Deep Learning - Term paper

Nicolas Blache
Victor Garçonnot
Bruno Sciascia

April 2025

Introduction

Financial markets react rapidly to geopolitical tensions, yet existing measures like the Geopolitical Risk (GPR) index are global, low-frequency, and retrospective. In this project, we develop an automated weekly "Conflict-Score" that quantifies bilateral financial tensions by analyzing English article descriptions from Reuters and CNBC between 2017 and 2020. Our method combines named recognition, transformer-based sentiment classification with FinBert, and a weighting scheme to build a replicable and extensible indicator.

Research question

Can we anticipate week-to-week swings in geopolitical risk by tracking the volume-weighted share of negative financial news between country pairs, scaled by their trade exposure? Our goal is to assess whether this method can provide earlier and more granular signals of emerging tensions than traditional global indices.

Background section - Relevant & Related papers

Our project builds upon four main references that guide both our methodological choices and the construction of the Conflict-Score:

- **Gentzkow, Kelly, and Taddy (2019)**: Introduced text-as-data methods in economics by showing how unstructured text can be transformed into measurable economic signals. We follow this approach by treating news articles as an economic dataset to extract latent geopolitical tensions.
- **Caldara and Iacoviello (2022)**: Created the Geopolitical Risk (GPR) index based on newspaper coverage, highlighting the link between geopolitical events and market movements. Their work motivates our idea to

develop a similar conflict index, but at the pair-specific level rather than global.

- **Huang, Wang, and Yang (2023)**: Developed FinBERT, a transformer-based model fine-tuned on financial texts to better capture sentiment nuances. We leverage FinBERT to classify the tone of articles mentioning country pairs.
- **Van der Maaten and Hinton (2008)**: Proposed t-SNE, a visualization technique for high-dimensional data. We use t-SNE to explore and better understand the structure of the financial sentiment embeddings extracted from the news corpus.

Data set

Following the advice given during our Term Poster preparation, the analysis is built on the Financial-News-Headlines corpus released on Kaggle (dataset ID: notlucasp/financial-news-headlines). The file bundles every English headline carried by Reuters, CNBC, and The Guardian between 2017 and mid 2020, together with the short "description" that appears immediately under the headline on each website. After removing duplicates and discarding rows with missing time-stamps, the corpus contains about 33k news. The Guardian articles had no description so we dropped it. Each record provides a publication date , the headline text, the single-sentence lead, and a source tag, making the dataset compatible with our transformer pipeline.

Trade exposure comes from the CEPII Gravity database, which reports bilateral annual merchandise exports in current U.S. dollars for more than twenty-thousand country pairs from 1948 to 2020. We map the iso3_o and iso3_d columns to ISO-3 codes that match the output of the named-entity recogniser, take the mean of 2017-2020 flows to smooth idiosyncratic spikes, and rescale the resulting values linearly to the unit interval so that they serve as interpretable weights rather than dominate the score mechanically. The description table and the normalised trade matrix are the only two inputs used downstream: the former feeds spaCy's country extraction and FinBERT's sentiment model, while the latter supplies the pair-specific weight w_{ij} in the Conflict-Score formula.

Methodological Framework

Term Recognition

Country identification is based on spaCy's large English model. Each article description is passed through the named-entity recognizer, and tokens labeled as geopolitical entities (GPE) are mapped to ISO-3 country codes. To improve coverage, we cross-referenced spaCy's output with an external database of world capital cities and manually added a mapping of major political leaders (e.g., "Trump" → United States, "Macron" → France), since leaders are often mentioned without explicitly naming their country. Special cases such as "US" and "UK" were also manually corrected.

This function was applied to the article descriptions rather than headlines to maximize the detection of country pairs, which motivated the exclusion of The Guardian articles.

We also considered using a set of predefined conflict-related keywords, such as `conflict_terms = ["sanctions", "tariffs", "SWIFT ban", "debt default", "export controls", "currency manipulation", "asset freeze"]` then augmented it with embedding. However, relying solely on keyword labeling would have significantly reduced the number of identified country pairs, and would have made topic modeling and embedding methods ineffective.

Conflict-Score Construction

Each article description is assigned to an ordered country pair when exactly two distinct countries are detected. Sentiment is classified by FinBERT into positive, neutral, or negative categories.

For each country pair (i, j) and week t, the Conflict-Score is computed as:

$$\text{Score}_{ij,t} = \underbrace{\left(\frac{\text{Negative Articles}_t}{\text{Total Articles}_t}\right)}_{\text{Sentiment}} \times \underbrace{\log(\text{Total Articles}_t + 1)}_{\text{Media Attention}} \times \underbrace{\text{Trade Volume}_{ij}}_{\text{Economic Exposure}}$$

where:

- The **sentiment component** captures the proportion of articles classified as negative for a given country pair. To avoid extreme values in weeks with few articles, we apply Laplace smoothing: $n_{ij,t}^{\text{neg}} \to n_{ij,t}^{\text{neg}} + 1$ and $n_{ij,t} \to n_{ij,t} + 2$.
- The **media attention component** $\log(n_{ij,t} + 1)$ reflects the idea that market reactions intensify when a topic receives broader coverage.
- The **economic exposure component** is given by w_{ij}, the normalized trade weight between the two countries, ensuring that tensions involving major trading partners are weighted more heavily.

Finally, to reduce noise and highlight meaningful trends, a centered three-week moving average is applied to the weekly Conflict-Score series.

Topic Modeling and Embedding

To complement the Conflict-Score, we applied unsupervised topic modeling and embedding visualization techniques. Using the FinBERT embeddings extracted from article descriptions, we projected the high-dimensional vectors into two dimensions with t-SNE to explore the underlying semantic structure. In parallel, we performed Latent Dirichlet Allocation (LDA) on the descriptions to identify dominant themes across the corpus, helping to link clusters of tensions to specific economic or geopolitical narratives.

Empirical Results & interpretation

After applying our country-pair extraction pipeline, we successfully identified around 4,000 country pairs across 33,000 articles. For obvious reasons of focus and interpretability, we concentrated our analysis both on the overall network of country pairs and on two specific relationships:

- **United States–China (USA–CHN)**, which accounts for the most detected pairs.
- **Iran–United States (IRN–USA)**, which experienced key geopolitical events between 2017 and 2020.

It is important to note that, since we removed the Guardian articles (which only contained headlines) and given that Reuters articles represent only about 14% of the total corpus, our dataset has a strong U.S.-centric bias.

Below, we present the top five most frequently detected country pairs:

Country Pair	Number of Occurrences
CHN–USA	1767
JPN–USA	192
CHN–MEX	138
IRN–USA	136
CAN–USA	100

Table 1: Top 5 most represented country pairs extracted from the dataset.

Conflict Score

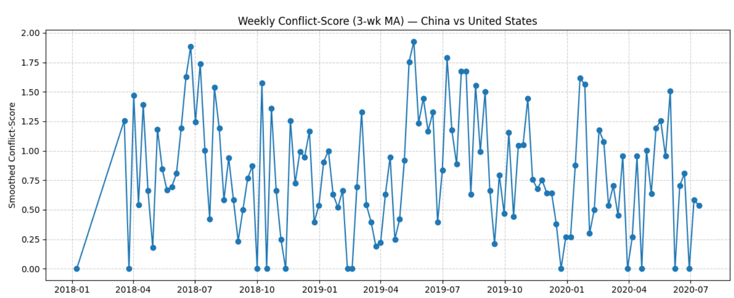

Figure 1: Conflict score of CHN-USA over time

The graph shows the weekly, three-week-smoothed Conflict-Score for the U.S.–China dyad[1] from January 2018 to July 2020. Two main patterns emerge.

[1]A dyad refers to a specific ordered pair of countries between which tensions are measured.

First, tension rises in distinct waves aligned with key trade war events. The initial surge starts in late March 2018, following Washington's Section 301 tariff memorandum, with the score climbing from near zero to 1.25 before collapsing after the G-20 truce in December 2018.

A second, sharper escalation appears from May to August 2019, when negotiations fail and tariff hikes escalate, pushing the Conflict-Score close to its maximum near 2.0. A smaller spike occurs in January 2020, coinciding with the "Phase-One" deal, after which the score temporarily declines.

These alternating spikes and collapses confirm that the weighting scheme effectively amplifies broad media coverage and suppresses noise from low-article weeks, avoiding the flat-line behavior seen in monthly GPR indices. The series also remains stationary over time, indicating that Laplace smoothing and trade normalization prevent mechanical drift. Overall, the Conflict-Score appears both interpretable—its peaks match meaningful confrontations—and sufficiently volatile for use in predictive regressions.

GPR Index

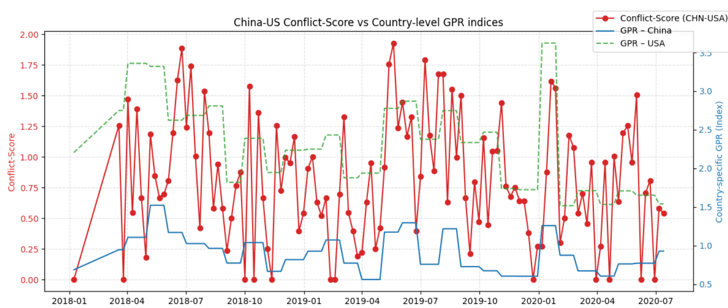

Figure 2: China and USA GPR index Vs Conflict Score

	Conflict–Score	GPRC–CHN	GPRC–USA
Conflict–Score	1.000	0.309	0.311
GPRC–CHN	0.309	1.000	0.799
GPRC–USA	0.311	0.799	1.000

Table 2: Pairwise Pearson correlations, weekly frequency, 2018–2020.

The country-level comparison confirms that the Conflict-Score behaves exactly as a bilateral tension barometer should. The China-specific series jumps most visibly in April 2018 during the first tariff volley, again in May–July 2019 when negotiations collapsed, and in January 2020 around the Phase-One signing—the Conflict-Score responds with an immediate burst of weekly spikes. The same pattern holds for the U.S. component: its surges line up with rises in the red series because both nations' newspapers are reporting on the very same confrontation.

Yet the Conflict-Score is not a simple replica of either index. Its amplitude is often greater, because it is scaled by contemporaneous headline volume and by the bilateral trade weight; and its week-to-week variation is largely independent, as shown by the modest correlations of 0.31 with $GPRC_{CHN}$ and $GPRC_{USA}$. Those numbers imply the our conflict score tracks the timing of the two national news indices while delivering a sharper, pair-specific signal that is essential for investors who care about exposure to Sino-American conflict rather than to the entire spectrum of geopolitical events captured by each country's press.

Topic Modeling and Embedding Visualization

We believe that topic modeling can be a powerful tool to track growing trends in geopolitical tensions even if we had difficulties making it really useful in our Term Paper. Initially, we attempted to replicate the type of visualization shown in the language modeling slides, by projecting the most important TF-IDF-weighted words into a 3D space using PCA followed by t-SNE. Although this method produced a rough semantic mapping, the absence of dynamic weighting across topics limited its interpretability.

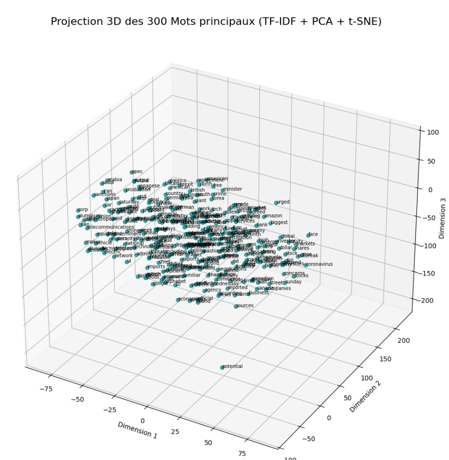

Figure 3: 300 most important words

We found a clearer structure by focusing on individual country pairs. Specifically, we selected the ten most relevant words for a given dyad, embedded them using a SentenceTransformer model, normalized the embeddings, and reduced their dimensions via PCA. This approach yielded much more intuitive 3D projections, where key thematic clusters (such as "oil", "sanction", "export" for

IRN-USA) became immediately visible and aligned with the underlying economic and geopolitical narratives.

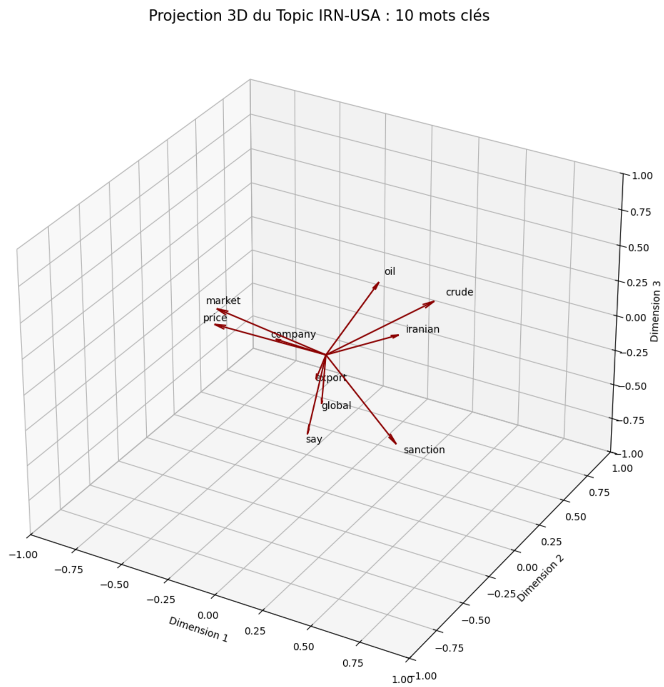

Figure 4: Top 10 3D topics IRN-USA

Suggestion for further research

Limitations

Our project faced two main limitations:

First, our dataset was relatively limited: about 33,000 articles from CNBC and Reuters, all before the COVID-19 pandemic. A broader and more recent dataset would have improved the diversity of our analysis. Additionally, computational constraints slowed down key tasks like country-pair extraction and sentiment classification, which sometimes required over 40 minutes to execute, forcing us to simplify parts of the pipeline.

Second, while constructing a textual-based tension score is an appealing idea, a deeper conceptual limitation remains: can such a score truly provide early warning signals before conflicts become widely recognized? For instance, if we had applied our method this year, could we have detected the growing tensions around tariffs between the U.S., China and the global economy several months before they became a major global issue during Trump's election campaign? Assessing the predictive power of our score relative to major geopolitical developments would require further validation and historical backtesting.

Further research and implementations

Building on the limitations identified above, future work could focus on applying our methodology to a larger and more diversified real-time data stream, with less emphasis on U.S.-centric news sources. Expanding the geographical and temporal coverage would provide a more global perspective on emerging financial tensions.

Moreover, our project shows that a real-time conflict perception score, derived from textual analysis, could become a valuable decision-support tool. Major asset managers such as BlackRock[2] and Amundi[3] already monitor geopolitical risks systematically. Similarly, trading desks could benefit from a dynamic and automated framework capable of detecting underlying tensions that might otherwise go unnoticed, providing timely signals beyond conventional news tracking services like Bloomberg.

[2] https://www.blackrock.com/corporate/insights/blackrock-investment-institute/interactive-charts/geopolitical-risk-dashboard

[3] https://research-center.amundi.com/article/geopolitical-risk-will-grow-here-how-we-track-it