UTILIZING GPT 3.5 FOR ARABIC INTENT CLASSIFICATION WITH PROMPTING

FATMA JAMAL HABIB HABIB [1], SALLY S.ISMAIL [2], ABEER M.MAHMOUD[3]
[1,2,3] Faculty of Computer and Information Science Ain Shams University, Department of Computer Science, Cairo
E-mail: [1] Fatma.Jamal.Habib@cis.asu.edu.eg, [2] Sallysaad@cis.asu.edu.eg, [3]Abeer.Mahmoud@cis.asu.edu.eg

ABSTRACT

Intent classification in natural language processing (NLP) is crucial for identifying user intents from their utterances. This task is particularly challenging for Arabic due to its complex morphology and high ambiguity. Recent advancements in NLP, including deep learning and transfer learning, have improved Arabic intent classification. However, a significant gap still exists compared to major languages like English. This paper proposes a new Arabic intent classification based on 1-shot and 3-shot techniques for investigating prompting strategies for Arabic in specific . The ’Temperature’ parameter is employed to modulate probability distributions and guide the model in classifying Arabic sentences into predefined categories. Results show that prompting, especially contrastive with dynamic reasoning, outperforms fine-tuning in both accuracy and resource efficiency, high- lighting the effectiveness of prompting techniques in enhancing Arabic NLP applications. Contrastive with Dynamic reasoning reported accuracy of 0.98% for 1- shot and 100% for 3-shots while fine tuning recorded 0.87% for training of the used Arabic intent dataset.

Keywords: *Natural Language, Prompting, Intent classification, Few shots, Deep learning, Transfer learning, Softmax, Temperature Scaling, Morphology, Ambiguity*

.

1. INTRODUCTION

Intent classification is a fundamental task in natural language processing (NLP) that aims to identify the intent of a user’s utterance [1]. This task is particularly challenging for Arabic, as it is a morphologically rich language with a high degree of ambiguity. Arabic is a semitic language with very rich morphological structures [2], thus posing a challenge to intent classifiers due to its wide specters of words depending on grammatical and other contextual characteristics. This tends to cause problems in attempts to interpret the actual intention of the given messages because the same words may mean different things at different times. Furthermore, there are some peculiarities in the Arabic language: the language is very ambiguous, which means it is not difficult to encounter multiple interpretations of the given expression based on the great intention of the author and the humble level of the reader’s language proficiency; The aforementioned factors further complicate the intent classification process.[3] However, in recent years, enormous amount of work has been done to classify Arabic intent[4][5][6]. This progress has been due to advances in new methods of NLP, such as deep learning [7][8][9] and transfer learning [10][11][12]. The use of deep learning models has been very helpful in the classification of intents in Arabic since the ability of the model to understand the relation- ship between words and intents is realized. Transfer learning has also been shown to work for Arabic intent classification as it enables models to be trained on large volumes of English text and later on the volumes of Arabic text are finetuned [13]. In natural language processing applications, the known task of intent classification has revolutions in how humans can interact with AI systems [14][15]. Much progress has been achieved in intent classification for large languages, most of which are in developed countries such as English, but little has been done for languages which have different characteristics from the rest, most of which are in Arabic. The issues related to the cursive nature of languages like Arabic can be understood within a larger context: The language’s complex grammar and deep cultural nuances make it especially challenging to understand and interpret intent accurately [16][17].

Despite these advances, most existing approaches to Arabic intent classification rely on supervised deep learning and transfer learning techniques that require large annotated datasets and substantial computational resources. Moreover, limited attention has been given to the use of large language models and prompt-based learning for Arabic, particularly under few-shot settings. As a result, the effectiveness of prompt engineering strategies for Arabic intent classification remains largely unexplored.

This study addresses this gap by investigating prompt-based intent classification for Arabic using GPT-3.5 under few-shot learning conditions. Specifically, the study evaluates multiple prompting strategies, including normal prompting, contrastive prompting, static reasoning prompting, dynamic reasoning prompting, and contrastive prompting with dynamic reasoning, using both Arabic and English reasoning formulations. The performance of 1-shot and 3-shot prompting strategies is systematically compared against fine-tuning-based approaches to assess both classification accuracy and resource efficiency.

While previous studies on Arabic intent classification have primarily focused on supervised fine-tuning approaches requiring substantial labeled data and computational resources, this work departs from that paradigm by investigating prompt-based learning using large language models under few-shot settings. By leveraging the same dataset used in prior fine-tuning studies, this work enables a direct and fair comparison, highlighting the advantages of prompting strategies as a more resource-efficient alternative for Arabic intent classification.

The main contributions of this work are as follows: (1) a comprehensive evaluation of GPT-3.5-based prompting strategies for Arabic intent classification, (2) an empirical comparison between prompting and fine-tuning approaches for Arabic under limited data conditions, and (3) an analysis demonstrating the effectiveness of contrastive prompting with dynamic reasoning as a resource-efficient alternative for Arabic intent classification.

The remainder of this paper is organized as follows. Section 2 reviews related work on intent classification. Section 3 describes the proposed methodology, including the different prompting strategies and few-shot configurations. Section 4 presents the experimental results and analysis. Section 5, present the experimental cases. Section 6 discusses the findings, study limitations. Finally, Section 7 concludes the paper and outlines future research directions.

2. RELATED WORK

In this section, an overview of the existing research and literature related to Arabic intent classification in natural language processing (NLP) is provided, highlighting the key studies that contributed to understanding this specialized domain.

M. F. Alruily et al. [18] optimized a deep learning system for Arabic, with a focus on intent categorization in Arabic text. It emphasized the difficulty of effectively recognizing user intent in Arabic chat-bots due to restricted tools. To solve this issue, they presented the ArRASA framework, which used tokenization, featurization, intent categorization, and entity extraction to enhance accuracy while resolving language difficulties. It outperformed the previous 85% record by achieving 92% accuracy on a test dataset and effectively handling noise and ambiguity in Arabic text. This result demonstrated ArRASA's ability to improve Arabic language processing tasks, offering more precise and context- aware responses to Arabic-speaking consumers.

Mahdi et al. [19], presented various machine learning algorithms and focused on Arabic intent classification using, aiming to automatically categorize Arabic text based on its un- derlying intent. The study compared four algorithms—Naive Bayes, K-Nearest Neighbors, Decision Tree, and Support Vector Machine (SVM)—using Modern Standard Arabic texts. Among these, SVM emerged as the most effective, achieving an impressive 96.7% accuracy. The authors highlighted SVM's superiority in handling com- plex Arabic text features, including non-linear relation- ships and high-dimensional spaces. These findings held significant promise for improving various Arabic natural language processing (NLP) applications, such as chatbots and sentiment analysis tools. Despite limitations such as reliance on a single dataset, the study provided valuable insights into enhancing Arabic NLP capabilities, paving the way for further advancements in the field. Abdallah.

M. Bashir et al. [20], a novel approach to Arabic intent classification was presented, focusing on the domain of home automation. The authors highlighted the limited advancement in Arabic NLU compared to English, which largely benefited from deep learning techniques. The study introduced a neural network-based solution using Long Short-Term Memory (LSTM) and Convolutional Neural Networks (CNNs) for intent classification, along with a Slot Tagger for entity recognition. The models were trained and evaluated on a dataset collected through an online survey, specifically tailored for home automation commands. Results showed that the LSTM model slightly outperformed the CNN model, achieving an F-Score of 92.01 for intent classification. For entity extraction, a Bidirectional LSTM with character-based word embed- dings achieved an impressive F-Score of 94.0, comparable to benchmarks in English. These findings suggest that neural network approaches hold promise for enhancing Arabic NLU, particularly in task-oriented dialogue systems such as those for home automation.

Xu Han et al. [21], focused on improving intent classification using natural language prompts and rule-based constraints. The Prompt Tuning with Rules (PTR) method combines prompt-based techniques with rule-based approaches to guide models toward desired behaviors, like avoiding toxic language or adhering to specific guidelines, while maintaining high accuracy. In intent classification, PTR generates prompts designed to elicit text intents, constructed based on pre- defined rules capturing key intent features. Evaluations on ATIS and SNIPS datasets in English showed PTR achieved high intent classification accuracies: 97.52% on ATIS, outperforming existing methods, and 98.2% on SNIPS, demonstrating its effectiveness. PTR's effective- ness was notable in scenarios requiring domain knowledge incorporation or specific constraint enforcement, making it promising for improving intent classification tasks.

Yi Zhu, et al. [22], proposed the Prompt-Learning approach for Short Text classification (PLST) to enhance learning tasks and address challenges in short text classification. The research aimed to improve short text classification by combining short text and external knowledge from Probase. The method, called PLST, involved concept retrieval, verbalizer construction, and text classification. PLST outperformed baseline methods in accuracy and F1-score on five benchmark datasets. It consistently outperformed other fine-tuning methods, including Knowledgeable Prompt- tuning (KPT), showcasing the effectiveness of expanding the verbalizer with external knowledge. This research emphasized the importance of incorporating external knowledge for enhancing short text classification performance.

Seham Basabin, et al. [23], proposed a method to improve feature extraction for Arabic text in few/zero-shot learning scenarios by augmenting input text with label semantics. The approach utilized the AraBERT pre-trained language model to extract features from input text and enhance them with label information, aiming to better capture the semantic meaning of the text. Experimental results on Arabic datasets for sentiment analysis, sarcasm detection, and stance detection showed that the proposed method outperformed baseline models and models trained without label augmentation. The approach achieved better performance in sentiment analysis tasks, with an accuracy of up to 0.874 in Arabic sarcasm detection in a zero-shot learn- ing setting, demonstrating its effectiveness in enhancing feature extraction for Arabic text classification tasks with limited labeled data.

Manoj Kumar, et al. [24], proposed a novel approach, called ProtoDA, for intent classification in natural language processing with limited training data. The approach combined meta-learning with data augmentation to improve classification performance. By training a conditional generator network with prototypical networks, ProtoDA generated task-specific samples, reducing sampling bias and improving generalization. The study demonstrated that increasing variability in training tasks significantly enhanced classification performance. In experiments, ProtoDA achieved up to 6.49% and 8.53% relative F1-score improvements over the best-performing systems in 5-shot and 10-shot learning scenarios, respectively, showcasing its effectiveness in few-shot learning settings.

Muhammad Khalifa, et al. [25], introduced a method to improve Arabic sequence labeling tasks when only a small amount of labeled data was available or when the target dialect differed from the pre-trained model's dialect. The approach combined self-training, a semi-supervised learning technique, with a pre-trained language model. By iteratively refining the model on pseudo-labeled data, the method aimed to enhance the model's ability to generalize to different Arabic dialects and perform well with limited labeled data. Experimental results demonstrated that the proposed approach outperformed baseline models on various Arabic sequence labeling tasks, achieving significant improvements in F1 scores, especially in zero- shot and few-shot learning scenarios, and across different dialects.

2.1 Synthesis and Research Gap

The existing literature on Arabic intent classification has progressed significantly, evolving from traditional statistical approaches such as Support Vector Machines (SVMs) to deep learning architectures including LSTMs and CNNs. Despite these advancements, several critical gaps remain.

First, data dependency remains a major challenge. Most high-performing Arabic models, such as the ArRASA framework, rely on large amounts of task-specific annotated data, which is difficult to obtain, particularly for diverse Arabic dialects.

Second, methodological limitations persist. Current Arabic-centric studies primarily adopt fine-tuning or feature augmentation strategies. While effective, these approaches often suffer from catastrophic forgetting and require retraining for each new intent domain, leading to high computational costs and limited scalability.

Third, there is a lack of research on Arabic prompting. Although prompt-based techniques such as PTR and PLST have demonstrated strong performance in English, limited work has explored how large language models can be effectively prompted to handle Arabic morphological complexity and dialectal variation for intent classification.

These limitations indicate a clear research gap in developing prompt-based learning strategies specifically tailored for Arabic intent recognition.

2.2 The Proposed Approach: Arabic-Specific Prompting

In contrast to the studies summarized above, this paper proposes a prompting-based framework specifically optimized for Arabic intent classification. Rather than modifying model parameters through extensive fine-tuning, the proposed approach leverages the inherent knowledge of large language models by designing natural language prompt templates aligned with Arabic linguistic structures.

By adopting few-shot prompting, the proposed method addresses the data scarcity issues reported in prior studies, achieving high classification accuracy with minimal labeled examples. Furthermore, the approach incorporates structured prompt formulations that map model outputs directly to predefined intent categories, bridging the gap between general-purpose large language models and the specialized task of Arabic intent detection.

.

2.3 Comparative Analysis

A synthesis of the reviewed literature is presented in Table I, which categorizes existing methodologies based on their architecture, target data, and reported performance. Several important trends emerge.

First, there has been a methodological evolution from classical machine learning models, such as SVMs achieving up to 96.7% accuracy on Modern Standard Arabic, to deep learning architectures such as ArRASA and LSTMs, which demonstrate strong performance in both intent classification and entity extraction.

Second, while prompt-based methods such as PTR and PLST report state-of-the-art results on English benchmarks (e.g., 98.2% accuracy on the SNIPS dataset), these methods remain largely unexplored for Arabic text, revealing a significant opportunity for adaptation and investigation.

Third, recent Arabic-focused studies increasingly address low-resource and multi-dialectal scenarios using techniques such as label-semantic augmentation and self-training. However, these approaches still rely heavily on supervised learning paradigms.

Overall, this comparative analysis reveals a clear research gap: the absence of a dedicated prompting framework that combines the effectiveness of prompt-based learning with the linguistic and data-scarce characteristics of the Arabic language. The proposed work directly addresses this gap by bridging prompt-based learning and Arabic intent classification.

Table I: Summary Of Related Work And Existing Methodologies

Paper	Year	Methodology	Data	Results	Key Findings
ArRASA: Channel optimization for deep learning-based Arabic NLU chatbot framework [18]	2023	Deep learning with channel optimization	Arabic text	- 96% accuracy for intent classification - 94% accuracy for entity extraction	- Improved accuracy over previous methods - Effective handling of noise and ambiguity
Intent Arabic text categorization based on different machine learning and term frequency [19]	2022	Machine learning (SVM)	Arabic text	96.7% accuracy	- SVM effective for complex Arabic text features - Improved Arabic NLP capabilities
Implementation of a neural network component for Arabic dialogue systems for home automation [20]	2018	Neural Networks (LSTM, CNNs)	Arabic dia- logue data	- F-Score of 92.01 for Intent classification (LSTM) - F -Score o f 94.0 for entity extraction (Bidirectional LSTM)	-Neural networks effective for Arabic NLU in home automation - LSTM outperforms CNN for intent classification
PTR: Prompt Tuning with Rules for Text Classification [21]	2021	Prompt Tuning with Rules	English text (ATIS, SNIPS)	- 97.52% accuracy on ATIS - 98.2% accuracy on SNIPS	- Improved intent classification accuracy - Effective for incorporating do- main knowledge and enforcing constraints

Prompt-Learning fo r Short Text Classification [22]	2022	Prompt Learning	Short text data	- High accuracy on ATIS and SNIPS datasets (English)	-Effective for incorporating domain knowledge - Improved short text classification
Enhancing Arabic-text feature extraction utilizing label-semantic augmentation in few/zero-shot learning [23]	2023	Label-semantic augmentation for few/zero-shot learning	Arabic text	-Improved performance On sentiment analysis, sarcasm detection, and stance detection	- Effective for improving feature extraction with limited data
ProtoDA: Efficient Transfer Learning for Few-Shot Intent Classification [24]	2021	Transfer Learning for Few-Shot Intent Classification	Not spec-ified (intent classifi-cation)	- Improved F1-score over baselines in 5-shot and 10-shot learning	- Effective for intent classification with limited data
Self-Training Pre - Trained Language Models for Zero- and Few-Shot Multi-Dialectal Arabic Sequence Labeling [25]	2021	Self-training pre-trained language models for zero/few-shot learning	Multi-dialectal Arabic text	- Improved F1 scores on various sequence labeling tasks	- Effective for handling limited data and different Arabic dialects

3. PROPOSED METHODOLOGY

The use of prompting in the context of Large Language Models (LLMs) has gained prominence due to its potential to enhance model performance and adaptability. Prompting involves providing specific instructions or queries to guide the model's generation of responses, making it a versatile approach for various tasks.

3.1 Overview of Prompt-Based Methodology

The research methodology involved a comprehensive exploration of distinct prompting strategies for Arabic intent classification, employing both 1-shot and 3-shot techniques. Different prompt types, including normal prompting, contrastive prompting, static reasoning prompting, contrastive, Dynamic reasoning and Contrastive with dynamic reasoning prompting (Arabic, English), were systematically examined. These prompts were strategically designed to assist the model in categorizing Arabic sentences into specific classes, such as (Question, Not).

3.2 Conceptual Framework of Prompt-Based Arabic Intent Classification

Figure 1 illustrates the conceptual framework of the proposed prompt-based Arabic intent classification approach. The model represents the flow of information from input utterances to predicted intent labels and highlights the role of different prompting strategies in guiding the language model's behavior. Arabic user utterances serve as the input to the system. These utterances are combined with a prompt template that includes task instructions, few-shot examples, and optional reasoning explanations. Depending on the experimental configuration, the prompt may include normal examples, contrastive examples, static reasoning descriptions, dynamic reasoning instructions, or a combination of contrastive and dynamic reasoning in either Arabic or English. The constructed prompt is then provided to the GPT-3.5 model, where inference is performed using a fixed temperature setting to ensure deterministic output. The model generates an intent prediction by classifying the input utterance into either the target intent or the *Not* class. The predicted intent is finally compared with the ground truth label to evaluate classification accuracy. This conceptual framework

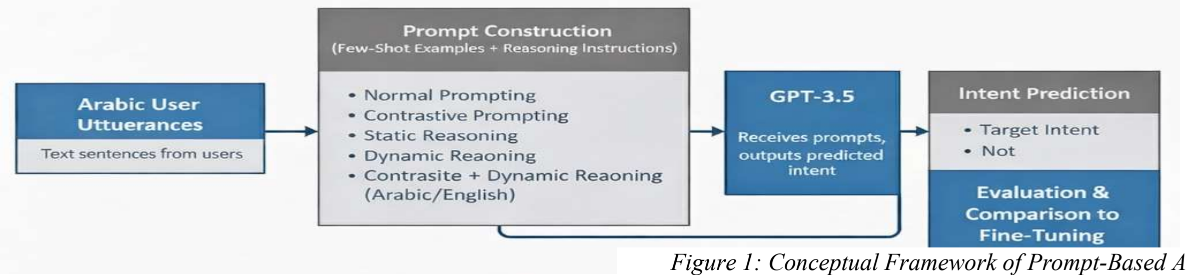

Arabic User Uttuerances
Text sentences from users
Prompt Construction
(Few-Shot Examples + Reasoning Instructions)
• Normal Prompting
• Contrastive Prompting
• Static Reasoning
• Dynamic Reaoning
• Contrasite + Dynamic Reaoning (Arabic/English)
GPT-3.5
Receives prompts, outputs predicted intent
Intent Prediction
• Target Intent
• Not
Evaluation & Comparison to Fine-Tuning

Figure 1: Conceptual Framework of Prompt-Based Arabic Intent Classification

highlights the central hypothesis of this study: that carefully designed prompting strategies can effectively guide large language models to perform Arabic intent classification with high accuracy while reducing the need for fine-tuning and large labeled datasets.

3.3 Model Configuration and Temperature Settings

The ’Temperature’ parameter plays a crucial role in this context. By scaling the randomness in the distribution, it affects the certainty of the model’s predictions. When the temperature is set to 0, it ”hardens” the distribution, making it more deterministic and less random, as shown in figure 2[26]. This leads to increased confidence in the model’s predictions. With the intentional incorporation of ”Temperature = 0” to modulate probability distributions, the prompts were designed to guide the model in classifying Arabic sentences into predefined categories.

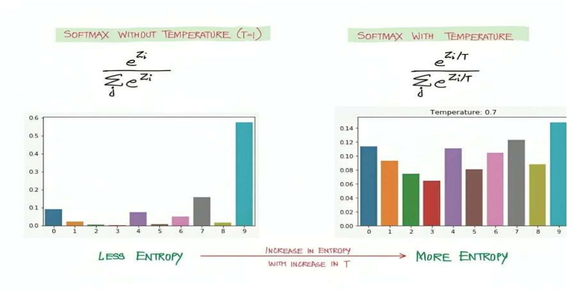


Figure 2: visualization of temperature effect

$$p(x_i) = \frac{e^{\frac{x_i}{T}}}{\sum_{j=1}^{V} e^{\frac{x_j}{T}}} \qquad Eq\ (1)$$

When the temperature T is close to 0, the exponential in the equation(1) becomes very large, especially for the elements with higher values. This is because e(xi/T) grows rapidly as T approaches 0. As a result, the probability distribution becomes ”hardened” or more deterministic because the probability of the highest element in the vector dominates, making it nearly certain that the model will select the element with the highest value as the output. On the other hand, when the temperature is higher, the exponentials are smaller and more evenly distributed, leading to a more uniform distribution of probabilities across all elements in the vector. This higher temperature allows for more randomness in the model’s predictions.

3.4 Dataset Description and Experimental Setup

In this study on Arabic intent classification, the task was initially undertaken through fine-tuning in a prior work. Building on this foundation, the present research introduced a novel approach by applying different prompting types using ChatGPT 3.5. For this study, the proposed method leverages a dataset previously employed in a fine-tuning project [27], serving as a benchmark for comparison. In the prior fine-tuning study [27], a supervised deep learning model based on a transformer architecture was trained on the same dataset containing 16 predefined intents. The model was fine-tuned using a standard train–test split (90% training, 10% testing) and achieved a baseline accuracy of 0.87% on the intent classification task. This fine-tuned model serves as the reference baseline in the present study, enabling a direct and fair comparison between traditional supervised fine-tuning and prompt-based few-shot learning approaches. This dataset encompasses 16 distinct in- tents, including but not limited to questions, translations, and others. Each intent within the dataset consists of a number of sentences, exemplifying the linguistic variations within the intent category. This inclusion of diverse sentences enhances the model’s exposure to the intricacies of each intent type, contributing to a more nuanced and accurate classification. To ensure a robust evaluation, the proposed method partitioned the dataset into training and test sets. The training set comprises 90% of the data, while the remaining 10% forms the test set which is the same as the prior work. The primary objective was to evaluate and compare the accuracy of the proposed prompting techniques against the baseline accuracy established through fine-tuning in the previous work. the comparison aimed to assess the effectiveness of the prompting in enhancing the model’s performance in Arabic intent classification. To quantitatively illustrate these findings, an accuracy comparison table [2] was generated, presenting a clear and comprehensive overview of how the prompting using 3 shots outperformed the baseline accuracy achieved through fine-tuning.

Table II: accuracy comparison table Prompting Vs Fine- Tuning

Methods	Accuracy
Fine Tuning	0.8788%
Normal Prompting (1 shot)	0.85 %
Normal Prompting (3 shots)	0.92 %

Prompting not only achieves higher accuracy compared to fine-tuning but also proves to be superior in terms of resource efficiency. It operates as a form of implicit fine-tuning, notably through the innovative concept of In-Context Learning (ICL), as proposed in the paper [28]. Unlike traditional fine-tuning [29], which explicitly modifies the model parameters through backpropagation, in- context learning achieves a similar effect by introducing additional keys and values into the attention mechanism through prompts (demonstration examples) as shown in the equation(2). Attention mechanisms, on the other hand, areS a way for LLMs to selectively focus on relevant parts of the input text, allowing them to better understand the context and generate more relevant output. This effectively alters the model's attention behavior without directly changing its parameters.

$$F^{\tilde{}} = (W_v + \Delta W_v)XX^T(W_k + \Delta W_k)^T \text{q}$$
$$= (W_{ZSL} + \Delta W_{FT})\text{q}$$
Eq(2)

The first term, (WV+AW)XXT (WK + AWK)q, represents the standard FFT attention computation. The second term, AWFT, represents the ICL updates to the attention weights. By updating the attention keys and values, ICL is able to quickly and efficiently adapt large language models to new tasks without requiring any retraining of the underlying parameters. This makes it a powerful and flexible method for transfer learning in natural language processing.

This approach not only contributes to the understanding of effective intent classification methods in Arabic but also highlights the advances made by incorporating various prompting techniques. In the case of normal prompting, explicit classification instructions were carefully provided , offering clear directives on target classes and providing associated sentences as illustrative examples. This approach aimed to guide the model's understanding and response within predefined categories, establishing a structured foundation for classification. Contrastive prompting took a comparative approach in which sentences from the positive class which represent the target intent or category and sentences from the negative class are injected into the prompt. The contrastive prompt achieved higher accuracy compared to the normal prompting in both 1 and 3 shots, as shown in Table 3, since this method encouraged the model to discern and distinguish between contrasting categories, boost a nuanced understanding of distinctions, leading to a deeper understanding of intent boundaries and making it more adept at handling diverse and unseen data. On the other hand, Reasoning plays a crucial role in enhancing the model's ability to understand and classify information. Unlike normal or contrastive prompting, which may rely on explicit examples or comparative distinctions, reasoning provides a deeper layer of understanding by incorporating contextual details and logical connections. The contrastive and reasoning methods for Arabic intent classification achieved similar accuracy when given a single training example, showcasing their potential for efficient learning. However, the reasoning approach revealed its ability to leverage richer information for more accurate predictions, surpassing the contrastive method's performance when provided with three examples shots as shown in table 3. The proposed method uses two reasoning techniques Static reasoning and Dynamic reasoning, table [4].

Static reasoning prompting, inject additional generic contextual details into the prompt to guide the model's classification process with static reason which is applicable for all the ,”لان معنى الجمل يدل على” classes, such as (as the meaning of the sentences indicates). This approach provides a baseline understanding but struggling with subtle distinctions. Dynamic reasoning prompting provides additional contextual details to guide the model's classification process with dynamically explaining the classification of sentences in each class such as ” لأن الجمل تسأل عن شيء ما,”, With a clearer grasp of each intent, the model becomes better equipped to differentiate between subtle distinctions and make accurate classifications The proposed method also applies the Contrastive prompting with dynamic reasoning prompting in both English and Arabic reasons, to investigate their combined impact on the model's comprehension, table [5]. An accuracy comparison table was generated; to show the difference between each type, the most effective type was Contrastive with dynamic reasoning (Arabic) as shown in the table.

Table III: Normal Vs Contrastive Vs Reasoning Prompting

Methods	Number Of training shots	
	1 shot	*3 shots*
Normal Prompt	0.85	0.92
Contrastive Prompt	0.96	0.96
Reasoning Prompt	0.96	0.98

Table IV: Static Vs Dynamic Reasoning Prompt

Methods	Number Of training shots	
	1 shot	*3 shots*
Static Reason	0.89 %	0.98 %
Dynamic Reason	0.96 %	0.98 %

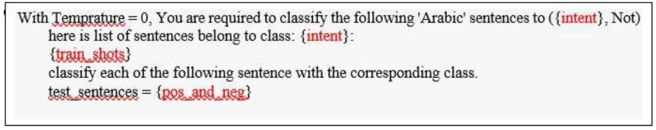
Table V: Contrastive Plus Dynamic Reason prompt (AR Vs EN)

Methods	Number Of training shots	
	1 shot	*3 shots*
Contrastive Plus Dynamic Reason AR	0.98 %	100 %
Contrastive Plus Dynamic Reason EN	0.92 %	0.98 %

4. EXPERIMENTAL RESULTS

The prompts : ("intent" is the class for example ques- tion, "train_shots" OR "train_shots_pos" is the training

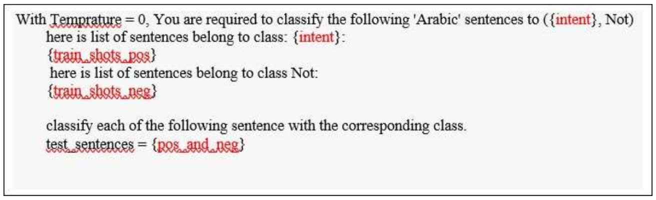
positives sentences," train_shots_neg" is the training neg- ative sentences,"pos_and_neg" is the positive and nega- tive test sentences) Given a temperature parameter T=0, you are required to classify the following Arabic sentences into either intent or Not. You have a list of sentences belonging to the class intent, denoted as train_shot. You should classify each sentence in the set pos_and_neg (which contains both positive and negative test sentences) with the corresponding class intent or Not, based on the training ,3



```
With Temprature = 0, You are required to classify the following 'Arabic' sentences to ({intent}, Not)
    here is list of sentences belong to class: {intent}:
    {train_shots}
    classify each of the following sentence with the corresponding class.
    test_sentences = {pos_and_neg}
```

Figure 3: Normal Prompt

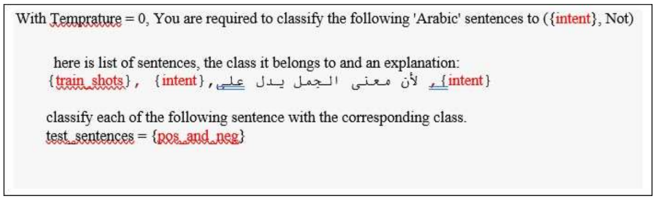
Given a temperature parameter T=0, the task is to classify the following Arabic sentences into either intent or Not. The list of sentences belonging to the class intent is provided as train_shots_pos. The list of sentences belong- ing to the class Not is provided as train_shots_neg. Each sentence in the set pos_and_neg (which contains both positive and negative test sentences) should be classified with the corresponding class. The goal is to classify each test sentence based on the training examples provided in train_shots_pos and train_shots_neg, 4



```
With Temprature = 0, You are required to classify the following 'Arabic' sentences to ({intent}, Not)
    here is list of sentences belong to class: {intent}:
    {train_shots_pos}
    here is list of sentences belong to class Not:
    {train_shots_neg}

    classify each of the following sentence with the corresponding class.
    test_sentences = {pos_and_neg}
```

Figure 4: Contrastive Prompt

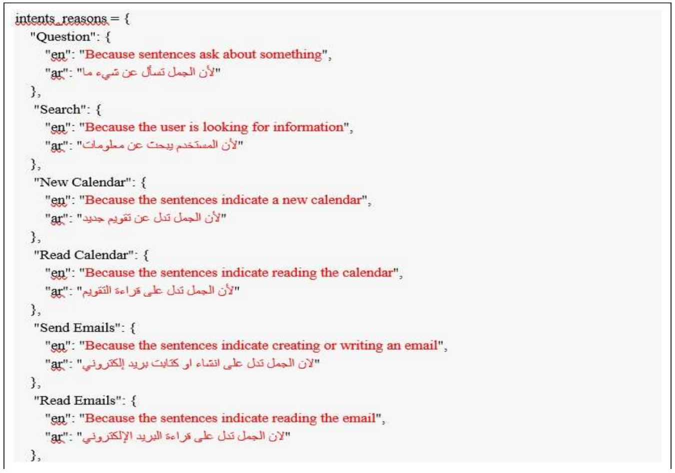
With a temperature parameter T=0 and the Static reason"لأن معنى الجمل يدل على,", the task is to classify the following Arabic sentences into either intent or Not .The pro- vided information includes a list of sentences train_shots, the class each sentence belongs to intent, and an ex- planation in Arabic "لأن معنى الجمل يدل على" that supports the class intent. Each sentence in the set pos_and_neg should be classified with the corresponding class, based on the training examples provided in train_shots and the explanation"لأن معنى الجمل يدل على " that supports the class intent 5.



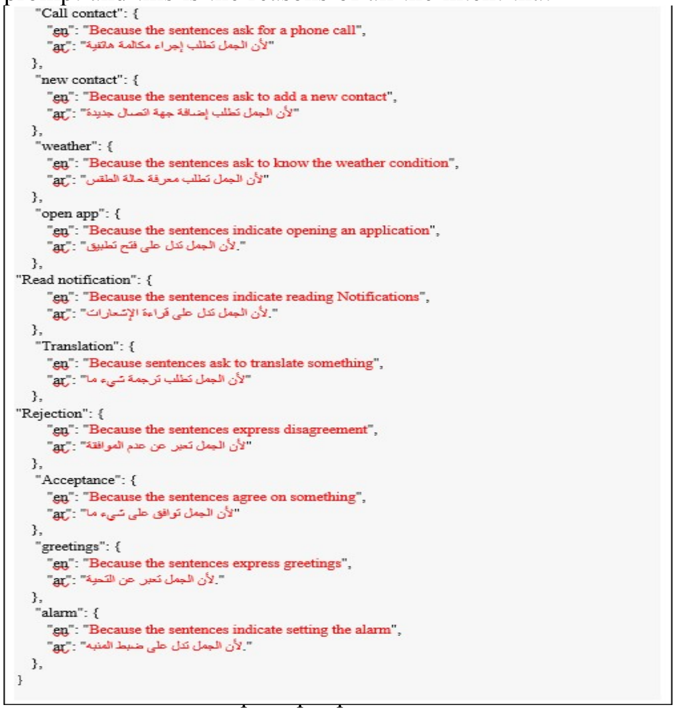
```
With Temprature = 0, You are required to classify the following 'Arabic' sentences to ({intent}, Not)

     here is list of sentences, the class it belongs to and an explanation:
    {train_shots} ,  {intent},لأن معنى الجمل يدل على ,{intent}

    classify each of the following sentence with the corresponding class.
    test_sentences = {pos_and_neg}
```

Figure 5: Static Reasoning Prompt

The "reason" is the dynamic reason injected in the prompt and this is the reasons of all the intent that

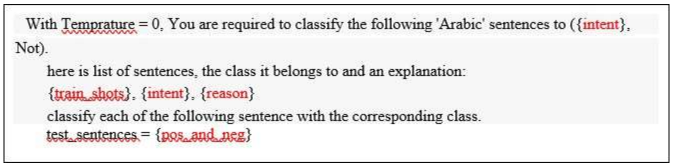


```
intents_reasons = {
  "Question": {
    "en": "Because sentences ask about something",
    "ar": "لأن الجمل تسأل عن شيء ما"
  },
   "Search": {
    "en": "Because the user is looking for information",
    "ar": "لأن المستخدم يبحث عن معلومات"
  },
   "New Calendar": {
    "en": "Because the sentences indicate a new calendar",
    "ar": "لأن الجمل تدل عن تقويم جديد"
  },
   "Read Calendar": {
    "en": "Because the sentences indicate reading the calendar",
    "ar": "لأن الجمل تدل على قراءة التقويم"
  },
   "Send Emails": {
    "en": "Because the sentences indicate creating or writing an email",
    "ar": "لان الجمل تدل على انشاء او كتابت بريد إلكتروني"
  },
   "Read Emails": {
    "en": "Because the sentences indicate reading the email",
    "ar": "لان الجمل تدل على قراءة البريد الإلكتروني"
  },
```



```
    "Call contact": {
      "en": "Because the sentences ask for a phone call",
      "ar": "لأن الجمل تطلب إجراء مكالمة هاتفية"
   },
    "new contact": {
      "en": "Because the sentences ask to add a new contact",
      "ar": "لأن الجمل تطلب إضافة جهة اتصال جديدة"
   },
    "weather": {
      "en": "Because the sentences ask to know the weather condition",
      "ar": "لأن الجمل تطلب معرفة حالة الطقس"
   },
    "open app": {
      "en": "Because the sentences indicate opening an application",
      "ar": ".لأن الجمل تدل على فتح تطبيق"
   },
"Read notification": {
      "en": "Because the sentences indicate reading Notifications",
      "ar": ".لأن الجمل تدل على قراءة الإشعارات"
   },
    "Translation": {
      "en": "Because sentences ask to translate something",
      "ar": "لأن الجمل تطلب ترجمة شيء ما"
   },
"Rejection": {
      "en": "Because the sentences express disagreement",
      "ar": "لأن الجمل تعبر عن عدم الموافقة"
   },
    "Acceptance": {
      "en": "Because the sentences agree on something",
      "ar": "لأن الجمل توافق على شيء ما"
   },
    "greetings": {
      "en": "Because the sentences express greetings",
      "ar": ".لأن الجمل تعبر عن التحية"
   },
    "alarm": {
      "en": "Because the sentences indicate setting the alarm",
      "ar": ".لأن الجمل تدل على ضبط المنبه"
   },
}
```

sentences train_shots, along with the class each sentence belongs to intent, and an explanation reason for why the sentence belongs to the class intent. Each sentence in the set pos_and_neg should be classified with the corresponding class based on the training examples provided in train_shots and the reason for why sentences belong to the class intent.7



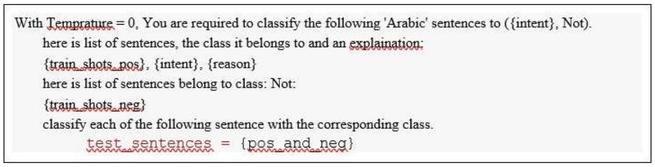
```
 With Temprature = 0, You are required to classify the following 'Arabic' sentences to ({intent},
Not).
      here is list of sentences, the class it belongs to and an explanation:
      {train_shots}, {intent}, {reason}
      classify each of the following sentence with the corresponding class.
      test_sentences = {pos_and_neg}
```

Figure 7: Dynamic Reasoning Prompt

With a temperature parameter T=0, the task is to classify the following Arabic sentences into either intent or Not. The provided information includes:

- A list of sentences belonging to the class intent (train_shots_pos),
- The class each sentence belongs to (intent),
- An explanation for why the sentence belongs to the class intent (reason),
- A list of sentences belonging to the class Not (train_shots_neg).

Each sentence in the set pos_and_neg should be classified with the corresponding class based on the training examples provided in train_shots_pos and train_shots_neg, along with the reason for why sentences belong to the class intent.8



```
With Temprature = 0, You are required to classify the following 'Arabic' sentences to ({intent}, Not).
      here is list of sentences, the class it belongs to and an explaination:
      {train_shots_pos}, {intent}, {reason}
      here is list of sentences belong to class: Not:
      {train_shots_neg}
      classify each of the following sentence with the corresponding class.
            test_sentences = {pos_and_neg}
```

Figure 8: Dynamic Reasoning with Contrastive Prompt

5. EXPERIMENT CASES

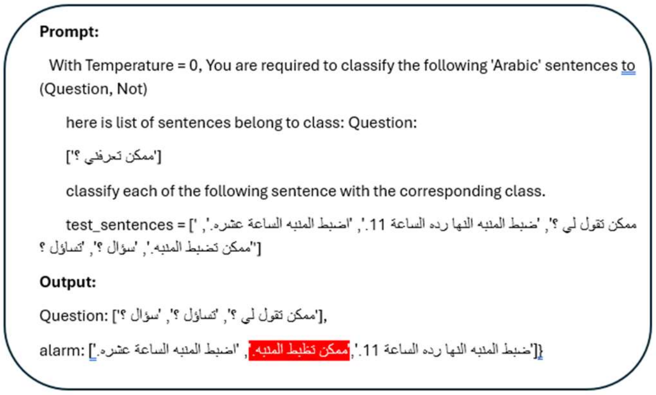
As shown in 9, with a temperature parameter T=0, the task is to classify the following Arabic sentences into either Question or Not. The provided information includes:

- A list of sentences belonging to the class Question: "ممكن تعرفني ؟"
- The test sentences to be classified:" لي
؟, ضبط المنبه النهارده الساعه 11., اضبط المنبه الساعه عشره.,
ممكن تظبط المنبه., سؤال ؟, تساؤل ؟"

The goal is to classify each test sentence into either Question or Not, based on the training example "ممكن تعرفني ؟" and the provided test sentences. The sentences are classified into two categories: "Question" and "alarm."

- Sentences classified as "Question": "ممكن تقول لي ؟, تساؤل ؟, سؤال"
- Sentences classified as "alarm":" ضبط
اضبط ., ممكن تظبط المنبه11.,المنبه النهارده الساعه
المنبه الساعه عشره".

there was a misclassification for the sentence "ممكن تظبط المنبه.", which was classified as a question but should belong to the "alarm" intent.



Prompt:

With Temperature = 0, You are required to classify the following 'Arabic' sentences to (Question, Not)

here is list of sentences belong to class: Question:

['ممكن تعرفني ؟']

classify each of the following sentence with the corresponding class.

test_sentences = ['ممكن تقول لي ؟', 'ضبط المنبه النها رده الساعة 11.', 'اضبط المنبه الساعة عشره.', '"ممكن تضبط المنبه.', 'سؤال ؟', 'تساؤل ؟']

Output:

Question: ['ممكن تقول لي ؟', 'تساؤل ؟', 'سؤال ؟'],

alarm: ['ضبط المنبه النها رده الساعة 11.', 'ممكن تظبط المنبه.', 'اضبط المنبه الساعة عشره.']}

Figure 9: Normal Prompt Example

As shown in 10, with a temperature parameter T=0, the task is to classify the following Arabic sentences into either the "greetings" category or "Not" category, The provided information includes:

- Sentences belonging to the class "greetings":
كيفك., تصبح علي خير. , صباح الخير.

- Sentences belonging to the class ”Not”: استقصاء عن., دور على., ابحث في جوجل عن.

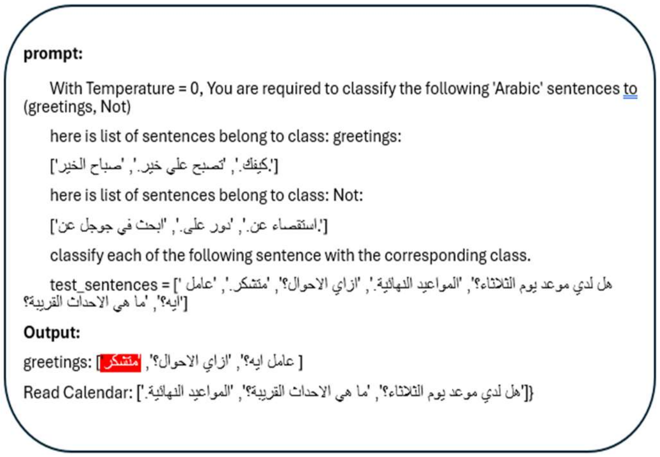
- The test sentences to be classified: هل لدي موعد يوم الثلاثاء؟, المواعيد النهائيه., ازاي الاحوال؟ , متشكر., عامل ايه؟, ما هي الاحداث القريبه؟

The output shows the classification of each test sentence into either ”greetings” or another category.

- Sentences classified as ”greetings”: عامل ايه؟, ازاي الاحوال؟, متشكر

- Sentences classified as ”Read Calendar”: هل لدي موعد يوم الثلاثاء؟, ما هي الاحداث القريبه,المواعيد النهائيه؟

There is a misclassification for the sentence ,متشكر , which should belong to the intent of the ’greetings’, but was not recognized by ChatGPT 3.5.



prompt:

With Temperature = 0, You are required to classify the following 'Arabic' sentences to (greetings, Not)

here is list of sentences belong to class: greetings:

['كيفك.', 'تصبح على خير.', 'صباح الخير']

here is list of sentences belong to class: Not:

['استقصاء عن.', 'دور على.', 'ابحث في جوجل عن']

classify each of the following sentence with the corresponding class.

test_sentences = ['هل لدي موعد يوم الثلاثاء؟', 'المواعيد النهائية.', 'ازاي الاحوال؟', 'متشكر.', 'عامل ايه؟', 'ما هي الاحداث القريبة؟']

Output:

greetings: ['عامل ايه؟', 'ازاي الاحوال؟', 'متشكر']

Read Calendar: ['هل لدي موعد يوم الثلاثاء؟', 'ما هي الاحداث القريبة؟', 'المواعيد النهائية.']}

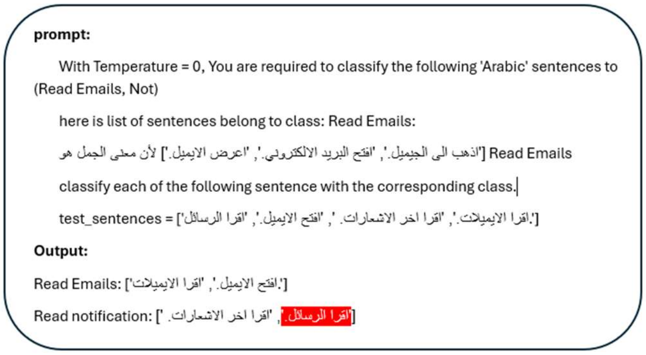
Figure 10: Contrastive Prompt Example

As shown in 11, with a temperature parameter T=0, the task is to classify the following Arabic sentences into the ”Read Emails” category or ”Not” category,. The provided information includes:

- Sentences belonging to the class ”Read Emails”: اعرض الايميل., افتح البريد الالكتروني.,اذهب الى الجيميل
- An explanation indicating that the meaning of these sentences is ”Read Emails.”: ’ الجمل هو لأن معنى ’

- The test sentences to be classified: اقرا الايميلات., اقرا اخر الاشعارات. , افتح الايميل., اقرا الرسائل.

The output shows the classification of each test sentence into either ”Read Emails” or another category.



prompt:

With Temperature = 0, You are required to classify the following 'Arabic' sentences to (Read Emails, Not)

here is list of sentences belong to class: Read Emails:

Read Emails ['اذهب الى الجيميل.', 'افتح البريد الالكترونى.', 'اعرض الايميل.'] لأن معنى الجمل هو

classify each of the following sentence with the corresponding class.

test_sentences = ['اقرا الايميلات.', 'اقرا اخر الاشعارات. ', 'افتح الايميل.', 'اقرا الرسائل']

Output:

Read Emails: ['افتح الايميل.', 'اقرا الايميلات']

Read notification: ['اقرا الرسائل.', 'اقرا اخر الاشعارات. ']

Figure 11: Static reasoning prompt example

As shown in figure 12,13 with a temperature parameter T=0, the task is to classify the following Arabic sentences into the category ”Read Notification” or ”Not” category, The provided information includes for figure 12:

- A sentence belonging to the class ”Read notification”: ’اقرا الرسائل.’, with an explanation that the sentence indicates reading notifications.

- A list of sentences belonging to the class ”Not”: ’مهاتفه شيماء.’ (meaning ”call Shaimaa”).

- The test sentences to be classified: سيرش اقرا الرسائل الاخيره., سيرش على., الاشعارات.,في جوجل عن .

The information provided includes, for figure 13:

- Sentences classified as ”Read Emails”: افتح الايميل. , اقرا الايميلات.

- Sentences classified as ”Read notification”: اقرا الرسائل., اقرا اخر الاشعارات

There is a misclassification for the sentence الرسائل اقرا, which should belong to the intent of the ”Read notification” but was classified as the ”Read emails” by ChatGPT 3.5.

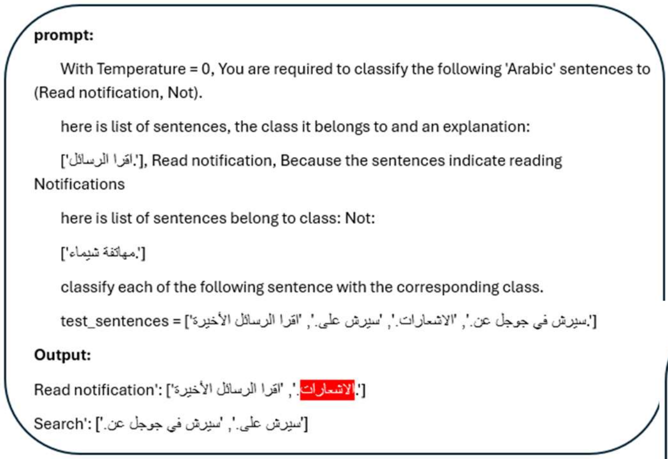
- A sentences belonging to the class “Read notification”:اقرا .، قول لي نوتفكيشن الموبايل ايه اللي في الاشعارات؟.,الرسائل with an explanation that these sentences indicate reading notification

- A list of sentences belonging to the class ”Not”: الايميلات., افتح الجيميل., رسائل البريد الالكتروني. قائمة (meaning ”email list,” ”open Gmail,” ”email messages.”

The output shows the classification of each test sentence into either ”Read notification” or another category.

- Sentences classified as ”Read notification”: الاشعارات., اقرا الرسائل الاخيره.
- Sentences classified as ”Search”:.في جوجل عن سيرش على., سيرش

In figure 12 there is a misclassification for the sentence ’.الاشعارات’, which should belong to the ”Read notification” but was not recognized by ChatGPT 3.5. However, as Figure 13 demonstrates, the model correctly detected, demonstrating how the quantity of shots influences the model’s behavior.



prompt:

With Temperature = 0, You are required to classify the following 'Arabic' sentences to (Read notification, Not).

here is list of sentences, the class it belongs to and an explanation:

['.اقرا الرسائل'], Read notification, Because the sentences indicate reading Notifications

here is list of sentences belong to class: Not:

['.مهاتفة شيماء']

classify each of the following sentence with the corresponding class.

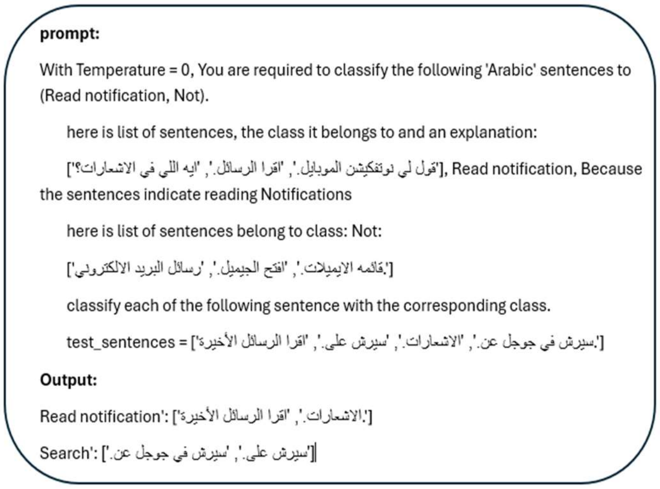
test_sentences = ['.سيرش في جوجل عن', '.الاشعارات', '.سيرش على', 'اقرا الرسائل الأخيرة']

Output:

Read notification': ['.الاشعارات', 'اقرا الرسائل الأخيرة']

Search': ['سيرش على.', 'سيرش في جوجل عن.']

Figure 12: 1 shot Dynamic Reasoning with contrastive (English reason) Prompt Example



prompt:

With Temperature = 0, You are required to classify the following 'Arabic' sentences to (Read notification, Not).

here is list of sentences, the class it belongs to and an explanation:

['قول لي نوتفكيشن الموبايل.', 'اقرا الرسائل.', 'ايه اللي في الاشعارات؟'], Read notification, Because the sentences indicate reading Notifications

here is list of sentences belong to class: Not:

['.قائمه الايميلات', '.افتح الجيميل', 'رسائل البريد الالكتروني']

classify each of the following sentence with the corresponding class.

test_sentences = ['.سيرش في جوجل عن', '.الاشعارات', '.سيرش على', 'اقرا الرسائل الأخيرة']

Output:

Read notification': ['.الاشعارات', 'اقرا الرسائل الأخيرة']

Search': ['سيرش على.', 'سيرش في جوجل عن.']]

Figure 13: 3 shot Dynamic Reasoning with contrastive (English reason) Prompt Example

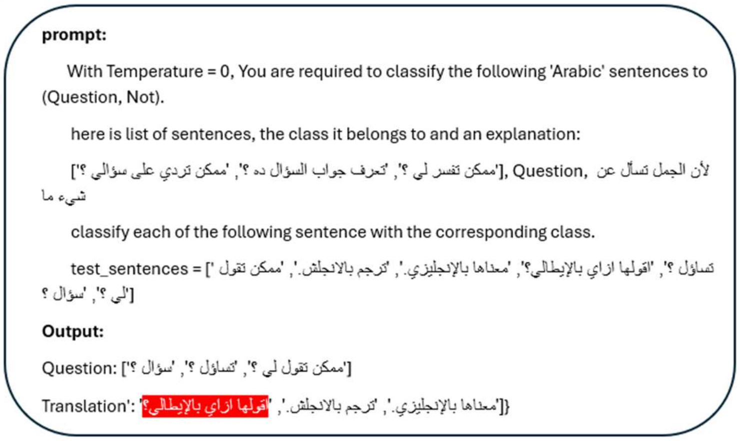
As shown in figure 14, with a temperature parameter T=0, the task is to classify the following Arabic sentences into either the ”Question” category or ”Not” category. The provided information includes:

- Sentences belonging to the class ”Question”: تعرف جواب السؤال ده ؟,, ممكن تفسر لي ؟ , ممكن تردي على سؤالي ؟ , with an explanation that these sentences ask about something ’لأن الجمل تسأل عن شيء ما’..
- The test sentences to be classified: تساؤل ترجم بالانجلش., معناها بالانجليزي,اقولها ازاي بالايطالي؟,؟,. سؤال ؟,ممكن تقول لي ؟

The output shows the classification of each test sentence into either ”Question” or another category.

- Sentences classified as ”Question؛ ممكن تقول لي ؟,تساؤل ؟,سؤال
- Sentences classified as ”Translation”:ترجم بالانجلش., معناها بالانجليزي ,. اقولها ازاي بالايطالي؟

There is a misclassification for the sentence ’اقولها ازاي بالايطالي؟’, which should belong to the ”Translation” intent but was not recognized by ChatGPT 3.5.



prompt:

With Temperature = 0, You are required to classify the following 'Arabic' sentences to (Question, Not).

here is list of sentences, the class it belongs to and an explanation:

['ممكن تفسر لي ؟', 'تعرف جواب السؤال ده ؟', 'ممكن تردي على سؤالي ؟'], Question, لأن الجمل تسأل عن شيء ما

classify each of the following sentence with the corresponding class.

test_sentences = ['تساؤل ؟', 'اقولها ازاي بالإيطالي؟', 'معناها بالإنجليزي.', 'ترجم بالانجلش.', 'ممكن تقول لي ؟', 'سؤال ؟']

Output:

Question: ['ممكن تقول لي ؟', 'تساؤل ؟', 'سؤال ؟']

Translation': 'اقولها ازاي بالإيطالي؟', 'ترجم بالانجلش.', 'معناها بالإنجليزي.']}

Figure 14: 3 shot Dynamic Reasoning with contrastive (Arabic reason) Prompt Example

6. DISCUSSION

The experimental results demonstrate that prompt-based learning is an effective approach for Arabic intent classification under few-shot settings.

6.1 Interpretation & implications

Among the evaluated strategies, contrastive prompting combined with dynamic reasoning consistently achieved the highest accuracy, outperforming both standard prompting techniques and the fine-tuning baseline. These findings indicate that carefully designed prompts can significantly enhance the performance of large language models on Arabic intent classification tasks, even with limited labeled data.

The strong performance of contrastive prompting with dynamic reasoning can be attributed to its ability to explicitly model distinctions between similar intents while guiding the model through structured reasoning steps. Contrastive examples help the model differentiate between closely related intent categories, which is particularly important for Arabic due to its rich morphology and contextual ambiguity. Dynamic reasoning further supports this process by encouraging the model to explain or reason about its predictions, leading to more accurate and consistent classifications.

Compared to traditional fine-tuning approaches, the proposed prompt-based methods offer a more resource-efficient alternative. Fine-tuning requires large annotated datasets and substantial computational resources, which may not always be available for Arabic or other low-resource languages. In contrast, prompting strategies enable competitive or superior performance using only a small number of examples, making them especially suitable for rapid deployment and low-resource scenarios.

These findings have important implications for Arabic natural language processing applications. Prompt-based learning provides a flexible and scalable solution for intent classification tasks without the need for extensive model retraining. This approach can be particularly beneficial for real-world systems where data availability is limited or where frequent updates to intent definitions are required

6.2 Limitations

Despite the strong performance of the proposed prompting strategies, this study has some limitations. First, the dataset contains only 16 predefined intents, which may limit the generalizability of the results to other intents or larger datasets. Second, the approach relies on GPT-3.5, which requires access to proprietary APIs and may not be directly reproducible in all environments. Third, the evaluation primarily focuses on accuracy; incorporating additional metrics such as precision, recall, and F1-score could provide a more comprehensive assessment. Finally, the approach has not been extensively evaluated across diverse Arabic dialects, which may affect its applicability in real-world scenarios.

7. Conclusion

This study demonstrates the effectiveness of prompting techniques in enhancing Arabic intent classification, addressing the challenges posed by the language's complex morphology and high ambiguity. The results highlight the superiority of prompting, particularly contrastive with dynamic reasoning over traditional fine-tuning methods in terms of accuracy and resource efficiency. These findings contribute to the advancement of Arabic NLP applications, showcasing the potential of prompting techniques to improve the understanding and classification of Arabic text.

8. Future Work

Future research can explore the application of prompt- ing techniques in other areas of Arabic NLP, such as sentiment analysis, machine translation, and named entity recognition. Additionally, investigating the impact of different prompt designs and parameters on model performance could provide further insights into optimizing the use of prompting for Arabic language processing tasks. Furthermore, extending this study to include more diverse datasets and languages could enhance the generalizability of the findings and contribute to the development of robust multilingual NLP models.

REFERENCES

[1] Daniele Comi, Dimitrios Christofidellis, Pier Francesco Piazza, and Matteo Manica. Z-bert-a: a zero-shot pipeline for unknown intent detection. *arXiv preprint arXiv:2208.07084*, 2022.

[2] Ali Farghaly and Khaled Shaalan. Arabic natural language processing: Challenges and solutions. *ACM Transactions on Asian Language Information Processing (TALIP)*, 8(4):1–22, 2009.

[3] Jong Myoung Kim, Young-Jun Lee, Sangkeun Jung, and Ho-Jin Choi. Semantic ambiguity detection in sentence classification using task-specific embeddings. In *Proceedings of the 61st An- nual Meeting of the Association for Computational Linguistics (Volume 5: Industry Track)*, pages 425–437, 2023.

[4] Mahmoud Al-Ayyoub, Aya Nuseir, Kholoud Alsmearat, Yaser Jararweh, and Brij Gupta. Deep learning for arabic nlp: A survey. *Journal of computational science*, 26:522–531, 2018.

[5] Mohamed A Galal, Ahmed Hassan Yousef, Hala H Zayed, and Walaa Medhat. Arabic sarcasm detection: An enhanced fine-tuned language model approach. *Ain Shams Engineering Journal*, 15(6):102736, 2024.

[6] Ali Saleh Alammary. Bert models for arabic text classification: a systematic review. *Applied Sciences*, 12(11):5720, 2022.

[7] Hanen Karamti, Maha MA Lashin, Fadwa M Alrowais, and Abeer M Mahmoud. A new deep stacked architecture for multi- fault machinery identification with imbalanced samples. *Ieee Access*, 9:58838–58851, 2021.

[8] Hanen Karamti and Abeer M Mahmoud. A pre-protective objective in mining females social contents for identification of early signs of depression using soft computing deep framework. *Scientific Reports*, 13(1):14899, 2023.

[9] Abeer M Mahmoud, Hanen Karamti, and Fadwa Alrowais. A two consequent multi-layers deep discriminative approach for classifying fmri images. *International Journal on Artificial Intelligence Tools*, 29(06):2030001, 2020.

[10] Fuzhen Zhuang, Zhiyuan Qi, Keyu Duan, Dongbo Xi, Yongchun Zhu, Hengshu Zhu, Hui Xiong, and Qing He. A comprehensive survey on transfer learning. *Proceedings of the IEEE*, 109(1):43–76, 2020.

[11] Meshrif Alruily, Abdul Manaf Fazal, Ayman Mohamed Mostafa, and Mohamed Ezz. Automated arabic long-tweet clas- sification using transfer learning with bert. *Applied Sciences*, 13(6):3482, 2023.

[12] Tahani N Alruqi and Salha M Alzahrani. Arabic chatbot evalu- ation based on extractive question-answering transfer learning and language transformers. *Preprints*, 2023.

[13] Khadige Abboud, Olga Golovneva, and Christopher DiPersio. Cross-lingual transfer for low-resource arabic language understanding. In Proceedings of the Seventh Arabic Natural Lan- guage Processing Workshop (WANLP), pages 225–237, 2022.

[14] Henry Weld, Xiaoqi Huang, Siqu Long, Josiah Poon, and Soyeon Caren Han. A survey of joint intent detection and slot filling models in natural language understanding. ACM Computing Surveys, 55(8):1–38, 2022.

[15] Sofia Rizou, Angelos Theofilatos, Antonia Paflioti, Eleni Pissari, Iraklis Varlamis, George Sarigiannidis, and K Ch Chatzisavvas. Efficient intent classification and entity recogni- tion for university administrative services employing deep learning models . *Intelligent Systems with Applications*, 19:200247, 2023.

[16] Imad A Al-Sughaiyer and Ibrahim A Al-Kharashi. Arabic morphological analysis techniques: A comprehensive survey. *Journal of the American society for information science and technology*, 55(3):189–213, 2004.

[17] Mohammed A Attia. *Handling Arabic morphological and syn- tactic ambiguity within the LFG framework with a view to machine translation*. The University of Manchester United Kingdom, 2008.

[18] Meshrif Alruily. Arrasa: channel optimization for deep learning-based arabic nlu chatbot framework. *Electronics*, 11(22):3745, 2022

[19] Mohammad Fadhil Mahdi and Mahmoud Shuker Mahmoud. Intent arabic text categorization based on different machine learning and term frequency. *IET Networks*, 2022.

[20] Abdallah M Bashir, Abubakr Hassan, Benjamin Rosman, Daniel Duma, and Mohanad Ahmed. Implementation of a neural natural language understanding component for arabic dialogue systems. *Procedia computer science*, 142:222–229, 2018.

[21] Xu Han, Weilin Zhao, Ning Ding, Zhiyuan Liu, and Maosong Sun. Ptr: Prompt tuning with rules for text classification. *AI Open*, 3:182–192, 2022.

[22] Yi Zhu, Ye Wang, Jipeng Qiang, and Xindong Wu. Prompt- learning for short text classification. *IEEE Transactions on Knowledge and Data Engineering*, 2023.

[23] Seham Basabain, Erik Cambria, Khalid Alomar, and Amir Hussain. Enhancing arabic-text feature extraction utilizing label-semantic augmentation in few/zero-shot learning. *Expert Systems*, 40(8):e13329, 2023.

[24] Manoj Kumar, Varun Kumar, Hadrien Glaude, Cyprien de Lichy, Aman Alok, and Rahul Gupta. Protoda: Efficient transfer learning for few-shot intent classification. In *2021 IEEE Spoken Language Technology Workshop (SLT)*, pages 966–972. IEEE, 2021.

[25] Muhammad Khalifa, Muhammad Abdul-Mageed, and Khaled Shaalan. Self-training pre-trained language models for zero- and few-shot multi-dialectal arabic sequence labeling. *arXiv preprint arXiv:2101.04758*,

2021.

[26] Harshit Sharma. Softmax temperature, 2022. https://medium. com/@harshit158/softmax-temperature-5492e4007f71.

[27] Ahmed AbdelGawad. arabic-intent-classification, 2022. https://www.kaggle.com/code/ahmedabd elgawad/ arabic-intent-classification/notebook.

[28] Damai Dai, Yutao Sun, Li Dong, Yaru Hao, Shuming Ma, Zhifang Sui, and Furu Wei. Why can gpt learn in-context? language models implicitly perform gradient descent as meta- optimizers. *arXiv preprint arXiv:2212.10559*, 2022.

[29] Jesse Dodge, Gabriel Ilharco, Roy Schwartz, Ali Farhadi, Han- naneh Hajishirzi, and Noah Smith. Fine-tuning pretrained language models: Weight initializations, data orders, and early stopping. *arXiv preprint arXiv:2002.06305*, 2020.