REU Projects 1 & 2

Project 1

EmoCause-PIC: Emotion-Cause Pair Identification during Crises

Abstract

In disaster scenarios, social media platforms like Twitter serve as real-time communication channels for individuals experiencing and reporting crises. While prior research has leveraged tweets for event detection and sentiment analysis, existing approaches often isolate emotion classification from the contextual triggers that provoke them—limiting situational understanding and impeding targeted emergency response. This work will introduce EPIC (Emotion-Cause Pairs Identification and Classification), a novel framework aimed at jointly identifying emotions and their corresponding causal triggers from user-generated content during natural disasters.

We will formulate this as a multi-label classification and joint sequence labeling task to extract fine-grained emotion-cause pairs from noisy, short-form texts. Leveraging fine-tuned large language models (LLMs), we will propose a dual-head architecture where one branch predicts emotional categories (e.g., fear, anger, sadness), and the other will extract corresponding cause spans (e.g., “floodwater rising”, “power outage”). Our approach will be specifically designed for early crisis stages when situational awareness and emotional distress rapidly escalate.

By capturing the *why* behind emotional expressions, EPIC will enable more accurate psychological profiling and prioritization of disaster response resources based on emotional severity and urgency. For instance, distinguishing between general anger and panic caused by immediate physical threats can inform more nuanced triage strategies for rescue coordination and mental health outreach. This work aims to bridge affective computing and crisis informatics to deliver a scalable, interpretable tool for enhancing disaster resilience.

REU Projects 1 & 2

Project 2

CELAQ-G: Ground-Truth-Free Quality Assessment via LLM Explanation Consistency

Abstract

Assessing the factual reliability of social media content during disasters is critical for mitigating the spread of misinformation and enabling effective emergency response. However, most existing fact-checking systems rely heavily on ground-truth labels or curated knowledge bases, which are often unavailable or delayed in rapidly evolving crisis contexts. This work will introduce a novel, ground-truth-free framework for tweet reliability assessment by leveraging large language model (LLM) explanation consistency.

We will propose a self-supervised method in which an LLM (e.g., GPT-4) is prompted to generate multiple natural language rationales for the factual claims made in a tweet. These rationales are elicited across diverse semantic paraphrases and prompt variations, and their internal coherence is measured to produce a consistency score. Intuitively, tweets with stable, logically aligned explanations across multiple reasoning paths are more likely to be grounded in fact, while incoherent or contradictory rationales indicate potential misinformation.

To enrich the model's reasoning, we will integrate retrieval-augmented generation (RAG), enabling the LLM to incorporate real-time disaster-related evidence from trusted sources during inference. The resulting system will classify tweets as likely true or false, entirely without the need for labeled training data or external validation sets.

This research will pioneer a new class of reliability detection tools that are scalable, interpretable, and resilient to data scarcity. It has broad applications for disaster informatics, misinformation mitigation, and real-time trust assessment on social platforms, particularly in environments where traditional fact-checking pipelines fall short.