

# AraFinNLP 2024: The First Arabic Financial NLP Shared Task

Sanad Malaysha<sup>1</sup> Mo El-Haj<sup>2</sup> Saad Ezzini<sup>2</sup> Mohammed Khalilia<sup>1</sup> Mustafa Jarrar<sup>1</sup>  
Sultan Almujaivel<sup>3</sup> Ismail Berrada<sup>4</sup> Houda Bouamor<sup>5</sup>

<sup>1</sup>Birzeit University, Palestine <sup>2</sup>Lancaster University, United Kingdom

<sup>3</sup>King Saud University, Saudi Arabia <sup>4</sup>Mohammed VI Polytechnic University, Morocco

<sup>5</sup>Carnegie Mellon University, Qatar

## Abstract

The expanding financial markets of the Arab world require sophisticated Arabic NLP tools. To address this need within the banking domain, the Arabic Financial NLP (AraFinNLP) shared task proposes two subtasks: (i) Multi-dialect Intent Detection and (ii) Cross-dialect Translation and Intent Preservation. This shared task uses the updated ArBanking77 dataset, which includes about 39k parallel queries in MSA and four dialects. Each query is labeled with one or more of a common 77 intents in the banking domain. These resources aim to foster the development of robust financial Arabic NLP, particularly in the areas of machine translation and banking chat-bots. A total of 45 unique teams registered for this shared task, with 11 of them actively participated in the test phase. Specifically, 11 teams participated in Subtask 1, while only 1 team participated in Subtask 2. The winning team of Subtask 1 achieved  $F_1$  score of 0.8773, and the only team submitted in Subtask 2 achieved a 1.667 BLEU score.

## 1 Introduction

Financial Natural Language Processing (FinNLP) is revolutionising the financial sector, offering unmatched potential to enhance decision-making, manage risks, and drive operational efficiency (Zavitsanos et al., 2023). By leveraging advanced linguistic analysis (Malaysha et al., 2024) and NLP algorithms (Barbon Junior et al., 2024), FinNLP optimises processes and streamlines workflows, delivering a myriad of benefits (Darwish et al., 2021). FinNLP enables the extraction of key information, including events (Aljabari et al., 2024), relationships (Jarrar, 2021), and named entities (Liqreina et al., 2023), from diverse sources such as financial reports, news articles, invoices, and social media posts.

Practical applications of Financial NLP include textual analysis in accounting and finance (Loughran and McDonald, 2016; El-Haj et al.,

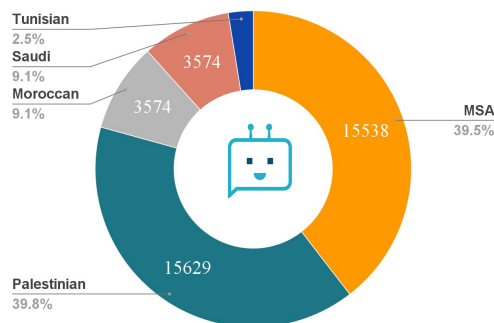


Figure 1: AraBanking2024 Datasets for Intent Detection

2019, 2020), analysis of financial transactions (Jørgensen and Igel, 2021), customer complaints (Jarrar, 2008), and text classification (Arslan et al., 2021; El-Haj et al., 2014). Nevertheless, as recently highlighted by Jørgensen et al. (2023); El-Haj et al. (2021), the majority of financial NLP research is conducted in English.

On the other hand, the Arab world’s financial landscape is experiencing robust growth, attracting global attention and investment across diverse sectors. This expansion underscores the critical role of Financial NLP in understanding and interpreting the intricacies of Arabic financial communications. As highlighted by Zmandar et al. (2021, 2023), the dynamic nature of Middle Eastern stock markets reflects the region’s evolving financial environment, necessitating advanced NLP tools tailored to local linguistic nuances.

In this paper, we provide an overview of the AraFinNLP-2024 Shared Task<sup>1</sup>, which represents a significant step forward in advancing the development of Arabic NLP capabilities within the finance domain. We propose two subtasks aimed at addressing key challenges in the banking sector: (i) Multi-dialect Intent Detection and (ii) Cross-dialect Translation and Intent Preservation. For this shared task, we provided participants access

<sup>1</sup>Task: <https://sina.birzeit.edu/arbking77/arafinnlp/>

to the ArBanking77<sup>2</sup> corpus (Jarrar et al., 2023b), for training purposes, which contains about 31k queries in Modern Standard Arabic (MSA) and Palestinian dialect, each labeled by at least one of the 77 intents in the banking domain. The data excluded any examples covering the Moroccan, Tunisian and Saudi dialects. We discuss the test data preparation in Section 4. The two subtasks are designed to tackle the complexities inherent in interpreting and managing diverse banking data prevalent in Arabic-speaking regions, catering to the linguistic diversity across the Arab world (Haff et al., 2022). By focusing on intent detection and dialectal translation, AraFinNLP aims to enhance customer service, automate query handling, and facilitate seamless communication across various Arabic dialects, thus fostering inclusivity and efficiency in financial services.

The rest of the paper is organized as: Section 2 presents the related work; Section 3 describes the tasks; Section 4 presents the dataset; Section 5 reports on the performance of the participating systems; and, Section 6 concludes.

## 2 Related Work

In the rapidly evolving field of NLP, integrating financial data has become a significant area of research, particularly in the context of multilingual and dialectal variations.

Recent studies in the financial NLP (FinNLP) domain have investigated machine translation between MSA and dialectal Arabic (Zmandar et al., 2021). Noteworthy contributions to FinNLP have been made by Zmandar et al. (2023), who examined the application of NLP techniques to analyse Arabic financial texts. Additionally, Jarrar et al. (2023c) introduced the ArBanking77 dataset, derived from the English Banking77 dataset (Casanueva et al., 2020), which has been pivotal in advancing research in this area. The related works can be divided into four main categories: financial document extraction, lexical diversity and sentiment analysis, cross-lingual and dialectal studies, and related machine learning techniques.

Within the realm of FinNLP, the process of Financial Document Extraction emerges as a critical component in the financial industry's digital transformation. This intricate process entails the automated retrieval and meticulous analysis

of pertinent data from a diverse array of financial documents, including but not limited to invoices, receipts, and financial statements. By harnessing cutting-edge methodologies such as semantic role-labeling schemes and integrating deep NLP techniques with knowledge graphs (Hammouda et al., 2024), researchers have significantly elevated the accuracy and efficiency of financial narrative processing (Lamm et al., 2018; Cavar and Josefy, 2018; Abreu et al., 2019). These advanced techniques not only streamline data extraction processes but also enable deeper insights into complex financial structures and trends, thereby empowering decision-makers with actionable intelligence to drive strategic initiatives, mitigate risks, and seize opportunities in a rapidly evolving financial landscape. Through the seamless integration of technology and financial expertise, FinNLP continues to revolutionise traditional workflows, facilitating agile decision-making and fostering innovation across the financial sector.

In addition to document extraction, FinNLP extensively explores the realms of lexical diversity and sentiment analysis. The assessment of lexical diversity delves into the richness and variety of vocabulary employed within financial texts. Concurrently, sentiment analysis endeavours to decipher the emotional nuances embedded in financial communications. For instance, researchers have scrutinised sentiment patterns in Arabic financial tweets, shedding light on the linguistic intricacies prevalent in such contexts (Alshahrani et al., 2018). Similarly, investigations into expressions of trust and doubt within financial reports underscore the emotional dimensions inherent in financial discourse (Žnidaršič et al., 2018; Akl et al., 2019). By elucidating these linguistic subtleties, FinNLP not only enhances our understanding of financial communications, but also empowers stakeholders to navigate and interpret complex data with confidence and clarity.

Although MSA and Arabic dialects have been studied, mainly from morphological (and semantic) annotations and dialect identification perspectives (Nayouf et al., 2023; Jarrar et al., 2023e; Abdul-Mageed et al., 2023; Al-Hajj and Jarrar, 2021), less attention was given to intent detection. In terms of cross-lingual and dialectal studies, important works examine how languages or dialects interact and translate across cultural and linguistic boundaries, focusing on understanding and processing variations in language usage and meaning. (Kwong,

<sup>2</sup>Corpus: <https://sina.birzeit.edu/arbanking77/>

2018) addresses the challenges of cross-lingual financial communications by focusing on analyzing and annotating English-Chinese financial terms. The proposal of a multilingual financial narrative processing system by (El-Haj et al., 2018) highlights the necessity for tools that function across diverse languages.

Štihec et al. (2021) used forward-looking sentence detection techniques to anticipate future statements in financial texts. These efforts highlight the varied methodologies in financial NLP, from structural data extraction to predictive analytics, while noting a gap in addressing Arabic linguistic nuances. (Naskar et al., 2022) unveiled a transformer-based architecture to detect causality in financial texts, and (Tsutsumi and Utsuro, 2022) employed a BERT-based model to discern stock price triggers from news, enhancing the understanding of nuanced financial changes. These studies advance automatic text processing, although they primarily focus on English datasets and traditional NLP tasks. Another study by (Zmandar et al., 2022) focused on the French financial sector, developing CoFiF Plus, a corpus for summarizing French financial narratives, highlighting the need for tools that cater to non-English financial data. In the last year, NLP has seen significant advances in the understanding and generation of financial narratives across various languages (Zmandar et al., 2023) introduced FinAraT5, a text-to-text model tailored for Arabic financial text understanding and generation, emphasizing the importance of language-specific models in financial contexts.

While exploring the frontier of FinNLP, it is essential to acknowledge the broader landscape of shared tasks and initiatives aimed at understanding Arabic MSA and dialects. These include notable endeavours such as ArabicNLU for word-sense disambiguation (Khalilia et al., 2024; Jarrar et al., 2023d), NADI for dialect identification (Abdul-Mageed et al., 2023), and WojoodNER for named entity recognition (Jarrar et al., 2024, 2023a). Notably, these tasks extend their focus to entities within the finance and banking domains (Jarrar et al., 2022; Liqreina et al., 2023), aligning closely with the objectives of FinNLP. Through collaborative efforts and interdisciplinary research, these shared tasks complement the advancements in FinNLP, fostering a holistic understanding of linguistic nuances and enhancing the applicability of NLP techniques within financial contexts. As FinNLP continues to evolve, leveraging insights

from these shared tasks can further enrich its capabilities, ultimately driving innovation and efficacy in financial analysis, decision-making, and risk management.

### 3 Task Description

The AraFinNLP shared task comprises two sub-tasks aimed at advancing Financial Arabic NLP: Subtask 1 focuses on Multi-dialect Intent Detection, while Subtask 2 addresses Cross-dialect Translation and Intent Preservation within the banking domain. Nonetheless, AraFinNLP is the first shared task in Arabic financial NLP, as well as the first Arabic multi and cross dialects, where banking intents predicted and preserved across four dialects translated from MSA and English. In this section we break into details of each sub-task.

#### 3.1 Subtask 1: Multi-dialect Intent Detection

Subtask 1 of the AraFinNLP shared task revolves around Multi-dialect Intent Detection in the banking domain. Participants are tasked with developing NLP models capable of accurately classifying customer intents from queries in diverse Arabic dialects, taking into consideration that dialects classes are hidden from participants. The challenge lies in training models that can understand both MSA and regional dialects, such as Gulf, Levantine, and North African, to enhance customer service and automate query handling. Figure 2 shows example of queries in different dialects with their corresponding intents.

#### 3.2 Subtask 2: Cross-dialect Translation and Intent Preservation

In Subtask 2 of the AraFinNLP shared task, participants focus on Cross-dialect Translation and Intent Preservation. The objective is to translate queries from MSA language to various Arabic dialects, while ensuring the preservation of the original intent as shown in Figure 2. The target dialects are limited to Gulf (Saudi), Moroccan (Darija), Palestinian, and Tunisian. Participants are provided with datasets containing MSA queries and their corresponding intents, tasked with accurately translating them into dialectal Arabic while maintaining semantic integrity.

#### 3.3 Restrictions

The followed norm in the shared tasks is to set a lot of strict restrictions on the usage of external data and online resources. However, since these

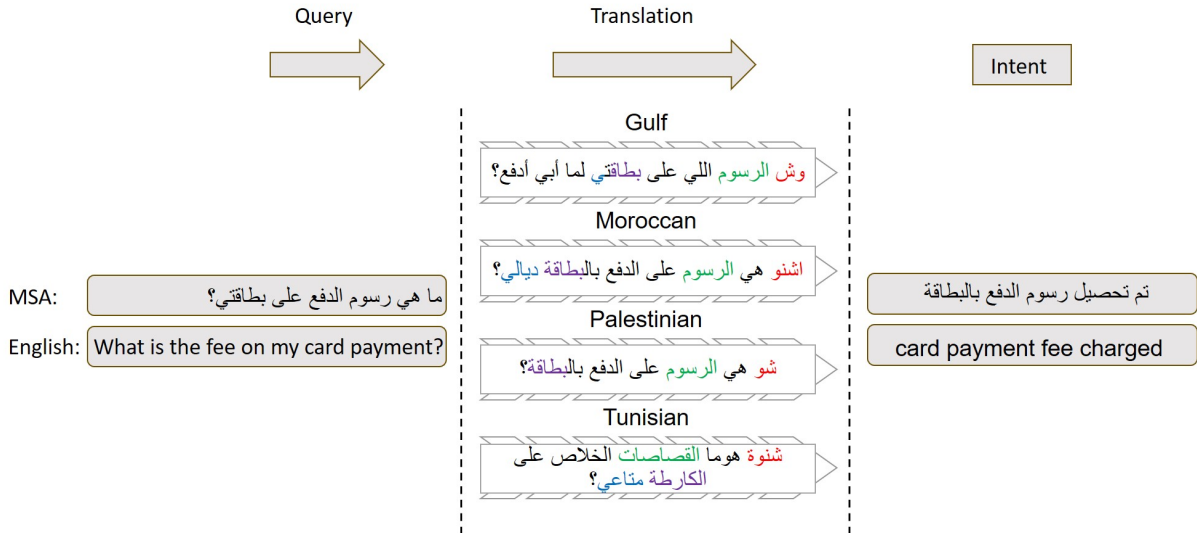


Figure 2: Translation of MSA into four Arabic dialects.

tasks related to Arabic language and dialects which have low resources (Malaysha et al., 2023), we relaxed the restrictions and allowed participants to exploit any resources at their disposal including pre-trained encoders, generative models, and augmented datasets. Participants have the freedom to incorporate diverse online resources to enhance their models further, fostering a broader exploration of methodologies and potentially leading to more innovative solutions in intent detection and dialectal translation within the banking domain and beyond.

However, we dictated the submission format. Because we utilized CodaLab<sup>3</sup> framework for scoring the participants’ results, we define a JSON-based structure for representing their submission in both sub-tasks, submission details are posted in the official shared task website<sup>4</sup> of the shared task.

## 4 Dataset

The shared task provided a multi and cross dialectal Arabic banking-related dataset, which consists of queries translated into MSA and various Arabic dialects, including Palestinian, Saudi, Tunisian, and Moroccan (Figure 1). The queries are classified into intent classes selected from 77 intents, where each query could match to one or more intent. The initial data was obtained from ArBanking77 dataset (Jarrar et al., 2023c), available in MSA and Palestinian dialects. For this shared task, we augmented the ArBanking77 dataset with three additional di-

lects. We cooperated with teams of linguists each specialized in each dialect. Table 1 outlines the statistics of the final dataset.

Dialect	Train	Development	Test
MSA	10,733	1231	3,574
Moroccan	–	–	3,574
Palestinian	10,821	1234	3,574
Saudi	–	–	3,574
Tunisian	–	–	1,000
<b>Total</b>	<b>21,554</b>	<b>2,465</b>	<b>15,296</b>

Table 1: Statistics of the AraFinNLP shared task dataset.

**Moroccan Dialect:** the team first translated the original sentences from English and MSA to Moroccan Darija using GPT-4 and Meta’s seamless4t-v2 models (Barrault et al., 2023). A team of seven native Moroccan Darija speakers from various regions manually reviewed and corrected the translations for accuracy. They divided the dataset into seven parts, each checked by one annotator. The annotators were asked to refer to both MSA and English queries when the translation was ambiguous. Each translated sentence was carefully reviewed, resulting in approximately 67% of them being edited. Lastly, two additional annotators conducted a final review of all the data to ensure consistency.

**Saudi and Tunisian Dialects:** one specialized linguist in each of these two dialects worked on translating data from MSA to the Saudi and Tunisian dialects. As they are native speakers of these dialects, they ensured the translations captured all

<sup>3</sup><https://codalab.lisn.upsaclay.fr/>

<sup>4</sup><https://sina.birzeit.edu/arbanking77/arafinnlp/>



the linguistic nuances of the Najdi dialect for Saudi and local terms for Tunisian, preserving the true meaning of the text. During the process, they carefully double-checked the translation to address any typos or ambiguity.

## 5 Results and Discussion

The AraFinNLP shared subtasks attracted a good number of teams, with a diverse array of participants employing various techniques and methodologies tailored to the challenges of multi-dialectal intent detection and cross-dialect translation in the banking domain. Leveraging a range of NLP models and approaches, including fine-tuning pre-trained transformers and utilising deep learning models, participating teams navigated the complexities of Arabic dialects and the nuances inherent in financial communications (Khalilia et al., 2024; Jarrar et al., 2023d; Abdul-Mageed et al., 2023; Jarrar et al., 2024, 2023a, 2022; Liqreina et al., 2023).

With queries in MSA and several Arabic dialects, extractive techniques were predominantly employed. The AraFinNLP shared task results (Table 2) provide insights into the effectiveness of different methodologies and approaches in addressing the unique challenges of Financial Arabic NLP.

The evaluation metrics for each subtask offer a comprehensive overview of system performance across various dialects and tasks, facilitating comparisons and informing future research directions in Arabic NLP, particularly within the finance domain.

### 5.1 Participating Teams and Results

A total of 45 unique teams registered for the AraFinNLP shared task, with 11 teams effectively participating and submitting systems to Subtask 1 and Subtask 2, leading to 30 submissions during the development phase and 168 submissions during the test phase.

For Subtask 1, which involves multi-class classification, the  $F_1$  micro score was the primary metric for ranking and comparison. Additionally, secondary metrics such as macro precision and recall were provided for further reference in the evaluation process. Subtask 2 utilised the BiLingual Evaluation Understudy (BLEU) score as the primary metric for ranking and comparison, with secondary metrics like CHaRacter-level F-score (chrF) and Translation Error Rate (TER).

The results of the AraFinNLP shared task, pre-

sented in Table 2, reveal that the MA team achieved the highest Micro- $F_1$  score of 0.8773 during the test phase, securing the top rank. This impressive performance can be attributed to their methodology, which involved using an ensemble of fine-tuned BERT-based models and integrating contrastive loss for training. This approach likely provided a robust way to handle the nuanced differences in Arabic dialects and ensured better generalisation across the diverse dataset. The MA team’s method also included augmenting the ArBanking77 dataset with additional Arabic dialects, which could have enriched the training data and improved the model’s ability to generalise better on unseen data. Their approach stands out as it meticulously addressed both model architecture and data augmentation, crucial factors in achieving superior performance in NLP tasks involving complex and diverse languages like Arabic. A detailed breakdown of the Subtask 1 results for each team, including performance during both the development and test phases, is presented in Table 2. Six out of 11 teams participated in the development phase as the participation was optional. It is shown that the SemanticCUETSync and SMASH teams ranked 1<sup>st</sup> and 2<sup>nd</sup> during the development phase, but dropped to 5<sup>th</sup> and 9<sup>th</sup> place in the final evaluation phase. This indicates that the developed solutions overfitted on the development data and could not generalise to the additional dialects in the test set.

### 5.2 Teams Description

The AraFinNLP shared subtasks attracted a diverse array of participants, each employing unique methodologies tailored to the specific challenges of multi-dialectal intent detection and cross-dialect translation within the banking domain. Here are descriptions of each team that participated in the shared task, highlighting their team name, task, methodology, and techniques used:

1. **BFCI**: This team participated in the AraFinNLP2024 shared task, specifically in the subtask of Multi-dialect Intent Detection. They employed traditional machine learning approaches integrated with basic vectorization for feature extraction. The primary algorithms used were Multi-layer Perceptron, Stochastic Gradient Descent, and Support Vector Machines (SVM), with SVM outperforming the others. The approach

Submission ID	Codalab Username	Team Name	Test Phase		Dev Phase	
			Micro- $F_1$	Rank	Micro- $F_1$	Rank
758563	AsmaaRamadan	MA	0.8773	1		
757837	Hossam-Elkordi	AlexuNLP24	0.8762	2	0.9614	3
757408	murhaf	BabelBot	0.8709	3	0.9458	5
749856	sultan	-NPS-	0.8342	4		
755226	SemanticCUETSync	SemanticCUETSync	0.8208	5	0.9852	1
755293	abdelmomenbennasr	SENIT	0.8204	6		
755331	Fired_from_NLP	Fired_from_NLP	0.8014	7	0.9466	4
753646	Haithem	-NPS-	0.7894	8		
747668	yalhariri	SMASH	0.7866	9	0.9639	2
748164	licvol	dzFinNlp	0.6721	10	0.9302	6
747581	Nsrin_Ashraf	BFCI	0.4907	11		

-NPS-: No Paper Submission.

Table 2: Subtask 1 results breakdown by team for both development and test phases.

achieved a micro  $F_1$  score of 0.4907.

2. **SENIT**: Hailing from the National Engineering School of Tunisia, the SENIT Team tackled the Multi-dialect Intent Detection subtask by fine-tuning several pre-trained contextualised text representation models, including multilingual BERT and Arabic-specific models like MARBERTv1 (Abdul-Mageed et al., 2021), MARBERTv2<sup>5</sup>, and CAMEL-BERT (Inoue et al., 2021). They also employed an ensemble technique, combining MARBERTv2 and CAMEL-BERT embeddings, with MARBERTv2 achieving a micro  $F_1$  score of 0.8204.
3. **dzFinNlp**: Focused on improving intent detection in financial conversational agents, the dzFinNlp Team experimented with various models and feature configurations. They explored traditional machine learning methods like LinearSVC with Term Frequency-Inverse Document Frequency (TF-IDF) and deep learning models like Long Short-Term Memory (LSTM) and bidirectional LSTM (BiLSTM). Additionally, transformer-based models were employed, achieving a micro  $F_1$ -score of 0.6721.
4. **MA**: From Alexandria University, the MA Team participated in the cross-dialectal Arabic intent detection subtask using an ensemble of fine-tuned BERT-based models. They in-

tegrated contrastive loss for training and augmented the ArBanking77 dataset with additional Arabic dialects, achieving an  $F_1$ -score of 0.8773, ranking first in the task.

5. **BabelBot**: Participating in the Multi-dialect Intent Detection subtask, the BabelBot Team employed an encoder-only T5 model fine-tuned for the task. They generated synthetic data and used model ensembling to address cross-dialect challenges, securing third place with a micro- $F_1$  score of 0.8709.
6. **SemanticCuETSsync**: This team worked on intent detection using a combination of traditional machine learning and deep learning techniques. They implemented models like LSTM and transformer-based models, focusing on enhancing feature extraction and classification performance in the context of Arabic financial text analysis. The approach achieved a micro  $F_1$  score of 0.8208.
7. **AlexUNLP24**: Hailing from the University of Edinburgh, the AlexUNLP24 Team tackled the intent detection task using various BERT and BART-based models. Their approach involved direct fine-tuning across all intents, with QARiB and MARBERTv2 achieving the a micro  $F_1$  score of 0.8762 and ranking second place in the task. They found that translating the data to MSA impaired model performance in multi-dialect settings.
8. **Fired from NLP**: This team employed a bidi-

<sup>5</sup><https://huggingface.co/UBC-NLP/MARBERTv2>

rectional interrelated model for joint intent detection and slot filling, leveraging both machine learning and deep learning approaches. They focused on enhancing the accuracy of intent detection in noisy and unstructured text data in financial conversational settings. The approach achieved a micro  $F_1$  score of 0.8014.

9. **SMASH**: Utilising several BERT and BART-based models for the Multi-dialect Intent Detection task, the SMASH Team’s experiments showed that MARBERTv2 outperformed other models using a two-step approach, achieving an  $F_1$  score of 0.7866. Their work highlighted the challenges and opportunities in Arabic financial NLP.

## 6 Conclusions and Future Work

In conclusion, the AraFinNLP shared task represents a significant effort in advancing Financial Arabic Natural Language Processing by addressing critical challenges in multi-dialect intent detection and cross-dialect translation within the banking domain. The participation of diverse teams employing various methodologies has yielded valuable insights into the effectiveness of different approaches. These findings can inform future research, guiding the development of more robust and accurate NLP models for handling the complexities of Arabic dialects in financial contexts.

The results highlight the impressive performance of the MA team, whose use of fine-tuned BERT-based models and contrastive loss for training, coupled with data augmentation techniques, proved particularly effective. This underscores the importance of both sophisticated model architectures and enriched training datasets in achieving high performance in NLP tasks involving diverse languages.

Looking ahead, future work in Financial Arabic NLP could explore several avenues for further improvement and innovation. Enhancing the performance and adaptability of NLP models across a wider range of Arabic dialects could lead to more inclusive and effective solutions for financial communication. Additionally, the development of specialized resources and datasets tailored to specific financial domains and dialectal variations could facilitate more targeted and accurate analysis of financial texts. Furthermore, ongoing advancements in NLP technologies, such as the integration of multi-modal inputs and the incorporation of domain-specific knowledge, offer promising oppor-

tunities for improving the capabilities of Financial Arabic NLP systems.

The AraFinNLP shared task serves as a catalyst for advancing research and development in Financial Arabic NLP, paving the way for more sophisticated and versatile systems capable of addressing the diverse linguistic and communicative needs of Arabic-speaking communities in the financial domain. By fostering collaboration and innovation, the shared task contributes to the broader goal of enhancing accessibility, efficiency, and inclusivity in financial services through the application of natural language processing technologies.

## Limitations

It is important to acknowledge the limitations inherent in the AraFinNLP shared task and its associated datasets. One potential limitation is the representation of Arabic dialects in the provided data, as dialectal variations may not be fully captured or balanced across different regions. Additionally, the complexity of financial communications and the nuances of Arabic language usage may pose challenges for accurate interpretation and analysis, particularly in cross-dialect translation tasks. Furthermore, the availability of annotated data and resources for training NLP models in Arabic dialects may be limited, potentially impacting the scalability and generalisation of systems developed within the shared task.

## Ethics Statement

The datasets provided for this shared task are derived from public sources, eliminating specific privacy concerns. The results of the shared task will be made publicly available to enable the research community to build upon them for the public good and peaceful purposes. Our data and ideas are strictly intended for non-malicious, peaceful, and non-military purposes. The AraFinNLP shared task is committed to upholding ethical standards and promoting responsible research practices in NLP. Participants are expected to adhere to principles of fairness, transparency, and accountability throughout the development and evaluation of their systems. Additionally, participants are encouraged to consider the broader societal implications of their research, including issues related to accessibility, inclusivity, and potential impacts on vulnerable populations. The organizers of the shared task are dedicated to fostering an inclusive and collabora-

tive research environment, where diverse perspectives and ethical considerations are valued and integrated into the development and dissemination of NLP technologies.

## Acknowledgements

This research is partially funded by the Palestinian Higher Council for Innovation and Excellence and by the research committee at Birzeit University. We extend our gratitude to Taymaa Hammouda for the technical support.

## References

- Muhammad Abdul-Mageed, AbdelRahim Elmadany, Chiyu Zhang, El Moatez Billah Nagoudi, Houda Bouamor, and Nizar Habash. 2023. [NADI 2023: The fourth nuanced Arabic dialect identification shared task](#). In *Proceedings of ArabicNLP 2023*, pages 600–613, Singapore (Hybrid). Association for Computational Linguistics.
- Muhammad Abdul-Mageed, AbdelRahim A. Elmadany, and El Moatez Billah Nagoudi. 2021. [ARBERT & MARBERT: deep bidirectional transformers for arabic](#). In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing, ACL/IJCNLP 2021, (Volume 1: Long Papers), Virtual Event, August 1-6, 2021*, pages 7088–7105. Association for Computational Linguistics.
- Carla Abreu, Henrique Cardoso, and Eugénio Oliveira. 2019. [FinDSE@FinTOC-2019 shared task](#). In *Proceedings of the Second Financial Narrative Processing Workshop (FNP 2019)*, pages 69–73, Turku, Finland. Linköping University Electronic Press.
- Hanna Abi Akl, Anubhav Gupta, and Dominique Mariko. 2019. [FinTOC-2019 shared task: Finding title in text blocks](#). In *Proceedings of the Second Financial Narrative Processing Workshop (FNP 2019)*, pages 58–62, Turku, Finland. Linköping University Electronic Press.
- Moustafa Al-Hajj and Mustafa Jarrar. 2021. [ArabGlossBERT: Fine-Tuning BERT on Context-Gloss Pairs for WSD](#). In *Proceedings of the International Conference on Recent Advances in Natural Language Processing (RANLP 2021)*, pages 40–48, Online. INCOMA Ltd.
- Alaa Aljabari, Lina Duaibes, Mustafa Jarrar, and Mohammed Khalilia. 2024. [Event-Arguments Extraction Corpus and Modeling using BERT for Arabic](#). In *Proceedings of the Second Arabic Natural Language Processing Conference (ArabicNLP 2024)*, Bangkok, Thailand. Association for Computational Linguistics.
- Mohammed Alshahrani, Fuxi Zhu, Mohammed Alghaili, Eshrag Refaee, and Mervat Bamiah. 2018. [Borsah: An arabic sentiment financial tweets corpus](#). In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, Paris, France. European Language Resources Association (ELRA).
- Yusuf Arslan, Kevin Allix, Lisa Veiber, Cedric Lothritz, Tegawendé F Bisseyandé, Jacques Klein, and Anne Goujon. 2021. [A comparison of pre-trained language models for multi-class text classification in the financial domain](#). In *Companion Proceedings of the Web Conference 2021*, pages 260–268.
- Sylvio Barbon Junior, Paolo Ceravolo, Sven Groppe, Mustafa Jarrar, Samira Maghool, Florence Sèdes, Soror Sahri, and Maurice Van Keulen. 2024. [Are Large Language Models the New Interface for Data Pipelines?](#) In *Proceedings of the International Workshop on Big Data in Emergent Distributed Environments, BiDEDE '24*, New York, NY, USA. Association for Computing Machinery.
- Loïc Barrault, Yu-An Chung, Mariano Coria Meglioli, David Dale, Ning Dong, Mark Duppenthaler, Paul-Ambroise Duquenne, Brian Ellis, Hady Elsahar, Justin Haaheim, John Hoffman, Min-Jae Hwang, Hirofumi Inaguma, Christopher Klaiber, Ilia Kulikov, Pengwei Li, Daniel Licht, Jean Maillard, Ruslan Mavlyutov, Alice Rakotoarison, Kaushik Ram Sadagopan, Abinesh Ramakrishnan, Tuan Tran, Guillaume Wenzek, Yilin Yang, Ethan Ye, Ivan Evtimov, Pierre Fernandez, Cynthia Gao, Prangthip Hansanti, Elahe Kalbassi, Amanda Kallet, Artyom Kozhevnikov, Gabriel Mejia Gonzalez, Robin San Roman, Christophe Touret, Corinne Wong, Carleigh Wood, Bokai Yu, Pierre Andrews, Can Balioglu, Peng-Jen Chen, Marta R. Costa-jussà, Maha Elbayad, Hongyu Gong, Francisco Guzmán, Kevin Heffernan, Somya Jain, Justine Kao, Ann Lee, Xutai Ma, Alexandre Mourachko, Benjamin Peloquin, Juan Pino, Sravya Popuri, Christophe Ropers, Safiyyah Saleem, Holger Schwenk, Anna Y. Sun, Paden Tomasello, Changhan Wang, Jeff Wang, Skyler Wang, and Mary Williamson. 2023. [Seamless: Multilingual expressive and streaming speech translation](#). *CoRR*, abs/2312.05187.
- Iñigo Casanueva, Tadas Temčinas, Daniela Gerz, Matthew Henderson, and Ivan Vulić. 2020. [Efficient intent detection with dual sentence encoders](#). In *Proceedings of the 2nd Workshop on Natural Language Processing for Conversational AI*, pages 38–45, Online. Association for Computational Linguistics.
- Damir Cavar and Matthew Josefy. 2018. [Mapping deep nlp to knowledge graphs: An enhanced approach to analyzing corporate filings with regulators](#). In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, Paris, France. European Language Resources Association (ELRA).
- Kareem Darwish, Nizar Habash, Mourad Abbas, Hend Al-Khalifa, Huseein T. Al-Natsheh, Houda



- Bouamor, Karim Bouzoubaa, Violetta Cavalli-Sforza, Samhaa R. El-Beltagy, Wassim El-Hajj, Mustafa Jarrar, and Hamdy Mubarak. 2021. [A Panoramic survey of Natural Language Processing in the Arab Worlds](#). *Commun. ACM*, 64(4):72–81.
- Mahmoud El-Haj, Paulo Alves, Paul Rayson, Martin Walker, and Steven Young. 2020. Retrieving, classifying and analysing narrative commentary in unstructured (glossy) annual reports published as pdf files. *Accounting and Business Research*, 50(1):6–34.
- Mahmoud El-Haj, Paul Rayson, Paulo Alves, and Steven Young. 2018. Towards a multilingual financial narrative processing system. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, Paris, France. European Language Resources Association (ELRA).
- Mahmoud El-Haj, Paul Rayson, Martin Walker, Steven Young, and Vasiliki Simaki. 2019. In search of meaning: Lessons, resources and next steps for computational analysis of financial discourse. *Journal of Business Finance & Accounting*, 46(3-4):265–306.
- Mahmoud El-Haj, Paul Rayson, Steven Young, and Martin Walker. 2014. Detecting document structure in a very large corpus of uk financial reports. In *LREC*, volume 14, pages 1335–1338.
- Mahmoud El-Haj, Antonio Moreno Sandoval, and José Antonio Jiménez Millán. 2021. Machine learning models for classifying spanish beaters and non-beaters financial reports. In *Financial narrative processing in spanish*, pages 179–198. Tirant Humanidades.
- Karim El Haff, Mustafa Jarrar, Tymaa Hammouda, and Fadi Zaraket. 2022. [Curras + Baladi: Towards a Levantine Corpus](#). In *Proceedings of the International Conference on Language Resources and Evaluation (LREC 2022)*, Marseille, France.
- Tymaa Hammouda, Mustafa Jarrar, and Mohammed Khalilia. 2024. [SinaTools: Open Source Toolkit for Arabic Natural Language Understanding](#). In *Proceedings of the 2024 AI in Computational Linguistics (ACLING 2024)*, Procedia Computer Science, Dubai. ELSEVIER.
- Go Inoue, Bashar Alhafni, Nurpeiis Baimukan, Houda Bouamor, and Nizar Habash. 2021. [The interplay of variant, size, and task type in arabic pre-trained language models](#). In *Proceedings of the Sixth Arabic Natural Language Processing Workshop, WANLP 2021, Kyiv, Ukraine (Virtual), April 9, 2021*, pages 92–104. Association for Computational Linguistics.
- Mustafa Jarrar. 2008. [Towards Effectiveness and Transparency in e-Business Transactions, An Ontology for Customer Complaint Management](#), chapter 7. IGI Global.
- Mustafa Jarrar. 2021. [The Arabic Ontology - An Arabic Wordnet with Ontologically Clean Content](#). *Applied Ontology Journal*, 16(1):1–26.
- Mustafa Jarrar, Muhammad Abdul-Mageed, Mohammed Khalilia, Bashar Talafha, AbdelRahim Elmadany, Nagham Hamad, and Alaa’ Omar. 2023a. [WojoodNER 2023: The First Arabic Named Entity Recognition Shared Task](#). In *Proceedings of the 1st Arabic Natural Language Processing Conference (ArabicNLP), Part of the EMNLP 2023*, pages 748–758. ACL.
- Mustafa Jarrar, Ahmet Birim, Mohammed Khalilia, Mustafa Erden, and Sana Ghanem. 2023b. [ArBanking77: Intent Detection Neural Model and a New Dataset in Modern and Dialectical Arabic](#). In *Proceedings of the 1st Arabic Natural Language Processing Conference (ArabicNLP), Part of the EMNLP 2023*, pages 276–287. ACL.
- Mustafa Jarrar, Ahmet Birim, Mohammed Khalilia, Mustafa Erden, and Sana Ghanem. 2023c. [ArBanking77: Intent detection neural model and a new dataset in modern and dialectical Arabic](#). In *Proceedings of ArabicNLP 2023*, pages 276–287, Singapore (Hybrid). Association for Computational Linguistics.
- Mustafa Jarrar, Nagham Hamad, Mohammed Khalilia, Bashar Talafha, AbdelRahim Elmadany, and Muhammad Abdul-Mageed. 2024. [WojoodNER 2024: The Second Arabic Named Entity Recognition Shared Task](#). In *Proceedings of the Second Arabic Natural Language Processing Conference (ArabicNLP 2024)*, Bangkok, Thailand. Association for Computational Linguistics.
- Mustafa Jarrar, Mohammed Khalilia, and Sana Ghanem. 2022. [Wojood: Nested Arabic Named Entity Corpus and Recognition using BERT](#). In *Proceedings of the International Conference on Language Resources and Evaluation (LREC 2022)*, Marseille, France.
- Mustafa Jarrar, Sanad Malaysha, Tymaa Hammouda, and Mohammed Khalilia. 2023d. [SALMA: Arabic Sense-annotated Corpus and WSD Benchmarks](#). In *Proceedings of the 1st Arabic Natural Language Processing Conference (ArabicNLP), Part of the EMNLP 2023*, pages 359–369. ACL.
- Mustafa Jarrar, Fadi Zaraket, Tymaa Hammouda, Daanish Masood Alavi, and Martin Waahlich. 2023e. [Lisan: Yemeni, Irqi, Libyan, and Sudanese Arabic Dialect Copora with Morphological Annotations](#). In *The 20th IEEE/ACS International Conference on Computer Systems and Applications (AICCSA)*. IEEE.
- Rasmus Jørgensen, Oliver Brandt, Mareike Hartmann, Xiang Dai, Christian Igel, and Desmond Elliott. 2023. [Multifin: A dataset for multilingual financial nlp](#). In *Findings of the Association for Computational Linguistics: EACL 2023*, pages 894–909.
- Rasmus Kær Jørgensen and Christian Igel. 2021. Machine learning for financial transaction classification across companies using character-level word embeddings of text fields. *Intelligent Systems in Accounting, Finance and Management*, 28(3):159–172.

- Mohammed Khalilia, Sanad Malaysha, Reem Suwaileh, Mustafa Jarrar, Alaa Aljabari, Tamer Elsayed, and Imed Zitouni. 2024. [ArabicNLU 2024: The First Arabic Natural Language Understanding Shared Task](#). In *Proceedings of the Second Arabic Natural Language Processing Conference (ArabicNLP 2024)*, Bangkok, Thailand. Association for Computational Linguistics.
- Oi Yee Kwong. 2018. Analysis and annotation of english-chinese financial terms for benchmarking and language processing. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, Paris, France. European Language Resources Association (ELRA).
- Matthew Lamm, Arun Chaganty, Dan Jurafsky, Christopher D. Manning, and Percy Liang. 2018. Qsrl: A semantic role-labeling schema for quantitative facts. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, Paris, France. European Language Resources Association (ELRA).
- Haneen Liqreina, Mustafa Jarrar, Mohammed Khalilia, Ahmed Oumar El-Shangiti, and Muhammad Abdul-Mageed. 2023. [Arabic Fine-Grained Entity Recognition](#). In *Proceedings of the 1st Arabic Natural Language Processing Conference (ArabicNLP), Part of the EMNLP 2023*, pages 310–323. ACL.
- Tim Loughran and Bill McDonald. 2016. Textual analysis in accounting and finance: A survey. *Journal of Accounting Research*, 54(4):1187–1230.
- Sanad Malaysha, Mustafa Jarrar, and Mohammed Khalilia. 2023. [Context-Gloss Augmentation for Improving Arabic Target Sense Verification](#). In *Proceedings of the 12th International Global Wordnet Conference (GWC2023)*. Global Wordnet Association.
- Sanad Malaysha, Mustafa Jarrar, and Mohammed Khalilia. 2024. [NLU-STR at SemEval-2024 Task 1: Generative-based Augmentation and Encoder-based Scoring for Semantic Textual Relatedness](#). In *In Proceedings of the SemEval 2024 Shared Task 1 (Semantic Relatedness)*. ACL.
- Abir Naskar, Tirthankar Dasgupta, Sudeshna Jana, and Lipika Dey. 2022. Atl at fincausal 2022: Transformer based architecture for automatic causal sentence detection and cause-effect extraction. In *Proceedings of the 4th Financial Narrative Processing Workshop (FNP 2022)*, pages 131–134, Marseille, France. European Language Resources Association.
- Amal Nayouf, Mustafa Jarrar, Fadi zaraket, Tymaa Hammouda, and Mohamad-Bassam Kurdy. 2023. [Nābra: Syrian Arabic Dialects with Morphological Annotations](#). In *Proceedings of the 1st Arabic Natural Language Processing Conference (ArabicNLP), Part of the EMNLP 2023*, pages 12–23. ACL.
- Gakuto Tsutsumi and Takehito Utsuro. 2022. Detecting causes of stock price rise and decline by machine reading comprehension with bert. In *Proceedings of the 4th Financial Narrative Processing Workshop (FNP 2022)*, pages 27–35, Marseille, France. European Language Resources Association.
- Elias Zavitsanos, Aris Kosmopoulos, George Giannakopoulos, Marina Litvak, Blanca Carbajo-Coronado, Antonio Moreno-Sandoval, and Mahmoud El-Haj. 2023. The financial narrative summarisation shared task (fns 2023). In *Proceedings of the 2023 IEEE International Conference on Big Data (BigData)*, pages 2890–2896, Sorrento, Italy.
- Nadhém Zmandar, Tobias Daudert, Sina Ahmadi, Mahmoud El-Haj, and Paul Rayson. 2022. [Cofif plus: A french financial narrative summarisation corpus](#). In *Proceedings of the 4th Financial Narrative Processing Workshop (FNP 2022)*, pages 1622–1639, Marseille, France. European Language Resources Association.
- Nadhém Zmandar, Mahmoud El-Haj, and Paul Rayson. 2021. [Multilingual financial word embeddings for arabic, english and french](#). In *2021 IEEE International Conference on Big Data (Big Data)*, pages 4584–4589.
- Nadhém Zmandar, Mo El-Haj, and Paul Rayson. 2023. [FinAraT5: A text to text model for financial Arabic text understanding and generation](#). In *Proceedings of the 4th Conference on Language, Data and Knowledge*, pages 262–273, Vienna, Austria. NOVA CLUNL, Portugal.
- Jan Štihec, Senja Pollak, and Martin Žnidaršič. 2021. [Preliminary experimentation with combinations and extensions of forward-looking sentence detection wordlists](#). In *Proceedings of the 3rd Financial Narrative Processing Workshop*, pages 26–30, Lancaster, United Kingdom. Association for Computational Linguistics.
- Martin Žnidaršič, Jasmina Smailović, Jan Gorše, Miha Grčar, Igor Mozetič, and Senja Pollak. 2018. Trust and doubt terms in financial tweets and periodic reports. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, Paris, France. European Language Resources Association (ELRA).