An NLP Approach to Weaponization of Migration Flows

Baris Alan
Tulane University
balan@tulane.edu

April 2025

Abstract

1 Introduction

In the International Relations (IR) literature, conventional wisdom holds that stronger nations tend to prevail in asymmetric conflicts. Powerful states primarily leverage their military and economic resources to coerce or deter weaker actors. In contrast, weaker nations are typically viewed as vulnerable, co-opted, or subject to external pressure. However, recent developments in international migration have created new strategic opportunities for these weaker states. When migrants are unable to reach developed nations and instead become stranded in transit countries—such as Turkey, Egypt, Libya, and Mexico—these transit states can exploit migrant populations as bargaining chips against destination countries. By (1)**threatening** to open their borders and send migrants or (2)**criticizing** the recipient countries based on their anti-migration policies—a tactic referred to here as the *weaponization of migration flows (WOM)*—transit nations have successfully extracted billions of dollars in financial aid from entities such as the European Union and the United States.

Despite its increasing prevalence, the existing political science literature lacks a systematic tool to quantitatively measure or analyze threats related to migration flows. While qualitative case studies offer valuable insights into specific successful cases, they fall short of providing a comprehensive, longitudinal understanding of this tactic.

To address this gap, the current project focuses on a foundational question: **How can we operationalize WOM?** This study aims to develop a quantitative measure of WOM, using Turkey as a pilot case. Specifically I will webscrape all the speeches given by Turkish President since 2014, and then create a random sampling of these speeches to manually label them. Then, I aim to apply (1) a Logistic Regression with TF-IDF and (2) a Bert-based fine-tuned transformer and evaluate the performances of these two models. The model with the a satisfactory performance will be used to predict the classes of remaining texts, and also will be applied to other cases. Hence, the resulting measure will serve as the dependent variable in the broader research agenda, which investigates how migration network centrality influences the likelihood and intensity of WOM.

2 Literature Review

WOM is a relatively new area of research, and most scholars have adopted qualitative methods to study it.

Greenhill (2010) is a pioneer in the field, using comparative case studies and process tracing to identify over 60 WOM cases since 1951. Her work relies on archival evidence and qualitative methods. Tsourapas (2018) uses a least-likely case study design to show how Jordan and Libya threatened to expel Egyptian workers to coerce the Egyptian government. He supports his claims using newspaper articles and elite interviews. Tsourapas & Malit (2021) adopt a similar approach to examine how the Philippine government threatened Gulf states (Qatar and UAE) to relax migration policies in favor of Filipino workers. Petty (2022) and Baser (2022) also rely on qualitative case studies to explore how migration has become a bargaining chip for weaker states in asymmetric conflicts. To the best of the researcher's knowledge, existing WOM literature has not used NLP methods to systematically or quantitatively analyze the textual evidence of threats. As a result, most studies focus on handpicked **successful** case

studies, which risks the methodological pitfall of "selecting on the dependent variable"—akin to p-hacking in quantitative research.

This project will address that gap by applying a Logistic Regression Model and a fine-tuned Transformer to a comprehensive set of public speeches by Turkish President. The goal is to create a quantitative measure of when, how many times, and how often Turkish leader uses a threatening or criticizing rhetoric around migration flows.

3 Data

Since WOM is operationalized in this study as (i) *threatening* to open borders and facilitate migrant flows, and (ii) *criticizing* recipient countries' anti-migration policies, this project will analyze public speeches by Turkish President Erdoğan.

The speeches by Turkish President are available on the website of Turkish presidency. All available speeches delivered between September 1, 2014 and April 29, 2025 were web-scraped using Python's `BeautifulSoup` library to retrieve the title, date, URL, and full transcript. Each transcript was then segmented into paragraphs under the assumption that paragraph boundaries typically coincide with topical shifts. During preprocessing, the following filters are applied:

- **Title filter:** Exclude speeches whose titles contain the keywords "visit" or "phone call," as these entries generally consist of one-sentence event notices (e.g., "The President visited [Country X].").
- **Paragraph filter:** Discard any paragraph that (a) ends with a comma, or (b) contains fewer than four words, since such segments are usually salutations (e.g., "My dear nation," or "Thank you, brothers!").

After applying these criteria, the final corpus comprises 3,976 paragraphs (from 242 distinct public speeches) which are the unit of analysis of this project. Turkish case will serve as a pilot study, and in the broader research agenda, the goal is to expand this approach by scraping similar documents from other transit migration countries.

4 Method

4.1 Manual Annotation of the Corpus

The original corpus is entirely unlabeled, necessitating a manual annotation process to generate training data. I began by drawing a stratified random sample representing approximately 10% of the full dataset: four paragraphs were randomly selected from each month–year bin, yielding 472 paragraphs. Then I labeled each paragraph as either WOM-positive (i.e., conveying a threat to open borders or criticism of migration policies) or WOM-negative. This initial pass resulted in only 16 WOM-positive cases (3.4%), an extreme class imbalance that would undermine supervised model training.

To enrich the positive class, I applied `facebook/bart-large-mnli` pretrained model's zero-shot classification approach. Every paragraph in the full corpus was assigned to one of six provisional categories: *"explicitely threatening to send migrants", "implicitly threatening to send migrants", "criticizing based on migration policy","criticizing based on migration", "cooperative migration discourse", "irrelevant/neutral"*. I then manually reviewed all paragraphs labeled as one of the four threat/criticism categories, and labeled an additional 338 WOM-positive instances. From the remainder paragraphs (i.e., paragraphs not flagged as threats or criticisms by the zero-shot classification model), I randomly sampled 500 as WOM-negative examples. The resulting hand-coded dataset comprises the final dataset comprised:

- **Total examples:** 838
- **Non-threat (class 0):** 500 (59.7 %)
- **Threat (class 1):** 338 (40.3 %)

4.2 Classification Methods

I implemented and compared two supervised classification approaches: a traditional Logistic Regression baseline and a fine-tuned transformer-based model. Both methods employ a multi-stage NLP pipeline comprising text preprocessing, feature extraction, and classifier training.

Baseline Model

The baseline classifier is a Logistic Regression trained on TF–IDF features. Text preprocessing is performed with spaCy and NLTK libraries, including tokenization, lemmatization, stop-word removal, and named-entity recognition. I reserved 20% of the corpus as a true hold-out set for final evaluation, using the remaining 80% for model development and training. Feature extraction uses `TfidfVectorizer` with grid-searched hyperparameters to down-weight very frequent tokens and up-weight rare ones. To mitigate class imbalance, SMOTE is applied so that threat and non-threat classes are equally represented during training. Model performance is assessed via the traditional evaluation metrics (accuracy, precision, recall, F1-score) and a stratified 5-fold cross-validation on the development set and then validated on the hold-out set to obtain an unbiased estimate of real-world performance.

Advanced Model

My approach involves fine-tuning BERT model using the HuggingFace's Transformers library to classify paragraphs based on contextual understanding. This advanced classifier model includes a multi-step Natural Language Processing (NLP) pipeline. Preprocessing parallels the baseline pipeline (spaCy lemmatization, NLTK stop-word filtering), followed by BERT tokenization. I again split off 20% of the data as a hold-out set and fine-tune on the remaining 80% for three epochs with a learning rate of 2×10^{-5}. Evaluation is conducted both on the hold-out set and via stratified 5-fold cross-validation over the full dataset, using accuracy, precision, recall, and F_1 as performance metrics.

4.3 Evaluation

After model training, I assess performance using four standard classification metrics: accuracy, precision, recall, and F_1-score. Each metric captures a different aspect of prediction quality:

Accuracy The proportion of all instances (both threat and non-threat) that the model correctly classifies. While easy to interpret, accuracy can be misleading on imbalanced datasets.

Precision The fraction of predicted *threat* cases that are truly threats. High precision indicates few false positives, i.e., that the model does not frequently mislabel non-threats as threats.

Recall The fraction of actual *threat* cases that the model successfully identifies. Also known as sensitivity, high recall means the model misses few true threats (few false negatives).

F_1-score The harmonic mean of precision and recall:

$$F_1 = 2 \times \frac{Precision \times Recall}{Precision + Recall}.$$

This single measure balances the trade-off between precision and recall, and is particularly useful when the positive class (threat) is of primary interest.

These metrics are computed both on the held-out test set and via 5-fold cross-validation to provide a comprehensive picture of each model's generalization and stability.

5 Experiment Results

5.1 Evaluation Metrics

Table 1 reports the hold-out set performance for our two model variants: a Logistic Regression baseline built on TF–IDF features, and an "Advanced BERT" classifier fine-tuned end-to-end on contextualized embeddings.

Table 1: Hold-out Set Performance

Model	Accuracy	Precision	Recall	F_1
Logistic Regression with TF–IDF	0.7143	0.7143	0.7224	0.7117
Advanced BERT Model	0.8095	0.7195	0.8676	0.7867

Accuracy and Overall Error Reduction On the hold-out set, the Advanced BERT model attains an accuracy of 80.95%, compared to 71.43% for the Logistic Regression baseline. This 9.5 percentage-point improvement corresponds to a relative reduction in misclassification rate of over 27%, underscoring that the transformer's deep, contextual representations generalize substantially better to unseen examples.

Precision vs. Recall Trade-off While both models exhibit similar precision (71.43% vs. 71.95%), the real benefit of BERT emerges in recall: the baseline only correctly retrieves 72.24% of true "threat" instances, whereas BERT captures 86.76%. In practical threat-detection scenarios, raising recall by over 14 points dramatically reduces false negatives—i.e., dangerous content slipping through undetected—at only a very slight cost to precision.

Balanced Performance ($\mathbf{F}_1$) The F_1-score, which harmonizes precision and recall, climbs from 71.17% to 78.67%. This 7.5 point uplift confirms that BERT does not merely specialize in one metric at the expense of others, but delivers a robust improvement in the overall balance between correctly flagging threats and avoiding over-alerting.

Implications for Real-World Deployment These results suggest that leveraging a pre-trained language model is particularly advantageous when the cost of missing a threat is high. The substantial recall gain with BERT means far fewer critical instances will be overlooked, which is vital for applications such as automated content moderation, security monitoring, or early-warning systems. At the same time, the modest change in precision ensures I do not substantially increase the burden of handling false alarms. In sum, the Advanced BERT model's contextual understanding translates into markedly stronger generalization on novel data—a key requirement for reliable, real-time threat detection.

5.2 Cross-Validation Analysis

Table 2 presents the 5-fold cross-validation accuracies and their summary statistics for the Advanced BERT model.

Table 2: Cross-Validation Results

Fold	Accuracy
1	0.8690
2	0.8155
3	0.8452
4	0.8802
5	0.8024
Mean ± Std	0.8425 ± 0.0299

Fold-wise Stability The per-fold accuracies range from 0.8024 to 0.8802, indicating consistently strong performance across different data splits. Fold 4 yields the highest accuracy (0.8802), while Fold 5 is the lowest (0.8024), but even the weakest fold exceeds 80%.

Average Performance and Variability With a mean accuracy of 84.25% and a standard deviation of just 2.99 percentage points, the small spread demonstrates that the model's performance is robust to sampling variation. A standard deviation under 3% suggests reliable generalization and low sensitivity to which data are held out in each fold.

Implications for Model Generalization The combination of high average accuracy and low inter-fold variability confirms that the Advanced BERT model achieves stable and reproducible results. This stability reinforces confidence that the hold-out performance is representative of expected behavior on new, unseen data.

6 Conclusion

In this project, I developed and evaluated a scalable NLP pipeline to operationalize the weaponization of migration flows (WOM) by analyzing Turkish presidential speeches from 2014 to 2025. I first hand-annotated (after an initial zero-shot classification using Facebook's `bart-large-mnli` pretrained model) a balanced sample of 838 paragraphs (338 WOM-positive, 500 WOM-negative) and then compared two supervised classifiers:

- A *Logistic Regression* baseline trained on TF–IDF features (plus spaCy lemmatization, NLTK stop-word filtering, and SMOTE for class balancing), which achieved 71.43% accuracy, 71.43% precision, 72.24% recall, and 71.17% F_1 on the hold-out set.
- An *Advanced BERT* model (DistilBERT fine-tuned with the HuggingFace Trainer), which attained 80.95% accuracy, 71.95% precision, 86.76% recall, and 78.67% F_1 on the same hold-out set.

Moreover, 5-fold cross-validation of the BERT model yielded a mean accuracy of 84.25% ($\pm$ 2.99%), demonstrating both high performance and low variance across different data splits. The transformer-based classifier substantially outperformed the TF–IDF baseline—most notably in recall—indicating its superior ability to detect WOM rhetoric while maintaining precision.

By converting qualitative WOM indicators into a robust, quantitative measure, the pipeline addresses a critical methodological gap in the International Relations literature. The resulting WOM scores can serve as a dependent variable in downstream empirical analyses, for example to test how network centrality or domestic political factors influence a state's propensity to weaponize migration.

Building on these promising results, the broader research agenda will:

- **Expand the corpus** to include speeches and statements from other transit countries (e.g., Libya, Egypt, Mexico) and in multiple languages.
- **Incorporate temporal and network features** (e.g., migration flows, international bargaining events) to model the dynamics and external determinants of WOM.
- **Explore alternative architectures** (e.g., multilingual BERT, RoBERTa) and multi-task objectives (e.g., joint threat detection and stance classification) to further enhance robustness.

Overall, my findings demonstrate that transformer-based models offer a powerful and generalizable tool for the systematic study of coercive migration tactics, paving the way for large-scale quantitative investigations into how weaker states employ migration as a diplomatic lever.

7 Screenshot Demo

To visually demonstrate the full execution pipeline, I have included screenshots taken from the Jupyter notebooks `wom_data_webscraping.ipynb`, `wom_zeroshot_classification.ipynb`, and `wom_nlp_model.ipynb`. These figures illustrate the workflow from data acquisition through model evaluation.

The first step involves web-scraping the public speeches delivered by the President of Turkey.The `f_scrape_tr_speech` function shown in Figure 1 iterates over each page of Erdoğan's English speech listings, applies a regex filter to skip "visit" or "phone call" entries, and then uses BeautifulSoup to extract and clean only substantive paragraphs (dropping greetings and fragments under four words). Finally, it aggregates each paragraph's date, title, link, and text into a pandas DataFrame for downstream NLP processing.

Figure 2 demonstrates the zero-shot classification step, where each paragraph in the scraped corpus is automatically assigned to one of six semantic categories—ranging from "explicitly threatening to send migrants" to "neutral or irrelevant"—using the facebook/bart-large-mnli model. This automated annotation yields a set of candidate threat and criticism segments that are then manually reviewed to enrich the positive WOM examples.

```python
# Turkey Speeches in English (22 seconds)
df_tr_speech_en = f_scrape_tr_speech("/en/receptayyiperdogan/speeches/")
#df_tr_articles_en.to_csv("../data/raw/tr_articles_en.csv", index=False, encoding="utf-8-sig")
df_tr_speech_en.head()
```

```
Scraping page 1 → https://www.tccb.gov.tr/en/receptayyiperdogan/speeches/
Scraping page 2 → https://www.tccb.gov.tr/en/receptayyiperdogan/speeches/?&page=2
Scraping page 3 → https://www.tccb.gov.tr/en/receptayyiperdogan/speeches/?&page=3
Scraping page 4 → https://www.tccb.gov.tr/en/receptayyiperdogan/speeches/?&page=4
Scraping page 5 → https://www.tccb.gov.tr/en/receptayyiperdogan/speeches/?&page=5
Scraping page 6 → https://www.tccb.gov.tr/en/receptayyiperdogan/speeches/?&page=6
Scraping page 7 → https://www.tccb.gov.tr/en/receptayyiperdogan/speeches/?&page=7
Scraping page 8 → https://www.tccb.gov.tr/en/receptayyiperdogan/speeches/?&page=8
Scraping page 9 → https://www.tccb.gov.tr/en/receptayyiperdogan/speeches/?&page=9
Scraping page 10 → https://www.tccb.gov.tr/en/receptayyiperdogan/speeches/?&page=10
Scraping page 11 → https://www.tccb.gov.tr/en/receptayyiperdogan/speeches/?&page=11
Scraping page 12 → https://www.tccb.gov.tr/en/receptayyiperdogan/speeches/?&page=12
Scraping page 13 → https://www.tccb.gov.tr/en/receptayyiperdogan/speeches/?&page=13
Scraping page 14 → https://www.tccb.gov.tr/en/receptayyiperdogan/speeches/?&page=14
Scraping page 15 → https://www.tccb.gov.tr/en/receptayyiperdogan/speeches/?&page=15
Scraping page 16 → https://www.tccb.gov.tr/en/receptayyiperdogan/speeches/?&page=16
Scraping page 17 → https://www.tccb.gov.tr/en/receptayyiperdogan/speeches/?&page=17
Scraping page 18 → https://www.tccb.gov.tr/en/receptayyiperdogan/speeches/?&page=18
```

	date	link	title	text
0	29.04.2025	https://www.tccb.gov.tr/en/speeches-statements...	President Erdoğan to Visit Italy	President Recep Tayyip Erdoğan will pay an off...
1	29.04.2025	https://www.tccb.gov.tr/en/speeches-statements...	President Erdoğan to Visit Italy	At the Summit to be co-chaired by President Er...
2	29.04.2025	https://www.tccb.gov.tr/en/speeches-statements...	President Erdoğan to Visit Italy	Various documents, which will strengthen the c...
3	20.04.2025	https://www.tccb.gov.tr/en/speeches-statements...	President Erdoğan's Message on Easter	Following is the message President Recep Tayyi...
4	20.04.2025	https://www.tccb.gov.tr/en/speeches-statements...	President Erdoğan's Message on Easter	"I wish happy Easter to all our Christian citi...

Figure 1: Web-scraping script

```python
# Apply Function Zeroshot Classification for English

model_names = ['facebook/bart-large-mnli']
labels = [
    "explicitely threatening to send migrants",
    "implicitly threatening to send migrants",
    "criticizing based on migration policy",
    "criticizing based on migration",
    "cooperative migration discourse",
    "neutral or irrelevant",
]
batch_size = 32

for model in model_names:
    df_results = f_zeroshot_threat_tr(df_tr_speech_en, "text", model, labels, batch_size=batch_size)
```

```
Device set to use mps:0
Classifying texts in batches: 100%|██████████| 126/126 [19:13<00:00,  9.15s/it]
```

Figure 2: Zero-Shot Classification script

Figure 3 illustrates the classification performance of the Logistic Regression pipeline at each stage. On the training set, the model achieves 86.42% accuracy (with F_1 scores of 0.8826 for the non-threat class and 0.8389 for the threat class), which declines to 71.43% accuracy on the hold-out set (F_1=0.7391 for non-threat and 0.6842 for threat). The out-of-fold analysis further identifies "important" and "allah" as the most predictive TF–IDF features for the non-threat class, as evidenced by their largest negative coefficients.

Figure 4 shows that fine-tuned DistilBERT classifier delivers substantially stronger results than the TF–IDF baseline. On the hold-out test set, it achieves an accuracy of 80.95%, precision of 71.95%, recall of 86.76%, and an F_1-score of 78.67%. Furthermore, in stratified 5-fold cross-validation over the full dataset, it attains a mean accuracy of 84.25% with a standard deviation of just 2.99%, alongside an average precision of 77.33%, recall of 86.68%, and F_1 of 81.68%, demonstrating both high overall performance and low variability across folds.

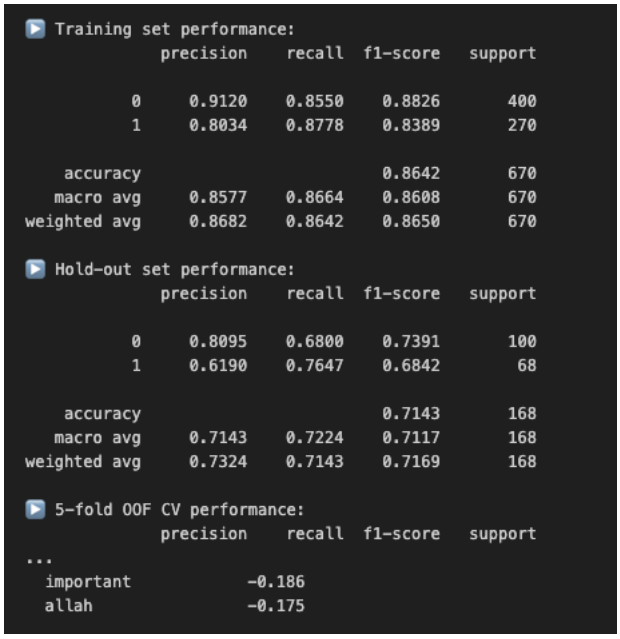

Figure 3: Classification performance of Logistic Regression

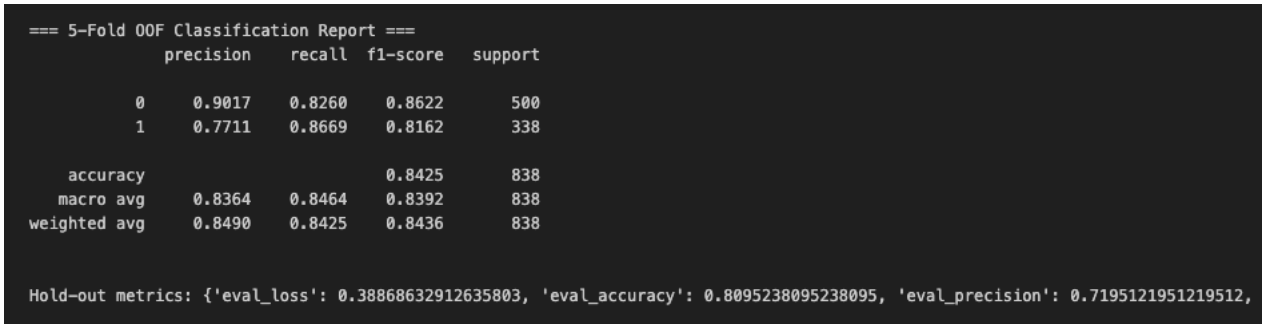

Figure 4: Classification performance of BERT Model

References

- Baser, Selin. 2022. "The Most Insidious Weapon of the Changing World: Migration." *Bilge Strateji*, 13(24): 167–185. https://doi.org/10.35705/bs.1198447
- Greenhill, Kelly. 2010. *Weapons of Mass Migration: Forced Displacement, Coercion and Foreign Policy*. Ithaca and London: Cornell University Press.
- Malit, Froilan T., and Gerasimos Tsourapas. 2021. "Weapons of the Weak? South–South Migration and Power Politics in the Philippines–GCC Corridor." *Global Studies Quarterly*, 1(3): ksab010. https://doi.org/10.1093/isagsq/ksab010
- Tsourapas, Gerasimos. 2018. "Labor Migrants as Political Leverage: Migration Interdependence and Coercion in the Mediterranean." *International Studies Quarterly*, 62(2): 383–395. https://doi.org/10.1093/isq/sqx088
- Petty, Aaron R. 2022. "Migrants as a Weapons System." *Journal of National Security Law and Policy*, 13: 113–139.