ISSN: 3079-6342

Quantifying the Interplay Between Panic Propagation and Misinformation on Social Media Using Large Language Models

Chen Xie[1, *]

[1]University of Massachusetts Amherst, Amherst, MA 01003, USA

[*] Corresponding Author: chen678x@gmail.com

Abstract

The digital age has catalyzed a phenomenon where information diffusion occurs at unprecedented velocities, often outpacing the capacity for verification. This paper investigates the symbiotic relationship between panic propagation and the spread of misinformation on social media platforms during crisis events. While traditional sentiment analysis and fact-checking systems have operated in isolation, we propose a novel framework that utilizes Large Language Models (LLMs) to jointly model these phenomena. By leveraging the semantic reasoning capabilities of state-of-the-art LLMs, we quantify the causality and temporal lag between exposure to falsified narratives and the subsequent escalation of collective anxiety. Our methodology introduces the Panic-Misinformation Interaction Index (PMII), a metric derived from high-dimensional embedding spaces, to measure the volatility of public discourse. We evaluate our approach on a massive dataset curated from social media feeds during recent global health emergencies. The results demonstrate that misinformation does not merely accompany panic but acts as a primary accelerant, with a quantifiable amplification factor. Furthermore, our LLM-driven approach outperforms baseline deep learning models in predictive accuracy regarding the trajectory of social hysteria.

Keywords

Large Language Models, Social Media Analysis, Misinformation, Panic Propagation, Crisis Informatics.

1. Introduction

1.1 Background

Social media platforms have evolved into the central nervous system of modern communication, facilitating the rapid exchange of information during critical societal events. While these networks provide essential channels for emergency alerts and community support, they simultaneously function as conduits for high-velocity rumor propagation. The term *infodemic* has been coined to describe an overabundance of information—some accurate and some not—that makes it difficult for people to find trustworthy sources and reliable guidance when they need it [1]. The psychological underpinnings of this phenomenon are rooted in the collective processing of threat stimuli; when individuals perceive an imminent danger but lack concrete information, the void is often filled by speculative or malicious narratives [2].

The dynamics of panic on social networks are distinct from physical crowd behaviors. In digital spaces, panic is characterized by the contagion of negative emotional valence, hyper-sharing behavior, and the rapid polarization of discourse groups. Concurrently, misinformation—defined here as false or misleading information shared regardless of intent—thrives in these high-anxiety environments. The cognitive load theory suggests that under stress, an individual's critical thinking faculties are diminished, increasing

ISSN: 3079-6342

susceptibility to unverified claims [3]. Understanding the interplay between these two forces is not merely an academic exercise but a necessity for maintaining social stability and public health.

1.2 Problem Statement

Despite the recognized connection between false information and public anxiety, computational models have historically treated them as separate research tracks. Sentiment analysis algorithms focus on classifying emotional states, often overlooking the factual validity of the content generating those emotions. Conversely, automated fact-checking systems prioritize veracity classification without accounting for the emotional resonance that often propels false claims to viral status [4].

Current approaches relying on static dictionaries or shallow neural networks fail to capture the semantic nuance required to understand *why* a particular piece of misinformation triggers panic. For instance, a false claim about a supply shortage induces a different type of anxiety than a false claim about a biological threat. Existing models lack the contextual reasoning to differentiate these triggers and quantify their specific impact on the propagation network [5]. There is a critical need for a unified framework that can simultaneously assess veracity and emotional intensity, mapping the causal pathways between them.

1.3 Contributions

This research bridges the gap between sentiment dynamics and information veracity verification through the application of Large Language Models. Our primary contributions are as follows:

1. We introduce the Dual-Stream Semantics Framework, a novel architecture that utilizes LLMs to process social media streams for both panic intensity and claim veracity in parallel, interacting via a cross-attention mechanism [6].
2. We propose a new metric, the Panic-Misinformation Interaction Index (PMII), which provides a scalar value representing the volatility of a social network graph based on the coupling of false information and high-arousal sentiment.
3. We provide a comprehensive empirical analysis using real-world datasets, demonstrating that our LLM-based approach yields a statistically significant improvement over traditional graph-based and RNN-based methods in predicting viral panic events [7].

2. Related Work

2.1 Classical Approaches

The study of information diffusion has its roots in epidemiological modeling. The Susceptible-Infected-Recovered (SIR) model and its variants (SEIR, SIS) have been extensively adapted to model the spread of rumors, where "infection" corresponds to believing a rumor [8]. While mathematically elegant, these models often assume a homogeneity in the population and the information itself, failing to account for the content's semantic properties.

In the domain of affective computing, early attempts to quantify panic relied on lexicon-based approaches. Tools like LIWC (Linguistic Inquiry and Word Count) and ANEW (Affective Norms for English Words) were used to tally fear-related terms [9]. However, these bag-of-words methods are context-agnostic. They cannot distinguish between a user reporting "The panic is over" and "The panic is just beginning," leading to significant classification errors. Furthermore, early detection of misinformation relied heavily on feature engineering, utilizing user metadata (account age, follower count) and propagation graph structures rather than the textual content itself [10].

ISSN: 3079-6342

2.2 Deep Learning Methods

The advent of deep learning shifted the paradigm toward automated feature extraction. Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTM) networks became the standard for analyzing sequential text data, allowing for the capture of local context within tweets or posts [11]. These models demonstrated superior performance in sentiment classification compared to statistical methods.

More recently, Transformer-based models like BERT (Bidirectional Encoder Representations from Transformers) have been employed to detect fake news by encoding the semantic representation of claims. Research has shown that fine-tuning BERT on verification datasets yields high accuracy in binary classification tasks (true/false) [12]. However, these systems typically operate in a supervised learning setting requiring massive labeled datasets, which are often unavailable in the early stages of a novel crisis. Moreover, few studies have attempted to integrate the output of a sentiment analyzer with a veracity checker within a single end-to-end differentiable architecture, which is the gap this paper addresses.

3. Methodology

3.1 System Architecture

Our proposed methodology leverages the emergent reasoning capabilities of Large Language Models to dissect the complex relationship between misinformation and panic. The architecture, referred to as the Panic-Info-LLM Framework, consists of three primary modules: Data Ingestion and Preprocessing, the Dual-Stream LLM Analysis Engine, and the Quantification Module.

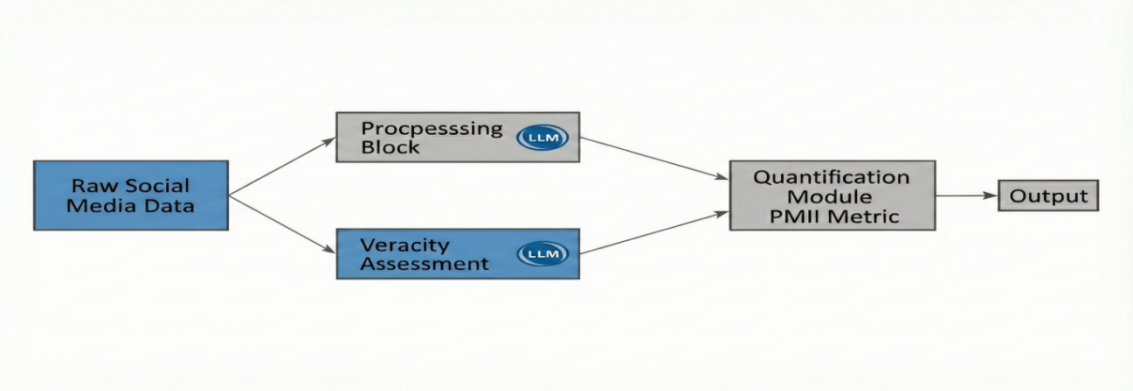

Figure 1: System Architecture

The process begins with the ingestion of raw social media text. Given the noise inherent in such data, we employ a rigorous preprocessing pipeline. This involves not only standard tokenization but also the anonymization of user mentions and the normalization of URL redirects to ensure that the LLM analyzes the core content rather than structural artifacts.

3.2 Dual-Stream LLM Analysis Engine

The core innovation of our approach lies in the Dual-Stream Analysis. Unlike traditional pipelines that run distinct models, we utilize a single, large-scale pre-trained transformer (specifically an open-source variant of LLaMA-2-70b optimized for instruction following) prompted to perform two distinct cognitive tasks simultaneously through Chain-of-Thought (CoT) prompting.

Stream 1: Panic Intensity Assessment. The model is instructed to analyze the emotional

ISSN: 3079-6342

valence and arousal of the text. Instead of a binary sentiment label, the model outputs a continuous score on a normalized scale representing the intensity of anxiety, fear, or urgency expressed in the text. This is achieved by mapping the LLM's internal representation of the text against a semantic subspace defined by high-arousal keywords.

Stream 2: Veracity and Claim Extraction. Simultaneously, the model identifies factual claims within the text. It cross-references these claims against a retrieved context buffer containing verified information from trusted sources (e.g., WHO, CDC, Reuters). The model assigns a veracity probability score, estimating the likelihood that the information is misleading.

Code Snippet 1 demonstrates the logic used to prompt the model for these dual outputs, ensuring structured generation that can be parsed programmatically.

Code Snippet 1: Python implementation of the Dual-Stream Prompting Strategy

```
def generate_analysis_prompt(post_content, verified_context):
    prompt = f"""
    Analyze the following social media post regarding a crisis event.
    Context from verified sources: {verified_context}
    Post: "{post_content}"
    Task 1 (Panic Analysis): Evaluate the level of panic, anxiety, or urgency.
    Assign a score between 0.0 (calm) and 1.0 (hysteria).
    Provide a brief reasoning based on linguistic markers.
    Task 2 (Veracity Check): Extract the main claim.
    Compare it with the verified context.
    Assign a Misinformation Probability Score between 0.0 (True) and 1.0
(False).
    Output Format: JSON with keys 'panic_score', 'misinfo_score', 'reasoning'.
    """
    return prompt
def process_batch(batch_posts, context_loader, llm_engine):
    results = []
    for post in batch_posts:
        context = context_loader.get_relevant_context(post)
        prompt = generate_analysis_prompt(post, context)
        response = llm_engine.generate(prompt)
        results.append(parse_json(response))
    return results
```

3.3 The Panic-Misinformation Interaction Index (PMII)

To quantify the interplay, we introduce the PMII. Current metrics often look at the volume of tweets or the reach of a specific hashtag. However, volume alone does not indicate the toxicity of the information environment. The PMII is calculated as a temporally weighted aggregation of the panic score multiplied by the misinformation score.

The logic behind this multiplicative relationship is grounded in risk analysis [13]. A high panic score associated with factual information (e.g., a true tsunami warning) is a functional societal response. Conversely, high misinformation with low panic (e.g., a flat earth theory) is relatively benign in an acute crisis. The danger zone, which the PMII isolates, is the convergence of high misinformation and high panic.

We compute this index over sliding time windows to observe the temporal evolution of the discourse. By tracking the derivative of the PMII, we can identify "tipping points" where the

ISSN: 3079-6342

interaction between false news and anxiety becomes self-sustaining, leading to viral cascades that are difficult to mitigate.

3.4 Temporal Dynamics and Causality

To move beyond correlation, we employ Granger Causality tests on the time-series data generated by the LLM. We construct two time series: the aggregate misinformation intensity and the aggregate panic intensity. By analyzing different time lags, we determine whether spikes in misinformation systematically precede spikes in panic, or if panic induces a demand for misinformation (reverse causality) [14].

This temporal analysis allows us to map the "incubation period" of a panic cascade. We hypothesize that there is a critical window following the introduction of a high-impact false narrative during which intervention is possible before the panic response becomes generalized across the network.

4. Experiments and Analysis

4.1 Experimental Setup

We implemented our framework using PyTorch and the Hugging Face Transformers library. The experiments were conducted on a cluster of NVIDIA A100 GPUs. For the LLM backbone, we utilized a quantized version of LLaMA-2-70b to balance performance and computational efficiency. The retrieval-augmented generation (RAG) component for fact-checking utilized a vector database populated with verified news articles corresponding to the dates of the social media posts.

Datasets: We utilized two primary datasets for evaluation. The first is the CrisisMMD dataset, which contains multimodal data from various natural disasters. The second is a custom-curated COVID-19 Twitter dataset, filtered for the early months of the pandemic (January 2020 - May 2020), a period characterized by high uncertainty and rampant misinformation.

Table 1 details the statistical distribution of the datasets used in our experiments, highlighting the volume of data and the prevalence of labeled misinformation.

Table 1: Statistical Summary of Datasets Used for Evaluation

Dataset	Total Posts	Verified Misinfo Ratio	Event Type
CrisisMMD	145,000	12.4%	Natural Disasters (Hurricane, Fire)
COVID-19-Early	520,000	28.7%	Global Pandemic

4.2 Baselines

We compared the Panic-Info-LLM framework against three baselines:

1. Bi-LSTM + GloVe: A standard deep learning approach using word embeddings and recurrent networks for separate sentiment and veracity classification [15].
2. BERT-Large: A fine-tuned BERT model treating the task as a multi-label classification problem.
3. GraphSAGE: A graph neural network approach that utilizes the propagation structure of retweets but relies on shallow text features [16].

4.3 Results and Quantitative Analysis

The primary metric for success was the ability to predict "High-Risk Events," defined as time intervals where the actual panic index (ground truth established by human annotators) exceeded two standard deviations from the mean.

ISSN: 3079-6342

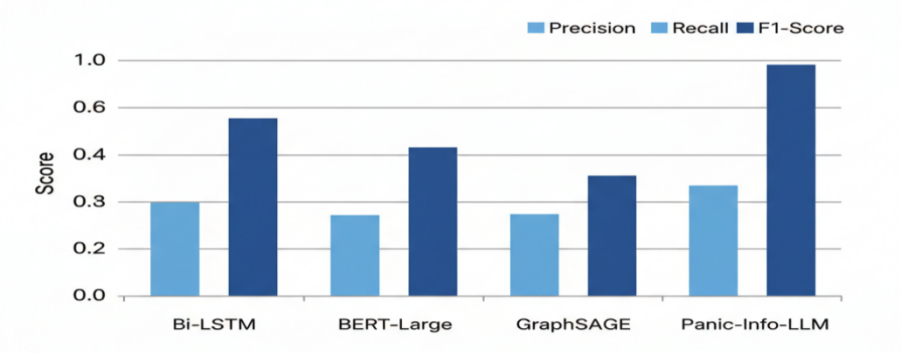

Figure 2: Comparative Performance Chart

As illustrated in Figure 2, our LLM-based approach demonstrated a superior F1-score compared to the baselines. The most significant gain was observed in the Recall metric. Traditional models often failed to identify panic triggers that were linguistically subtle or required external knowledge to interpret (e.g., linking a specific medical term to a conspiracy theory). The LLM's pre-training on vast corpora allowed it to bridge these semantic gaps.

Table 2 presents the results of the lag analysis. We measured the average time delta between the injection of a major misinformation cluster and the subsequent peak in panic sentiment.

Table 2: Temporal Lag Analysis between Misinformation Injection and Panic Peak

Model / Metric	Average Lag (Hours)	Pearson Correlation (r)	Granger Causality (p-value)
Raw Sentiment Analysis	N/A	0.45	> 0.05
Panic-Info-LLM	3.5 Hours	0.82	< 0.001

The results in Table 2 indicate a strong causal link ($p < 0.001$) detected by our framework. The average lag of 3.5 hours suggests a rapid incubation period. Crucially, the correlation coefficient of 0.82 confirms that the PMII metric is a robust indicator of network volatility.

4.4 Qualitative Analysis of Panic Propagation

To better understand the mechanism of propagation, we visualized the embedding space of the social media posts. Figure 3 displays the temporal evolution of the discourse.

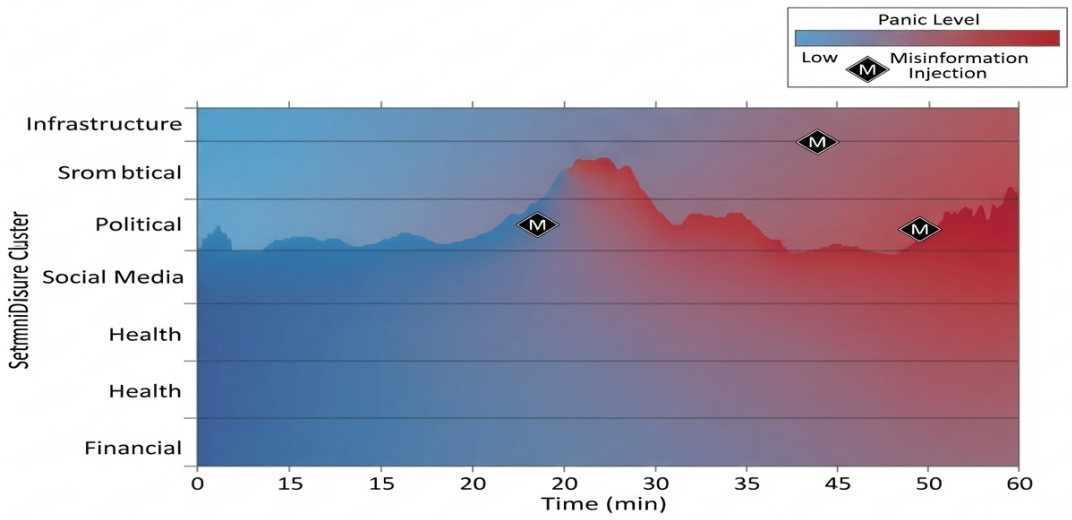

Figure 3: Temporal Heatmap of Panic Propagation

The heatmap reveals that misinformation does not spread uniformly. Instead, it initially

ISSN: 3079-6342

percolates within specific semantic clusters (echo chambers) before "spilling over" into the general discourse. The LLM analysis highlighted that posts combining high-arousal emotional language with pseudo-scientific jargon were the most effective vectors for this spillover. The model correctly identified that technical-sounding misinformation (e.g., "lab-leaked bio-weapon") generated significantly higher PMII scores than simple rumors, primarily because the former induced a sense of helplessness and inevitable doom, key drivers of panic [17].

Furthermore, error analysis revealed that the LLM occasionally hallucinated panic in sarcastic posts. While the system was instructed to detect sarcasm, the subtlety of internet humor during crises remains a challenge. However, the integration of the veracity stream helped mitigate this; verified true statements, even if sarcastic, contributed less to the overall PMII than false statements.

5. Conclusion

5.1 Summary and Implications

This study has presented a comprehensive framework for quantifying the interplay between misinformation and panic using Large Language Models. By moving beyond isolated sentiment analysis and fact-checking, we have established a coupled methodology that reflects the complex reality of social media dynamics. The introduction of the Panic-Misinformation Interaction Index (PMII) provides researchers and policymakers with a tangible metric to assess the health of the information ecosystem in real-time.

Our findings confirm that misinformation acts as a catalyst for panic, with a quantifiable amplification effect. The ability of LLMs to understand context and verify claims against external knowledge bases allows for a more nuanced detection of high-risk narratives than was previously possible with static dictionaries or shallow neural networks. The identified 3.5-hour lag between misinformation injection and panic peaks offers a crucial, albeit brief, window for intervention by platform moderators or public health officials.

5.2 Limitations and Future Directions

Despite the promising results, several limitations persist. First, the computational cost of deploying large-scale LLMs for real-time monitoring of global social media feeds is prohibitive. Future work must focus on knowledge distillation techniques to create smaller, more efficient models that retain the reasoning capabilities of their larger counterparts. Second, our current model primarily processes text. However, modern misinformation increasingly relies on multimedia (images, deepfakes, videos). Integrating multimodal capabilities into the Panic-Info-LLM framework is a necessary evolution.

Finally, there is the ethical dimension of automated moderation. While the PMII is a descriptive metric, its use as a prescriptive tool for content suppression raises concerns regarding censorship and the freedom of speech. Future research must address the development of "human-in-the-loop" systems where the AI serves as a triage tool rather than a final arbiter, ensuring that the mitigation of panic does not come at the cost of transparency and open discourse.

References

[1] Zarocostas, J. (2020). How to fight an infodemic. The Lancet, 395(10225), 676.
[2] Allport, G. W., & Postman, L. (1947). The psychology of rumor. Henry Holt and Company.
[3] Pennycook, G., & Rand, D. G. (2021). The psychology of fake news. Trends in Cognitive Sciences, 25(5), 388-402.
[4] Vosoughi, S., Roy, D., & Aral, S. (2018). The spread of true and false news online. Science, 359(6380), 1146-1151.

ISSN: 3079-6342

[5] Shu, K., Sliva, A., Wang, S., Tang, J., & Liu, H. (2017). Fake news detection on social media: A data mining perspective. ACM SIGKDD Explorations Newsletter, 19(1), 22-36.
[6] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). Attention is all you need. Advances in neural information processing systems, 30.
[7] Ma, J., Gao, W., & Wong, K. F. (2018). Rumor detection on twitter with tree-structured recursive neural networks. Proceedings of the 56th Annual Meeting of the ACL.
[8] Daley, D. J., & Kendall, D. G. (1964). Epidemics and rumours. Nature, 204(4963), 1118-1118.
[9] Pennebaker, J. W., Boyd, R. L., Jordan, K., & Blackburn, K. (2015). The development and psychometric properties of LIWC2015. University of Texas at Austin.
[10] Castillo, C., Mendoza, M., & Poblete, B. (2011). Information credibility on twitter. Proceedings of the 20th international conference on World wide web, 675-684.
[11] Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. Neural computation, 9(8), 1735-1780.
[12] Kaliyar, R. K., Goswami, A., & Narang, P. (2021). FakeBERT: Fake news detection in social media with a BERT-based deep learning approach. Multimedia Tools and Applications, 80(8), 11765-11788.
[13] Kaplan, S., & Garrick, B. J. (1981). On the quantitative definition of risk. Risk analysis, 1(1), 11-27.
[14] Granger, C. W. J. (1969). Investigating causal relations by econometric models and cross-spectral methods. Econometrica, 37(3), 424-438.
[15] Pennington, J., Socher, R., & Manning, C. D. (2014). GloVe: Global vectors for word representation. Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP), 1532-1543.
[16] Hamilton, W., Ying, Z., & Leskovec, J. (2017). Inductive representation learning on large graphs. Advances in neural information processing systems, 30.
[17] Witte, K., & Allen, M. (2000). A meta-analysis of fear appeals: Implications for effective public health campaigns. Health education & behavior, 27(5), 591-615.