GeoSentinel: Real-Time Geopolitical Tension Analysis Using Data-Driven Approaches

Mohd Arshad
Student
Dr. APJ Abdul Kalam Technical University
Lucknow, India
arshadmohd8574@gmail.com

***Abstract*—This paper proposes a novel framework for quantifying geopolitical tension in real-time, specifically tailored to the India-Pakistan and Russia-Ukraine dyads. Addressing the limitations of traditional risk indices, which often lack the nuance to capture the multifaceted nature of modern conflict, this research introduces the Geopolitical Tension Index (GPTI). The index is built upon a two-pillar conceptual architecture: *Military & Conflict Tension (MCT)*, which measures objective kinetic events, and *Intelligent Narrative Tension (INT)*, which gauges the sentiment of the information environment. The core innovation lies in a hybrid AI analysis engine that employs a symbiotic division of labor: a Large Language Model (LLM) performs complex relevance filtering, while a fine-tuned DistilBERT model executes high-throughput sentiment analysis. Furthermore, the index utilizes an adaptive dynamic weighting mechanism based on Principal Component Analysis (PCA). Crucially, this study validates the "lead-lag" hypothesis of modern conflict. Using Granger Causality tests, we demonstrate that narrative hostility is a statistically significant predictor of kinetic violence ($p < 0.05$), offering a validated early-warning instrument for policymakers.**

***Index Terms*—Geopolitical Risk, Natural Language Processing, Hybrid AI, Granger Causality, Time-Series Analysis**

I. Introduction

Modern conflict is no longer defined solely by physical violence; it is preceded, shaped, and amplified by a "war of words" in the information domain. Yet, traditional methods for measuring geopolitical risk often rely on qualitative expert assessments, which suffer from low frequency and subjective bias. While automated indices like the Caldara & Iacoviello GPR Index have pioneered text-based measurement, they typically rely on keyword counting, failing to capture the semantic nuance of pre-conflict rhetoric.

This paper presents **GeoSentinel**, a computational framework designed to measure, visualize, and scientifically validate the relationship between hostile narratives ("Words") and physical violence ("War"). We propose a dyad-specific approach that fuses multi-source data streams to generate a high-frequency, real-time tension score.

II. The Two-Pillar Architecture

The theoretical foundation of the GPTI is the disaggregation of tension into two distinct but interrelated dimensions.

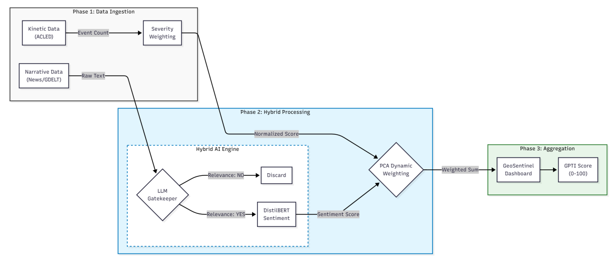

Fig. 1. The GeoSentinel System Architecture: Data flows from Kinetic (ACLED) and Narrative (News) sources through a Hybrid AI Engine before aggregation via PCA.

A. Pillar 1: Military & Conflict Tension (MCT)

The MCT score serves as the objective anchor of the framework, measuring "ground truth" kinetic activity. It is constructed by processing event data (e.g., from ACLED), filtering for dyadic events between the target nations. Events are weighted by a severity rubric (e.g., an airstrike receives a higher weight than a border protest).

B. Pillar 2: Intelligent Narrative Tension (INT)

The INT score measures the "temperature" of the information environment. It fuses data from global news APIs and official government communications. To process this unstructured text, we employ a **Hybrid AI Engine**:

1) Stage 1: The Gatekeeper (LLM): Raw news data is noisy. A general-purpose Large Language Model (LLM) acts as a zero-shot classifier to filter relevance. The prompt instructs the model to distinguish between genuine geopolitical tension and irrelevant topics such as cultural events or sports rivalries.

2) Stage 2: The Analyst (DistilBERT): Text identified as relevant is passed to a fine-tuned DistilBERT model. Selected for its computational efficiency, this model classifies the sentiment of the text. The INT score is derived from the volume of negative-sentiment content, reflecting the intensity of hostile rhetoric.

III. Index Aggregation and Weighting

To combine the heterogeneous data streams from the Kinetic and Narrative pillars, a robust mathematical aggregation strategy is required.

A. Normalization

Raw scores are normalized using a rolling Min-Max technique over a 36-month window. This ensures comparability between the count-based kinetic data and the volume-based narrative data while adapting to the non-stationary nature of conflict time series.

B. Dynamic Weighting via PCA

Unlike static indices, GPTI employs Principal Component Analysis (PCA) on a rolling basis. Let X_t be the standardized data matrix of the two pillars at time t. We compute the covariance matrix Σ and solve for the eigenvectors v:

$$\Sigma v = \lambda v \tag{1}$$

The weights w_{MCT} and w_{INT} are derived from the squared loadings of the first principal component (PC_1), normalized such that $w_{MCT} + w_{INT} = 1$. This ensures the index adaptively emphasizes the domain—kinetic or narrative—that is driving the system's variance at any given time.

$$GPTI_t = w_{MCT,t} \cdot MCT_t^{norm} + w_{INT,t} \cdot INT_t^{norm} \tag{2}$$

IV. Validation and Case Studies

The framework was validated against two distinct historical crises to assess its accuracy and generalizability.

A. Case Study A: India-Pakistan (2019)

The index successfully reconstructed the timeline of the February 2019 Pulwama-Balakot crisis. The dashboard visualized a distinct pattern:

- **Feb 14 (Pulwama Attack):** A sharp spike in the INT pillar, reflecting immediate media outrage.
- **Feb 26-27 (Balakot & Dogfight):** The peak of the GPTI, driven by a massive surge in the MCT pillar as the conflict transitioned to the kinetic domain.

B. Case Study B: Russia-Ukraine (2022)

To test generalizability, the model was applied to the Russia-Ukraine dyad. The GPTI correctly identified the "Troop Buildup" phase (Nov 2021–Feb 2022) as a period of high narrative tension but low kinetic activity. The invasion on February 24, 2022, was registered as a structural break, creating a "wall" of sustained kinetic intensity.

TABLE I
Comparative Analysis of Conflict Dyads

Dyad	Conflict Type	Dominant Pillar	Key Signal
India-Pakistan	Kinetic Pulse	Kinetic (MCT)	Sharp Spike
Russia-Ukraine	Hybrid Invasion	Narrative (INT)	Structural Break

V. Analytical Insights: Granger Causality

Beyond visualization, this research seeks to verify the causal pathways of escalation. We applied the Granger Causality test to investigate the lead-lag relationship between the Narrative and Kinetic pillars.

A. Statistical Results

We tested the null hypothesis (H_0) that Narrative Tension does not Granger-cause Kinetic Tension. The analysis yielded a statistically significant result:

$$p\text{-value} = 0.0294 \tag{3}$$

Since $p < 0.05$, we reject the null hypothesis. This provides statistical evidence that escalations in the information environment serve as a leading indicator for physical violence in the studied datasets. This validates the predictive utility of the INT pillar as an early-warning mechanism.

VI. Conclusion

GeoSentinel represents a significant advancement in computational political science. By fusing multi-domain data with an adaptive hybrid AI architecture, it provides policymakers with a validated, real-time instrument for conflict analysis. The successful application to both India-Pakistan and Russia-Ukraine demonstrates the framework's versatility, while the Granger Causality results offer empirical proof that in modern hybrid warfare, words are indeed a precursor to war.

Future iterations of GeoSentinel will aim to incorporate a third "Economic Pillar" tracking market volatility and expand the LLM gatekeeper to support multilingual datasets beyond English.

References

[1] D. Caldara and M. Iacoviello, "Measuring Geopolitical Risk," *American Economic Review*, vol. 112, no. 4, pp. 1194-1225, 2022.

[2] Raleigh, C., Linke, A., Hegre, H. and Karlsen, J., "Introducing ACLED: An Armed Conflict Location and Event Dataset," *Journal of Peace Research*, vol. 47, no. 5, pp. 651-660, 2010.

[3] Vaswani, A. et al., "Attention Is All You Need," *Advances in Neural Information Processing Systems*, 2017.

[4] Sanh, V. et al., "DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter," *arXiv preprint arXiv:1910.01108*, 2019.

[5] Granger, C. W. J., "Investigating Causal Relations by Econometric Models and Cross-spectral Methods," *Econometrica*, vol. 37, no. 3, pp. 424–438, 1969.