

AMERICAN UNIVERSITY OF BEIRUT

MULTI-MODAL ARABIC NEGOTIATING BOT

by

SAMAH MOHAMAD ALBAST

A thesis  
submitted in partial fulfillment of the requirements  
for the degree of Master of Science  
to the Department of Computer Science  
of the Faculty of Arts and Sciences  
at the American University of Beirut

Beirut, Lebanon  
January 2024

AMERICAN UNIVERSITY OF BEIRUT

MULTI-MODAL ARABIC NEGOTIATING BOT

by  
SAMAH MOHAMAD ALBAST

Approved by:



Signature

---

Dr. Wassim El Hajj, Professor  
Department of Computer Science

Advisor

*Haidar Safa*

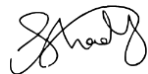


Signature

---

Dr. Haidar Safa, Professor  
Department of Computer Science

Member of Committee



Signature

---

Dr. Shady Elbassuouni, Associate Professor  
Department of Computer Science

Member of Committee

Date of thesis defense: January 23, 2024

# AMERICAN UNIVERSITY OF BEIRUT

## THESIS RELEASE FORM

Student Name: AlBast Samah Mohamad  
Last First Middle

I authorize the American University of Beirut, to: (a) reproduce hard or electronic copies of my thesis; (b) include such copies in the archives and digital repositories of the University; and (c) make freely available such copies to third parties for research or educational purposes:

- As of the date of submission
- One year from the date of submission of my thesis.
- Two years from the date of submission of my thesis.
- Three years from the date of submission of my thesis.



February 5, 2024

---

Signature

Date

## ACKNOWLEDGEMENTS

I want to start by expressing my deep thanks to the U.S. Middle Eastern Partnership Initiative's scholarship program for giving me the chance to pursue my master's studies at AUB.

I want to extend my thanks to my advisor Dr. Wassim El Hajj, for his guidance and support. My appreciation also goes to my committee members, Dr. Haidar Safa and Dr. Shady Elbassuouni, for their invaluable help and comments.

To my family, I dedicate this thesis with deep gratitude. Your constant presence and support have been my pillars, and I could not have come this far without you.

To my dearest friends, Hala, Jihane, and Mona thank you for always being there for me. Your support has been my source of strength throughout this journey. I am grateful for your presence in my life.

Lastly, this journey brought amazing people with beautiful hearts into my life. Nour and Rana, your presence has been a blessing. Thanks for making it lighter and brighter.

# ABSTRACT OF THE THESIS OF

Samah AlBast

for

Master of Science

Major: Computer Science

Title: Multi-Modal Arabic Negotiating Bot

Negotiation is a fundamental aspect of human interaction, involving a dynamic exchange of communication between two or more parties to reach mutually agreeable outcomes. With recent advancements in chatbots, leveraging artificial intelligence (AI) for negotiation has emerged as an ideal application. Despite significant progress in English negotiation bots using deep learning and reinforcement learning, such advancements are notably absent in other languages, particularly Arabic. Furthermore, while previous research has primarily focused on developing high-performing neural response generation systems for negotiation bots, the integration of multimodality into these automated agents remains unexplored. The incorporation of multimodality is represented in image analysis, and it contributes to a more comprehensive and user-friendly negotiation model. This thesis presents the first Arabic negotiation model, distinguished by incorporating multimodality into negotiation models. The integration of multimodality, particularly through image analysis, provides a more comprehensive and user-centric approach to negotiation. Our primary objective is to develop an Arabic multimodal negotiating bot, a seller agent capable of engaging in negotiations with buyers in the context of item sales. This seller agent is designed to understand the buyer's Arabic utterances and to interpret the negotiation context through images provided by the buyer. To achieve this, we trained a Generative Pre-trained Transformer (GPT-2) model on an Arabic dataset, integrating it with a Convolutional Neural Network (CNN) for image analysis. The model's automatic evaluation yielded a BLEU-4 score of 0.21 and a cross-entropy loss of 0.55, metrics that are promising for the first model of its kind in Arabic. Our experiments and analyses reveal both the successes and limitations of the designed multi-modal Arabic negotiating model, offering insights into the inherent challenges and setting directions for future research.

## TABLE OF CONTENTS

ACKNOWLEDGEMENTS .....	1
ABSTRACT .....	2
ILLUSTRATIONS .....	5
TABLES .....	6
ABBREVIATIONS .....	7
INTRODUCTION .....	8
1.1. Background & Motivation .....	8
1.2. Problem Description .....	9
1.3. Challenges & Contributions.....	10
1.4. Thesis Organization .....	12
LITRATURE REVIEW .....	13
2.1. Automated Arabic ChatBots.....	13
2.2. Automated Negotiation ChatBots.....	14
PROPOSED METHOD .....	18
3.1. Baseline Workflow .....	18
3.1.1. Loss Function.....	20
3.1.2. Rewards .....	21

3.2.	Proposed Model for Multi-Modal Arabic Negotiation .....	22
3.2.1.	Multi-Modal Arabic Negotiation Workflow.....	24
<b>EXPERIMENTS &amp; RESULTS .....</b>		<b>26</b>
4.1.	Arabic Dataset.....	26
4.2.	Negotiation Models.....	28
4.2.1.	Experiment Setup.....	28
4.2.2.	English Negotiation Model .....	28
4.2.3.	Arabic Negotiation Model .....	29
4.2.4.	Arabic Negotiation Model Enhanced with Augmented Data .....	30
4.2.5.	Human Evaluation of the Adopted Arabic Negotiation Model .....	31
4.3.	Image-Based Color and Brand Recognition Model.....	33
4.3.1.	Dataset .....	33
4.3.2.	Color and Brand Recognition Model.....	34
4.4.	Arabic Multi-Modal Negotiation Model.....	34
4.5.	Statistical Significance.....	36
4.6.	Discussion.....	36
4.7.	Limitation.....	37
<b>CONCLUSION &amp; FUTURE WORK .....</b>		<b>38</b>
5.1.	Conclusion .....	38
5.2.	Future Work.....	39
<b>APPENDIX .....</b>		<b>40</b>
<b>BIBLIOGRAPHY .....</b>		<b>42</b>

# ILLUSTRATIONS

## Figure

1.	Blackbox representation of the Arabic multi-modal negotiation bot.....	10
2.	Example of a Belief-State Pair.....	18
3.	Negotiation Model Pipeline.....	19
4.	Multi-Modal Arabic Negotiation Bot Pipeline.....	22
5.	Buyer and Seller Negotiation Sample.....	23
6.	Manual evaluation results of the Arabic negotiation model.....	33
7.	Manual evaluation results of the Multi-Modal Arabic negotiation model.....	35



## TABLES

### Table

1. Examples of unreasonable translations.....	27
2. Examples of good translations.....	27
3. Evaluation of the trained negotiation models using Bleu-4 and cross-entropy loss.....	31
4. Evaluation Results of CNN Model for Color and Brand Prediction .....	34
5. Negotiation Example with the Arabic negotiation model enhanced with augmented data as the seller.....	40
6. Negotiation Example with the Arabic multi-modal negotiation model as the seller.....	41

## ABBREVIATIONS

AI	Artificial Intelligence
AIML	Artificial Intelligence Markup Language
NLP	Natural Language Processing
NLG	Natural Language Generation
RNN	Recurrent Neural Network
RL	Reinforcement Learning
Seq2Seq	Sequence-To-Sequence
LM	Language Model
GPT	Generative Pre-trained Transformer
CNN	Convolutional Neural Network
LSTM	Long Short-Term Memory
BERT	Bidirectional Encoder Representations from Transformers
BLEU	Bilingual Evaluation Understudy

# CHAPTER 1

## INTRODUCTION

In this chapter, we introduce the topic, highlight the existing challenges, and list the contributions of this thesis.

### **1.1. Background & Motivation**

Negotiation is a common phenomenon that occurs in almost every aspect of personal and professional life, ranging from individual discussions over a purchase to corporate contract negotiations. It involves a dynamic exchange of communication between two or more parties with the objective of achieving mutually agreeable outcomes and facilitating the realization of shared objectives [1]. With the recent advances in chatbots, negotiation appears to be an ideal application for artificial intelligence (AI), empowering intelligent agents to engage in bargaining and collaborative discussions. Automating the negotiation process has significant potential for smoothing conflict resolution and enhancing outcomes across various domains, such as marketplaces and autonomous driving [2].

The field of English negotiation bot research has undergone significant advancements, initially relying on game theory and rule-based approaches [7][8][9][10]. This was followed by the adoption of deep learning techniques, which have demonstrated remarkable efficacy in natural language processing (NLP). The work of [11] established the foundation for product negotiation through the utilization of two sequence-to-sequence (SEQ2SEQ) recurrent neural network (RNN) models, representing the buyer and the seller, respectively. These models underwent

supervised training, with subsequent refinement via reinforcement learning (RL). Building upon this foundation, subsequent studies, such as [12][13][14][15][16], have introduced enhancements including strategies for incorporating emotions, persuasion, and politeness into negotiation dialogues. In their work in [16], the authors employed a generative pre-trained transformer-based model, GPT-2, and reinforcement learning techniques to develop a negotiating bot acting as a seller in e-commerce transactions with buyers. This bot is equipped with persuasion strategies, enabling it to present alternative product options to users when their desired goals are unavailable, thereby broadening its capabilities in guiding and completing successful sales transactions.

While prior research mainly focused on developing high-performing neural response generation systems for negotiation bots, the integration of multimodality into these automated agents remains unexplored. Specifically, in the case of [16], multimodality involves users providing images of desired products, enabling the negotiation model to extract product features from images and gain a deeper understanding of user requirements. In addition to the gap in multimodal negotiation bot research, little research has been done in the development of non-English negotiation bots, particularly in languages such as Arabic. Our work seeks to address these research gaps by aiming to develop a multimodal Arabic negotiation bot, inspired by the methodology outlined in [16]. This bot is designed to conduct negotiations in Arabic while also being able to comprehend user requirements from images.

## **1.2. Problem Description**

In this work, we aim to address the gaps of prior works. We tackle the problem of building a multimodal Arabic automated negotiating bot. Our primary focus lies in

the context of item sales, where a buyer engages in communication with a seller agent to buy a specific product with defined features. The buyer may choose to provide an image of the desired product. Meanwhile, the seller agent tries to convince the buyer to make a purchase, employing various persuasive strategies, even in cases where the buyer's desired product might be unavailable. Our objective is to automate the functions of the seller agent, ensuring it responds persuasively to guide the buyer toward completing the sales transaction. Additionally, we integrate image analysis capabilities to better understand user needs, further enhancing the agent's ability to provide tailored assistance. We make the developed datasets and resources publicly available for further research in this area<sup>1</sup>.

Figure 1 provides a visual representation of the desired bot in action.



Figure 1: Blackbox representation of the Arabic multi-modal negotiation bot.

### 1.3. Challenges & Contributions

Designing and building Arabic multimodal negotiating bots presents several challenges that need to be addressed for effective performance. Here are some of the key challenges:

<sup>1</sup> <https://github.com/samahAlbast/Thesis-Code>

- **Exploring Negotiation Dialogue in Real-world Contexts:** Most negotiation dialogue systems are currently generated through human crowdsourcing, which may not reflect real-world scenarios. There is a need to collect real negotiation dialogues that reflect real-life scenarios instead of relying on crowd-sourced data.
- **Effectively communicating in Arabic:** Arabic is a language rich in nuances and informal expressions. While Arabic is widely spoken, there is a scarcity of native Arabic datasets specifically tailored to negotiation scenarios. Consequently, many Arabic chatbots rely on translated datasets, which can lead to misunderstandings and hinder the bot's ability to engage in natural, effective negotiations.
- **Limited Availability of Multimodal Arabic Datasets:** Developing a multimodal Arabic chatbot requires access to a large and diverse dataset of Arabic text and visual data. However, such datasets are currently scarce, making it difficult to train the chatbot effectively.
- **Ability to negotiate and persuade:** This involves the capability to create, evaluate, and modify offers based on the ongoing conversation and though the comprehension of the negotiation subject.

Our objective is to address the exiting challenges, with a specific focus on targeting the following contributions:

- Introducing the first Arabic negotiation chatbot.
- Developing the first multi-modal negotiation chatbot.

#### **1.4. Thesis Organization**

The remainder of this thesis is structured as follows: In Chapter 2, we provide a comprehensive review of the current state of the literature. In Chapter 3, we present our proposed method. In Chapter 4, we outline experimental results. In Chapter 5, we conclude.

## CHAPTER 2

### LITRATURE REVIEW

While considerable research has been conducted on Arabic chatbots and English automated negotiating bots independently, the intersection of these two remains significantly understudied. To the best of our knowledge and as of the present date, there is a notable absence of research specifically dedicated to Arabic negotiating chatbots. Therefore, this section aims to address advancements made in both the Arabic chatbot and English automated negotiating bot domains.

#### **2.1. Automated Arabic ChatBots**

Chatbots are machine-learning-based conversational systems that generate text responses given human inputs or questions. Over the past few years, many chatbots have been developed in languages such as English, Spanish, and French for use in education, medicine, and entertainment. However, work on Arabic chatbots is still limited due to the complex linguistic features of the Arabic language and the lack of large native Arabic datasets covering different domains.

The initial work in Arabic chatbot development is represented by ArabChat [3], published in 2014. ArabChat utilizes pattern-matching techniques to generate appropriate responses to user queries. The architecture was based on a rule-based scripting language and a scripting engine. ArabChat proved its ability to show flexibility in conversations with users and to align with the users' topics of interest. Another example of a retrieval-based chatbot designed for the Egyptian dialect is BOTTA [4]. This chatbot's knowledge base was constructed using Artificial



Intelligence Markup Language (AIML), a specialized markup language specifically tailored for creating conversational agents. Notably, the AIML files employed in BOTTA were translated from ROSIE (variation of ALICE [29], a chatbot that communicates in English), a well-established rule-based dialogue system, to ensure a robust foundation for the chatbot's knowledge and responses.

Recently, Arabic chatbot development has started incorporating deep learning techniques. However, the scarcity of Arabic datasets has been a limiting factor. Despite this, some researchers have tackled the challenge by using translation methods to train deep learning-based Arabic chatbots. AraBERT [5], for instance, is an Arabic pre-trained language model (LM) based on Google's Bidirectional Encoder Representations from Transformers (BERT) architecture. Another example is discussed in [6], where the researchers aimed to generate empathetic responses for Arabic expressions. They employed a Bi-Long Short-Term Memory (LSTM) Seq2Seq model integrated with attention mechanisms. In contrast, the work in [30] utilized the same dataset but adopted a transformer model with pretrained checkpoints. When evaluated on this shared dataset, the transformer model demonstrated superior performance, outperforming the LSTM model.

While advancements in Arabic chatbots have been remarkable, the development of an Arabic negotiating chatbot remains unexplored. Our work stands as the first attempt to create an Arabic chatbot tailored for negotiation purposes.

## **2.2. Automated Negotiation ChatBots**

The field of automated negotiation has witnessed the investigation of diverse methodologies [17]. One traditional approach is the use of rule-based methods, which

relies on a set of if-then statements to determine the strategy of the negotiating agent. These rules specify how the agent should react in different situations encountered during the negotiation process [7][8].

Additionally, game theory has been employed in the development of negotiating bots because it can be applied to interactions involving self-interested agents, which aim to achieve outcomes that favor their goals, and negotiation participants are driven by their own interests and objectives. Approaches utilizing game theory focus on solving the problem of what is the most rational action an agent can take regardless of how others might act. One example of these approaches is the monotonic concession, where one or both negotiation parties must concede each time an agreement is not reached, following specific methods to determine the conceding agent and the extent of concession. However, incorporating game theory into negotiation considers that the agent must select the best strategy from the space of all possible strategies. This involves considering all possible interactions, which often leads to computationally expensive calculations [9].

Another approach employed in the development of negotiating agents is the heuristics approaches such as the introduction of two heuristics in [10] to guide the decision-making process. The first heuristic relies on the observation that there is an inverse relationship between the preference for an issue and how frequently its values change throughout negotiations. The second heuristic is based on the association between the preference for a particular value and the frequency with which it is being offered. Utilizing these heuristics helps in comprehending the opponent's preferences and adapting the negotiation strategy accordingly. [8].

All the previously mentioned approaches to construct a negotiating bot rely on manually crafted sets of rules and lack the ability to engage in negotiation using natural language.

Recent efforts have turned to leveraging deep learning techniques for the development of negotiating bots. The work in [11] is the first to apply deep learning techniques to build an end-to-end model for natural language negotiation. An end-to-end model is a machine learning system that directly maps raw input data to desired outputs without relying on intermediate stages or representations. This model acquires both linguistic and reasoning skills through a combination of supervised and RL techniques. The model's initial training leverages supervised learning, employing a Gated Recurrent Unit (GRU) recurrent neural network (RNN) to generate responses and offers. Subsequently, the researchers fine-tune the RNN-trained model using RL to maximize its utility. However, this fine-tuning step with RL introduced degeneracy issues, leading to the generation of inaccurate responses. This inaccuracy stems from the RL model's singular focus on utility maximization, irrespective of the generated text's reliability and relevance. In [12], they addressed the problem of degeneracy of the work in [11] by decoupling the negotiating strategy from language generation. In this work, the authors tackle the challenge of creating a negotiating bot by decoupling the negotiation strategy from language generation. Negotiation strategy refers to the bot's plan to achieve its goals within a negotiation. This approach allows the bot to employ diverse strategies for various negotiation objectives while simultaneously generating natural and persuasive language tailored to the specific context. In [13], the authors focused specifically on persuasion strategies. They collected a rich dataset of human-human persuasion dialogues annotated with persuasion strategies. They then leveraged

this annotated dataset to develop a high-performing classifier capable of accurately predicting persuasion strategies. In [14], they developed a negotiation coach offering tactics to enhance deals for the seller, utilizing an LSTM-based model to generate tactic suggestions. In [15], they trained and fine-tuned a LM using RL while considering various sub-rewards for persuasion, emotion, politeness, coherence, and repetitiveness. In [16], they developed a persuasive sales agent to persuade a buyer to buy a target item. The authors used a GPT-2 model combined with RL that has 2 sub-rewards (repetitiveness and consistency). This is the first automated negotiation dialogue that tackles the challenge of creating a virtual assistant that works well even when user goals are unavailable. The goal is to reduce task failures caused by conflicting goals.

The study presented in [16] represents the initial end-to-end dialogue agent capable of persuading users in goal unavailability situations. The outcomes demonstrate that employing multiple sub-rewards and the proposed framework enhances the quality of dialogue generation, particularly in terms of coherence and repetitiveness. Considering these findings, we have chosen to leverage this work as the foundation for constructing our multi-modal Arabic negotiating bot.

To the best of our knowledge, our work is the first to explore the use of deep learning in the area of multi-modal Arabic negotiation bots.

# CHAPTER 3

## PROPOSED METHOD

In this section, we present our methodology, which centers on developing a neural-based multimodal negotiation dialogue generation framework. The first section elaborates on the pipeline outlined in [16], while subsequent sections discuss the workings of our Arabic negotiation model, emphasizing the integration of image processing to enhance multimodal capabilities.

### 3.1. Baseline Workflow

In this section, we will present the workflow of the module introduced in [16]. After receiving the user utterance  $U_0$  at turn 0, the model proceeds as follows:

1. First the model extracts the belief state  $B_0$ , consisting of slot-value pairs. Each pair (slot, value), denoted as  $(s_{l_i}, v_i)$ , encapsulates slot and value information extracted from the current utterance. An illustrative example of such a belief state is presented in Figure 2. In this example, we have two belief states. The first one has these values: slot = brand and value = Samsung, and the second one has these values: slot = RAM and value = 4.

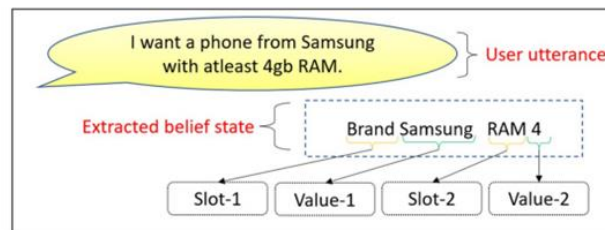


Figure 2: Example of a Belief-State Pair

2. The model then performs a database query employing the belief state  $B_0$ , determining the count of database instances ( $D_0$ ) that align with  $B_0$ .
3. Lastly, based on the context  $[U_0, B_0, D_0]$ , the model utilizes GPT-2 [19] to generate the necessary action  $A_0$  and the response  $R_0$ , benefiting from the effectiveness of large pre-trained language models that have surpassed their smaller counterparts across various NLP tasks.

Similarly, at turn  $t$ , the model considers  $[U_0, B_0, D_0, A_0, R_0, \dots, U_{t-1}, B_{t-1}, D_{t-1}, A_{t-1}, R_{t-1}, U_t]$  as context and generates  $B_t, D_t, A_t,$  and  $R_t$ , respectively. The full model pipeline is provided in Figure 3.

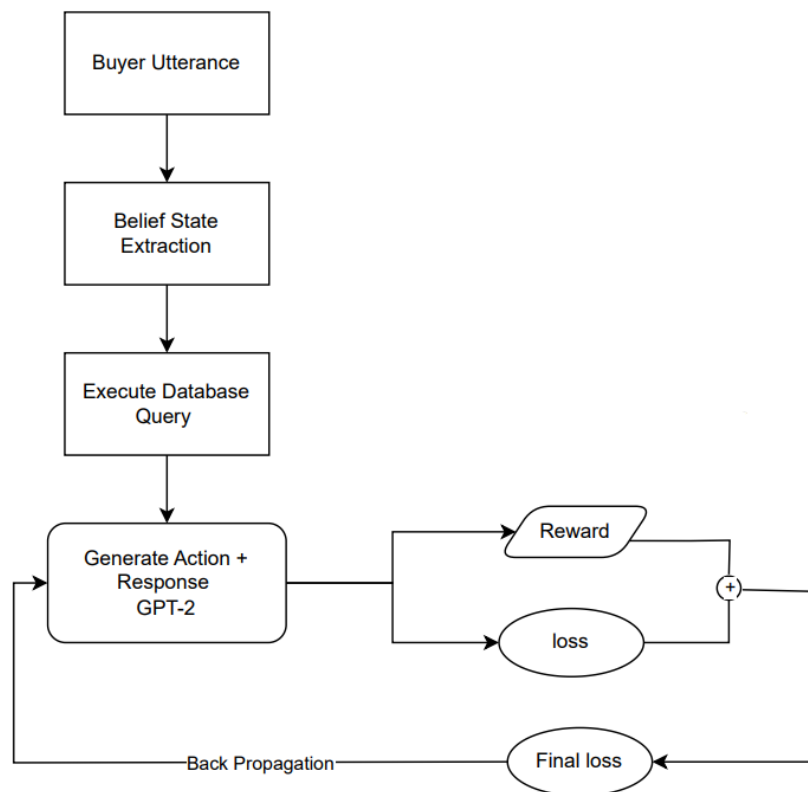


Figure 3: Negotiation Model Pipeline.

The pipeline starts by receiving the buyer utterance, followed by a belief state extraction. Based on this belief state, a database query is then extracted, and finally, the seller response is generated.

### 3.1.1. Loss Function

At the end of every turn  $t$  a reward  $r_t$  is calculated based on 2 sub-rewards: repetitiveness and consistency; this reward will make the model aware of the quality of the generated response. At the end of the whole conversation, an average  $r$  of the rewards obtained at the end of each turn is calculated in addition to the cross-entropy loss  $l$  and those will be considered as the end loss  $L$  and will be used to update the models' parameters.

$$r = \frac{\sum_i^n r_{t=i}}{n} \quad (1)$$

$$l = \text{cross-entropy loss} \quad (2)$$

$$L = l + r \quad (3)$$

Combining the aggregated rewards  $r$  with the cross-entropy loss  $l$  will create a composite loss function that balances the need for fluency and coherence (cross-entropy) with the need for high-quality responses (reward-based metrics). This combination guides the model to generate responses that are not only grammatically correct but also contextually appropriate and engaging.

### 3.1.2. Rewards

#### 3.1.2.1. Repetitiveness Reward

According to [20], language models tend to produce repetitive utterances, meaning they generate phrases or sentences that occur frequently in the training dataset. Consequently, the dialogue becomes dull, potentially diminishing its persuasiveness over time. To tackle this issue, the solution involves the use of Jaccard Score [21], a metric based on unigrams, to assess the similarity between the previous utterances in the conversation and the current response generated by the model. The process begins by normalizing the sentences, which typically involves standardizing their format, structure, or representation for consistency and ease of comparison. This normalization is carried out using the spaCy library. Following the normalization step, the Jaccard Score is computed, and the resulting score is subsequently utilized as a sub-reward.

$$r_1^t = \frac{R_0 \cap R_1 \cap \dots \cap R_t}{R_0 \cup R_1 \cup \dots \cup R_t} \quad (4)$$

#### 3.1.2.2. Consistency Reward

Consistency here is used to measure how closely the model-generated responses resemble a human-generated response in terms of word usage, structure, and overall quality. The goal is to generate responses that are not only contextually appropriate but also maintain a level of consistency with what a human might produce in a similar dialogue scenario. To measure this consistency, Meteor score [22] is employed between the generated responses (hypothesis) and a gold human response, which serves as the reference. The selection of the gold human response as a benchmark is based on the belief that it optimally aligns with the desired consistency in the dialogue. Meteor calculates a score for this comparison, taking



into consideration unigram precision, unigram recall, and a fragmentation measure. This comprehensive evaluation assesses how well-ordered the matched words in the machine-generated responses are compared to the reference, offering insights into the overall quality and consistency of the generated content.

### 3.2. Proposed Model for Multi-Modal Arabic Negotiation

This workflow is similar to what's discussed in section 3.1. The main difference is that our model will be trained on an Arabic dataset, and it will be able to understand the belief state from an image of the product provided by the buyer. The new pipeline is visualized in Figure 4.

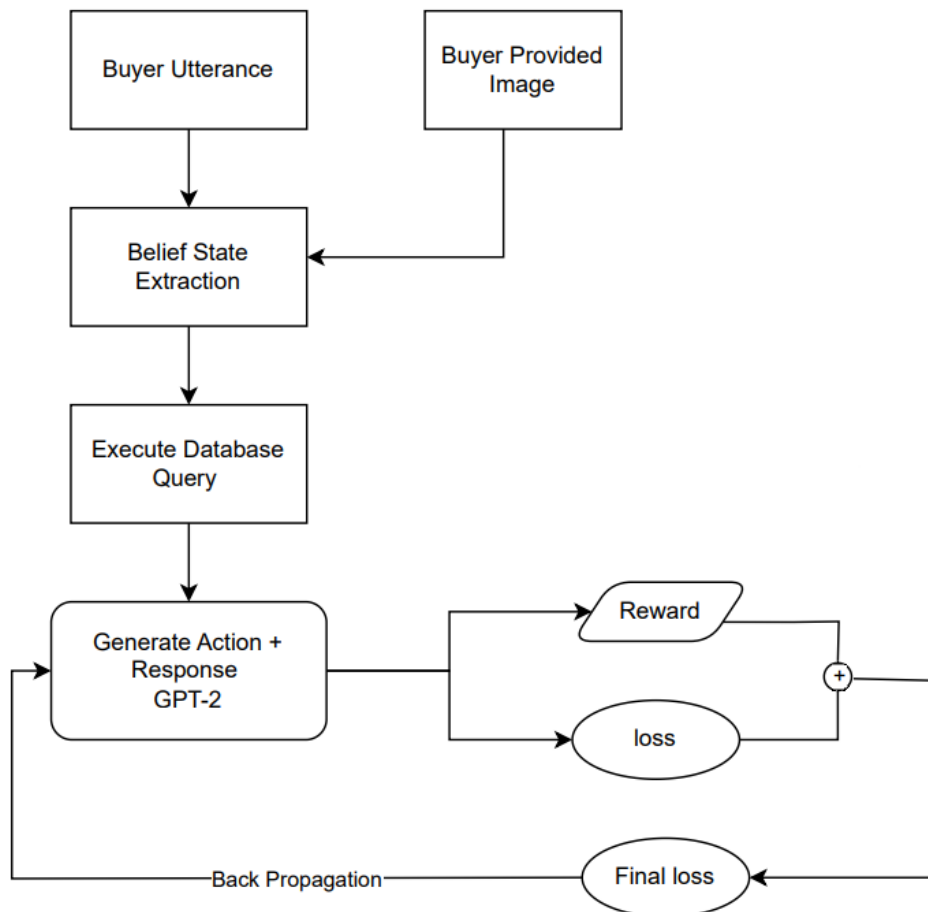


Figure 4: Multi-Modal Arabic Negotiation Bot Pipeline

The pipeline begins by receiving a buyer's utterance, which may include an HTTP link for an image of the requested product. If an image link is provided, the pipeline performs belief state extraction on both the provided image and the buyer's text. Subsequently, the pipeline proceeds by executing a database query based on the extracted belief states. Finally, the seller's response is generated.

The task of performing image feature extraction from product images was identified as a future endeavor in [16]. The original English negotiation dataset used in this study consists of buyer utterances where the buyer negotiates about products presented in HTTP links containing item images provided by the buyer. However, in their work, these links were treated as if they were text. An example of a data sample with image links is illustrated in Figure 6. To address this, an image processing step will be introduced to appropriately handle these image links.

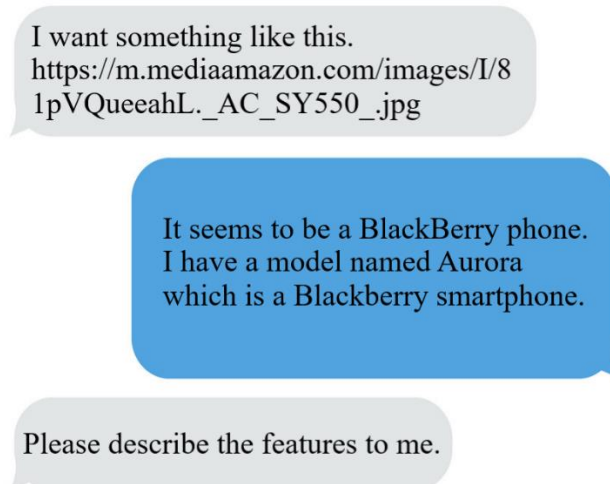


Figure 5: Buyer and Seller Negotiation Sample.

### 3.2.1. Multi-Modal Arabic Negotiation Workflow

The dataset used in [16] undergoes a translation from English to Arabic, and this translated dataset is then utilized to build our multi-modal Arabic negotiation model.

The model begins by receiving the buyer utterance  $U_0$ .

1. **Multimodal Integration:** If the buyer provides an image HTTP link  $I_0$  as part of the utterance  $U_0$ , the model processes the image using separate machine learning techniques designed for image analysis. This network extracts relevant information from the image and identifies slot-value pairs related to the user's query, adding this information to the belief state.
2. The model combines the textual belief state from  $U_0$  and the extracted slot-value pairs from the image (if provided) to form an extended belief state  $B_0$ .
3. The model executes a database query using the extended belief state  $B_0$  and identifies the count of database instances ( $D_0$ ) that align with the combined belief state.
4. Based on the combined information  $[U_0, B_0, D_0]$ , the model produces the necessary action for the agent  $A_0$  and generates an Arabic response  $R_0$ .
5. Similarly, at turn  $t$ , the model considers  $[U_0, B_0, D_0, A_0, R_0, \dots, U_{t-1}, B_{t-1}, D_{t-1}, A_{t-1}, R_{t-1}, U_t]$  as context and generates  $B_t$  (from  $U_t$  and  $I_t$  if provided),  $D_t$ ,  $A_t$ , and  $R_t$ , respectively.

The model parameters are updated using the same loss function and rewards used in Section 3.1. Baseline Workflow.

This workflow leverages multimodal capabilities to enhance the system's understanding of user queries by incorporating information from both text and images. It enables the model to extract valuable slot-value pairs from images and use them in conjunction with textual input to improve belief state extraction and response generation.

# CHAPTER 4

## EXPERIMENTS & RESULTS

### 4.1. Arabic Dataset

The primary challenge faced by every Arabic natural language processing project is the scarcity of datasets that accurately represent the data requirements for these applications [18]. Given the absence of native Arabic datasets, the primary approach for generating Arabic datasets involves the translation of English resources.

To construct the dataset for our model, we utilized the one introduced in [16]. In [16], the researchers created their own dataset by collaborating with electronic devices retailers. They generated 1000 instances of conversations between sellers and buyers (approximately 12,632 instances of pairs consisting of buyer utterances and corresponding seller responses), with a specific focus on the process of acquiring these devices. This dataset is annotated, capturing various aspects such as the user intent, slot-value, user sentiment, persuasive strategy, and dialogue act for each speech in the interaction. The dialogues in this dataset were crafted with attention to the Goal Unavailability task. This means that if a user is asking about a specific item that is unavailable, the seller tends to suggest an alternative item to fulfill the user's needs. Additionally, the dataset emphasizes personalized persuasion, where the seller's responses are tailored to the user's persona throughout the dialogue process.

To train our Arabic negotiation model, we translate the dataset from English to Arabic using the python Library Googletrans<sup>2</sup> that implements the google translate API.

---

<sup>2</sup> <https://pypi.org/project/googletrans/>

To evaluate the quality of the translation, we followed the work of [6]. We sample 100 utterances from the original English dataset and their respective translations to assess the quality of the translation. The goal is to evaluate the meaningfulness of translations rather than only focusing on how perfect they are. Upon evaluating the translation quality, it was observed that 6 out of 10 instances were deemed unreasonable, while the remaining 94 were considered reasonable. The reasons behind the unreasonable translations can be attributed to the informal nature of certain utterances and the inherent complexities in accurately translating English slang into Arabic. Examples of both accurate and unreasonable translations are presented in Table 1 and Table 2.

English Utterance	Arabic Translation
How much are you willing to spend?	كم أنت على استعداد لقضاء؟
No, Thank you for now. I will wait till save some cash.	لا، شكرا لك الآن. سأنتظر حتى أنقذ بعض النقود.

Table 1: Examples of unreasonable translations.

English Utterance	Arabic Translation
Hello and good morning. I am here to help you. Do you have any specification requirements?	مرحبا وصباح الخير. هل يمكن أن تساعدني في شراء جهاز كمبيوتر محمول؟
Yes, I have wide variety of smartpone. Could you please tell me your requirements?	نعم، لدي مجموعة واسعة من الهواتف الذكية. هل يمكن أن تخبرني من فضلك متطلباتك؟

Table 2: Examples of good translations.

## **4.2. Negotiation Models**

In this section, we present the evaluation results for the proposed negotiation model. During the training of this negotiation model, no image analysis was conducted on the HTTP image links provided by the buyer. This analysis will be carried out in the subsequent sections.

### ***4.2.1. Experiment Setup***

In our experimental setup, we employed the same configuration as outlined in [16]. The GPT-2 model from HuggingFace’s Transformers [23] served as the foundational framework for our experiments. For optimization, we used AdamW as our optimizer. We employed top-p sampling [24] as our decoding method with a temperature setting of 0.7. The loss function, as detailed in Section 0, was applied to penalize the model and update its parameters. This configuration was applied in both the English negotiation model and Arabic negotiation model experiments.

### ***4.2.2. English Negotiation Model***

In this section, we re-trained and evaluated the model outlined in [16] using the English dataset. The dataset was divided into training (80%), validation (10%), and test (10%) sets. The model was trained on the training and validation data, and its performance was evaluated using the test data.

Our evaluation metrics consist of Bilingual Evaluation Understudy [25] (Bleu-4) and cross-entropy loss. Bleu-4 assesses the quality of generated text by comparing it to a set of human reference texts. This metric emphasizes the accuracy and fluency of the generated text, evaluating the overlap of sequences of four consecutive words (4-grams)

between the generated text and the reference texts. The Bleu-4 score ranges between zero and one, where zero indicates no overlap at all, and one means that the generated text and the reference text are identical. Meanwhile, cross-entropy loss measures the dissimilarity between the predicted probability distribution over the vocabulary and the true distribution of the target text. Cross-entropy loss is represented as a number between zero and one, with lower values indicating more confident predictions.

The evaluation results for the English negotiation model show a cross-entropy loss of **0.45** and a BLEU-4 score of **0.28**. These findings align with the results reported in [16] and underscore the high quality of the model-generated responses.

#### ***4.2.3. Arabic Negotiation Model***

We conducted training on the Arabic translated data using the model outlined in [16]. Likewise, the dataset was partitioned into training (80%), validation (10%), and test (10%) sets. Training and validation data were used to train the model, and its performance was subsequently assessed using the test data.

The evaluation of the Arabic negotiation model shows a cross-entropy of **0.7** and a BLEU-4 score of **0.07**. The automatic evaluation results for the Arabic negotiation model indicate suboptimal performance in generating Arabic text compared to the English model, as the results are lower than those obtained for the English model. These findings indicate that the English models exhibited significantly superior performance despite using the same amount of data. This performance discrepancy reflects the complexity of the Arabic language and its morphological intricacies, highlighting the need for a larger dataset to attain learning levels comparable to English.



#### ***4.2.4. Arabic Negotiation Model Enhanced with Augmented Data***

To enhance the performance of the Arabic negotiation model, we augmented the Arabic dataset by incorporating data generated through GPT-3.5 Turbo. We specifically prompted GPT-3.5 Turbo to generate seller and buyer Arabic negotiation utterances within the context of item sales negotiations. Ensuring that the negotiation buyer utterances include HTTP image links was important for subsequent image analysis steps. The model database instances include a set of phone and laptop images HTTP links, which were provided for GPT-3.5 Turbo to be used while generating the buyer utterances. This process resulted in the creation of approximately 2000 additional instances, each comprising pairs of Arabic buyer utterances and their corresponding seller responses.

We conducted the experiment again on the augmented data, and the results demonstrate an improvement in the performance of the Arabic negotiation model. The cross-entropy loss was recorded at **0.58**, and the BLEU-4 score reached **0.2**, surpassing the performance of the Arabic negotiation model trained in the previous section. By incorporating data generated through GPT-3.5 Turbo and expanding the dataset with more instances, we provided the model with a richer and more diverse set of examples enabling it to generalize better to unseen data. The augmented dataset allowed the model to capture a wider spectrum of linguistic nuances, negotiation strategies, and contextual variations which helped in enhancing the Arabic model performance and allowing it to make more accurate predictions and generate more contextually relevant responses during the negotiation process.

Since the model derived from this experiment demonstrated improved performance, it will be adopted for further integration with the image analysis model.

The evaluation results of three trained negotiation models, presented in Table 3, indicate that the English Negotiation Model outperforms the Arabic counterparts in terms of BLEU-4 and cross-entropy scores. The lower BLEU-4 score for Arabic suggests challenges in generating text that closely aligns with reference sequences, potentially due to the language's complex syntax and larger vocabulary. Furthermore, the Arabic Negotiation Model, when enhanced with augmented data, demonstrates improved performance compared to the original Arabic model, emphasizing the positive impact of a more extensive and diverse training dataset. The findings suggest that language-specific characteristics and the inclusion of augmented data play crucial roles in determining the effectiveness of the trained negotiation models.

<b>Model</b>	<b>BLEU-4</b>	<b>Cross-Entropy Loss</b>
English Negotiation Model	<b>0.28</b>	<b>0.45</b>
Arabic Negotiation Model	0.07	0.7
Arabic Negotiation Model Enhanced with Augmented Data	0.2	0.58

Table 3: Evaluation of the trained negotiation models using Bleu-4 and cross-entropy loss.

#### ***4.2.5. Human Evaluation of the Adopted Arabic Negotiation Model***

Automatic evaluation metrics alone are insufficient for assessing machine-generated texts. According to [26], automated evaluations for machine-generated texts do not correlate with human evaluation. This emphasizes the importance of conducting human evaluations to accurately assess the quality of the generated texts from the Arabic negotiation model.

To manually evaluate our model's performance, we selected 100 examples from the testing datasets (comprising 20 full conversations, each with 5 turns). We then fed

these examples into the trained Arabic negotiation model to generate responses. Arabic-speaking rater then assessed these generated responses. The rater was asked to evaluate the model's persuasion skills, response consistency, and fluency. For this assessment, rater was required to answer three questions regarding the model's persuasion skills, response consistency, and fluency:

- Fluency: Were the generated responses understandable from a language perspective?
- Consistency: Was the model's response consistent with the negotiation context?
- Model Persuasion: Were the model's responses persuasive?

Each response has been assigned three ratings, with each rating ranging from 1 to 5, where 1 signifies the lowest and 5 represents the highest score. The manual evaluation revealed the following average ratings: 3.5 for fluency, 3.06 for consistency, and 3.0 for persuasion.

The detailed results are provided in Figure 6. The manual evaluation indicates that the model has succeeded in generating responses that are fluent, consistent, and persuasive. The manual evaluation of this model confirms its success in producing responses characterized by fluency, consistency, and persuasiveness. Fluency is observed in the smooth and natural flow of language, ensuring coherent and easily comprehensible communication. Consistency is maintained across responses, providing users with reliable and predictable interaction. The model's persuasiveness is highlighted by its ability to engage and convince users effectively, influencing their actions. In summary, the evaluation demonstrates the chatbot's ability in generating linguistically sound, consistent, and compelling responses.

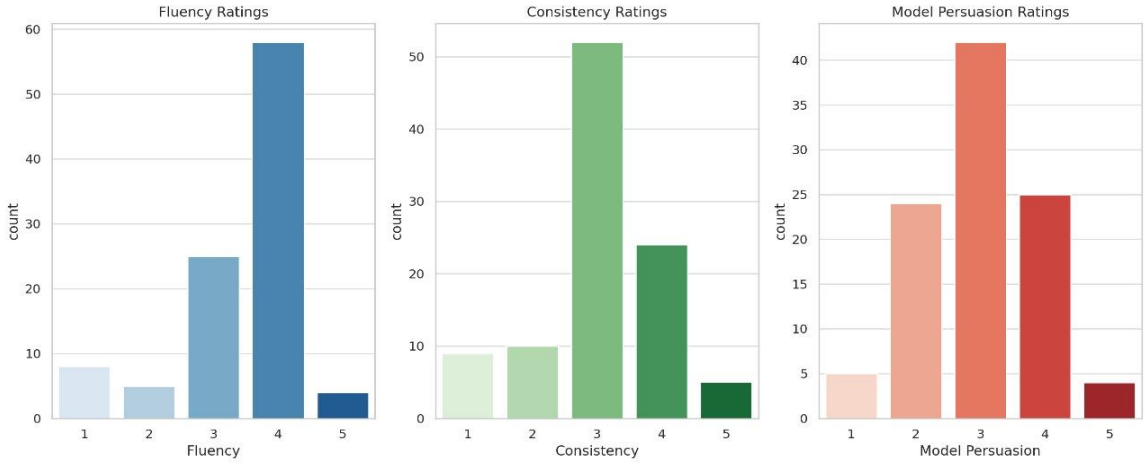


Figure 6: Manual evaluation results of the Arabic negotiation model.

### 4.3. Image-Based Color and Brand Recognition Model

In this experiment, we train a convolutional neural network (CNN) model to predict the color and brand of a product from an input product image.

#### 4.3.1. Dataset

To train our CNN model, we gathered a dataset of product images along with corresponding brand and color annotations sourced from the product databases referenced in [16]. The dataset comprises approximately 5000 product images, encompassing laptops and phone products. The images are labeled with various color categories, including Black, White, Silver, Gray, Gold, and Pink, and belong to distinct brands such as Apple, Samsung, Huawei, BlackBerry, and Dell. The dataset is divided, with 80% allocated for training, 10% for validation, and 10% for testing purposes.

### 4.3.2. Color and Brand Recognition Model

The goal of this experiment was to predict both the color and brand of products from images, leveraging a multi-output CNN model. Inspired by existing research [27][28], we built our model upon the VGG16 architecture. The VGG-16 network consists of 13 convolution layers, each with a 3\*3 size and a step length of 1. All convolution layers share the same configuration. Additionally, the network includes five max-pooling layers with a 2\*2 size and a step length of 2, along with three fully connected layers. The pre-trained convolutional layers of VGG16 excel at feature classification and extraction, making it a powerful base for our model. This pre-trained base not only reduces the need for extensive training from scratch but also provides a solid foundation for our multi-output model, enhancing overall performance. Building on this VGG16 base, we constructed a custom multi-output model. We defined an input layer, followed by the VGG16 base layers. Then, we added two separate output layers - one for predicting color and one for brand.

The evaluation results of this CNN model are provided in Table 4. The model demonstrates an accuracy of **84%** for color prediction and **75%** for brand prediction.

Evaluation Metrics	Color Prediction	Brand Prediction
Model Accuracy	<b>84%</b>	<b>75%</b>

Table 4: Evaluation Results of CNN Model for Color and Brand Prediction

### 4.4. Arabic Multi-Modal Negotiation Model

In this experiment, we integrated the color and brand recognition model with the Arabic Negotiation model to predict the buyer's requested product requirements from the images they provide. In this process, after the buyer's utterance is input into the

negotiation model, if an image HTTP link is included in the utterance, the link is extracted, the image is downloaded, and then fed into the color and brand recognition model for color and brand identification. The recognized brand and color information are added into the model context, and subsequently, the response generation task is performed.

The Arabic Multi-Modal Negotiation model undergoes evaluation on the test data, resulting in an automatic evaluation score of BEU-4 at **0.21** and a cross-entropy loss of **0.55**. Additionally, a manual evaluation for the integrated model was also conducted, yielding average ratings of 3.55 for fluency, 3.31 for consistency, and 3.27 for persuasion. Details of the manual evaluation are provided in Figure 7. The performance of the Arabic Multi-Modal Negotiation model surpasses that of the standard Arabic Negotiation model, likely due to the richer context provided to the trained GPT-2 model.

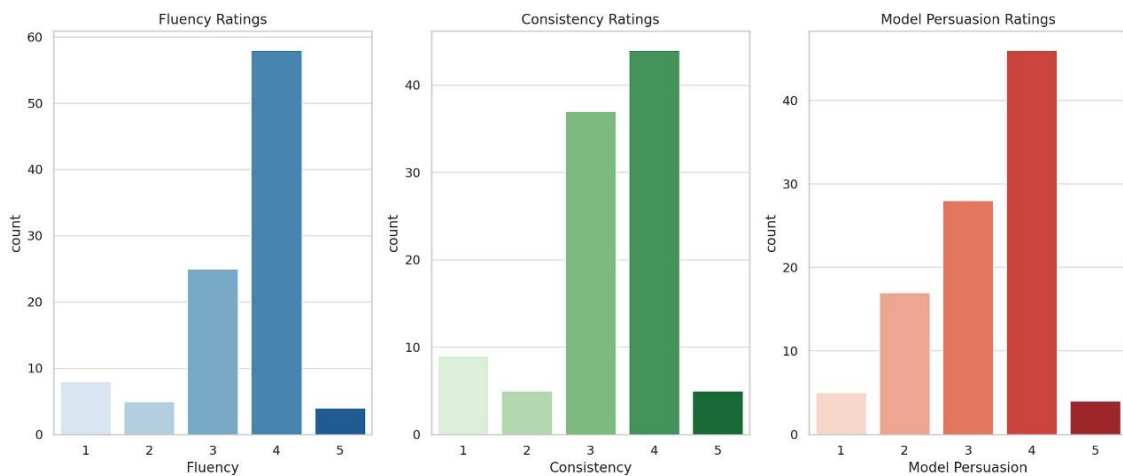


Figure 7: Manual evaluation results of the Multi-Modal Arabic negotiation model.

#### **4.5. Statistical Significance**

To gain deeper insights into the significance of improvements in BLEU scores between the Arabic negotiation model and the Arabic multi-modal negotiation model, a paired bootstrap resampling test was employed inspired by the work in [31]. This non-parametric statistical method is known for its robustness to violations of classical statistical assumptions, making it suitable for scenarios with limited data.

In this study, 500 buyer utterances were randomly sampled from the dataset and seller responses were generated using both models. The BLEU-4 scores were compared, and the paired bootstrap procedure was used to calculate the p-value and win-loss ratio of the models.

The results indicate that the multimodal Arabic negotiation model outperforms the Arabic negotiation model, with an average p-value of 0.28. This p-value suggests a 72% probability that the observed improvement in scores is not due to chance. This statistical analysis provides robust evidence supporting the superiority of the multimodal Arabic negotiation model in terms of BLEU scores.

#### **4.6. Discussion**

The output of the Arabic multi-modal negotiation model shows that the model has achieved a high quality of generated responses in terms of fluency, consistency, and persuasion. Notably, the English model demonstrated superior performance in the automatic evaluation results. Even when compared to the Arabic Negotiation Model Enhanced with Augmented Data, the English model outperformed, suggesting that the English language benefits from improved performance even with a smaller amount of consistent data. This discrepancy may be attributed to potential error margins in the translation of training and test data, as well as the complexity and extensive vocabulary

of the Arabic language compared to English. Furthermore, when evaluating the performance of the Arabic negotiation model against the multi-modal negotiation model, the automatic evaluation results appear similar. However, the integration of feature recognition models has positively influenced manual evaluation results, indicating improved human-rated fluency, consistency, and persuasion. This suggests that the model gained a better understanding of buyer inquiries, leading to more consistent responses within the negotiation context. The findings highlight the necessity to enhance the Arabic model's performance through additional training data. Moreover, improvements to the CNN model are essential to extract more product features from images, thereby providing richer information about the negotiation context and enhancing the model's performance in terms of consistency and persuasive abilities.

#### **4.7. Limitation**

The limitations of this work extend across various areas, with a significant challenge encountered in most Arabic natural language generation (NLG) tasks being the absence of native Arabic datasets. The available datasets for Arabic natural generation tasks are typically translations from the English language, introducing potential errors in translation and impacting the overall performance of the Arabic language model. Additionally, the scarcity of data for multimodal negotiation dialogue tasks further restricts the development of comprehensive models, limiting exposure to diverse negotiation contexts and hindering generalization in real-world scenarios.



## CHAPTER 5

### CONCLUSION & FUTURE WORK

In this section, we summarize the main findings of the thesis and present future work.

#### **5.1. Conclusion**

In this study, we introduce a research project focused on training a multimodal Arabic negotiation bot. To achieve this, we utilized the model presented in [16] as our foundational framework. Initially, we retrained the model from [16] using an English dataset for comparative analysis. For training the model in the Arabic language, we translated the English dataset into Arabic and conducted a retraining process. Due to the suboptimal performance of the Arabic model compared to the English counterpart, attempts were undertaken to improve it by augmenting the Arabic dataset using data generated by GPT-3 Turbo. Subsequent retraining of the Arabic model on the augmented dataset resulted in improved metrics, specifically, better BLEU-4 scores and reduced cross-entropy loss. To incorporate multimodality, we developed a CNN-based model designed to predict product colors and brands. Integration of this CNN model with the Arabic negotiation model yielded the best automatic and manual results in terms of fluency, consistency, and persuasion compared to all other Arabic negotiation trained models. This work not only marks the development of the first multimodal Arabic negotiation model but also lays the foundation for subsequent research into multimodal and multilingual models.

## 5.2. Future Work

As we consider the future direction of this research, there are several opportunities for improvement and exploration. Firstly, enhancing the performance of our model can be achieved by using a larger dataset generated by GPT-3. Employing a more extensive dataset will contribute to refining the linguistic nuances of the Arabic negotiation bot and improving its overall performance. A critical step toward enhancing the authenticity and effectiveness of the model involves acquiring and incorporating a native Arabic dataset. The integration of culturally diverse and linguistically rich data will lead to a more nuanced understanding of Arabic negotiation dynamics, refining the bot's responses and adaptability to real-world scenarios. Moreover, although the multimodal aspect of our model has shown promising results, there is room for improvement by extracting additional features from images. Future work may involve a thorough exploration of advanced techniques for image analysis, such as enhancing the capabilities of the CNN model or considering the integration of Large Language Models (LLMs) specifically designed for image analysis. These approaches could significantly enhance the extraction of meaningful information from the visual components of the negotiation process.

## APPENDIX

### EXAMPLES ON RESPONSE GENERATION

In this section, we provide examples of seller-generated utterances. We present two examples, one for the Arabic negotiation model enhanced with augmented data (Table 5) and the other for the Arabic multi-modal negotiation model (Table 6).

Turn	Utterance
Buyer	مرحبًا، أنا أبحث عن هاتف جديد
Seller	مرحبًا، هل يمكنك أن تخبرني ما هي الميزات التي تريدها؟
Buyer	Apple أريد شراء هاتف
Seller	هل لديك أي تفضيلات أخرى؟
Buyer	ما هي تكلفتها؟
Seller	يكلف 250 يورو
Buyer	حسنًا، هل يمكنك حجزها من أجلي
Seller	بالتأكيد، سأقوم بحجزها لك
Buyer	شكرًا جزيلًا

Table 5: Negotiation Example with the Arabic negotiation model enhanced with augmented data as the seller.

Turn	Utterance
Buyer	مرحباً، أنا أبحث عن هاتف ذكي من سامسونج
Seller	مرحباً، هل يمكنك أن تخبرني ما هي الميزات التي تريدها؟
Buyer	أنا مهتم بصورة هذا الهاتف <a href="http://cdn2.gsmarena.com/vv/bigpic/samsung-i7500.jpg">http://cdn2.gsmarena.com/vv/bigpic/samsung-i7500.jpg</a>
Seller	أسود. Samsung هذا هو هاتف ما هي ميزانيتك؟
Buyer	أرغب في البقاء ضمن ميزانية 400 دولار
Seller	إنه ضمن ميزانيتك. هل تريد المتابعة؟
Buyer	نعم، أعجبنى
Seller	بالتأكيد، سأقوم بحجزها لك
Buyer	شكراً جزيلاً

Table 6: Negotiation Example with the Arabic multi-modal negotiation model as the seller.

## BIBLIOGRAPHY

- [1]- Fisher, R., Ury, W. L., & Patton, B. (2011). *Getting to yes: Negotiating agreement without giving in*. Penguin.
- [2]- Baarslag, T., Kaisers, M., Gerding, E., Jonker, C. M., & Gratch, J. (2017). When will negotiation agents be able to represent us? The challenges and opportunities for autonomous negotiators. *International Joint Conferences on Artificial Intelligence*.
- [3] - Hijjawi, M., Bandar, Z., Crockett, K., & Mclean, D. (2014, March). ArabChat: An arabic conversational agent. In *2014 6th International Conference on Computer Science and Information Technology (CSIT)* (pp. 227-237). IEEE.
- [4] - Ali, D. A., & Habash, N. (2016, December). Botta: An arabic dialect chatbot. In *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: System Demonstrations* (pp. 208-212).
- [5] - Antoun, W., Baly, F., & Hajj, H. (2020). Arabert: Transformer-based model for arabic language understanding. *arXiv preprint arXiv:2003.00104*.
- [6] - Naous, T., Hokayem, C., & Hajj, H. (2020, December). Empathy-driven Arabic conversational chatbot. In *Proceedings of the Fifth Arabic Natural Language Processing Workshop* (pp. 58-68).
- [7]- Diamah, A., Mohammadian, M., & Balachandran, B. (2012, August). Fuzzy utility and inference system for bilateral negotiation. In *2012 2nd International Conference on Uncertainty Reasoning and Knowledge Engineering* (pp. 115-118). IEEE.
- [8]- Hussain, M. J. (2014, April). Parallel processing and intercommunication of multiple negotiation processes in E-commerce for optimal business. In *2014 Fourth International Conference on Communication Systems and Network Technologies* (pp. 583-586). IEEE.
- [9]- Jennings, N. R., Faratin, P., Lomuscio, A. R., Parsons, S., Sierra, C., & Wooldridge, M. (2001). Automated negotiation: prospects, methods and challenges. *International Journal of Group Decision and Negotiation*, 10(2), 199-215.
- [10]- Koley, G., & Rao, S. (2018, October). Adaptive human-agent multi-issue bilateral negotiation using the Thomas-Kilmann conflict mode instrument. In *2018 IEEE/ACM 22nd International Symposium on Distributed Simulation and Real Time Applications (DS-RT)* (pp. 1-5). IEEE.
- [11]- Lewis, M., Yarats, D., Dauphin, Y. N., Parikh, D., & Batra, D. (2017). Deal or no deal? end-to-end learning for negotiation dialogues. *arXiv preprint arXiv:1706.05125*.
- [12]- He, H., Chen, D., Balakrishnan, A., & Liang, P. (2018). Decoupling strategy and generation in negotiation dialogues. *arXiv preprint arXiv:1808.09637*.

- [13]- Wang, X., Shi, W., Kim, R., Oh, Y., Yang, S., Zhang, J., & Yu, Z. (2019). Persuasion for good: Towards a personalized persuasive dialogue system for social good. *arXiv preprint arXiv:1906.06725*.
- [14]- Zhou, Y., He, H., Black, A. W., & Tsvetkov, Y. (2019). A dynamic strategy coach for effective negotiation. *arXiv preprint arXiv:1909.13426*.
- [15]- Mishra, K., Samad, A. M., Totala, P., & Ekbal, A. (2022, October). PEPDS: A polite and empathetic persuasive dialogue system for charity donation. In *Proceedings of the 29th International Conference on Computational Linguistics* (pp. 424-440).
- [16]- Raut, A., Tiwari, A., Das, S., Saha, S., Maitra, A., Ramnani, R., & Sengupta, S. (2023). Reinforcing personalized persuasion in task-oriented virtual sales assistant. *Plos one*, 18(1), e0275750.
- [17]- Zhan, H., Wang, Y., Feng, T., Hua, Y., Sharma, S., Li, Z., ... & Haffari, G. (2022). Let's Negotiate! A Survey of Negotiation Dialogue Systems. *arXiv preprint arXiv:2212.09072*.
- [18]- Shaalan, K., Siddiqui, S., Alkhatib, M., & Abdel Monem, A. (2019). Challenges in Arabic natural language processing. In *Computational linguistics, speech and image processing for arabic language* (pp. 59-83).
- [19]- Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., & Sutskever, I. (2019). Language models are unsupervised multitask learners. *OpenAI blog*, 1(8), 9.
- [20]- Shi, W., Li, Y., Sahay, S., & Yu, Z. (2020). Refine and imitate: Reducing repetition and inconsistency in persuasion dialogues via reinforcement learning and human demonstration. *arXiv preprint arXiv:2012.15375*.
- [21]- Jaccard, P. (1901). Comparative study of floral distribution in a portion of the Alps and Jura. *The Company Vaudoise Bulletin of Natural Sciences*, 37(5), 547-579.
- [22]- Banerjee, S., & Lavie, A. (2005, June). METEOR: An automatic metric for MT evaluation with improved correlation with human judgments. In *Proceedings of the acl workshop on intrinsic and extrinsic evaluation measures for machine translation and/or summarization* (pp. 65-72).
- [23]- Wolf, T., Debut, L., Sanh, V., Chaumond, J., Delangue, C., Moi, A., ... & Rush, A. M. (2019). Huggingface's transformers: State-of-the-art natural language processing. *arXiv preprint arXiv:1910.03771*.
- [24]- Holtzman, A., Buys, J., Du, L., Forbes, M., & Choi, Y. (2019). The curious case of neural text degeneration. *arXiv preprint arXiv:1904.09751*.
- [25]- Papineni, K., Roukos, S., Ward, T., & Zhu, W. J. (2002, July). Bleu: a method for automatic evaluation of machine translation. In *Proceedings of the 40th annual meeting of the Association for Computational Linguistics* (pp. 311-318).

- [26]- Liu, C. W., Lowe, R., Serban, I. V., Noseworthy, M., Charlin, L., & Pineau, J. (2016). How not to evaluate your dialogue system: An empirical study of unsupervised evaluation metrics for dialogue response generation. *arXiv preprint arXiv:1603.08023*.
- [27]- Zhang, M., Wang, P., & Zhang, X. (2019, July). Vehicle color recognition using deep convolutional neural networks. In *Proceedings of the 2019 International Conference on Artificial Intelligence and Computer Science* (pp. 236-238).
- [28]- Rachmadi, R. F., & Purnama, I. (2015). Vehicle color recognition using convolutional neural network. *arXiv preprint arXiv:1510.07391*.
- [29]- Wallace, R. (1995). Artificial linguistic internet computer entity (alice). *City*.
- [30]- Naous, T., Antoun, W., Mahmoud, R. A., & Hajj, H. (2021). Empathetic BERT2BERT conversational model: Learning Arabic language generation with little data. *arXiv preprint arXiv:2103.04353*.
- [31]- P. Koehn, "Statistical significance tests for machine translation evaluation," in *Proceedings of the 2004 conference on empirical methods in natural language processing, 2004*, pp. 388-395.