity Recognition (NER) was additionally performed using spaCy-based linguistic parsing to extract contextual entities including locations, organizations, geopolitical regions, and persons relevant to emergency communication scenarios.

The enriched citizen-context dataset was generated through relational integration of the extracted contextual components:

$$D_{citizen} = D_{message} \bowtie D_{emotion} \bowtie D_{keywords} \bowtie D_{taxonomy} \tag{23}$$

where $\bowtie$ denotes relational join operations using the message identifier as the primary key.

Additional contextual indicators were generated to quantify operational disaster complexity. Event-category density was defined as:

$$C_i^{event} = \sum_{j=1}^{|\mathcal{D}|} d_{ij} \tag{24}$$

while communication-subcategory density was computed as:

$$C_i^{sub} = \sum_{j=1}^{|\mathcal{S}|} s_{ij} \tag{25}$$

A critical-context indicator was further defined as:

$$critical_i = \begin{cases} 1, & s_i > 0.8 \wedge C_i^{event} > 1 \\ & \wedge\ entities_i \neq \emptyset \\ 0, & \text{otherwise} \end{cases} \tag{26}$$

The resulting citizen-context dataset provides a multi-dimensional representation of disaster communication by jointly modeling disaster type, operational sub-category, emotional state, semantic keywords, and contextual entities for downstream dissemination modeling and emergency communication analysis.

Table 3 presents representative instances extracted directly from the enriched citizen-context disaster communication dataset.

Table 3: Representative examples from the citizen-context disaster communication dataset

Citizen Message	Emotion	Extracted Keywords	Disaster Type
"Weather update - a cold front from Cuba that could pass over Haiti"	neutral	Weather update; pass over Haiti; front from Cuba	weather_related
"Is the Hurricane over or is it not over"	neutral	Hurricane	storm
"UN reports Leogane 80-90 destroyed. Only Hospital St. Croix functioning."	sadness	Leogane destroyed; Hospital St. Croix; UN reports	infrastructure
"says: west side of Haiti, rest of the country today and tonight"	neutral	west side; Haiti; today and tonight	weather_related
"Information about the National Palace"	neutral	National Palace; Information	infrastructure

A.1.2 Official IPAWS Alerts

Each raw alert instance is represented as:

$$o_i = \{id_i, metadata_i, xml_i\} \tag{27}$$

where xml_i denotes the embedded CAP dissemination message. The XML alerts were parsed under the CAP namespace:

$$\mathcal{N} = \texttt{urn:oasis:names:tc:emergency:cap:1.2} \tag{28}$$

A structured operational representation was extracted from each alert:

$$\begin{aligned} A_i = \{&event_i,\ urgency_i,\ severity_i,\\ &certainty_i,\ headline_i,\\ &description_i,\ instruction_i,\\ &area_i\} \end{aligned} \tag{29}$$

where the extracted fields represent the disaster event type, dissemination urgency, impact severity, event certainty, official headline, dissemination description, protective instructions, and affected geographic region.

Text normalization was subsequently applied to remove formatting artifacts and redundant whitespace:

$$T_i^{clean} = \psi(T_i) \tag{30}$$

where $\psi(\cdot)$ denotes the dissemination-text cleaning function.

To align official alerts with citizen-generated disaster communication, heterogeneous IPAWS event types were mapped into a shared disaster taxonomy using a manually curated mapping dictionary (`ipaws_to_disaster_type.json`). The standardized disaster space is defined as:

$$\begin{aligned} \mathcal{D} = \{&storm,\ flood,\ fire,\ earthquake,\\ &cold,\ weather_related,\\ &public_safety,\ infrastructure\} \end{aligned} \tag{31}$$

A deterministic mapping function was applied:

$$f : event_i \rightarrow d_k \tag{32}$$

where $d_k \in \mathcal{D}$ denotes the standardized disaster category. For example, operational events such as *Severe Thunderstorm Warning*, *Hurricane Warning*, and *Tornado Warning* were grouped under the *storm* category.

Administrative and testing alerts were removed using:

$$A_i^{valid} = \begin{cases} 1, & d_k \neq test \wedge d_k \neq administrative \\ 0, & \text{otherwise} \end{cases} \tag{33}$$

The final structured dissemination dataset is represented as:

$$\begin{aligned} D_{IPAWS} = \{&event_i, urgency_i, severity_i,\\ &certainty_i, headline_i, description_i,\\ &instruction_i, area_i, d_k\} \end{aligned} \tag{34}$$

The resulting pipeline transformed the original 12 GB IPAWS archive into a lightweight structured dissemination corpus while preserving operational emergency semantics, protective guidance, and geographic dissemination context required for downstream disaster communication modeling.

A.1.3 Structured Context Extraction

To construct a controlled disaster-dissemination training corpus, citizen-generated disaster reports were semantically aligned with operational IPAWS emergency alerts using a structured contextual-matching framework. Let the citizen-context dataset be represented as:

$$D_C = \{c_i\}_{i=1}^{N_c} \tag{35}$$

where each citizen-context instance is defined as:

$$c_i = \{t_i, d_i, s_i, k_i, e_i\} \tag{36}$$

Here, t_i denotes the citizen disaster message, d_i the disaster category, s_i the operational subcategory representation, k_i the extracted semantic keywords, and e_i the emotion label associated with the message.

Similarly, the structured IPAWS dissemination dataset is represented as:

$$D_A = \{a_j\}_{j=1}^{N_a} \tag{37}$$

where each official dissemination alert is represented as:

$$\begin{aligned} a_j = \{&event_j,\ urgency_j,\ severity_j, \\ &certainty_j,\ headline_j, \\ &description_j\} \end{aligned} \tag{38}$$

A lightweight semantic cleaning function was applied to the IPAWS dissemination messages to prevent the model from overfitting to highly specific operational details:

$$a_j^{clean} = \psi(a_j) \tag{39}$$

where $\psi(\cdot)$ removes explicit timestamps, exact dates, long geographic enumerations, and highly specific numerical measurements while preserving operational dissemination semantics, emergency hierarchy, and protective guidance structure.

To establish semantic correspondence between citizen communication and official dissemination messages, both datasets were embedded into a shared latent semantic space using the pretrained SentenceTransformer encoder `all-MiniLM-L6-v2`. The embedding representations are defined as:

$$\mathbf{u}_i = f_\theta(t_i) \tag{40}$$

$$\mathbf{v}_j = f_\theta(a_j^{clean}) \tag{41}$$

where f_θ denotes the pretrained transformer encoder and $\mathbf{u}_i, \mathbf{v}_j \in \mathbb{R}^{384}$ represent dense semantic vector representations.

Semantic similarity between citizen messages and IPAWS alerts was computed using cosine similarity:

$$Sim(c_i, a_j) = \frac{\mathbf{u}_i \cdot \mathbf{v}_j}{\|\mathbf{u}_i\|\|\mathbf{v}_j\|} \tag{42}$$

For each citizen disaster message, the most semantically relevant IPAWS dissemination alert was selected as:

$$a_i^* = \arg\max_{a_j \in D_A} Sim(c_i, a_j) \tag{43}$$

To ensure operational relevance, only alignments satisfying the following:

$$Sim(c_i, a_i^*) \geq \tau \tag{44}$$

Only semantically aligned citizen-official pairs with cosine similarity scores greater than the semantic filtering threshold were retained, where the threshold was empirically set to $\tau = 0.35$ following preliminary qualitative inspection to balance retrieval relevance and dataset coverage, similar to prior embedding-based weak supervision and semantic retrieval approaches (Reimers and Gurevych, 2019; Gillick et al., 2018). The aligned citizen and official dissemination contexts were integrated into a structured dissemination representation:

$$x_i = \{d_i, s_i, k_i, e_i, t_i, a_i^*\} \tag{45}$$

The final dissemination target was generated through contextual augmentation of the matched IPAWS dissemination message:

$$y_i = \phi(a_i^*, k_i, s_i) \tag{46}$$

where $\phi(\cdot)$ denotes the dissemination augmentation function that incorporates citizen-level contextual indicators into the official dissemination structure while preserving operational realism.

The final dissemination training dataset is represented as:

$$D_{train} = \{x_i, y_i\}_{i=1}^{N} \tag{47}$$

After semantic alignment and filtering, the final dissemination corpus contained $N = 6914$ training examples. The semantic quality of the alignment process was evaluated using the average cosine similarity score:

$$\bar{S} = \frac{1}{N} \sum_{i=1}^{N} Sim(c_i, a_i^*) \tag{48}$$

resulting in an average semantic alignment score of $\bar{S} = 0.46$

with observed similarity values distributed within the interval:

$$0.35 \leq Sim(c_i, a_i^*) \leq 0.76 \quad (49)$$

The resulting structured alignment framework enables controlled integration of citizen-generated disaster communication with operational emergency dissemination patterns while preserving semantic consistency, contextual realism, and dissemination reliability for downstream disaster communication modeling. Figure 6 illustrates a representative instruction-tuning example from the weakly aligned dissemination corpus. The training instance combines a structured citizen-disaster context with semantically matched IPAWS dissemination metadata, enabling the model to learn the transformation from noisy citizen-generated crisis signals to structured official emergency communication outputs.

A.2 Context-Control Adapter Architecture

This study uses a context-control adapter architecture to condition the language model with explicit disaster, emotion, and style signals before generating official-style dissemination messages. Instead of relying only on raw citizen messages, the model receives structured control tokens that describe the crisis context and the expected communication behavior. This design follows the broader idea of controllable generation, where special control codes guide the output style, domain, and task behavior (Keskar et al., 2019). It also follows the text-to-text learning paradigm, where task-specific inputs are serialized into a unified textual format for generation (Raffel et al., 2020). In this study, the control tokens act as a lightweight adapter between noisy citizen-generated disaster discourse and formal IPAWS-style emergency communication.

A.2.1 Structured Context Representation

The structured context representation is created by extracting disaster-relevant signals from each citizen message or training example. The goal is to convert noisy and unstructured disaster messages into a compact crisis-state representation before generation. In the implementation, each example is checked for explicit metadata fields such as `disaster_type`, `disaster`, `event_type`, `category`, and `label`. If such fields are available, the disaster type is directly extracted and normalized. If these fields are absent, the function searches the user message or input text for disaster-specific control tokens such as `<DISASTER_FLOOD>`, `<DISASTER_STORM>`, `<DISASTER_FIRE>`, `<DISASTER_EARTHQUAKE>`, `<DISASTER_COLD>`, and `<DISASTER_WEATHER_RELATED>`. As a final fallback, the function searches for structured textual patterns such as "Disaster type: flood" and maps them into a normalized disaster label.

This layered extraction strategy makes the dataset construction process robust to different input formats, including chat-style `messages`, instruction-style `input`/`output` pairs, and examples that already contain structured labels. The final disaster categories used in the model are `fire`, `weather_related`, `cold`, `storm`, `flood`, and `earthquake`. These categories are mapped to numerical identifiers through a fixed disaster-type dictionary, which supports reproducible training and downstream analysis. This representation allows the model to learn from citizen-generated disaster signals while preserving a clear connection to operational disaster categories used in emergency communication.

A.2.2 Context-Control Tokens

The implementation defines three groups of context-control tokens and one group of style-control tokens. The disaster tokens represent the type of hazard in the citizen message. The emotion tokens represent the intensity of emotional or distress signals in the message. The style tokens guide the model toward official, factual, and actionable language for dissemination.

The disaster-control tokens are:

```
<DISASTER_FLOOD>
<DISASTER_STORM>
<DISASTER_FIRE>
<DISASTER_EARTHQUAKE>
<DISASTER_COLD>
<DISASTER_WEATHER_RELATED>
```

The emotion-control tokens are:

```
<EMOTION_LOW>
<EMOTION_MEDIUM>
<EMOTION_HIGH>
<EMOTION_CRITICAL>
```

The style-control tokens are:

```
<OFFICIAL_STYLE>
<NO_UNSUPPORTED_FACTS>
<ACTIONABLE_MESSAGE>
```

Together, these special tokens define the conditioning space of the proposed model. They are added to the tokenizer as additional special

Structured Dissemination Training Example

Instruction

`Generate an official emergency dissemination message from structured citizen disaster context.`

Structured Citizen Context

Disaster type	storm
Sub-category	Aid & Assistance; Meta
Keywords	Hurricane
Emotion label	neutral
Emotion score	0.539
Citizen signal	"Is the Hurricane over or is it not over"

Generated Official Dissemination Output

[OFFICIAL_ALERT]

Event: Storm Warning

Urgency: Past

Severity: Minor

Headline: The Storm Warning has been cancelled.

Description: The Storm Warning has been cancelled and is no longer in effect.

Semantic Match Metadata

Matched IPAWS event	Storm Warning
Semantic match score	0.517

Figure 6: Example instruction-tuning instance from the final dissemination training corpus. The model receives a structured citizen disaster context and generates an IPAWS-style official dissemination message aligned through semantic matching.

tokens before fine-tuning, allowing the model to treat them as explicit semantic markers rather than ordinary text fragments. This is important because the model must learn not only the content of a disaster message but also the expected output behavior. For example, `<OFFICIAL_STYLE>` encourages formal dissemination language, `<NO_UNSUPPORTED_FACTS>` discourages unsupported claims, and `<ACTIONABLE_MESSAGE>` encourages protective-action-oriented outputs. This design is consistent with prior work on prompt-based and instruction-based conditioning, where task and style markers are used to improve controllability and generalization (Sanh et al., 2022; Ouyang et al., 2022).

A.2.3 Controlled Context Conditioning

Controlled context conditioning is performed by prepending the selected disaster, emotion, and style tokens to the citizen message or structured crisis context. The resulting input sequence informs the model about both the observed crisis signal and the desired dissemination behavior. For example, a citizen message about flooded roads and rescue needs may be encoded with the following prefix:

```
<DISASTER_FLOOD>
<EMOTION_CRITICAL>
<OFFICIAL_STYLE>
<NO_UNSUPPORTED_FACTS>
<ACTIONABLE_MESSAGE>
```

This prefix is followed by the citizen-generated message or extracted crisis-state context. During fine-tuning, the model learns to generate an official-style dissemination message conditioned on these control signals. The IPAWS official alert archive provides the target dissemination style, including structured public-warning language, urgency, severity, certainty, and protective-action patterns. Therefore, the control tokens function as an intermediate

bridge between the citizen communication domain and the official dissemination domain.

This conditioning design supports the central weak-supervision setting of the study. Since direct citizen-to-alert pairs are not naturally available, the model uses extracted disaster context and IPAWS-style targets to learn a weakly aligned transformation from citizen chaos to official-ready communication. The special tokens reduce ambiguity in the input, improve consistency across training examples, and make the generation process more controllable. In this way, the model is encouraged to generate messages that are factual, concise, actionable, and aligned with official dissemination conventions rather than producing generic summaries or unsupported emergency claims.

A.3 Inference-Time Controlled Dissemination

During inference, the proposed framework generates structured emergency dissemination messages by combining structured crisis-state representations, operator-provided information, and retrieved official IPAWS dissemination examples. The system first constructs a retrieval query using contextual attributes such as disaster type, urgency indicators, and operational metadata and retrieves semantically similar official alerts to provide dissemination grounding and stylistic guidance. The final prompt supplied to the LoRA fine-tuned language model integrates system-level dissemination instructions, structured user and operator context, and retrieved dissemination examples, enabling the generation of operationally grounded, controllable, and standardized emergency dissemination messages while reducing unsupported or hallucinated information.

A.3.1 Human-in-the-Loop Operational Grounding

The proposed framework incorporates a Human-in-the-Loop (HITL) operational grounding mechanism to improve dissemination reliability, factual consistency, and controllability during inference by integrating verified operator-provided information through a dedicated `[OPERATOR_CONTEXT]` section. This operator context includes validated attributes such as location, urgency, severity, evacuation guidance, and protective actions, enabling the language model to prioritize authoritative dissemination guidance over noisy citizen-generated signals. By separating citizen-derived situational information from verified operational instructions, the framework improves dissemination grounding, operational alignment, and controllability while allowing emergency personnel to guide urgency framing and protective recommendations without manually composing the final dissemination message. An example prompt structure used during inference is shown below:

```
[STRUCTURED_CONTEXT]
Disaster type: storm
Sub-category: Aid & Assistance
Keywords: hurricane; coastal area;
evacuation guidance
Emotion label: urgent
Citizen signal: Hurricane approaching
the coastal area
[/STRUCTURED_CONTEXT]

[OPERATOR_CONTEXT]
Location: Coastal area
Urgency: Immediate
Severity: Severe
Protective action: Follow evacuation
instructions
[/OPERATOR_CONTEXT]
```

A.3.2 Controlled Dissemination Generation

The final dissemination message is generated using the LoRA fine-tuned language model conditioned on structured crisis-state representations, context-control tokens, and retrieved official IPAWS dissemination examples. The framework enforces a controlled dissemination structure containing predefined fields such as event type, urgency, severity, headline, and description, while disaster-specific, emotion-control, and style-control tokens guide the model toward generating calm, actionable, and policy-aligned dissemination messages. During inference, the model receives structured disaster context together with constrained dissemination-style instructions designed to enforce factual consistency, official communication behavior, and operational reliability while discouraging unsupported claims, hallucinated locations, fabricated casualty numbers, and unverified operational details. Retrieval augmentation further improves dissemination grounding and structural consistency by conditioning generation on semantically similar official alerts, while optional post-processing and validation steps ensure dissemination compliance and reliability. This dual-conditioning strategy enables the framework to generate concise, operationally grounded, and structurally consistent IPAWS-style emergency dissemination messages. The system-level instruction used during inference is shown below:

```
SYSTEM_PROMPT = (
    "You are an emergency communication assistant. "
```

```
    "Generate official, calm, clear, actionable public
         dissemination "
    "messages. "
    "Use only the provided structured context and verified
         operator "
    "context. "
    "Do not invent unsupported locations, times, casualty
         numbers, "
    "measurements, agency names, or event status. "
    "Check the operator information when deciding the
         description "
    "for the message. "

    "The output MUST strictly follow this structure:\n\n"

    "[OFFICIAL_ALERT]\n\n"

    "[EVENT]\n"
    "<event type>\n\n"

    "[URGENCY]\n"
    "<Immediate | Expected | Future | Past | Unknown>\n\n"

    "[SEVERITY]\n"
    "<Extreme | Severe | Moderate | Minor | Unknown>\n\n"

    "[HEADLINE]\n"
    "<short official headline>\n\n"

    "[DESCRIPTION]\n"
    "<clear public dissemination message>\n\n"

    "Rules:\n"
    "- Keep the tone official and calm.\n"
    "- Use concise public-safety language.\n"
    "- Include protective guidance if provided.\n"
    "- If information is missing, use generic wording.\n"
    "- Do NOT generate unsupported facts.\n"
    "- Do NOT invent locations, times, measurements, "
    "zones, casualties, or agencies.\n"
    "- The DESCRIPTION should be short and actionable.\n\n"
)
```

A.4 Experimental Results

To evaluate the quality of the generated official-style disaster dissemination messages, this study employs multiple automatic text generation evaluation metrics widely used in natural language generation and summarization tasks. Since the proposed framework aims to transform noisy citizen-generated disaster reports into operationally reliable dissemination messages, the evaluation focuses on lexical overlap, semantic similarity, and contextual fluency between generated outputs and reference IPAWS-style dissemination messages. The evaluation metrics used in this study include ROUGE-1, ROUGE-2, ROUGE-L, BLEU, and METEOR. To evaluate the operational reliability and dissemination quality of the generated emergency messages, a rule-based evaluation framework was developed to measure hallucination risk, constraint adherence, actionability, and official communication tone. Let the generated dissemination prediction for the sample i be denoted by $\hat{y}_i$, and the corresponding structured citizen-context input by x_i.

Listing 1 presents an example training instance from the final instruction-tuning corpus. The model receives a structured citizen-disaster context together with dissemination control tokens and generates an official IPAWS-style emergency dissemination message, illustrating the transformation from noisy citizen communication to structured operational alert generation.

Listing 1: Example structured dissemination generation instance showing the transformation of citizen-generated disaster communication into an official IPAWS-style dissemination message.

```
[INPUT PROMPT STRUCTURE]

<DISASTER_WEATHER_RELATED>
<EMOTION_HIGH>
<OFFICIAL_STYLE>
<NO_UNSUPPORTED_FACTS>
<ACTIONABLE_MESSAGE>

Disaster type: weather_related
Sub-category: Aid & Assistance; Direct Report;
    Request/Offer; Meta
Keywords: Titus Magloire Ambroise; Magloire
    Ambroise street; Titus Magloire; Magloire
    Ambroise; Ambroise street
Emotion label: neutral
Emotion score: 0.6787301898002625
Citizen signal: We need help please at Titus
    Magloire Ambroise street we need every help
    group to contact the emergency opersonnele
    thank you

Generate an official, calm, clear, actionable
    emergency dissemination message.

[GENERATED OFFICIAL DISSEMINATION MESSAGE]

[OFFICIAL_ALERT]

[EVENT]
Special Weather Statement

[URGENCY]
Expected

[SEVERITY]
Moderate

[HEADLINE]
Special Weather Statement issued during the
    warning period by official weather
    authorities

[DESCRIPTION]
MOIST AND UNSTABLE ENVIRONMENT WILL FAVOR THE
    DEVELOPMENT OF NUMEROUS SHOWERS AND
    THUNDERSTORMS ACROSS MOST OF PUERTO RICO
    TODAY. SOILS ACROSS THE ISLAND REMAIN
    SATURATED AS SIGNIFICANT RAINFALL
    ACCUMULATIONS HAVE BEEN OBSERVED ACROSS THE
    ISLAND IN THE PAST FEW DAYS. ADDITIONAL
    HEAVY RAIN OVER THE ALREADY SATURATED SOILS
    WILL RESULT IN FLOODING AT RIVERS...STREAMS
    ...URBAN AND LOW-LYING AREAS..
```

A.4.1 Implementation and Experimental Details

The experiments were conducted using the instruction-tuned large language models `mistralai/Mistral-7B-Instruct-v0.3` and `Qwen/Qwen2.5-3B-Instruct`. The final dissemination training corpus consisted of weakly aligned citizen-to-alert dissemination examples constructed from semantically matched citizen-generated disaster reports and IPAWS dissemination archives. A stratified train-validation split of 90:10 was used to preserve disaster-category distribution across the dataset. For evaluation, a stratified sample of 100 dissemination instances was used across multiple disaster categories, including flood, storm, fire, earthquake, cold-weather, public-safety, and infrastructure-related events.

Parameter-efficient fine-tuning was performed using LoRA with rank $r = 8$, LoRA alpha $= 16$, and dropout $= 0.05$. Training was conducted for 3 epochs using the TRL `SFTTrainer` framework with a learning rate of 1.5×10^{-4}, a batch size of 1, gradient accumulation steps of 4, a sequence length of 512 tokens, and 8-bit paged AdamW optimization. Gradient checkpointing and mixed-precision FP16 training were enabled for memory efficiency. Training and inference were performed on Google Colab GPU infrastructure using PyTorch, Hugging Face Transformers, PEFT, TRL, BitsAndBytes, and Datasets libraries. The trainable parameter count corresponds only to the injected LoRA adapter layers, while the frozen base models retained their original parameter scales of approximately 7B parameters for Mistral and 3B parameters for Qwen2.5.

Several existing NLP and evaluation packages were used during preprocessing, training, retrieval, and evaluation. The implementation used HuggingFace Transformers for model loading and inference, PEFT for LoRA adaptation, TRL for supervised fine-tuning, SentenceTransformers for semantic retrieval, Scikit-learn for dataset splitting, and NLTK for tokenization support. Automatic evaluation metrics were implemented using the HuggingFace `evaluate` framework, including ROUGE, METEOR, and SacreBLEU. During generation, temperature was set to 0.2 and top-p sampling to 0.85 to improve dissemination consistency and reduce hallucinated outputs.

A.4.2 Keyword and Entity Analysis

To further analyze the semantic characteristics of citizen-generated disaster communication, a comprehensive keyword and entity-level analysis was performed on the enriched citizen-context dataset. The analysis focused on identifying dominant disaster-related terms, named entity distributions, emotional keyword associations, urgent communication patterns, and resource-demand semantics.

Semantic keywords were extracted using the YAKE unsupervised keyword extraction framework, while contextual entities were identified using spaCy-based Named Entity Recognition (NER). Let the extracted keyword set for message i be represented as:

$$K_i = \{w_1, w_2, \ldots, w_n\} \tag{50}$$

where w_n denotes the extracted candidate keyword associated with message i. The global keyword frequency across the dataset is defined as:

$$F(w_k) = \sum_{i=1}^{N} \mathbb{I}(w_k \in K_i) \tag{51}$$

where $\mathbb{I}(\cdot)$ denotes the indicator function and N represents the total number of disaster messages.

The analysis identified 87,886 unique semantic keywords across the dataset, indicating substantial lexical diversity within citizen disaster communication. Table 4 presents the most frequently occurring disaster-related keywords.

Table 4: Top semantic keywords extracted from citizen disaster messages

Keyword	Frequency	Percentage
food	548	2.1%
information	485	1.8%
earthquake	475	1.8%
people	380	1.4%
find	373	1.4%
water	341	1.3%
Haiti	312	1.2%
message	328	1.2%
Santiago	206	0.8%
Sandy	175	0.7%

The extracted keywords reveal strong associations with disaster response requirements, situational awareness, and humanitarian resource demands. Terms such as *food*, *water*, and *information* indicate recurring operational concerns within citizen communication streams during disaster events.

Named entity extraction was subsequently performed to identify contextual references associated with geographic regions, organizations, temporal expressions, and affected populations. Let the extracted entity set for message i be represented as:

$$E_i = \{(e_1, t_1), (e_2, t_2), \ldots, (e_n, t_n)\} \tag{52}$$

where e_n denotes the extracted entity and t_n represents the corresponding entity type.

The entity-type distribution across the dataset is summarized in Table 5.

Table 5: Distribution of extracted named entity types

Entity Type	Count	Percentage
ORG	7,239	27.6%
GPE	6,810	25.9%
DATE	5,598	21.3%
CARDINAL	5,529	21.1%
PERSON	3,995	15.2%
NORP	1,858	7.1%
LOC	1,129	4.3%
TIME	918	3.5%
QUANTITY	862	3.3%
MONEY	537	2.0%

The high occurrence of geopolitical entities (GPE) and organizational references (ORG) demonstrates that disaster communication is strongly associated with affected regions, humanitarian organizations, and emergency response agencies.

Table 6 presents the most frequently mentioned geographic locations identified from citizen disaster communication.

Table 6: Most frequently referenced geographic locations

Location	Mentions
Haiti	658
Pakistan	250
China	164
Afghanistan	159
Santiago	147
US	145
India	144
Indonesia	138
Chile	133
Port-au-Prince	87

Similarly, the most frequently referenced organizations are summarized in Table 7.

To analyze the relationship between emotional states and semantic communication patterns, keywords were additionally grouped according to the dominant emotion label assigned to each disaster message. Let the emotional keyword distribution for emotion class e_k be represented as:

Table 7: Most frequently referenced organizations

Organization	Mentions
WFP	196
UN	171
UNICEF	114
United Nations	79
UNHCR	79
Reuters	59

$$F(w|e_k) = \sum_{i=1}^{N} \mathbb{I}(w \in K_i \wedge \hat{e}_i = e_k) \tag{53}$$

where $\hat{e}_i$ denotes the predicted dominant emotion associated with message i.

The analysis revealed distinct semantic patterns across emotional categories. Fear-related messages were strongly associated with disaster-event terminology such as *earthquake*, *aftershocks*, and *Sandy*, while sadness-oriented communication emphasized humanitarian concerns including *family*, *food*, and *water*. Anger-related communication frequently contained infrastructure and aid-related terminology, including *destroyed*, *hungry*, and *house*.

High-priority disaster communication was identified using a combined emotional intensity and urgency criterion. Let the urgent message indicator be defined as:

$$urgent_i = \begin{cases} 1, & (\hat{e}_i \in \{fear, anger\}) \wedge (s_i > 0.99) \\ 0, & \text{otherwise} \end{cases} \tag{54}$$

where s_i represents the emotion-confidence score. Using this criterion, 2,367 messages (9.0% of the dataset) were categorized as high-priority urgent communication instances.

Resource-demand analysis further revealed strong operational requirements related to humanitarian aid and infrastructure disruption. Table 8 summarizes the frequency of major resource-related semantic categories.

Table 8: Resource and supply-demand mentions in citizen disaster communication

Resource Type	Mentions	Urgent Requests
Food	3,295	182
Shelter	1,979	148
Water	1,837	75
Medicine	743	35
Electricity	741	80
Rescue	437	60

The analysis demonstrates that citizen-generated disaster communication contains rich operational semantics associated with emotional response, humanitarian needs, geographic context, and emergency resource demands. These contextual signals provide valuable information for downstream disaster dissemination modeling, emergency prioritization, and situational awareness analysis.

A.4.3 Emotion Analysis

To analyze the emotional characteristics of citizen-generated disaster communication, an emotion-aware semantic analysis was performed using the pretrained transformer model *Emotion English DistilRoBERTa-base* (Hartmann, 2023). Each disaster message was classified into one of seven emotional categories:

$$\mathcal{E} = \{neutral,\ fear,\ sadness,\ disgust,\ anger,\ joy,\ surprise\} \tag{55}$$

Given a disaster message t_i, the transformer model predicts an emotion probability distribution:

$$P(e_k|t_i) = \text{softmax}(z_k) \tag{56}$$

where z_k represents the transformer output logits associated with emotion class e_k. The dominant emotion label assigned to message i is computed as:

$$\hat{e}_i = \arg\max_k P(e_k|t_i) \tag{57}$$

while the associated emotional confidence score is defined as:

$$s_i = \max_k P(e_k|t_i) \tag{58}$$

where $s_i \in [0, 1]$ denotes the emotional confidence associated with the predicted emotion.

The emotion analysis was performed on 26,248 citizen disaster messages. Table 9 presents the overall emotional distribution across the dataset.

Table 9: Distribution of detected emotions in citizen disaster communication

Emotion	Count	Percentage
neutral	12,331	47.0%
fear	6,030	23.0%
sadness	2,940	11.2%
disgust	1,904	7.3%
anger	1,396	5.3%
joy	1,238	4.7%
surprise	409	1.5%

The results indicate that neutral communication dominated the dataset, primarily corresponding to informational reports, news dissemination, and operational updates. Fear-related communication constituted the second-largest category, highlighting the prevalence of uncertainty, panic, and situational anxiety during disaster events.

Statistical analysis of the emotion-confidence scores demonstrated relatively strong emotional separability across the dataset. Let the global emotional confidence distribution be represented as:

$$S = \{s_1, s_2, \ldots, s_N\} \tag{59}$$

The descriptive statistics associated with the emotion-confidence scores are summarized in Table 10.

Table 10: Statistical distribution of emotion-confidence scores

Statistic	Value
Mean	0.673
Standard deviation	0.192
Minimum	0.000
25th percentile	0.512
Median	0.684
75th percentile	0.843
Maximum	0.995

To identify highly expressive emotional communication, high-confidence emotional messages were defined using the threshold criterion:

$$HC_i = \begin{cases} 1, & s_i > 0.8 \\ 0, & \text{otherwise} \end{cases} \tag{60}$$

Using this criterion, 8,557 messages (32.6% of the dataset) were categorized as high-confidence emotional communication instances.

Table 11 presents the distribution of high-confidence emotional messages.

Table 11: Distribution of high-confidence emotional messages ($s_i > 0.8$)

Emotion	Count	Percentage
neutral	4,460	52.1%
fear	2,117	24.7%
sadness	763	8.9%
disgust	473	5.5%
joy	421	4.9%
anger	250	2.9%
surprise	73	0.9%

Emotion-aware analysis was further conducted across communication genres to examine variations in emotional expression among direct communication, social media content, and news reports. Let

the genre-specific emotional distribution be represented as:

$$P(e_k|g_j) = \frac{count(e_k, g_j)}{count(g_j)} \tag{61}$$

where $g_j \in \{direct, social, news\}$ denotes the communication genre.

Table 12 summarizes the dominant emotional distributions across genres.

The results demonstrate that social-media communication exhibited the highest average emotional intensity, with fear emerging as the dominant emotional category. In contrast, direct and news-oriented communication primarily contained neutral informational content.

Given the operational importance of fear-related communication during disaster events, an additional fear-specific urgency analysis was performed. Fear-oriented disaster communication was identified as:

$$F_i = \begin{cases} 1, & \hat{e}_i = fear \\ 0, & \text{otherwise} \end{cases} \tag{62}$$

The analysis identified 6,030 fear-related messages with an average confidence score of 0.683. High-confidence fear communication frequently contained references to aftershocks, hurricanes, infrastructure instability, evacuation behavior, and uncertainty regarding ongoing disaster conditions.

To investigate the relationship between emotional intensity and communication complexity, message-length analysis was additionally performed. Let the message length for disaster message i be represented as:

$$L_i = |t_i| \tag{63}$$

where $|t_i|$ denotes the number of characters in message i.

The correlation between emotional confidence and message length was computed using Pearson correlation:

$$\rho_{L,S} = \frac{Cov(L, S)}{\sigma_L \sigma_S} \tag{64}$$

The resulting correlation coefficient was $\rho_{L,S} = -0.008$, indicating negligible linear correlation between message length and emotional intensity. Table 13 summarizes the average emotional scores across message-length categories.

The emotion analysis demonstrates that citizen-generated disaster communication contains strong emotionally expressive patterns associated with fear, uncertainty, humanitarian suffering, and operational disruption. These emotional signals provide valuable contextual information for downstream emergency dissemination modeling, urgency prioritization, and disaster situational-awareness analysis.

A.5 Text Similarity Metrics

A.5.1 ROUGE Metrics

ROUGE (Recall-Oriented Understudy for Gisting Evaluation) measures lexical overlap between generated text and reference text (Lin, 2004). In this study, ROUGE-1, ROUGE-2, and ROUGE-L are used to evaluate unigram overlap, bigram overlap, and longest common subsequence similarity, respectively.

ROUGE-1 measures unigram overlap between the generated dissemination message and the reference message:

$$\text{ROUGE-1} = \frac{\sum_{w \in \text{Ref}} \min(\text{Count}_{gen}(w), \text{Count}_{ref}(w))}{\sum_{w \in \text{Ref}} \text{Count}_{ref}(w)} \tag{65}$$

where w represents unigram tokens in the reference dissemination message.

ROUGE-2 extends this formulation to bigram overlap:

$$\text{ROUGE-2} = \frac{\sum_{b \in \text{Ref}} \min(\text{Count}_{gen}(b), \text{Count}_{ref}(b))}{\sum_{b \in \text{Ref}} \text{Count}_{ref}(b)} \tag{66}$$

where b denotes bigram sequences.

ROUGE-L evaluates sequence-level similarity using the Longest Common Subsequence (LCS) between generated and reference dissemination messages:

$$\text{ROUGE-L} = \frac{LCS(X, Y)}{|Y|} \tag{67}$$

where X represents the generated message, Y represents the reference dissemination message, and $LCS(X, Y)$ denotes the longest common subsequence between the two sequences.

ROUGE metrics are particularly suitable for this study because official dissemination messages often follow structured operational phrasing patterns and standardized protective-action language.

Table 12: Emotion distribution across communication genres

Genre	Messages	Dominant Emotion	Fear Messages	Average Score
Direct	10,782	neutral (5,013)	2,754	0.651
Social	2,398	fear (975)	975	0.704
News	13,068	neutral (6,949)	2,301	0.686

Table 13: Average emotion scores across message-length categories

Length Category	Average Score	Messages
Very Short	0.659	2,868
Short	0.666	7,130
Medium	0.680	11,454
Long	0.681	4,653
Very Long	0.614	143

A.5.2 BLEU Score

BLEU (Bilingual Evaluation Understudy) evaluates the precision of n-gram overlap between generated and reference texts (Papineni et al., 2002). BLEU is widely used in machine translation and controlled text generation tasks to measure similarity between generated sequences and target outputs.

The BLEU score is computed as:

$$\text{BLEU} = BP \cdot \exp\left(\sum_{n=1}^{N} w_n \log p_n\right) \tag{68}$$

where p_n represents the modified n-gram precision, w_n denotes the weighting factor for each n-gram order, and BP is the brevity penalty defined as:

$$BP = \begin{cases} 1, & c > r \\ e^{(1-r/c)}, & c \leq r \end{cases} \tag{69}$$

where c is the generated sequence length and r is the reference sequence length.

BLEU is important in this study because it evaluates how closely the generated dissemination messages align with official IPAWS dissemination structures and phrase-level communication patterns.

A.5.3 METEOR Score

METEOR (Metric for Evaluation of Translation with Explicit Ordering) measures semantic similarity between generated and reference texts using unigram matching, stemming, and synonym alignment (Banerjee and Lavie, 2005). Unlike BLEU, METEOR incorporates both precision and recall while accounting for word ordering and semantic variation.

The METEOR score is computed as:

$$F_{mean} = \frac{10PR}{R + 9P} \tag{70}$$

where P denotes unigram precision and R denotes unigram recall. A fragmentation penalty is then applied:

$$Penalty = 0.5 \left(\frac{chunks}{matches}\right)^3 \tag{71}$$

The final METEOR score is:

$$\text{METEOR} = (1 - Penalty) \cdot F_{mean} \tag{72}$$

METEOR is particularly useful for evaluating disaster dissemination generation because official emergency communication may contain semantically equivalent but lexically different protective-action phrases and operational wording.

The combination of ROUGE, BLEU, and METEOR provides a comprehensive evaluation framework for measuring lexical overlap, structural consistency, contextual fluency, and semantic alignment between generated dissemination messages and reference IPAWS-style alerts.

A.6 Dissemination Quality Metrics

A.6.1 Hallucination Detection

Hallucination analysis was designed to identify unsupported operational details introduced by the model that were not present in the provided input context. A set of unsafe semantic categories was defined:

$$\mathcal{H} = \{time_date,\ casualties,\ measurements,\ agencies,\ event_status\} \tag{73}$$

Each category was detected using rule-based regular-expression matching over the generated dissemination output. For a hallucination category $h_k \in \mathcal{H}$, the hallucination indicator is defined as:

$$H_{ik} = \begin{cases} 1, & h_k \in \hat{y}_i \wedge h_k \notin x_i \\ 0, & \text{otherwise} \end{cases} \tag{74}$$

The final hallucination flag for sample i is computed as:

$$H_i = \mathbb{I}\left(\sum_{k=1}^{|\mathcal{H}|} H_{ik} > 0\right) \tag{75}$$

where $\mathbb{I}(\cdot)$ denotes the indicator function. The overall hallucination rate across the evaluation corpus is defined as:

$$HallucinationRate = \frac{1}{N}\sum_{i=1}^{N} H_i \tag{76}$$

where N represents the total number of evaluated dissemination outputs.

A.6.2 Constraint Following Evaluation

To measure adherence to the `NO_UNSUPPORTED_FACTS` dissemination constraint, the constraint-following score was computed directly from the hallucination flag:

$$C_i = 1 - H_i \tag{77}$$

The overall constraint-following rate is therefore:

$$ConstraintRate = \frac{1}{N}\sum_{i=1}^{N} C_i \tag{78}$$

A value close to 1 indicates that the generated dissemination messages consistently avoid unsupported operational claims.

A.6.3 Actionability Evaluation

Actionability analysis was designed to measure whether the generated dissemination messages contain actionable public-safety guidance. A predefined operational-action lexicon was constructed:

$$\mathcal{A} = \{evacuate,\ shelter,\ avoid,\ stay\ indoors,\ follow\ instructions,\ move\ to\ higher\ ground, prepare\} \tag{79}$$

For each dissemination output $\hat{y}_i$, the actionability score is computed as:

$$A_i = \min\left(\frac{|K_i^{action}|}{3}, 1\right) \tag{80}$$

where K_i^{action} denotes the set of matched operational-action keywords within the generated message.

The binary actionability indicator is defined as:

$$A_i^{bin} = \mathbb{I}\left(|K_i^{action}| > 0\right) \tag{81}$$

The overall actionability rate is computed as:

$$ActionabilityRate = \frac{1}{N}\sum_{i=1}^{N} A_i^{bin} \tag{82}$$

while the average actionability score is:

$$AvgActionability = \frac{1}{N}\sum_{i=1}^{N} A_i \tag{83}$$

A.6.4 Official Tone Evaluation

To evaluate dissemination professionalism and operational calmness, an official tone evaluation framework was introduced. The framework measures three properties:

1. absence of panic-oriented language,
2. concise dissemination length,
3. presence of operational emergency terminology.

Let the panic lexicon be represented as:

$$\mathcal{P} = \{catastrophic,\ terrifying,\ deadly,\ panic, massive\ destruction\} \tag{84}$$

The calmness score is defined as:

$$Calm_i = \begin{cases} 1, & \mathcal{P} \cap \hat{y}_i = \emptyset \\ 0, & \text{otherwise} \end{cases} \tag{85}$$

Conciseness was evaluated using message length constraints. Let L_i denote the word count of dissemination output $\hat{y}_i$. The conciseness score is computed as:

$$Concise_i = \begin{cases} 1, & 20 \le L_i \le 80 \\ 0.5, & \text{otherwise} \end{cases} \tag{86}$$

Operational-awareness scoring was additionally performed using operational emergency terminology such as *authorities*, *officials*, *instructions*, *guidance*, and *evacuation*. The operational-tone indicator is defined as:

$$Operational_i = \begin{cases} 1, & \exists\, t_k \in \hat{y}_i \\ 0, & \text{otherwise} \end{cases} \tag{87}$$

The final official-tone score is computed as:

$$Tone_i = \frac{Calm_i + Concise_i + Operational_i}{3} \tag{88}$$

The corpus-level official-tone score is then calculated as:

$$AvgTone = \frac{1}{N}\sum_{i=1}^{N} Tone_i \tag{89}$$

Finally, the average dissemination length is computed as:

$$AvgWords = \frac{1}{N}\sum_{i=1}^{N} L_i \tag{90}$$

This evaluation framework enables systematic measurement of dissemination safety, factual consistency, operational usefulness, and communication professionalism in generated emergency dissemination messages.

A.7 Future Works

The study also contributes to the broader discussion on bridging citizen communication and official dissemination domains within crisis informatics. Existing systems mainly use citizen-generated content for situational analytics, whereas emergency managers require structured dissemination outputs. The proposed weakly supervised citizen-to-alert framework demonstrates that semantic alignment between citizen-generated disaster communication and official IPAWS archives can partially bridge this gap and extend disaster NLP research beyond classification toward operational dissemination generation. Future work can further strengthen this framework through multimodal crisis-state modeling, adaptive retrieval strategies, reinforcement learning, multilingual dissemination generation, and Human-in-the-Loop (HITL) validation, where expert-guided refinement of weakly aligned citizen-to-alert pairs may improve dataset quality, contextual consistency, dissemination reliability, and controllability.

A.8 Limitations

Although the proposed framework demonstrates promising results for controllable disaster dissemination generation, several limitations remain. First, the dataset construction process relies on weakly supervised semantic alignment between citizen-generated disaster messages and official IPAWS dissemination alerts. Since no directly annotated citizen-to-alert dissemination dataset currently exists, the alignment process depends on contextual similarity, disaster-category matching, and retrieval-based semantic association rather than manually verified expert annotations. Consequently, some citizen-official dissemination pairs may contain contextual inconsistencies, incomplete operational mappings, or stylistic mismatches that can affect downstream generation quality and controllability.

Second, the current framework primarily focuses on textual disaster communication and does not incorporate multimodal disaster evidence such as images, videos, geospatial streams, sensor observations, or real-time environmental data. Modern disaster communication environments are inherently multimodal, and integrating visual and geospatial grounding may significantly improve crisis-state understanding, dissemination accuracy, and operational reliability. In addition, although the proposed context-control framework substantially reduces hallucination and improves dissemination consistency, large language models may still generate unsupported or operationally unsafe statements under highly ambiguous or rapidly evolving disaster scenarios. Therefore, the generated outputs should currently be considered decision-support dissemination drafts rather than fully autonomous operational emergency alerts.

Another limitation concerns evaluation and operational validation. While automatic metrics such as ROUGE, BLEU, and METEOR provide useful quantitative measurements, they do not fully capture dissemination effectiveness, public communication clarity, actionable quality, or real-world emergency usability. Although the study introduced additional safety and controllability evaluations, broader human evaluation involving emergency managers, dissemination authorities, public-warning experts, and crisis communication practitioners would provide stronger operational validation. Furthermore, the current experiments mainly focus on English-language disaster communication and IPAWS-style dissemination structures, which may limit generalizability to multilingual, culturally diverse, or region-specific emergency communication environments. Future work should therefore explore multilingual dissemination generation, stronger human-in-the-loop validation, reinforcement learning from expert feedback, adaptive re-

trieval grounding, and expert-curated citizen-to-alert alignment strategies to improve dissemination reliability, factual consistency, and operational applicability.