Propaganda Detection in Arabic Narratives

Course: AIS411 - Natural Language Processing

Date: January 3, 2026

Team Members: Ahmed Abdelmoneim

Introduction & Objective

- **Problem:** The spread of propaganda in Arabic social media, specifically regarding the war on Gaza (2023–2025), is a significant challenge.
- **Objective:** Develop an NLP pipeline to automatically detect and classify text as "Propaganda" or "Non-Propaganda."
- **Goal:** Build an accurate model that handles real-world challenges like noisy data and class imbalance.

Data & Preprocessing (Phase 1)

- **Dataset:** FigNews 2024 (CAMeL Lab).
  - 6,342 Arabic samples.
- **Preprocessing Pipeline:**
  - **Cleaning:** Removed URLs, English, emojis, and numbers.
  - **Fixing:** Stripped invisible Unicode control characters that caused visualization errors.
- **Exploratory Data Analysis (EDA):**
  - **Significant Imbalance:** 65% Propaganda vs. 35% Non-Propaganda.
  - **Word Cloud:** Frequent terms included "Gaza," "Hamas," and "Occupation."

Methodology – Comparative Study (Phase 2)

Requirement:

Compare different modeling approaches.

Model A: Baseline (Classical ML)

- Logistic Regression with TF-IDF features.
- Uses class_weight='balanced' to handle imbalance.

Model B: Deep Learning (AraBERT)

- Pre-trained Arabic Transformer (bert-base-arabertv02).
- Fine-tuned on an RTX 3050 GPU using Mixed Precision.

Initial Results – The "Laziness" Problem

AraBERT (Base):

Achieved high **Accuracy (65.3%)** but a low **F1-Macro (0.508)**.

- **Why?** The model learned to guess the majority class ("Propaganda") almost exclusively. It was "lazy."

Baseline (LogReg):

Achieved a higher **F1-Macro (0.526)**.

- **Why?** The explicit class weighting forced it to pay attention to the minority class.

Conclusion: Deep Learning requires optimization to handle imbalanced data effectively.

Optimization Strategy (Phase 3)

- **Solution:** Implement **Cost-Sensitive Learning** via **Weighted Loss**.
- **Mechanism:** Penalize the model more for making mistakes on the minority class ("Non-Propaganda").
- **Calculated Weights:**
  - **Non-Propaganda:** 1.45 (High penalty)
  - **Propaganda:** 0.76 (Low penalty)
- We implemented a custom WeightedTrainer to inject these penalties into the loss function during training.

Final Results & Comparison

The optimization was successful. AraBERT now outperforms the baseline.

Metric	Baseline (LogReg)	AraBERT (Optimized)	Improvement
F1 Macro	0.526	0.545	+3.6%
Accuracy	54.5%	58.6%	+7.5%

Key Takeaway: The optimized model is "fairer" and a better overall classifier, even if its raw accuracy is lower than the initial "lazy" model.

Bonus – Interactive Demo

- We deployed the final model using a **Gradio** web interface for real-time testing.
- **Test Case:**
  - **Input:** "The brutal Zionist enemy commits savage massacres..." (العدو الصهيوني الغاشم...)
  - **Prediction: Propaganda (56%)**
- This demonstrates the model's ability to detect emotionally loaded language, a key characteristic of propaganda.

Conclusion

- Successfully built an end-to-end NLP pipeline for Arabic propaganda detection.
- Demonstrated that **Cost-Sensitive Learning** is crucial for training deep learning models on imbalanced datasets.
- The final **AraBERT** model achieves an F1-Macro score of 0.545, surpassing the statistical baseline.
- **Future Work:** Explore advanced techniques like data augmentation or ensemble methods to further improve performance.

Q&A

Thank You!

Questions?