Multimodal Arabic Negotiation Bots

Samah Albast
American Univeristy of Beirut
Beirut, Lebanon
sma122@mail.aub.edu

Wassim Elhajj
American Univeristy of Beirut - Mediterraneo
Paphos, Cyprus
wassim@aubmed.ac.cy

Hazem Hajj
American University of Beirut
Beirut, Lebanon
hh63@aub.edu.lb

Khaled Bashir Shaban
Qatar University
Doha, Qatar
khaled.shaban@qu.edu.qa

Shady Elbassuoni
American Univeristy of Beirut
Beirut, Lebanon
se58@aub.edu.lb

ABSTRACT

Negotiation is a fundamental aspect of human interaction. With recent advancements in chatbots, leveraging artificial intelligence for negotiation has emerged as an ideal application. Despite significant progress in English negotiation bots, such advancements are notably absent in Arabic. Furthermore, while previous research has focused on developing high-performing neural response generation systems for negotiation bots, the integration of multi-modality remains unexplored. This work presents the first Arabic multi-modal negotiation bot presented by 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 fine-tuned a generative pre-trained transformer (GPT-2) model on an Arabic dataset, integrating it with reinforcement learning for more coherent and persuasive responses, and a convolutional neural network to support multimodality. To evaluate our model, we relied on both automatic evaluation using established metrics such as cross-entropy loss and the BLEU score, as well as human evaluation in terms of fluency, consistency and persuasion. Our evaluation results reveal both the successes and limitations of the designed multi-modal Arabic negotiation bot, offering insights into the inherent challenges and setting directions for future research.

CCS CONCEPTS

• **Computing methodologies → Discourse, dialogue and pragmatics**.

KEYWORDS

negotiation bot, automatic negotiations, virtual assistant, e-commerce chatbot, Arabic NLP

ACM Reference Format:
Samah Albast, Wassim Elhajj, Hazem Hajj, Khaled Bashir Shaban, and Shady Elbassuoni. 2024. Multimodal Arabic Negotiation Bots. In *Proceedings of the first workshop on Generative AI for E-Commerce 2024, October 25, 2024.* ACM, New York, NY, USA, 6 pages.

1 INTRODUCTION

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 [4]. With the recent advances in chatbots, negotiation appears to be an ideal application for artificial intelligence (AI). Automating the negotiation process has significant potential for smoothing conflict resolution and enhancing outcomes across various domains, such as marketplaces and autonomous driving [1]. The field of English negotiation bot research has undergone significant advancements, initially relying on game theory and rule-based approaches [3, 6, 7, 9]. This was followed by the adoption of deep learning techniques, which have demonstrated remarkable efficacy in natural language processing (NLP) applications.

One such pioneering work [10] 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. Building upon this foundation, subsequent studies, such as [5, 12, 16, 17, 21], have introduced enhancements including strategies for incorporating emotions, persuasion, and politeness into negotiation dialogues. In more recent work [16], the authors employed a generative pre-trained transformer-based model, GPT-2, and reinforcement learning techniques to develop a negotiation bot acting as a seller in e-commerce transactions with buyers.

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 mostly unexplored. In this context, 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 [16]. In addition to the scarcity of research on multimodal negotiation bots, to the best of our knowledge, research on Arabic negotiation bots in general is still unaddressed. Our work seeks to fill these research gaps by developing the first multimodal Arabic negotiation bot, inspired by the approach outlined in [16].

Our Arabic negotiation bot was developed by fine-tuning a generative pre-trained transformer (GPT-2) model using a translated version of the personalised persuasive dialogue (PPD) dataset [16]. The PPD dataset comprises 1000 conversations between sellers and buyers of various electronic devices (mobile phone, tablet, camera, etc.). The trained negotiation bot takes on the role of the seller, identifies a suitable device from an underlying database that matches the requirements of the buyer, and then tries to persuade the buyer to purchase the identified device or alternatives. To generate more coherent and fluent responses, and be more persuasive, reinforcement learning is also infused in the model. Finally, to incorporate multimodality, the buyer is able to also provide an image of the desired device, and we trained a custom convolutional neural network to detect the brand and color of the device based on the provided image, which are then used in conjunction with the buyer textual interactions to identify a suitable device and persuade the buyer to complete a purchase of the identified device or alternatives.

To evaluate our developed negotiation bot, we relied on automatic evaluation using the biLingual evaluation understudy (BLEU) score and the cross-entropy loss on a holdout test-set from the PPD dataset. Our negotiation bot achieved a BLEU-4 score of 0.21 and a cross-entropy loss of 0.55. Moreover, we also performed human evaluation in terms of fluency, consistency and persuasion, achieving an average rating of 3.55, 3.31, and 3.27, respectively. The automatic and human evaluation both demonstrate promising results for the first Arabic multimodal negotiation bot

2 RELATED WORK

The field of automated negotiation has witnessed the investigation of diverse methodologies [19]. One traditional approach is the use of rule-based methods. These rules specify how the agent should react in different situations encountered during the negotiation process [3, 6]. Additionally, game theory has been employed in the development of negotiation bots because it can be applied to interactions involving self-interested agents, and negotiation participants are driven by their interests and objectives. 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 [7]. Another family of approaches employed in the development of negotiating agents is the heuristics approaches such as the introduction of two new heuristics in [9] to guide the decision-making process. All the previously mentioned approaches to build a negotiation bot rely on manually crafted sets of rules and cannot engage in negotiation using natural language.

Recent efforts have turned to leveraging deep learning techniques for the development of negotiation bots. The work in [10] is the first to apply deep learning techniques to build an end-to-end model for natural language negotiations. This model acquires both linguistic and reasoning skills through a combination of supervised and reinforcement learning techniques. In [5], they addressed the problem of the degeneracy of the work in [10] by decoupling the negotiating strategy from language generation. 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 [17], 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 [21], the authors developed a negotiation coach offering tactics to enhance deals for the seller, utilizing an LSTM-based model to generate tactic suggestions. In [12], the authors trained and fine-tuned a language model using reinforcement learning while considering various sub-rewards for persuasion, emotion, politeness, coherence, and repetitiveness. In [16], the authors developed a persuasive sales agent to persuade a buyer to buy a target item. The authors used a GPT-2 model combined with reinforcement learning that has 2 sub-rewards (repetitiveness and consistency). This is the first work to tackle 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 work in [16] is considered the first effort to develop an end-to-end dialogue agent capable of persuading users in goal unavailability situations. The experimental results demonstrate that 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 negotiation bot. To the best of our knowledge, our work is the first to explore the use of deep learning in the area of multimodal Arabic negotiation bots.

3 APPROACH

Our Arabic negotiation bot is inspired by the framework in [16]. It is based on a fine-tuned GPT-2 model using a translated version of the PPD dataset, which consists of conversations between sellers and buyers across various electronic devices such as mobile phones, tablets, cameras, etc. Functioning as the seller, the trained negotiation bot discerns an appropriate device from an underlying database that aligns with the buyer's needs. Subsequently, it aims to convince the buyer to acquire the identified device or explore alternative options. To enhance the coherence and persuasiveness of responses, reinforcement learning (RL) techniques are integrated into the model. Furthermore, to support multimodality, buyers have the ability to provide an image of their desired device, which is in turn subjected to a convolutional neural network (CNN) to extract brand and color information of the desired device. This information, combined with textual interactions with the buyer, aids in identifying suitable devices and persuading the buyer to finalize a purchase. We first describe the basic framework we based our negotiation bot on. Next, we describe how we extended such framework to develop our own multimodal Arabic negotiation bot.

3.1 Basic Framework

The basic framework we based on our negotiation bot on works as follows. Given a first utterance of the buyer U_0 at turn 0, the system proceeds as follows:

(1) First, it extracts the belief state B_0, consisting of slot-value pairs. Each pair $(\text{slot}_i, \text{value}_i)$, denoted as (s_i, v_i), encapsulates slot and value information extracted from the current utterance. For example, a slot-value pair can be {brand: Samsung}, or {ram: 4G}.
(2) The system 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) Finally, based on the context $[U_0, B_0, D_0]$, the system utilizes GPT-2 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 system considers $[U_0, B_0, D_0, A_0, R_0, \ldots, D_{t-1}, A_{t-1}, R_{t-1}, U_t]$ as context and generates B_t, D_t, A_t, and R_t, respectively.

To enhance the coherence and persuasiveness of the generated responses, the framework employs reinforcement learning as follows. At the end of every turn t, a reward r_t is calculated based on two sub-rewards: repetitiveness and consistency. Jaccard coefficient is used to measure repetitiveness and Meteor score [2] is used for consistency. This reward guides the system to generate more refined responses. At the end of the whole conversation, an average r of the rewards obtained at the end of each turn is calculated as shown in Equation 1. In addition, the cross-entropy loss l is also calculated, and the sum of the reward r and the cross-entropy loss l is considered as the end loss L used to update the model's parameters as shown in Equation 2.

$$r = \frac{\sum_{t=1}^{n} r_t}{n} \tag{1}$$

$$L = l + r \tag{2}$$

Combining the aggregated rewards r with the cross-entropy loss l creates a composite loss function that balances the need for fluency and coherence (cross-entropy) with the need for high-quality responses (reward-based scores).

3.2 Multimodal Arabic Negotiation Bot

The basic framework outlined above serves as the basis for our proposed multimodal arabic negotiation bot. To adopt this framework for the Arabic language, the base GPT-2 model had to be fine-tuned using an Arabic dataset. On the other hand, to support multimodality, the framework was extended to be able to extract belief states from images of the desired device provided by the buyer through a customized CNN. The complete framework of our multimodal Arabic negotiation bot is depicted in Figure 1. We first describe how we fine-tuned the base GPT-2 model for Arabic negotiations, then we explain how we trained the customized CNN to be able to extract belief states from images.

3.2.1 Fine-tuning the GPT-2 Model. Similar to the basic framework, the GPT-2 model from HuggingFace's Transformers [18] served as the foundational model for our Arabic negotiation bot. To fine-tune the model, we resorted to automatic translation of the PPD dataset [16] using the Google Translate API [1]. The PPD dataset consists

[1]https://pypi.org/project/googletrans/

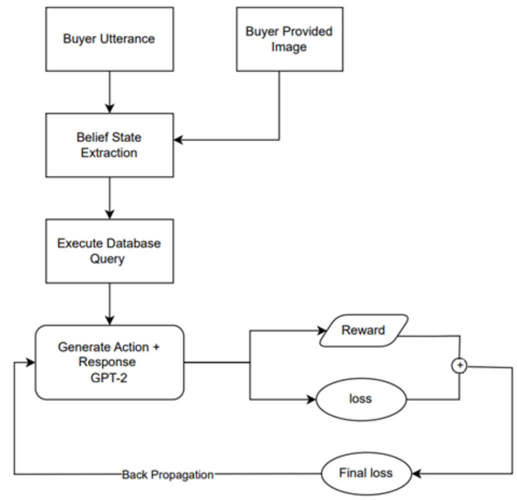

Figure 1: Multimodal Arabic Negotiation Bot Framework

of 1000 conversations between sellers and buyers about various electronic devices such as mobile phones, tablets, cameras, etc. The PPD dataset is annotated to capture various aspects such as the user intent, slot-value pairs, user sentiment, persuasive strategy, and dialogue act for each speech in the interaction.

To assess the quality of the automatic translation, we followed the approach in [13]. We sampled 100 utterances from the original English dataset and their corresponding translations and observed that only 6 out of the 100 sampled instances were deemed inadequate. The reasons behind the inadequacy of those 6 translations can be attributed to the informal nature of certain utterances and the inherent complexities in accurately translating English slang into Arabic.

Once an Arabic version of the PPD dataset was obtained through translation, we divided it into training (80%), validation (10%), and test (10%) sets. We then used AdamW to fine-tune the GPT-2 model using the training and validation sets, and employed top-p sampling as our decoding method with a temperature setting of 0.7.

3.2.2 CNN Model. To be able to support multimodality in our Arabic negotiation bot, we trained a CNN model to predict the color and brand of a device from an input image provided by the buyer. To train our CNN model, we generated a dataset of device images, along with corresponding brand and color annotations, which we obtained from the referenced databases in [16]. The dataset comprises approximately 5000 images of various devices, such as laptops and mobile phones. The images were labeled with various color categories, including Black, White, Silver, Gray, Gold, and Pink, and they belonged to distinct brands such as Apple, Samsung, Huawei, BlackBerry, and Dell. The dataset was divided into training (80%), validation (10%), and test (10%) sets.

Our CNN model is based on the VGG16 architecture [15, 20]. It consists of 13 convolution layers, each with a $3 * 3$ filter and a stride of 1. Additionally, the network includes five max-pooling layers with a $2 * 2$ filter and a stride of 2, and ending with three fully-connected layers. Building on this VGG16 architecture, we

Table 1: Automatic Evaluation Results

Model	Cross-Entropy Loss	BLEU-4 Score
GPT-2 (English)	0.45	0.28
GPT-2 (Arabic)	0.73	0.07
Augmented GPT-2 (Arabic)	0.58	0.20
Augmented GPT-2 + CNN (Arabic)	0.55	0.21

constructed a custom multi-class CNN. We defined an input layer, followed by the VGG16 base layers as described above and then we added two separate output layers - one for predicting color and one for brand. The CNN model achieved a test accuracy of 84% for color prediction and 75% for brand prediction.

3.2.3 Final Framework. Similar to the basic framework, our final framework for the multimodal Arabic negotiation bot begins by receiving the buyer utterance U_0. It then proceeds as follows.

(1) If the buyer provides an image HTTP link I_0 as part of the utterance U_0, the framework processes the image using the customized CNN to extract the color and brand information from the image and generate slot-value pairs corresponding to those, which are then added to the belief state.
(2) The system then 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 system then 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 $[U_0, B_0, D_0, A_0, R_0, \ldots, A_{t-1}, R_{t-1}, U_t]$ 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 in the basic framework (Section 3.1).

4 EXPERIMENTS

4.1 Automatic Evaluation

To evaluate the responses generated by our multimodal Arabic negotiation bot described in Section 3, we relied on two established metrics to evaluate text generation, namely the bilingual evaluation understudy (BLEU-4) [14] and cross-entropy loss. BLEU-4 assesses the quality of generated text by comparing it to a set of human reference texts. Meanwhile, cross-entropy loss measures the dissimilarity between the predicted probability distribution over the vocabulary and the true distribution of the target text.

To establish a baseline for evaluation, we fine-tuned the GPT-2 model outlined in [16] using the English PPD dataset, since it serves as our basic framework as explained in Section 3.1. 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. The first row of Table 1 shows the cross-entropy loss and the BLEU-4 score of this baseline model when evaluated using the test set. The test results for the English negotiation bot achieved a cross-entropy loss of 0.45 and a BLEU-4 score of 0.28. These findings align with the results reported in [14] and underscore the high quality of the basic framework's generated responses.

As a second baseline, we evaluated the Arabic GPT-2 model described in Section 3.2.1 that was fine-tuned using the training and validation sets of the Arabic version of the PPD dataset, without using the CNN model (i.e., without any multimodality support), by evaluating its responses for the test set. The test results of this second baseline, which we refer to as the GPT-2 (Arabic) model in Table 1 (second row) achieved a cross-entropy loss of 0.7 and a BLEU-4 score of 0.07. These results indicate that the English baseline exhibited significantly superior performance compared to its Arabic counterpart 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.

To enhance the performance of this second baseline when performing Arabic negotiations, we augmented the translated PPD dataset with GPT-3.5 Turbo-generated data, specifically focused on buyer and seller Arabic negotiation utterances in the context of item sales negotiations. The augmented dataset, included around 2000 additional instances. The GPT-2 model was then fine-tuned and validated using this augmented dataset, the same way the baseline GPT-2 model was fine-tuned as described in Section 3.2.1. The test results of this augmented GPT-2 model are shown in the third row of Table 1. It achieved a cross-entropy loss of 0.58 and a BLEU-4 score of 0.2, surpassing the performance of the GPT-2 model without data augmentation. This indeed confirms that by increasing the amount of training data, the model demonstrated better generalization, capturing linguistic nuances, negotiation strategies, and contextual variations, as hypothesized.

Finally, we evaluated the performance of our full-fledged multimodal Arabic negotiation bot, which consisted of the *augmented* GPT-2 model, infused with the CNN model described in Section 3.2.2 to extract the color and brand from the images the buyers provided, if any. Recall that our full-fledged approach works as follows. After the buyer's utterance is provided to the negotiation bot, if an image HTTP link is included in the utterance, the link is extracted, the image is downloaded and then fed into the CNN model for color and brand identification. The recognized brand and color information is added into the model context, and subsequently, the response generation task is performed. The test results of our full-fledged multimodal Arabic negotiation bot are shown as the fourth row in Table 1, and we refer to it as (Augmented GPT-2 + CNN). Our multimodal Arabic negotiation bot achieved a cross-entropy loss of 0.55 and a BLEU-4 score of 0.21.

The automatic evaluation results shown in Table 1 indicate that the English negotiation bot outperformed the two Arabic baselines (GPT-2 and Augmented GPT-2) in terms of BLEU-4 score and the cross-entropy loss. This highlights the challenges in generating Arabic text closely aligned with reference sequences, likely due to complex syntax and a larger vocabulary. However, the Augmented Arabic GPT-2 model, exhibited improved performance compared to the baseline Arabic GPT-2 model, emphasizing the positive impact of a more extensive and diverse training dataset. The findings underscore the importance of language-specific characteristics and augmented data in determining the effectiveness of trained negotiation models. Finally, the infusion of the CNN model with the Arabic Augmented GPT-2 model achieved the best performance in terms of both cross-entropy loss and BLEU-4 score in the case of Arabic, highlighting the merit of supporting multimodality when performing Arabic negotiations in the context of sales.

To ensure statistical significance in the improvements in BLEU-4 scores between the full-fledged multimodal Arabic negotiation model and its closest baseline, the Augmented GPT-2 model, a paired bootstrap resampling test was employed [8]. 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 test, 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 full-fledged multimodal Arabic negotiation bot (Augmented GPT-2 + CNN model in Table 1) outperforms the Augmented GPT-2 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 our proposed multimodal Arabic negotiation bot compared to its most superior baseline.

4.2 Human Evaluation

Automatic evaluation metrics alone are insufficient for assessing machine-generated texts. As suggested in [11], 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 our multimodal Arabic negotiation bot.

To manually evaluate our negotiation bot's performance, we selected 100 samples from the test set of the Arabic version of the PPD dataset (comprising 20 full conversations, each with 5 turns). We then provided these examples to the Arabic negotiation bot to generate responses and an Arabic-speaking rater assessed these responses using the following three metrics:

- **Fluency:** How understandable were the generated responses from a language perspective?
- **Consistency:** How consistent were the generated responses with the negotiation context?
- **Persuasion:** How persuasive were the generated responses?

Thus, each response was assigned three ratings, with each rating ranging from 1 to 5, where 1 represents the lowest score and 5 represents the highest score.

The results of the human evaluation for the full-fledged multimodal Arabic negotiation bot (Augmented GPT-2 + CNN model) achieved average ratings of 3.55 for fluency, 3.31 for consistency, and 3.27 for persuasion. The distribution of ratings for the full-fledged multimodal Arabic negotiation bot across the three criteria is shown in Figure 2. The human evaluation results indicate that the negotiation bot was indeed successful in generating responses that are fluent, consistent, and persuasive. On the other hand, the human evaluation for the best Arabic baseline (Augmented GPT-2 model without multimodality) achieved the following average ratings: 3.5 for fluency, 3.06 for consistency, and 3.0 for persuasion. Thus, the human evaluation confirms that supporting multimodality through the infusion of the CNN model into the negotiation bot does indeed result in the generation of better responses as perceived by humans, which coincides with the results obtained by the automatic evaluation.

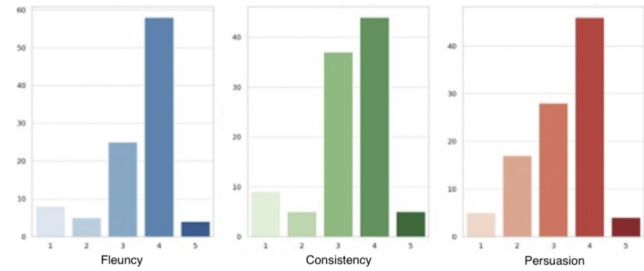

Figure 2: Human Evaluation Rating Distribution (x-axis: rating, y-axis: count)

5 CONCLUSION

In this paper, we laid the foundation for the first multimodal Arabic negotiation bot. Our negotiation bot relies on an infused model consisting of a fine-tuned GPT-2 model, enhanced with reinforcement learning to generate more coherent and persuasive responses, and a convolutional neural network to support multimodality. The GPT-2 model was fine-tuned using a translated version of an established dataset for negotiation simulation in the context of sales. To enhance the performance of the GPT-2 model, the training data was further augmented with automatically generated data using GPT-3.5 Turbo. Our full-fledged multimodal Arabic negotiation bot trained with augmented data showed very promising results compared to various baselines in terms of cross-entropy loss and BLEU-4 score. This was also echoed by human evaluation in terms of fluency, consistency, and persuasion. This work not only marks the development of the first multimodal Arabic negotiation bot, but also lays the foundation for subsequent research into multimodal and multilingual negotiation bots.

In terms of future work, there are several opportunities for improvement and exploration. First, enhancing the performance of our negotiation bot can be achieved by using larger training data, be it automatically generated using LLMs, or by utilizing 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 bot has shown promising results, there is room for improvement by extracting additional features from images. Thus, another future direction involves a thorough exploration of advanced techniques for computer vision, or the integration of multimodal LLMs. These explorations could significantly enhance the extraction of meaningful information from any visual input during the negotiation process.

ACKNOWLEDGMENTS

This work was made possible by NPRP13S-0112-200037 grant from Qatar National Research Fund (a member of Qatar Foundation). The statements made herein are solely the responsibility of the authors.

REFERENCES

[1] Tim Baarslag, Michael Kaisers, Enrico Gerding, Catholijn M Jonker, and Jonathan Gratch. 2017. When will negotiation agents be able to represent us? The challenges and opportunities for autonomous negotiators. International Joint Conferences on Artificial Intelligence.

[2] Satanjeev Banerjee and Alon Lavie. 2005. 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*. 65–72.

[3] Aodah Diamah, Masoud Mohammadian, and Bala Balachandran. 2012. Fuzzy utility and inference system for bilateral negotiation. In *2012 2nd International Conference on Uncertainty Reasoning and Knowledge Engineering*. IEEE, 115–118.

[4] Roger Fisher, William L Ury, and Bruce Patton. 2011. *Getting to yes: Negotiating agreement without giving in*. Penguin.

[5] He He, Derek Chen, Anusha Balakrishnan, and Percy Liang. 2018. Decoupling strategy and generation in negotiation dialogues. *arXiv preprint arXiv:1808.09637* (2018).

[6] Mohamed Jawahar Hussain. 2014. 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*. IEEE, 583–586.

[7] Nicholas R Jennings, Peyman Faratin, Alessio R Lomuscio, Simon Parsons, Carles Sierra, and Michael Wooldridge. 2001. Automated negotiation: prospects, methods and challenges. *International Journal of Group Decision and Negotiation* 10, 2 (2001), 199–215.

[8] Philipp Koehn. 2004. Statistical significance tests for machine translation evaluation. In *Proceedings of the 2004 conference on empirical methods in natural language processing*. 388–395.

[9] Gaurav Koley and Shrisha Rao. 2018. 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)*. IEEE, 1–5.

[10] Mike Lewis, Denis Yarats, Yann N Dauphin, Devi Parikh, and Dhruv Batra. 2017. Deal or no deal? end-to-end learning for negotiation dialogues. *arXiv preprint arXiv:1706.05125* (2017).

[11] Chia-Wei Liu, Ryan Lowe, Iulian V Serban, Michael Noseworthy, Laurent Charlin, and Joelle Pineau. 2016. How not to evaluate your dialogue system: An empirical study of unsupervised evaluation metrics for dialogue response generation. *arXiv preprint arXiv:1603.08023* (2016).

[12] Kshitij Mishra, Azlaan Mustafa Samad, Palak Totala, and Asif Ekbal. 2022. PEPDS: A polite and empathetic persuasive dialogue system for charity donation. In *Proceedings of the 29th International Conference on Computational Linguistics*. 424–440.

[13] Tarek Naous, Christian Hokayem, and Hazem Hajj. 2020. Empathy-driven Arabic conversational chatbot. In *Proceedings of the Fifth Arabic Natural Language Processing Workshop*. 58–68.

[14] Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu. 2002. Bleu: a method for automatic evaluation of machine translation. In *Proceedings of the 40th annual meeting of the Association for Computational Linguistics*. 311–318.

[15] Reza Fuad Rachmadi and I Purnama. 2015. Vehicle color recognition using convolutional neural network. *arXiv preprint arXiv:1510.07391* (2015).

[16] Aritra Raut, Abhisek Tiwari, Subrata Das, Sriparna Saha, Anutosh Maitra, Roshni Ramnani, and Shubhashis Sengupta. 2023. Reinforcing personalized persuasion in task-oriented virtual sales assistant. *Plos one* 18, 1 (2023), e0275750.

[17] Xuewei Wang, Weiyan Shi, Richard Kim, Yoojung Oh, Sijia Yang, Jingwen Zhang, and Zhou Yu. 2019. Persuasion for good: Towards a personalized persuasive dialogue system for social good. *arXiv preprint arXiv:1906.06725* (2019).

[18] Thomas Wolf, Lysandre Debut, Victor Sanh, Julien Chaumond, Clement Delangue, Anthony Moi, Pierric Cistac, Tim Rault, Rémi Louf, Morgan Funtowicz, et al. 2019. Huggingface's transformers: State-of-the-art natural language processing. *arXiv preprint arXiv:1910.03771* (2019).

[19] Haolan Zhan, Yufei Wang, Tao Feng, Yuncheng Hua, Suraj Sharma, Zhuang Li, Lizhen Qu, and Gholamreza Haffari. 2022. Let's Negotiate! A Survey of Negotiation Dialogue Systems. *arXiv preprint arXiv:2212.09072* (2022).

[20] Mingyang Zhang, Pengli Wang, and Xiaoman Zhang. 2019. Vehicle color recognition using deep convolutional neural networks. In *Proceedings of the 2019 International Conference on Artificial Intelligence and Computer Science*. 236–238.

[21] Yiheng Zhou, He He, Alan W Black, and Yulia Tsvetkov. 2019. A dynamic strategy coach for effective negotiation. *arXiv preprint arXiv:1909.13426* (2019).