A2NLP at StanceNakba Shared Task: Fine-Tuned AraBERT for Topic-Based Arabic Stance Detection

Alaa Nairat, Aysar Nairat
Birzeit University, Modern University College
Ramallah, Palestine
anairat@birzeit.edu, aysar.nairat@muc.edu.ps

Abstract

This paper describes A2NLP's system for Subtask B of the StanceNakba Shared Task, which addresses cross-topic Arabic stance detection. The goal is to classify sentence–topic pairs into pro, against, or neutral labels. We introduce a topic-conditioned prompting strategy built on `AraBERTv0.2-Twitter`, where each instance is reformulated into a structured prompt that explicitly models the interaction between the sentence and its target topic. The model is trained using 5-fold stratified cross-validation with class-weighted loss to ensure robustness under mild label imbalance. Our final submission achieves a Macro-F1 score of 0.8483 on the official test set, outperforming the `AraBERTv2` baseline (0.810) and ranking fifth overall. Ablation analysis confirms that topic-conditioned prompting substantially improves generalization across topics. The findings demonstrate the importance of structured input design and domain-aligned pretraining for reliable stance detection in dialectal Arabic social media discourse.

Keywords: Arabic NLP, Stance detection, transformer, topic-conditioned prompts, text classification

1. Introduction

Understanding public opinion in online discourse is crucial for social, political, and health-related decision-making, yet manually tracking nuanced stances at scale is infeasible. Stance detection is the automated task of determining whether the author of a text expresses a position *in favor of*, *against*, or *neutral* toward a specific subject of interest, commonly referred to as a *target* (Küçük and Fazli, 2020; AlDayel and Magdy, 2021). Unlike sentiment analysis, stance detection is inherently target-dependent and often requires inferring implicit opinions that are not explicitly stated. This capability is essential for applications such as social media monitoring, political analysis, and public health surveillance, where nuanced public positions are critical and traditional survey-based methods are insufficient (Küçük and Fazli, 2020).

The *StanceNakba Shared Task* 2026 (Aldous et al., 2026), co-located with LREC 2026, provides a challenging benchmark for Arabic stance detection. In particular, **Subtask B** focuses on *Cross-Topic Stance Detection*, requiring a unified model capable of predicting stance across multiple conflict-related topics. This setup reflects realistic deployment scenarios in which systems must generalize beyond a single topic and handle emerging or unseen targets. Arabic, spoken by over 420 million people, remains underrepresented in stance detection research (Guellil et al., 2021; Qu et al., 2024) and poses unique Natural Language Processing (NLP) challenges, including rich morphology, extensive dialectal variation, and informal social media writing with spelling inconsistencies, abbreviations, and emojis (Badaro et al., 2019).

Our Team (**A2NLP**) proposes a prompt-based `AraBERTv0.2-Twitter` model for topic-conditioned Arabic stance detection. Our approach outperforms the official baseline in terms of Macro-F1. Our main contributions are: (1) **Input Reformulation:** Topic-conditioned prompting to model target–text interactions; (2) **Robust Training:** 5-fold stratified cross-validation with class-weighted loss; and (3) **Public Resources:** Public model weights and an interactive *Hugging Face Space* demo. The code implementation can be reached via A2NLP-Code[1], and Interactive demo can be reached via A2NLP-SharedTask-Demo[2]. The remainder of this paper is organized as follows. Section 2 reviews related work. Section 3 describes the task and dataset. Section 4 presents our system. Section 5 outlines the experimental setup. Section 6 reports results and error analysis. Section 7 concludes and highlights future directions.

2. Related Work

2.1. Evolution of Arabic Stance Resources

Although Arabic stance detection has progressed in recent years, most early research was dominated by English datasets (Küçük and Fazli, 2020; AlDayel and Magdy, 2021; Guellil et al., 2021). Several Arabic resources have since

[1] https://github.com/eng-aomar/A2NLP-StanceNakba2026-SubtaskB-AraBERTv02-Twitter

[2] https://huggingface.co/spaces/aomar85/A2NLP-StanceNakba-Demo

Proceedings of the 2nd International Workshop on Nakba Narratives as Language Resources (Nakba-NLP 2026) @ LREC 2026, pages 147–159

emerged. MAWQIF introduced $4,121$ tweets annotated for stance and related phenomena (Alturayeif et al., 2022). AraStance provided $4,063$ claim–article pairs for fact-checking-oriented stance detection (Alhindi et al., 2021). ArabicStanceX expanded coverage to $14,477$ tweets across 17 topics, enabling multi-topic and zero-/few-shot settings (Alkhathlan et al., 2025). MARASTA further incorporated cross-domain and multi-dialectal variation (Charfi et al., 2024b). Despite these advances, robust cross-topic generalization in dialectal social media remains challenging.

2.2. Computational Methodologies

Early systems relied on traditional classifiers such as SVMs with TF-IDF and n-gram features (Darwish et al., 2017; Baly et al., 2018), later replaced by CNN and LSTM architectures. More recently, Multi-Task Learning (MTL) jointly modeled stance with related tasks (e.g., sentiment, sarcasm), improving shared representations and class balance (Alturayeif et al., 2023). Sequential MTL with hierarchical weighting (SMTL-HW) achieved 88.3% F1, highlighting the benefit of task interaction (Almasoudi et al., 2025).

2.3. Transformer Models

Transformer-based models now dominate Arabic NLP (Antoun et al., 2020). `AraBERTv0.2-Twitter`, pre-trained on 60M Arabic tweets, has shown strong performance in social media stance detection (Antoun et al., 2020; Badran et al., 2024). It underpinned the winning StanceEval 2024 system, enhanced with contrastive learning (Badran et al., 2024), and consistently outperforms MARBERT and CAMeLBERT-da on MAWQIF (Alturayeif et al., 2022, 2024). However, existing work primarily relies on standard fine-tuning, leaving input reformulation and topic-conditioned modeling underexplored.

3. Background

3.1. Task Setup

Subtask B of the StanceNakba 2026 Shared Task (Aldous et al., 2026) addresses cross-topic Arabic stance detection. Each instance consists of a sentence paired with one of two predefined target topics: وجود اللاجئين والمهاجرين في الأردن (من فلسطين، سوريا، العراق) Refugee/Immigrant Presence in Jordan from (Palestine, Syria and Iraq) or التطبيع مع اسرائيل (Normalization with Israel). The task is formulated as sentence pair classification problem with three labels, where the system predicts the stance expressed in the sentence toward the specified topic. The output label belongs to one of three categories: **pro**, **against**, or **neutral**. For example, given the sentence افضل شعب الشعب السوري اختكم من الأردن (The Syrian people are the best people, your sister from Jordan) and the topic وجود اللاجئين والمهاجرين في الأردن (من فلسطين، سوريا، العراق) (Refugee/Immigrant Presence in Jordan from (Palestine, Syria and Iraq)), the appropriate label is **pro**, as the sentence conveys support in relation to the target topic.

3.2. Dataset Details

The dataset used in this subtask is a subset of the MARASTA corpus (Charfi et al., 2024a), consisting of Arabic sentence–topic pairs primarily written in colloquial Arabic and regional dialects such as Palestinian, Jordanian, and Qatari. The genre mainly comprises tweets and short online posts, characterized by informal language, abbreviations, and frequent use of emojis.

The dataset contains $1,024$ instances in total. As shown in Table 1, the label distribution is approximately balanced across the three stance categories: *against* (361 instances, 35.25%), *pro* (348 instances, 33.98%), and *neutral* (315 instances, 30.76%). The prevalence of dialectal variation and informal expressions increases the complexity of the stance detection task, requiring models to generalize across linguistic and stylistic variability.

Topic / Label	Count	Percentage (%)
(التطبيع مع اسرائيل) (Normalization with Israel)		
Against	168	34.22
Neutral	177	36.05
Pro	146	29.73
Sub-total	491	
وجود اللاجئين والمهاجرين (فلسطين، سوريا، العراق) Refugee/Immigrant Presence in Jordan from (Palestine, Syria and Iraq)		
Against	193	36.21
Neutral	138	25.89
Pro	202	37.90
Sub-total	533	
Total (All Topics)	1024	

Table 1: Distribution of Stance Topics for each Label.

4. System Overview

Our system for Subtask B as presented in Figure 1 utilizes a fine-tuned `AraBERTv0.2-Twitter` model, chosen for its robust handling of Modern Standard Arabic (MSA) and the dialectal, emoji-rich nuances of social media text. To optimize performance, we employ dynamic batch-wise padding and class weighting to mitigate label

imbalance.

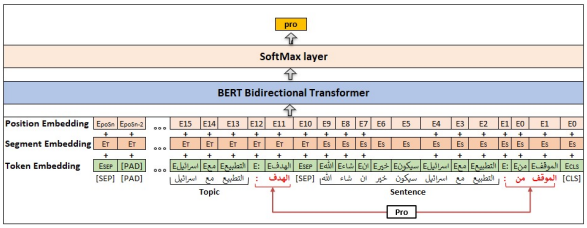

Figure 1: Architecture of the topic-conditioned AraBERT model.

Input Representation: To facilitate cross-topic generalization, we employ a prompt-based approach. Given a sentence x_s and topic x_t, we define static Arabic descriptors $p_1 =$ الهدف: (Target:) and $p_2 =$ الموقف من: (Stance toward:). The input sequence X is constructed as:

$$X = \text{[CLS]}\ p_1\ x_t\ \text{[SEP]}\ p_2\ x_s\ \text{[SEP]}$$

An example of the constructed input sequence is shown below:

```
[CLS] الموقف من: التطبيع مع اسرائيل [SEP] الهدف: التطبيع
مع اسرائيل سيكون خير ان شاء الله [SEP]
Target: Normalization with Israel
[SEP] Stance toward: Normalization
with Israel will be good, God willing
```

Classification Head: The encoder produces contextual embeddings $H \in \mathbb{R}^{L \times d}$. The sequence-level representation, $h_{\text{CLS}} \in \mathbb{R}^d$, is passed through a linear layer f_θ with dropout to compute the predicted probabilities $\hat{y} = \text{softmax}(f_\theta(h_{\text{CLS}}))$.

5. Experimental Setup

5.1. Dataset Details and Preprocessing

Experiments were conducted in Google Colaboratory using 5-fold Stratified Cross-Validation (SCV). For each phase ($N = 824$ and later $N = 1{,}024$), data was split into five stratified folds, with 80% for training and 20% for validation. Cross-validation guided model selection and hyperparameter tuning. Final evaluation was performed on the official Codabench platform. Tables A.5 and 3 report Macro-F1 on the unseen test set.

Preprocessing included Arabic-specific cleaning and normalization. Emojis were converted to text, non-Arabic characters, URLs, mentions, and hashtags were removed. We normalized orthographic variants (Alef forms, Alef Maqsura, Taa Marbuta), removed diacritics and tatweel, reduced extra whitespace, filtered empty samples, and eliminated duplicates.

5.2. Hyperparameter Fine-Tuning

The model was trained for up to 10 epochs with early stopping (patience $= 2$) based on Macro-F1. We applied weight decay (0.01) and dropout (0.2) to mitigate overfitting. Back-translation (Arabic→English→Arabic) generated 842 additional samples for augmentation. Hyperparameters are summarized in Table 2, with extended details in Appendix Table A.2.

Hyperparameter	Value
Max Sequence Length	128 tokens
Batch Size	16
Learning Rate	2×10^{-5}
Weight Decay	0.01
Dropout Rate	0.2
Optimization Metric	Macro-F1
Early Stopping Patience	2 epochs

Table 2: Fine-tuning hyperparameters for stance detection.

6. Results

We evaluated our system on the official test split. Our submission achieved a Macro-F1 score of 0.8483, ranking among the top teams in Subtask B and outperforming the baseline (AraBERT v2) by 0.0383. The full leaderboard results are provided in Appendix A.3 (Table A.5).

We followed a two-phase evaluation: an initial development track for model selection, followed by a final test track for performance verification. Table 3 summarizes our experiments. Prompt-based fine-tuning consistently improved F1 over baseline and augmentation, with the best score obtained using $1{,}024$ samples[3].

Exp	Model	F1	Technique	Size
1	M1[a]	0.777	Baseline	824
2	M1[a]	0.830	Prompt	824
3	M1[a]	0.811	Augmentation Back-translation	$1{,}684^*$
4	M2[a]	0.780	Prompt	824
5	M1[b]	0.836	Prompt	824
6	M1[b]	**0.848**	Prompt	$1{,}024$

Table 3: Subtask B experiments.

M1: AraBERTv02-Twitter; M2: MARBERT
*Augmented samples included.
[a]LR = 3×10^{-5}; [b]LR = 2×10^{-5}

[3]Only representative configurations are shown; additional runs were conducted during model selection.

6.1. Discussion of Key Findings

Our analysis reveals three primary factors that contributed to the model's success. First, the introduction of **topic-conditioned prompts** significantly enhanced the model's ability to generalize to the test set, effectively bridging the gap between training and unseen data. Second, optimization played a crucial role; reducing the learning rate from 3×10^{-5} to 2×10^{-5} improved Macro-F1 from 0.8328 to 0.8363 (absolute gain 0.0035). This consistent cross-validation improvement indicates that a slightly smaller optimization step stabilizes fine-tuning on small stance datasets.

Finally, the impact of data volume was evident in the final phase. Expanding the size of the training to $N = 1,024$ (including validation data released after development) increased Macro-F1 from 0.8364 to 0.8483, highlighting an absolute gain of 0.0119. This demonstrates that while architectural prompts and hyperparameter tuning are important, `AraBERTv0.2-Twitter` is highly data-responsive, benefiting substantially from the greater linguistic variety in the larger dataset.

6.2. Error Analysis

We analyzed misclassifications from out-of-fold (OOF) predictions across the 5-fold stratified cross-validation. Table 4 summarizes the most frequent confusion pairs, and Figure 2 shows the aggregated confusion matrix.

The dominant error is *Neutral* $\rightarrow$ *Against* (32.14%), followed by *Against* $\rightarrow$ *Neutral* (20.24%). This bidirectional confusion suggests difficulty distinguishing implicit opposition from weak or ambiguous sentiment, especially when negative tone is directed toward contextual entities rather than the target itself.

Confusions involving *Pro* are less frequent (*Pro* $\rightarrow$ *Against*: 10.71%, *Pro* $\rightarrow$ *Neutral*: 8.33%), indicating that supportive stances often contain clearer lexical cues. Overall, errors are relatively balanced across classes, suggesting that Macro-F1 gains are not driven by majority-class bias.

The prevalence of Neutral–Against confusions indicates that topic-conditioned prompting does not fully resolve stance ambiguity in cross-topic settings. High-confidence misclassified examples are provided in Appendix A.4(Table A.6).

Confusion Type	Count	Percentage
Neutral $\rightarrow$ Against	27	32.14%
Against $\rightarrow$ Neutral	17	20.24%
Neutral $\rightarrow$ Pro	12	14.29%
Against $\rightarrow$ Pro	12	14.29%
Pro $\rightarrow$ Against	9	10.71%
Pro $\rightarrow$ Neutral	7	8.33%

Table 4: Misclassification counts and percentages for each confusion type from OOF predictions.

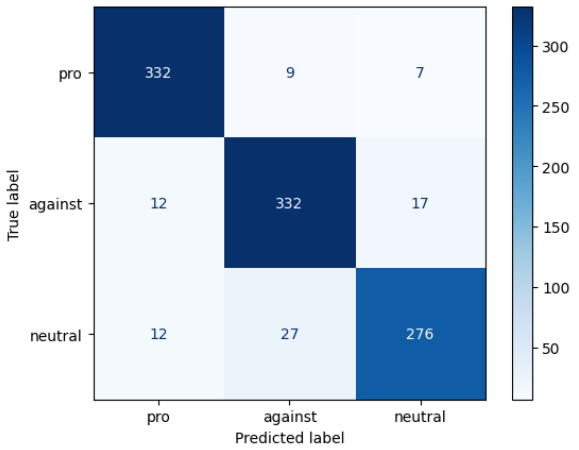

Figure 2: Aggregated confusion matrix for the 5-fold stratified cross-validation on the development set. Results are obtained via OOF inference using our fine-tuned `A2NLP-STANCE` model (based on `AraBERTv0.2-Twitter`) with topic-conditioned prompts.

7. Conclusion

We presented A2NLP's system for Subtask B of the StanceNakba 2026 Shared Task on cross-topic Arabic stance detection. Using topic-conditioned prompts with fine-tuned `AraBERTv0.2-Twitter`, our model reached a Macro-F1 of 0.8483, surpassing the `AraBERTv2` baseline by $+3.83$ points and ranking fifth. Our results show that structured input reformulation and domain-aligned pretraining boost cross-topic generalization, while distinguishing neutral from implicit opposing stances remains challenging. Future work will explore richer contextual modeling to improve stance disambiguation.

8. Limitations

Despite competitive performance, the A2NLP system is limited by a small dataset ($1,024$ instances) restricting generalization across Arabic dialects, difficulty distinguishing neutral from implicit opposition (32.14% misclassified), and evaluation on only two topics, while relying solely on `AraBERTv0.2-Twitter` without external knowledge or multi-task learning.

9. Acknowledgements

We thank the Arab NLP community and the StanceNakba 2026 organizers for providing a valuable benchmark for Arabic stance detection.

10. Bibliographical References

Abeer AlDayel and Walid Magdy. 2021. Stance detection on social media: State of the art and trends. *Information Processing & Management*, 58(4):102597.

Kholoud Khalil Aldous, Md Rafiul Biswas, Mabrouka Bessghaier, Shimaa Ibrahim, Kais Attia, and Wajdi Zaghouani. 2026. StanceNakba shared task: Actor and topic-aware stance detection in public discourse. In *Proceedings of the 15th International Conference on Language Resources and Evaluation (LREC'26)*, Palma, Spain.

Abeer Almasoudi, Muhammad Arif, Ahlam Hashem, and Esraa Samkari. 2025. Enhanced stance detection for arabic tweets. *Journal of Advanced Computational Intelligence and Intelligent Informatics*, 29(6):1262–1272.

Nora Alturayeif, Hamzah Luqman, Zaid Alyafeai, and Asma Yamani. 2024. Stanceeval 2024: The first arabic stance detection shared task. In *Proceedings of the Second Arabic Natural Language Processing Conference*, pages 774–782.

Nora Saleh Alturayeif, Hamzah Luqman, and Moataz A. Ahmed. 2023. Enhancing stance detection through sequential weighted multi-task learning. *Social Network Analysis and Mining*, 14(1):7.

Wissam Antoun, Fady Baly, and Hazem Hajj. 2020. Arabert: Transformer-based model for arabic language understanding. In *Proceedings of the 4th Workshop on Open-Source Arabic Corpora and Processing Tools*, pages 9–15.

Gilbert Badaro, Ramy Baly, Hazem Hajj, Wassim El-Hajj, Khaled Bashir Shaban, Nizar Habash, Ahmad Al-Sallab, and Ali Hamdi. 2019. A survey of opinion mining in arabic: A comprehensive system perspective covering challenges and advances in tools, resources, models, applications, and visualizations. *ACM Transactions on Asian and Low-Resource Language Information Processing (TALLIP)*, 18(3):1–52.

Mohamed Badran, Mo'men Hamdy, Marwan Torki, and Nagwa M El-Makky. 2024. Alexunlp-bh at stanceeval2024: Multiple contrastive losses ensemble strategy with multi-task learning for stance detection in arabic. In *Proceedings of the Second Arabic Natural Language Processing Conference*, pages 823–827.

Ramy Baly, Mitra Mohtarami, James Glass, Preslav Nakov, et al. 2018. Integrating stance detection and fact checking in a unified corpus. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 21–27.

Kareem Darwish, Walid Magdy, and Tarek Zanouda. 2017. Improved stance prediction in a user similarity feature space. In *Proceedings of the 2017 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*, pages 145–148.

Imane Guellil, Houda Saadane, Faiçal Azouaou, Billel Gueni, and Damien Nouvel. 2021. Arabic natural language processing: An overview. *Journal of King Saud University - Computer and Information Sciences*, 33(5):497–507.

Dilek Küçük and Cemal Aykut Nufus Fazli. 2020. Stance detection: A survey. *ACM Computing Surveys (CSUR)*, 53(1):1–37.

Xiaoye Qu, Yingjie Gu, Qingrong Xia, Zechang Li, Zhefeng Wang, and Baoxing Huai. 2024. A survey on arabic named entity recognition: Past, recent advances, and future trends. *IEEE Transactions on Knowledge and Data Engineering*, 36(3):943–959.

11. Language Resource References

Tariq Alhindi, Amal Alabdulkarim, Ali Alshehri, Muhammad Abdul-Mageed, and Preslav Nakov. 2021. Arastance: A multi-country and multi-domain dataset of arabic stance detection for fact checking. *arXiv preprint arXiv:2104.13559*.

Ali Alkhathlan, Faris Alahmadi, Faris Kateb, and Hend Al-Khalifa. 2025. Constructing and evaluating arabicstancex: a social media dataset for arabic stance detection. *Frontiers in Artificial Intelligence*, 8:1615800.

Nora Saleh Alturayeif, Hamzah Abdullah Luqman, and Moataz Aly Kamaleldin Ahmed. 2022. Mawqif: A multi-label arabic dataset for target-specific stance detection. In *Proceedings of the Seventh Arabic Natural Language Processing Workshop (WANLP)*, pages 174–184.

Anis Charfi, Mabrouka Ben-Sghaier, Andria Samy Raouf Atalla, Raghda Akasheh, Sara Al-Emadi, and Wajdi Zaghouani. 2024a. Marasta: A multi-dialectal arabic cross-domain stance corpus. In *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*, pages 11060–11069.

Anis Charfi, Mabrouka Bessghaier, Andria Atalla, Raghda Akasheh, Sara Al-Emadi, and Wajdi Zaghouani. 2024b. Stance detection in arabic with a multi-dialectal cross-domain stance corpus. *Social Network Analysis and Mining*, 14(1):161.

A. Appendix: Supplementary Materials, Software, and Data

This appendix provides implementation details and representative error examples to facilitate reproducibility and qualitative analysis.

A.1. Reproducibility Details

Framework / Library	Version	URL
Transformers	5.0.0	https:https://github.com/huggingface/transformers
PyTorch	2.9.0+cu128	https://pytorch.org
Datasets	4.0.0	https://github.com/huggingface/datasets
Tokenizers	0.22.2	https://github.com/huggingface/tokenizers

Table A.1: Environment setup.

Hyperparameter	Value
Number of epochs	10
Train batch size	16
Evaluation batch size	16
Learning rate	$2e-5$
Weight decay	0.01
Dropout	0.2 (hidden)
Evaluation strategy	Steps
Evaluation steps	50
Save strategy	Steps
Save steps	50
Push to Hugging Face Hub	True
Hub model ID	Dynamic per fold
Load best model at end	True
Metric for best model	Macro-F1
Greater is better	True
FP16	True
Logging strategy	Steps
Logging steps	50
Report to	None
Early stopping patience	2

Table A.2: Extended hyperparameter values for full model replication.

A.2. Additional Backbone Evaluation

To ensure reproducibility, we provide the specific model weights and configurations for each experiment conducted in Subtask B. Table A.3 details the performance metrics, fine-tuning techniques, and dataset sizes used, with corresponding Hugging Face repository links provided in the table notes for each unique model checkpoint.

Table A.4 reports additional pretrained Arabic models evaluated under identical training settings during development. These models were not selected for final submission due to lower Macro-F1 performance.

A.3. Leaderboard Results (Full Ranking)

For completeness, we include the full official leaderboard for Subtask B in the appendix. As shown in Table A.5, our system (A2NLP) ranks 5th overall with a Macro-F1 score of 0.8483, outperforming the AraBERT v2 baseline and placing among the top-performing teams.

Exp.	Model	F1	Technique	Size
1	M1[a]	0.777	Baseline	824
2	M1[a]	0.830	Prompt	824
3	M1[a]	0.811	Augmentation	1,684[*]
4	M2[a]	0.780	Prompt	824
5	M1[b]	0.836	Prompt	824
6	M1[b]	**0.848**	Prompt	1,024

Model URLs by Experiment:
Exp 1: https://huggingface.co/aomar85/StanceNakba-EXP1
Exp 2: https://huggingface.co/aomar85/StanceNakba-EXP2
Exp 3: https://huggingface.co/aomar85/StanceNakba-EXP3
Exp 4: https://huggingface.co/aomar85/StanceNakba-EXP4
Exp 5: https://huggingface.co/aomar85/StanceNakba-EXP5
Exp 6: https://huggingface.co/aomar85/A2NLP-STANCENAKBA2026-CROSS-TOPIC

M1: AraBERTv02-Twitter; M2: MARBERT
[*]Augmented samples included; [a]LR = 3×10^{-5}; [b]LR = 2×10^{-5}

Table A.3: Subtask B experiments and associated Hugging Face repositories.

Exp.	Model	Accuracy	Macro-F1
7	CAMeLBERT-Mix-Sentiment[†]	0.7459	0.7454
8	MARBERT[†]	0.6409	0.6354

Model URLs by Experiment:
Exp 7: https://huggingface.co/aomar85/StanceNakba-EXP7
Exp 8: https://huggingface.co/aomar85/StanceNakba-EXP8

Table A.4: Additional backbone models evaluated on the validation test set.

Rank	Team / Participant	Macro F1
1	wafaa	0.8724
2	mozalak	0.8607
3	U4RASD	0.8601
4	sarah_yassine	0.8562
5	**A2NLP (ours)**	0.8483
6	asmaa_q	0.8328
7	tokakhaled	0.8110
Baseline (AraBERT v2)		0.8100
8	AyahVerse	0.7904
9	jahidhasan	0.7407
10	walisa_alam	0.7317

Table A.5: Final leaderboard results for Subtask B. Our system (A2NLP) and the AraBERT v2 baseline are highlighted.

A.4. Representative Misclassifications

Sentence	Topic	Ground Truth	Prediction	Confidence
لا يزال اللاجئون يعانون من الآثار الاقتصادية الناجمة عن جائحة كورونا وارتفاع تكلفة المعيشة، والآن أيضا من ارتفاع تعرفة الخدمات Refugees still suffer from the economic impacts of the COVID-19 pandemic, the high cost of living, and now rising service fees.	وجود اللاجئين والمهاجرين في الأردن (من فلسطين ، سوريا ، العراق) Refugee/Immigrant Presence in Jordan from (Palestine, Syria and Iraq)	pro	neutral	97.95%
التطبيع مع الاسد اوسخ من التطبيع مع اسرائيل Normalization with Assad is filthier than normalization with Israel.	التطبيع مع اسرائيل Normalization with Israel	neutral	against	97.63%
أتوقع انو إخوانا السوريين مقدرين الوضع الي بتمر فيه الأردن وهم اخواننا ونسايبنا وقرايبنا لكن الوضع ما بتحمل (الي مش بإيدك بكيدك) الله يكون بعونهم وبعونا I believe our Syrian brothers understand the situation that Jordan is going through. They are our brothers, relatives, and family. However, the situation is becoming difficult to withstand what is beyond your control can still weigh heavily on you. May God help them and help us.	وجود اللاجئين والمهاجرين في الأردن (من فلسطين ، سوريا ، العراق) Refugee/Immigrant Presence in Jordan from (Palestine, Syria and Iraq)	neutral	pro	97.58%
يا سلام على الاخوة الاماراتيين التطبيع حتى في التعاليق ...ربنا معاكم O wow, the Emirati brothers, normalization even in the comments... God be with you.	التطبيع مع اسرائيل Normalization with Israel	neutral	pro	97.00%
الحديث لاجل فلسطين علامة مهمة لصحة الضمير والوعي سلام من ارض العراق لأصحاب الارض المقاومين سادة الرجال في زمن التطبيع المخنث Speaking up for Palestine is a clear sign of a living conscience and true awareness. Greetings from the land of Iraq to the rightful owners of the land the steadfast and resilient true men in a time of normalization.	التطبيع مع اسرائيل Normalization with Israel	against	pro	96.99%
نقدر فعلاً اخواننا السوريين و نتمنالهم الخير و الأمن و أن ترجع سوريا بخير ، بس الاردن مش قادرة تطعمي حالها عشان تطعمي غيرها .. We genuinely value our Syrian brothers and sisters and wish for Syria's safety and recovery, but Jordan is struggling to sustain its own people right now, making it hard to support others.	وجود اللاجئين والمهاجرين في الأردن (من فلسطين ، سوريا ، العراق) Refugee/Immigrant Presence in Jordan from (Palestine, Syria and Iraq)	against	pro	96.34%

Continued on next page

Sentence	Topic	Ground Truth	Prediction	Confidence
الأفضل ما حدا يرجع هناك لأنه مصيبة لو رجعوا عسوريا It's better that no one goes back there because it would be a disaster if they returned to Syria.	وجود اللاجئين والمهاجرين في الأردن (من فلسطين ، سوريا ، العراق) Refugee/Immigrant Presence in Jordan from (Palestine, Syria and Iraq)	neutral	against	96.10%
الأردن يكفيها إخواننا السوريين ومو ناقصة نتمنى أن يتم توزيع الجرحى التي تستدعي حالتهم النقل على مستشفيات دول الخليج Jordan has done its part for our Syrian brothers and can't take on any more. We hope that the wounded whose cases require transfer will be distributed among the hospitals of the Gulf countries	وجود اللاجئين والمهاجرين في الأردن (من فلسطين ، سوريا ، العراق) Refugee/Immigrant Presence in Jordan from (Palestine, Syria and Iraq)	against	pro	95.66%
الوقف إلى جانب إخواننا السوريين واجب قومي وطني و صدقت معاليك المملكة تحملت الكثير الكثير ولكن مصلحة الاردن العليا فوق اي اعتبارات Standing by our Syrian brothers is a national and patriotic duty, and you are right, Your Excellency the Kingdom has endured a great deal. However, Jordan's supreme interest stands above all other considerations.	وجود اللاجئين والمهاجرين في الأردن (من فلسطين ، سوريا ، العراق) Refugee/Immigrant Presence in Jordan from (Palestine, Syria and Iraq)	against	pro	95.64%
اسرائيل لا تحترم المطبعين ولا #التطبيع #فلسطين_ليست_قضيتي Israel respects neither those who normalize nor normalization #Palestine_is_not_my_cause	التطبيع مع اسرائيل Normalization with Israel	neutral	against	95.35%
طيب وين المشكله ؟ قلوب صادقه تبحث عن السلام وانشاء علاقه مع العرب ، وطبعاً التطبيع بعد انهاء الاحتلال وانشاء دوله فلسطينيه So where is the problem? These are sincere hearts seeking peace and building relations with the Arabs—and of course, normalization would follow the end of the occupation and the establishment of a Palestinian state.	التطبيع مع اسرائيل Normalization with Israel	neutral	pro	95.19%
إن شاء الله يثور الشعب على هذا الذل الإرهابيون الصهاينة يقتلون الفلسطينيين الدول العربية تطبيع مع هؤلاء الارهابيين الصهاينة بابتسامة باعوا دينهم ومكانهم المقدس God willing, the people will rise against this humiliation. The Zionist terrorists are killing Palestinians, while Arab countries normalize with these Zionist terrorists with a smile; they have sold their religion and their holy places.	التطبيع مع اسرائيل Normalization with Israel	neutral	against	93.78%

Continued on next page

Sentence	Topic	Ground Truth	Prediction	Confidence
عبئ على الأردن يا خنزير ؟؟ اقتصاد الأردن لم يرى الخير منذ قدومهم؟؟ اولا الأردن منهوبة ومن أفقر دول العالم قبل أزمة سوريا بقرووون وجميعنا ك أردنيين عارفين النهب والسرقة لأموال الشعب من حكومتنا الساقطة والساقط الي يسحج لها . بشو اثروا علينا السوريين؟؟ اخذوا منا بيوتنا ولا اخذوا من اهالينا؟؟؟ السوريين في منهم احسن منا بمراحل وسوريا يلي بتحكي عنها كانت بيوم عاصمتنا عاصمة الشام قبل التقسيمة الوسخة الي صارت . سوريا الي ما بحس فيها وبشعبها ما عنده لا ذمة ولا ضمير وحرام فيه يعيش في بلد عربي وبين العرب. تعرف ادعي دعوه وحده دايما ورح اضل ادعيها على كل شخص ما يقدر السوريين وما يحس فيهم ، الله يطعمك بحرب تشتت اهلك وتجعلك تتشتت انت وعيالك وتدور على مكان يحتويك وما تلاقيه ، هذه الدعوه مش لك شخصيا هذه لكل شخص ما يحس بيلي بيعانيه السوريين بس What 'good' was there since they arrived? First of all, Jordan has been looted and has been one of the poorest countries in the world for centuries, long before the Syrian crisis. All of us as Jordanians know the looting and theft of the people's money by our failed government and the sycophants who cheer for them. How have the Syrians affected us? Did they take our homes or take from our families? Many Syrians are far better than us, and the Syria you speak of was once our capital—the capital of the Levant—before the dirty partition that occurred. Anyone who doesn't feel for Syria and its people has neither integrity nor a conscience, and doesn't deserve to live in an Arab land among Arabs. I have one prayer that I will always repeat for anyone who doesn't appreciate or feel for the Syrians: May God afflict you with a war that scatters your family and forces you and your children to wander, searching for a place to take you in and finding none. This prayer isn't for you personally, but for anyone who doesn't feel the suffering of the Syrians.	وجود اللاجئين والمهاجرين في الأردن (من فلسطين ، سوريا ، العراق) Refugee/Immigrant Presence in Jordan from (Palestine, Syria and Iraq)	pro	against	92.76%

Continued on next page

Sentence	Topic	Ground Truth	Prediction	Confidence
احترام اليهود شي جيد المفروض نحترم جميع الأديان السماوية و المذاهب اما موضوع التطبيع مع اسرائيل مرفوض Respecting Jews is a good thing; we should respect all divinely revealed religions and sects. However, the issue of normalization with Israel is rejected.	التطبيع مع اسرائيل Normalization with Israel	against	pro	89.32%
التطبيع الحقيقي يبدأ بالتطبيع والسلام مع فلسطين والفلسطينيين، بما يمكنهم من العيش بكرامة على أرضهم ويزيل عن أذهاننا مشاهد الذل والإهانة والقمع التي تعرضوا لها على يد الجنود والمستوطنين الإسرائيليين لعقود، عندها سيكون هناك تطبيع مع الشعوب العربية وليس مع سلطات معزولة عن شعوبها True normalization begins with normalization and peace with Palestine and the Palestinians, enabling them to live with dignity on their land and erasing from our minds the scenes of humiliation, insult, and oppression they have endured at the hands of Israeli soldiers and settlers for decades. Only then will there be normalization with the Arab peoples, rather than with authorities isolated from their own populations.	التطبيع مع اسرائيل Normalization with Israel	against	pro	88.94%
@khaberni صاروا السوريين علاقة تعلقوا عليها سرقاتكم ورفاهيتكم ورواتبكم اللي ما بتستحقوها والله عيب الاردن بعاني لانوا في حرامية كثير وسرقات ومحسوبيات Syrians became an excuse to hang your thefts and your unearned salaries on.	وجود اللاجئين والمهاجرين في الأردن (من فلسطين ، سوريا ، العراق) Refugee/Immigrant Presence in Jordan from (Palestine, Syria and Iraq)	pro	against	88.53%
المهم. جاسم متي تزور اسرائيل. ان شاء الله. اخبارك مع التطبيع. امورك ماشيه. والدعم. مستمر Anyway Jassim, when will you visit Israel? How is normalization going for you?	التطبيع مع اسرائيل Normalization with Israel	neutral	pro	87.41%
بخصوص التطبيع هذي اسطوانه الغرب ومرتزقتها اتمنى من المحبين لبلاد الحرمين الشريفين عدم التطرق لها في التقرير Regarding normalization, this is a broken record of the West and its mercenaries.	التطبيع مع اسرائيل Normalization with Israel	neutral	against	87.37%
لست مع اعادتهم ليصبحوا ضحية لبشار الأسد وللحرب I am not for returning them only for them to become victims of Bashar al-Assad and the war.	وجود اللاجئين والمهاجرين في الأردن (من فلسطين ، سوريا ، العراق) Refugee/Immigrant Presence in Jordan from (Palestine, Syria and Iraq)	pro	against	86.22%

Continued on next page

Sentence	Topic	Ground Truth	Prediction	Confidence
خلوكم بس ليش ايمال بتسبونا Just stay, but why do you always insult us?	وجود اللاجئين والمهاجرين في الأردن (من فلسطين ، سوريا ، العراق) Refugee/Immigrant Presence in Jordan from (Palestine, Syria and Iraq)	neutral	against	93.58%

Table A.6: Representative high-confidence misclassified examples from out-of-fold predictions, illustrating systematic stance ambiguities.