

Agentic Multilingual NLP for Conflict Forecasting from Open-Source Text Streams

Motivation. Timely forecasting of political unrest, protests, and human rights violations from open-source text remains a major challenge in NLP. Key difficulties include multilinguality (diverse linguistic and cultural contexts), regional variability (differences in reporting density and tone), and ambiguity (sparse or conflicting signals). Most existing systems are static classifiers that output point predictions without handling uncertainty, bias, or shifting language, making them fragile in high-stakes applications like humanitarian monitoring and early warning.

Proposed Approach. We propose an *agentic multilingual forecasting framework* that adapts to uncertainty by refining its predictions through iterative reasoning. The framework integrates three main innovations:

- 1 - Agentic reasoning loop: The system detects when its confidence is low, retrieves additional evidence, revises prompts, and updates predictions. This moves beyond static classification toward adaptive decision-making.
- 2 - Multilingual and region-aware pipeline: Region-specific modules adjust thresholds, retrieval scope, and prompt styles to account for linguistic diversity and differences in reporting quality. This design explicitly addresses global inequality in data coverage.
- 3 - Hybrid retrieval-forecasting model: Forecasting models are enhanced with semantically retrieved documents, combining short-term temporal features with contextual historical information.

Implementation. Our prototype uses publicly available data only, including GDELT, ReliefWeb, and Google News, with ACLED events for evaluation. We employ multilingual embeddings (MPNet, LaBSE), region tagging, and time-window aggregation. Retrieval is implemented with FAISS, and the reasoning loop with a LangChain-style agent. All outputs include rationales and an “ethics log” summarizing source diversity, bias checks, and uncertainty flags.

Progress and Preliminary Results. We have completed the data pipeline (multilingual text ingestion from GDELT, ReliefWeb, and Google News) and developed static multilingual baselines using MPNet embeddings and regional features. Initial evaluations across seven countries reveal substantial regional variation: minority F1 scores range from 0.07–0.55 depending on country and model. For instance, simple TF-IDF classifiers achieve up to 0.55 in Iran but struggle in Syria and Yemen (<0.2). These disparities underscore the limitations of static models and highlight the need for adaptive, region-aware approaches. A prototype agentic forecasting loop using FAISS retrieval and LangChain reasoning is under development, with ongoing experiments planned to compare its calibration, fairness, and robustness against baseline systems across multiple languages and regions.

[1] Mueller & Rauh (2018). Reading between the lines: Prediction of political violence using newspaper text. APSR.

[2] Nemkova, Poli A., et al. "Towards Automated Situation Awareness: A RAG-Based Framework for Peacebuilding Reports." arXiv preprint arXiv:2505.10586 (2025).

[3] von der Maase, Simon P. "Next-Generation Conflict Forecasting: Unleashing Predictive Patterns through Spatiotemporal Learning." arXiv preprint arXiv:2506.14817 (2025).