

# Enhancing LLM-based Arabic Negotiation by Fine Tuning on Dialogue Shortcomings

**Yasmine Abu Adla, Hazem Hajj, Shady Elbassuoni, Wassim El Hajj, Khaled Shaban**

American University of Beirut, Lebanon; Qatar University, Qatar  
{yaa41, hh63, se58, we07}@aub.edu.lb, khaled.shaban@qu.edu.qa

## Abstract

This study advances Arabic dialogue negotiation by enriching the responses of Large Language Models (LLMs) with targeted fine-tuning that addresses key challenges such as arithmetic reasoning, chain-of-thought processing, and pronoun consistency. We validated the prevalence of these issues by testing responses from multiple LLMs including Gemini, LLaMA 2, and BARD. To tackle these shortcomings, we propose the use of fine-tuning to construct a comprehensive dataset that simulates varied negotiation scenarios. We apply our proposed fine-tuning technique to GPT-3.5 Turbo and showcase the LLM’s abilities to exceed current performance benchmarks. Our evaluation framework assesses improvements in dialogue length, strategic depth, and arithmetic accuracy, comparing the performance of fine-tuned models against original metrics and benchmarks from GPT-4. The results demonstrate significant advancements in negotiation performance by the fine-tuned LLM, surpassing the baseline model in engagement and strategic execution.

## 1 Introduction

Negotiations are fundamental to human interaction, occurring in various contexts from casual market transactions to complex diplomatic discussions (Schmid et al., 2021; Kramár et al., 2022). The rise of chatbots that mimic human negotiators has notably influenced negotiation practices. In this evolving landscape, intelligent agents often find themselves needing to cooperate with others who have different goals, primarily using natural language to reach agreements. Negotiation thus emerges as a complex interplay of linguistic skill and strategic reasoning, requiring agents to understand, plan, and strategically generate utterances to achieve objectives (Traum et al., 2008; Plüss and Piwek, 2016)

Despite advances in negotiation bots for English, utilizing game theory, rule-based systems (Jennings

et al., 2001; Hussain, 2014; Balachandran and Mohammadian, 2015; Koley and Rao, 2018), and deep learning (Bachrach et al., 2020), research into non-English languages, especially Arabic, remains limited. Pioneering efforts have largely focused on English, with advances in Seq2Seq Recurrent Neural Network models trained through supervised learning (Lewis et al., 2017). These models, refined with reinforcement learning (RL), set a benchmark for product negotiation bots. Subsequent studies have introduced enhancements in emotional intelligence and persuasion tactics (He et al., 2018; Wang et al., 2019; Zhou et al., 2019; Mishra et al., 2022; Raut et al., 2023).

More recently, the emergence of transformer-based models, notably the Generative Pre-trained Transformer (GPT) series, has revolutionized language processing (Chow et al., 2023). Large language models (LLMs) including Gemini, GPT-3.5 Turbo, LLaMA 2, and BARD have made significant strides in understanding and generating natural language. However, applying these to Arabic negotiation dialogues reveals critical challenges. First, the complexity of chain-of-thought (COT) reasoning in Arabic often leads to models failing to follow or generate coherent arguments. Second, LLMs exhibit a notable lack of precision in numerical reasoning and calculations. Finally, the intricacies of Arabic pronoun usage present a challenge, with models often demonstrating inconsistency in pronoun reference, affecting the coherence and naturalness of dialogues.

We identify specific areas where the aforementioned LLMs, despite their advanced capabilities, fall short in handling the nuanced requirements of Arabic negotiations. By examining the models’ performance across a range of negotiation scenarios, we uncover limitations in their reasoning, mathematical operations, and linguistic consistency. Our research addresses these shortcomings by focusing on GPT-3.5 Turbo as a case study for Arabic nego-

tiation dialogues. The contributions of our study are:

1. Identifying specific limitations of LLMs in Arabic negotiation scenarios.
2. Fine-tuning GPT-3.5-Turbo to enhance COT reasoning, arithmetic capabilities, and pronoun consistency in Arabic negotiation dialogues.
3. Developing a comprehensive Arabic negotiation dataset for model training.
4. Introducing an evaluation framework to measure the effectiveness of fine-tuning interventions on LLMs in negotiation contexts.

The rest of the paper is organized as follows: Section 2 reviews related work in deep learning-based negotiation bots. Section 3 describes our methodology, detailing the identification of LLM limitations and the fine-tuning process for GPT-3.5 Turbo, encompassing dataset creation and training. Section 4 details our experimental setup and results, highlighting improvements in negotiation dialogues. Section 5 discusses the implications and potential applications of our findings. Section 6 outlines limitations and future research directions. Section 7 concludes the paper.

## 2 Related Work

Numerous recent studies have employed deep learning to develop negotiation bots. The groundbreaking research (Lewis et al., 2017) was the first to utilize deep learning to create an end-to-end negotiation model in natural language. This model translates raw data directly into desired outcomes without intermediate steps, mastering both language and strategic reasoning through supervised and reinforcement learning. Addressing the limitations in (Lewis et al., 2017), the study in (He et al., 2018) separated negotiation strategy from language generation. The focus of (Wang et al., 2019) was on persuasive strategies, analyzing a dataset of human interactions annotated with persuasion tactics.

In (Zhou et al., 2019), the introduction of a negotiation coach based on LSTM technology aimed to enhance deal outcomes by providing strategic advice. The work in (Mishra et al., 2022) involved training a language model using RL, considering multiple sub-rewards for persuasion, emotion, politeness, coherence, and repetition.

(Raut et al., 2023) explored the development of a persuasive sales agent using a GPT-2 model augmented with RL and sub-rewards for repetitiveness, consistency, action alignment, and sentiment. This study also integrated meta-learning for adaptability in new negotiation sub-domains. Complementing these developments, recent research (Fu et al., 2023) has investigated human-compatible negotiation strategies. This study introduced an AlphaZero-like RL+search technique integrated with a pre-trained language model, aiming to develop negotiation tactics that are not only effective but also equitable and aligned with human fairness. The results showed notable success in fostering egalitarian outcomes and enhancing the overall quality of negotiations. Another key advancement (Abdelnabi et al., 2023) examines the use of LLMs in interactive, multi-agent negotiation games. This study created a testbed of text-based, multi-issue negotiation games to assess the arithmetic, exploration, and planning capabilities of LLMs, with a particular focus on their COT reasoning. The research highlighted a substantial performance gap between GPT-4 and earlier models and emphasized the utility of these games in evaluating critical aspects like interaction dynamics in the presence of adversarial or greedy players.

Upon review, it is evident that the current literature lacks investigations specifically targeting the Arabic language. This gap is particularly notable given the unique morphosyntactic characteristics and substantial global presence of Arabic. The scarcity of research may be attributed to the linguistic complexities of Arabic, including its rich morphological structure and dialectal variation, as well as the lack of accessible, high-quality datasets. Addressing this research gap is crucial for developing more comprehensive and linguistically inclusive computational models. Our work addresses the gap in negotiation bot development for the Arabic language, comparing various methodologies and outcomes with prior research. Our approach introduces unique features and strategies aimed at enhancing negotiation outcomes, demonstrating notable advancements over existing models.

## 3 Methodology

The challenge of effectively managing negotiation dialogues, particularly in languages as complex as Arabic, represents a common limitation across a broad spectrum of LLMs. This research was

inspired by the need to enhance the capabilities of LLMs, enabling them to navigate the intricacies of Arabic negotiations more adeptly.

Acknowledging the widespread nature of these limitations, our initial step involved conducting an extensive testing phase across several LLMs, including Gemini, GPT-3.5 Turbo, LLaMA 2, and BARD. This approach allowed us to uncover a range of common issues, from errors in mathematical reasoning and common-sense application to inconsistencies in pronoun usage. These issues, critical in the context of Arabic negotiation dialogues, underscore the essential need for focused enhancements across LLMs.

The methodology adopted for this research is structured to be universally applicable to LLMs, comprising two primary phases: testing and fine-tuning. By rigorously evaluating each model's performance in diverse Arabic negotiation scenarios, we were able to identify specific shortcomings that require attention. Detailed insights into the fine tuning and testing methodology are provided in the Appendix.

As part of our research, GPT-3.5 Turbo was selected for a detailed case study, given its prominence and potential for advanced language processing. This endeavor was motivated by the recognition of the model's existing limitations in handling the Arabic language, especially within the context of negotiation. This model, alongside others like Gemini, LLaMA 2, and BARD, served as practical examples to apply and refine our enhancement strategies. The findings from these case studies were pivotal in illustrating the effectiveness of our approach in real-world settings.

The fine-tuning phase was meticulously designed to address the identified challenges, extending beyond linguistic adjustments to encompass enhancements in reasoning and negotiation abilities. This comprehensive strategy aims to significantly improve the models' functionality in Arabic negotiation contexts, tackling the nuances and complexities that are characteristic of real-life interactions.

This study not only highlights a prevalent issue across LLMs but also proposes a scalable solution pathway, as evidenced through the application to models like GPT-3.5 Turbo, Gemini, LLaMA 2, and BARD. The enhancements developed through this research endeavor are expected to markedly boost the performance of these models, contributing to the advancement of AI applications in Arabic

language tasks. By showcasing the adaptability and effectiveness of our methodologies, this work encourages further exploration into improving LLMs for greater linguistic diversity and cultural sensitivity.

### **3.1 Experimental Evaluation of GPT-3.5 Turbo in Arabic Negotiation Scenarios**

To rigorously evaluate GPT-3.5 Turbo's proficiency in Arabic negotiation, we employed a structured simulation scenario. The prompt given to the model initiated a role-play negotiation game involving two characters, Amani and Samir, each with different valuations for a set of items: a book, three hats, and a ball. The challenge for the model was to simulate both characters' negotiations to maximize the value for each based on their individual item valuations.

The model was instructed to autonomously play out the negotiation between Amani and Samir. Each character's valuation for the items was distinctly set: Amani valued the book, hats, and ball differently from Samir. The negotiation involved a series of exchanges where each character alternated in making changes to their inventory based on the agreed-upon item distribution. The goal was to reach an agreement that maximizes value for both parties, with a final 'agreement' statement concluding the negotiation and a scoring system evaluating the outcome.

We conducted ten separate runs of this negotiation scenario. In each run, the model was tasked to simulate the dialogue, decision-making, and inventory adjustments for both characters.

Despite the model's advancements, it struggled in key areas during the negotiation simulations. The assessment focused on the model's capacity for logical mathematical operations, common-sense reasoning, and maintaining pronoun consistency in the context of Arabic negotiation, as shown in the Appendix. The results revealed that these targeted areas still presented significant challenges, underlining the need for further refinement in the model's handling of complex negotiation scenarios.

This experimental setup allowed for a comprehensive evaluation of GPT-3.5 Turbo's negotiation capabilities in a complex, multi-faceted Arabic negotiation scenario. It provided insights into the model's linguistic agility, reasoning proficiency, and decision-making abilities in a nuanced setting.

### 3.2 Dataset Creation

For the creation of a comprehensive dataset, we utilized GPT-4 to synthesize a diverse array of negotiation dialogues, which included a total of 1,400 negotiation scenarios. Although this dataset is smaller than some English-based negotiation datasets, our dataset was carefully designed to capture the linguistic and cultural subtleties of Arabic negotiations. These subtleties include indirect speech, rich morphological structures, and complex pronoun usage, which make Arabic negotiations inherently more intricate.

In generating the dataset, GPT-4 was prompted with carefully crafted instructions tailored to each dialogue category. For regular negotiation dialogues, prompts were designed to elicit typical buyer-seller interactions. In arithmetic-based scenarios, prompts included specific numerical constraints and calculations. Chain of Thought dialogues were prompted to encourage sequential reasoning, while pronoun consistency challenges included scenarios with varying gender contexts. The combined challenges were prompted with complex situations requiring the application of multiple skills simultaneously. These precise prompts ensured that the generated dialogues effectively targeted the desired negotiation aspects, thereby creating a rich and varied dataset for training and testing the model.

The dataset was stratified into distinct categories to systematically address various aspects of negotiation skills:

1. **Regular Negotiation Dialogues (200 Examples):** These are standard negotiation scenarios to establish a baseline of the model's performance.
2. **Arithmetic-Based Negotiation (200 Examples):** Scenarios that require the model to perform precise mathematical operations within the negotiation context.
3. **Chain of Thought (COT) Dialogues (200 Examples):** These dialogues are designed to assess the model's ability to follow and generate a sequence of reasoning steps.
4. **Pronoun Consistency Challenges (200 Examples):** Dialogues that specifically test the model's use of gender pronouns accurately and consistently.

5. **Combined Challenges:** We also included dialogues that combine these elements:

- (a) Arithmetic + COT (200 Examples)
- (b) Arithmetic + Pronoun (200 Examples)
- (c) COT + Pronoun (200 Examples)

From this synthesized dataset, we allocated 1,000 examples for training and reserved 400 for testing, ensuring a broad coverage of scenarios. The testing examples were randomly and equally sampled from the dataset to provide a balanced representation of each challenge during model evaluation.

This meticulous approach to dataset creation was fundamental in developing a nuanced understanding of the model's capabilities and guiding the fine-tuning process to specifically enhance performance in the identified areas of difficulty.

## 4 Fine-Tuning GPT-3.5 Turbo

The initial dataset, comprised of dialogues synthesized by GPT-4, underwent conversion into a format that mirrors chat-based negotiation conversations. This conversion was pivotal to align the dataset with the inherent conversational dynamics of negotiations, thereby facilitating a training regime that closely reflects real-world interaction patterns for the model.

Subsequently, the base model of GPT-3.5 Turbo was subjected to supervised fine-tuning on a specialized corpus encompassing 1.9 million tokens, spanning three epochs. This extensive training was vital for the model to acclimate to the complexities embedded within negotiation dialogues. A fine-tuning approach was adopted, employing a learning rate multiplier set at 2.0 and a dynamic batch size, which were integral to the adaptive optimization strategy aimed at enhancing the model's learning efficacy from a diverse array of dialogues. Upon completion of training, the model attained an average accuracy of 0.89 coupled with a training loss averaging 0.395. These metrics are indicative of the model's substantial acquisition and retention of the structural intricacies of negotiation dialogues.

## 5 Experiments and Evaluation

### 5.1 Evaluation of the Fine Tuned Model

The evaluation of the fine-tuned GPT-3.5 Turbo model entailed utilizing prompts from a curated testing dataset to elicit negotiation dialogues, which

were then generated by the model under a 3,000-token limit. These outputs were compared to the testing set’s dialogues based on predefined metrics — providing a multifaceted view of the model’s performance in simulated Arabic negotiation contexts.

For evaluation metrics that cannot be automatically quantified, we employed GPT-4 to assist in the assessment process. This approach allowed us to interpret and analyze components such as negotiation strategies and language nuances that are not readily evaluated through algorithmic means. This methodological choice adds a layer of qualitative analysis to our quantitative framework, enriching the evaluation of the model’s performance in conducting Arabic negotiations.

To comprehensively assess the fine-tuned GPT-3.5 Turbo model’s performance in Arabic negotiation dialogues, we established a multi-dimensional evaluation framework. This framework encompasses metrics across negotiation success, language proficiency, and arithmetic reasoning:

### 1. Negotiation Performance Metrics:

- (a) **Negotiation Success Rate:** Measures the frequency at which the model reaches a successful agreement within the simulated negotiation scenarios.
- (b) **Price Agreement Analysis:** Evaluates the prices from the model’s generated dialogues against those in the testing set. This analysis was crucial to determine if the model had effectively applied negotiation strategies as a buyer, seller, or both. The aim was to see whether the model’s generated prices showed an improvement or remained unchanged, indicating the effectiveness of the fine-tuning process in enhancing the model’s negotiation strategy implementation.
- (c) **Discrepancy in Negotiated Prices:** Assess the differences between the prices negotiated in the expected and generated dialogues. This metric scrutinizes the degree to which the fine-tuned model adjusts prices from initial offers to final agreements, in comparison with the outcomes anticipated by the testing set. It serves as an indicator of the model’s competency in executing financially sound negotiation strategies and its sensitivity to the contextual value assigned to items

within the negotiation scenarios.

- (d) **Average Response Length:** Assesses the verbosity of the model’s responses, indicating its ability to communicate efficiently and stay on point during negotiations.

### 2. Language Proficiency Metrics:

**Pronoun Consistency:** This metric evaluates the model’s use of gender pronouns in Arabic, ensuring they align accurately with the context and participants’ genders.

### 3. Arithmetic Reasoning Metrics:

**Arithmetic Capabilities:** Focuses on the model’s ability to perform and apply arithmetic operations correctly within the context of negotiation dialogues, an essential component for quantitative reasoning in negotiations.

## 5.2 Evaluation of Negotiation Metrics

### 5.2.1 Response Length and Efficiency

Our evaluation employed boxplot visualizations to compare dialogue lengths between the expected outcomes in the testing dataset and the generated dialogues by the fine-tuned GPT-3.5 Turbo model. The boxplots in Figure 1 depict a noticeable increase in the length of generated dialogues, where the model consistently produced a greater average word count compared to the testing dataset.

This increment suggests that the model has achieved a higher level of dialogue complexity, engaging in extended exchanges indicative of a more sophisticated negotiation strategy. The statistical spread in the boxplots, reflected by a wider interquartile range, points to the model’s diverse negotiation approaches. This diversity, along with significant outliers, hints at the model’s adaptive and multifaceted nature, akin to human negotiation tactics. The median of the generated dialogues, exceeding that of the expected responses, further reinforces the model’s inclination towards more complex negotiation constructs.

### 5.2.2 Negotiation Success Rate and Level of Agreement

To systematically evaluate the success rate and level of agreement in negotiations, our methodology leveraged the advanced natural language understanding capabilities of GPT-4. Each dialogue was parsed by the model, which was prompted to consider the dialogue’s content holistically to assess

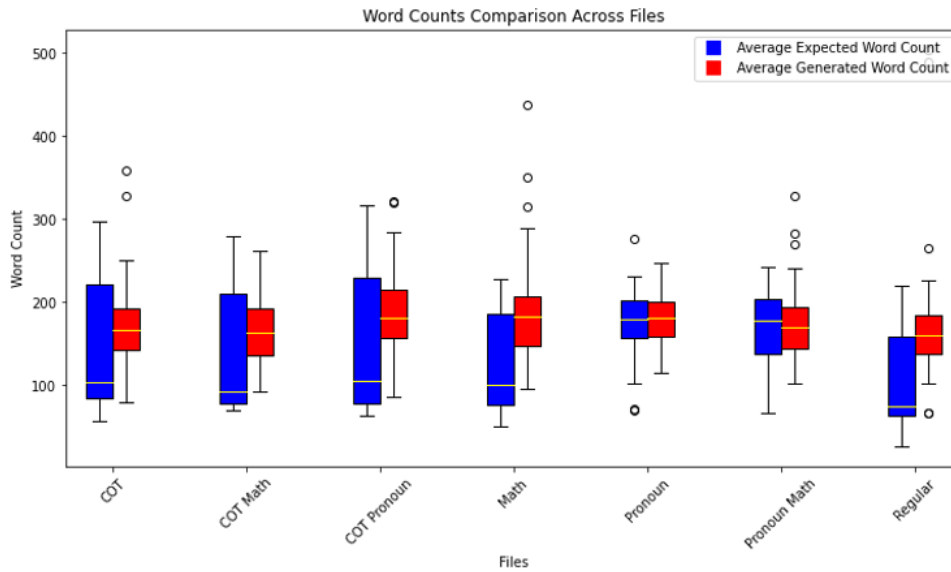


Figure 1: Boxplot Comparison of Dialogue Length in Negotiation Scenarios

negotiation success and agreement level. Success was determined by the presence of explicit affirmations or mutual concessions within the dialogue, while the level of agreement was appraised by examining the congruity between proposed terms and final outcomes. This evaluation was operationalized to automated the interaction with GPT-4, sending dialogues and parsing responses to quantify the negotiation outcomes. By automating this process, we ensured consistency and scalability in our evaluations, allowing for a robust analysis of the fine-tuned model’s performance across a large dataset.

In evaluating the fine-tuned GPT-3.5 Turbo model’s proficiency in Arabic negotiation dialogues, our analysis revealed significant findings. The negotiation success rate, as depicted in the Figure 2, showed that the fine-tuned model slightly outperformed the expected outcomes, indicating an improvement in successfully concluding negotiations. These results were statistically corroborated by McNemar’s test, which yielded a p-value of 0.0233, confirming that the differences observed were significant.

Conversely, the level of agreement between the generated and expected dialogues remained almost consistent (as shown in Figure 3), with McNemar’s test revealing a p-value of 1.0. This indicates that while the fine-tuning process enhanced the model’s ability to successfully negotiate, it maintained the quality of agreement as per the expected standards. These results collectively suggest that the fine-tuning not only fortified the model’s negoti-

ation capabilities but did so without compromising the depth and mutual satisfaction of the agreements reached.

### 5.2.3 Expected and Generated Negotiation Prices

In the analysis of negotiation outcomes, our methodology incorporated a careful examination of the price scales to ensure comparability between the generated and expected dialogues. This step was crucial in cases where bulk pricing could lead to scale discrepancies. Following this, GPT-4 assisted in extracting the final prices for a direct comparison. Manual review was conducted to verify the consistency of pricing scales and to identify the use of buyer or seller negotiation tactics within each dialogue. This meticulous approach allowed us to ascertain whether the fine-tuned model successfully applied negotiation strategies, and whether it achieved price outcomes that were more favorable compared to the expected dialogues, indicating an advanced strategic capability in negotiations.

Figure 4 showcases the frequency distribution of price prediction accuracy compared to the negotiation tactics used, categorized into instances where the generated price was higher, lower, or the same as the expected price. Notably, when seller tactics were employed, the generated price was more frequently higher than expected, which may indicate a model bias towards seller optimization strategies. In contrast, the use of buyer tactics correlates with a lower generated price, suggesting the model’s effectiveness in emulating buyer

### Negotiation Success Analysis

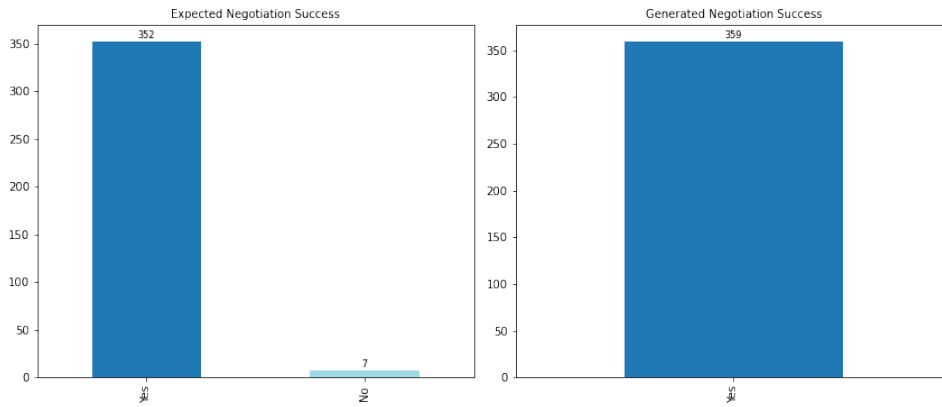


Figure 2: Bar Chart of Negotiation Success Rate

### Level of Agreement Analysis

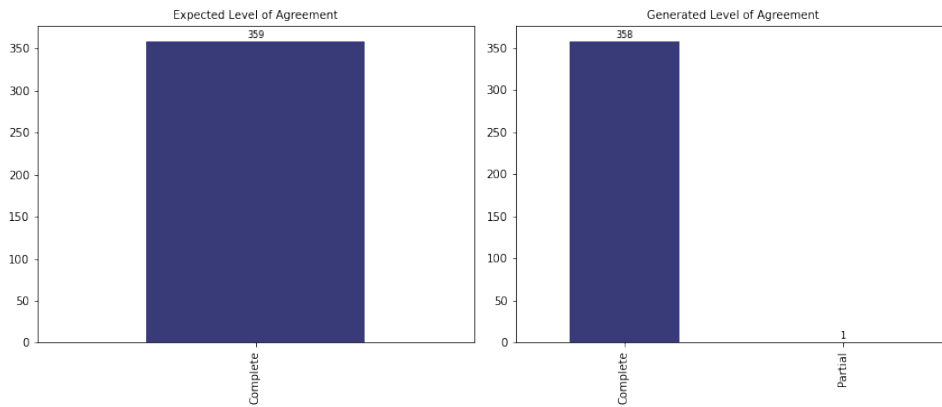


Figure 3: Bar Chart of Level of Agreement

negotiation behavior. When both tactics were considered, the generated prices tended to be the same as expected, illustrating the model’s balanced approach in scenarios where a blend of tactics was used. These outcomes reflect the model’s nuanced understanding of different negotiation roles and its ability to adjust prices accordingly

### 5.3 Evaluation of Pronoun Consistency Findings

To evaluate the model’s proficiency in Arabic pronoun consistency, we utilized GPT-4’s sophisticated language processing capabilities. GPT-4 was tasked with analyzing the dialogues generated by the fine-tuned GPT-3.5 Turbo model to ensure accurate use of gender pronouns.

As seen in Figure 5, the unanimous ‘Yes’ responses in all three bar charts dedicated to pronoun consistency — whether in dialogues involving Chain of Thought (COT), arithmetic, or stan-

dard exchanges — highlight the model’s exceptional adherence to correct pronoun use. This consistent accuracy across various dialogue scenarios is a testament to the success of the fine-tuning process, showcasing the model’s sophisticated grasp of the intricate pronoun rules in the Arabic language.

### 5.4 Evaluation of Arithmetic Capabilities

Figure 6 illustrate the outcomes of the model’s arithmetic correctness across different dialogue scenarios, assessed with the help of GPT-4. In the dialogues involving Chain of Thought (COT) and arithmetic, a vast majority of the model’s responses were mathematically correct, indicating a strong grasp of numerical concepts within the negotiation context. The near-perfect performance in Pronoun Math Dialogues suggests that the model’s fine-tuning process has effectively enhanced its computational accuracy. These results affirm the model’s improved arithmetic capabilities, which

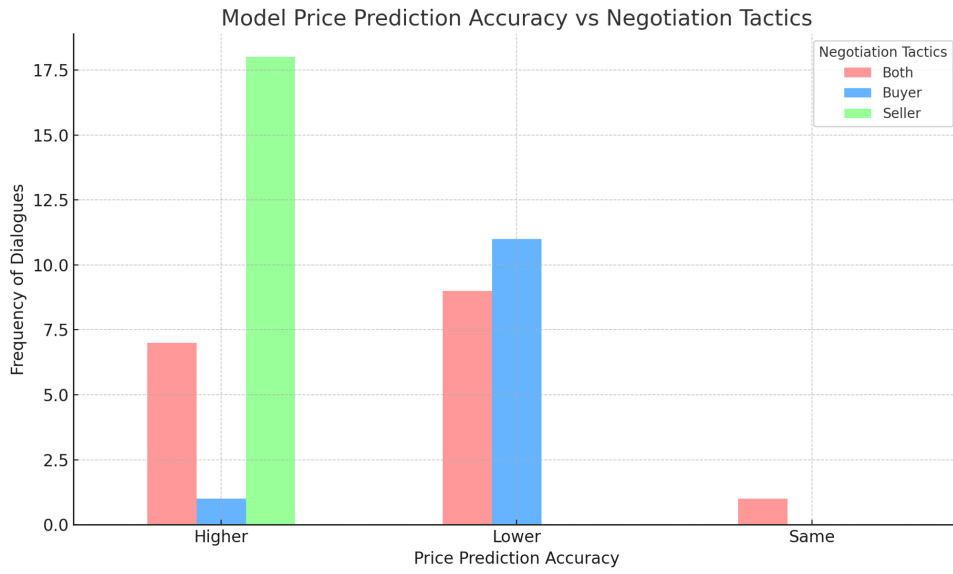


Figure 4: Model Price Prediction Accuracy vs Negotiation Tactics

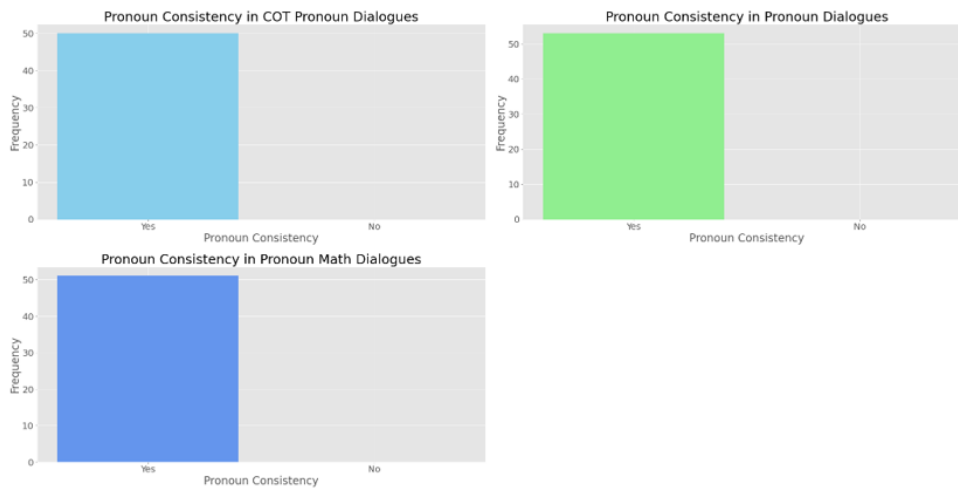


Figure 5: Comprehensive Pronoun Consistency Evaluation

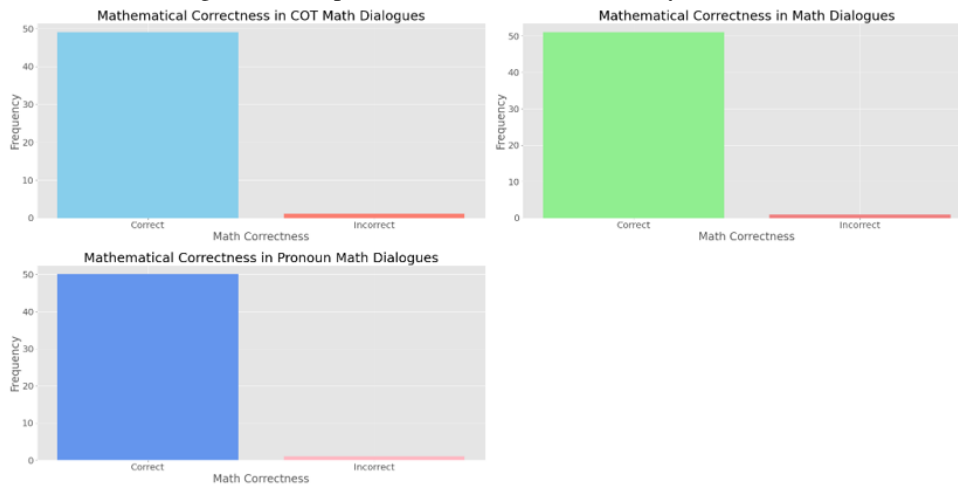


Figure 6: Evaluation of Arithmetic Correctness Across Dialogue Types

are critical for real-world negotiation tasks that often involve complex numerical reasoning



## 5.5 Comparative Analysis of Original vs. Fine-Tuned Model Outputs

Our comparative analysis between the base and fine-tuned versions of GPT-3.5 offers significant insights into the role of contextual accuracy and prompt adherence in AI-generated dialogues. The base model often strayed from the prescribed prompts, diverging into unrelated topics or inaccurately modifying the context of the item prices. Such deviations emphasize the limitations of the base model, making quantitative comparisons less effective. The fine-tuning process, however, has notably improved the model's performance, as seen in the enhanced precision in following prompts and maintaining the continuity and relevance of dialogues within the set negotiation scenarios.

The stark contrast observed in the outputs of the two models underscores the effectiveness of fine-tuning in elevating the model's understanding and adherence to complex prompts. This is vividly illustrated in the Appendix, where the fine-tuned model adeptly stays on topic, negotiating within the bounds of a specific scenario, while the base model shifts context erratically. Such qualitative differences highlight the refined capability of the fine-tuned model to produce dialogues that are not only contextually relevant but also coherent and closely aligned with the nuances of the given scenarios. The improvement in contextual understanding and dialogue relevance is a testament to the fine-tuning's success in enhancing the model's practical applicability in real-world settings, where adherence to contextual nuances is paramount.

In addition to the differences in context adherence, the base and fine-tuned models also diverged in their handling of currency, as shown in the Appendix. In one scenario, while the fine-tuned model maintained the specified Saudi Riyal currency in a negotiation dialogue about a custom-tailored suit, the base model unexpectedly shifted the conversation to a different currency, discussing prices in dollars. This shift not only deviated from the prompt but also demonstrated a lack of contextual awareness. Such an example underlines the enhanced precision of the fine-tuned model in adhering to specific details of the scenario, including currency, which is crucial for realistic and coherent negotiation dialogues.

## 6 Conclusions

In conclusion, this study contributes to the broader field of Arabic NLP and AI negotiation, demonstrating the effectiveness of fine-tuning techniques in overcoming language-specific challenges in LLMs. By improving Arabic dialogue negotiation, we not only enhance AI's utility in Arabic-speaking contexts but also pave the way for future advancements in LLM applications across other underrepresented languages. Our findings underscore the significant potential of targeted model refinement for developing more sophisticated and culturally aware AI negotiation tools.

Our research has successfully demonstrated the potential of fine-tuning GPT-3.5 for advancing Arabic dialogue negotiation, achieving significant improvements that surpass the capabilities of even GPT-4 in specific aspects of negotiation dialogues. Through meticulous methodology and rigorous evaluation, we have not only addressed the initial limitations of GPT-3.5 in arithmetic operations, common-sense reasoning, and pronoun consistency but have also enhanced the model's overall negotiation efficacy. The fine-tuned GPT-3.5 model now facilitates longer negotiation dialogues, indicative of its ability to engage in more complex and detailed discussions.

Moreover, our findings reveal that the model has mastered improved negotiation tactics, enabling it to strategically adjust offers to secure lower or higher prices when bargaining. This advancement represents a significant leap in the model's strategic reasoning and its understanding of negotiation dynamics, mirroring the nuanced tactics employed by skilled human negotiators.

Real-world applications of this fine-tuning approach are broad and impactful. In the context of e-commerce, fine-tuned models can be deployed to automate negotiations over product pricing, discounts, or custom service offerings, adapting dynamically to buyer preferences. In the legal and diplomatic sectors, these models have the potential to assist in drafting agreements or mediating negotiations, helping parties with conflicting interests reach mutually satisfactory resolutions. Additionally, the integration of these models into customer service platforms can enable automated yet nuanced interactions, such as resolving billing disputes or negotiating service plans. The ability of these models to understand and respect linguistic and cultural nuances positions them as espe-

cially valuable in cross-cultural business negotiations, where sensitivity to communication styles is crucial for success.

The contributions of this study extend beyond the immediate enhancements to the GPT-3.5 model, laying a foundation for future research in AI-driven negotiation in underrepresented languages. By developing a comprehensive Arabic negotiation dataset and demonstrating the effectiveness of targeted fine-tuning techniques, we have opened new avenues for exploring linguistic and cultural intricacies in AI negotiations. We are also in the process of preparing this dataset for public release, ensuring that it will be accessible to other research teams for reproducibility and further improvements. As we look to the future, we are encouraged by the potential for further advancements in this field, including the adaptation of these techniques to other languages and contexts, the integration of multi-modal negotiation capabilities, and the exploration of ethical considerations in automated negotiations.

This research underscores the importance of continued innovation in language model development, especially in enhancing non-English language capabilities. As we push the boundaries of what AI can achieve in understanding and facilitating human-like negotiations, we are not only expanding the technical capabilities of these models but also enriching the ways in which they can serve diverse global communities. The journey ahead is promising, with the prospect of creating AI systems that can navigate the complexities of human language and culture with unprecedented skill and sensitivity.

## 7 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

- Sahar Abdelnabi, Amr Goma, Sarath Sivaprasad, Lea Schönherr, and Mario Fritz. 2023. Llm-deliberation: Evaluating llms with interactive multi-agent negotiation games. *arXiv preprint arXiv:2309.17234*.
- Yoram Bachrach, Richard Everett, Edward Hughes, Angeliki Lazaridou, Joel Z. Leibo, Marc Lanctot, Michael Johanson, Wojciech M. Czarnecki, and Thore Graepel. 2020. [Negotiating team formation using deep reinforcement learning](#). *Artificial Intelligence*, 288:103356.
- Bala M Balachandran and Masoud Mohammadian. 2015. Development of a fuzzy-based multi-agent system for e-commerce settings. *Procedia Computer Science*, 60:593–602.
- James CL Chow, Leslie Sanders, and Kay Li. 2023. Impact of chatgpt on medical chatbots as a disruptive technology. *Frontiers in Artificial Intelligence*, 6:1166014.
- Yao Fu, Hao Peng, Tushar Khot, and Mirella Lapata. 2023. [Improving language model negotiation with self-play and in-context learning from ai feedback](#).
- He He, Derek Chen, Anusha Balakrishnan, and Percy Liang. 2018. Decoupling strategy and generation in negotiation dialogues. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 2333–2343.
- 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*, pages 583–586. IEEE.
- 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):199–215.
- Gaurav Koley and Shrisha Rao. 2018. Adaptive human-agent multi-issue bilateral negotiation using the thomas-kilman conflict mode instrument. In *2018 IEEE/ACM 22nd International Symposium on Distributed Simulation and Real Time Applications (DS-RT)*, pages 1–5. IEEE.
- János Kramár, Tom Eccles, Ian Gemp, Andrea Tacchetti, Kevin McKee, Mateusz Malinowski, Thore Graepel, and Yoram Bachrach. 2022. [Negotiation and honesty in artificial intelligence methods for the board game of diplomacy](#). *Nature Communications*, 13.
- Mike Lewis, Denis Yarats, Yann Dauphin, Devi Parikh, and Dhruv Batra. 2017. [Deal or no deal? end-to-end learning of negotiation dialogues](#). In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 2443–2453, Copenhagen, Denmark. Association for Computational Linguistics.
- 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*, pages 424–440.

- Brian Plüss and Paul Piwek. 2016. [Measuring non-cooperation in dialogue](#). In *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers*, pages 1925–1936, Osaka, Japan. The COLING 2016 Organizing Committee.
- 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):e0275750.
- Andreas Schmid, Oliver Kronberger, Niklas Vonderach, and Mareike Schoop. 2021. Are you for real? a negotiation bot for electronic negotiations.
- David Traum, Stacy Marsella, Jonathan Gratch, Jina Lee, and Arno Hartholt. 2008. [Multi-party, multi-issue, multi-strategy negotiation for multi-modal virtual agents](#). pages 117–130.
- Xuwei Wang, Weiyang 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](#). In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 5635–5649, Florence, Italy. Association for Computational Linguistics.
- Yiheng Zhou, He He, Alan W Black, and Yulia Tsvetkov. 2019. [A dynamic strategy coach for effective negotiation](#). In *Proceedings of the 20th Annual SIGdial Meeting on Discourse and Dialogue*, pages 367–378, Stockholm, Sweden. Association for Computational Linguistics.

## 8 Appendix

### Arabic Prompt:

أريد منك الدخول في وضع اللعب الذاتي ولعب هذه اللعبة التفاوض في هذه اللعبة التفاوض، يتم تقديم مبدئيًا لمشاركين، أماني وسمير، مع العناصر التالية: كتاب واحد، 3 قبعات، وكرة واحدة. ومع ذلك، تختلف القيم لهذه العناصر بالنسبة لكل مشارك:

- بالنسبة لأماني، القيم هي 1 للكتاب، و3 لكل من 3 قبعات، و0 للكرة.

- بالنسبة لسمير، القيم هي 2 للكتاب، و2 لكل من 2 قبعات، و2 للكرة.

خلال عملية التفاوض، يتعاون أماني وسمير لاتخاذ قرار بشأن كيفية تقسيم هذه العناصر بينهما بطريقة تزيد من القيمة بالنسبة لكل وكيل. يشاركون في سلسلة من التبادلات وفقًا لهذه القواعد:

- يبدأ كل مشارك بالمخزون الأولي للعناصر، بما في ذلك العدد والقيم كما هو موضح أعلاه.

- يتناوبون في إجراء تغييرات على مخزونهم عن طريق تحديد عدد العناصر الجديد لكل عنصر.

- يستمر التفاوض حتى يتفق كلا المشاركين على توزيع نهائي للعناصر.

يتبع التبادل النموذجي هذا النمط:

- يمكن لأماني وسمير تبادل بعض الرسائل.

- تشير أماني إلى اختيارها للتفاوض.

- يقوم كلا المشاركين بتعديل مخزونهم استنادًا إلى التغييرات المتفق عليها.

- ينتهي التفاوض ببيان 'الاتفاق'، ويتم استخدام نظام التسجيل لتحديد النتيجة. يحصل كل مشارك على نقاط استنادًا إلى التوزيع النهائي للعناصر.

مهمتك هي محاكاة كلاً من أماني وسمير خلال هذا التفاوض، واتخاذ القرارات، وتوجيه المحادثة. تذكر الهدف من زيادة القيمة بالنسبة لكل وكيل أثناء مشاركتك في التفاوض.

### English Prompt:

I want you to enter self-play mode and play this negotiation game: In this negotiation game, initially, participants Amani and Samir are presented with the following items: one book, three hats, and one ball. However, the values of these items differ for each participant:

- For Amani, the values are 1 for the book, 3 for each of the three hats, and 0 for the ball.
- For Samir, the values are 2 for the book, 2 for each of two hats, and 2 for the ball.

During the negotiation process, Amani and Samir cooperate to decide how to divide these items between them in a way that maximizes the value for each agent. They engage in a series of exchanges according to these rules:

- Each participant starts with an initial inventory of items, including the number and values as described above.
- They take turns making changes to their inventory by specifying new quantities for each item.
- The negotiation continues until both participants agree on a final distribution of the items.

The typical exchange follows this pattern:

- Amani and Samir can exchange some messages.
- Amani indicates her choice for negotiation.
- Both participants adjust their inventory based on the agreed changes.

The negotiation ends with an 'agreement' statement, and a scoring system is used to determine the outcome. Each participant scores points based on the final distribution of items.

## Bilingual Presentation of the Negotiation Game Prompt for GPT-3.5 Turbo Testing

### Arabic Prompt:

**أماني:** مرحبًا سمير، لنبدأ بالتفاوض. أعتقد أنني يمكنني البدء. أرغب في الحصول على الكتاب، وأعرض لك قبعتين. ما رأيك؟

**سمير:** مرحبًا أماني، أنا موافق على هذا الاقتراح. أقترح أن نمضي في الاتفاق. سأأخذ الكتاب وأمنحك قبعتين.

**أماني:** ممتاز، أنا موافقة أيضًا. نعلن أننا اتفقنا. الآن يجب علينا تحديث مخزوننا.

يتم تحديث مخزون العناصر: أماني - الكتاب (0)، القبعات (1)، الكرة (0) / سمير - الكتاب (1)، القبعات (2)، الكرة [(1)]

**سمير:** الآن، لدينا قبعة واحدة وكرة واحدة تبقين. ماذا تعتقد، أماني؟

**أماني:** لدي اقتراح، سمير. أعطني القبعة الواحدة والكرة، وسأمنحك الكتاب.

**سمير:** هذا اقتراح جيد. سأوافق على ذلك. نعلن أننا اتفقنا مرة أخرى.

يتم تحديث مخزون العناصر: أماني - الكتاب (1)، القبعات (1)، الكرة (1) / سمير - الكتاب (0)، القبعات (0)، الكرة [(0)]

**أماني:** يبدو أننا انتهينا، سمير. شكرًا على التفاوض. هل لديك أي تعليقات أخيرة؟

**سمير:** لا، ليس لدي تعليقات. استمتعت بالتفاوض معك. نتمنى أن تكون الصفقة مفيدة للجميع.

### English Prompt:

**Amani:** Hello Samir let's start negotiating. I think I can start. I want the book, and I offer you two hats. What do you think?

**Samir:** Hello Amani, I agree with this proposal. I suggest we proceed with the agreement. I'll take the book and give you two hats.

**Amani:** Excellent, I agree as well. We declare that we have reached an agreement. Now we need to update our inventory.

Inventory update: Amani - Book (0), Hats (1), Ball (0) / Samir - Book (1), Hats (2), Ball [(1)]

**Samir:** Now, we have one hat and one ball remaining. What do you think, Amani?

**Amani:** I have a suggestion, Samir. Give me the one hat and the ball, and I'll give you the book.

**Samir:** That's a good proposal. I agree to that. We declare that we have agreed once again.

Inventory update: Amani - Book (1), Hats (1), Ball (1) / Samir - Book (0), Hats (0), Ball [(0)]

**Amani:** It seems we are done, Samir. Thank you for negotiating. Do you have any final comments?

**Samir:** No, I have no comments. I enjoyed negotiating with you. We hope the deal is beneficial for

### Arabic Prompt:

**البايع:** مرحبًا، كيف يمكنني مساعدتك اليوم؟  
**المشتري:** أرغب في شراء سيارة جديدة، لكن لدي بعض الأسئلة.  
**البايع:** بالطبع، أنا هنا لأجيب على جميع استفساراتها.  
**المشتري:** هل السيارة متوفرة بألوان أخرى؟  
**البايع:** نعم، هي متوفرة باللون الأزرق والأحمر والأسود.  
**المشتري:** وما هي خيارات التمويل المتاحة لي؟  
**البايع:** لدينا خطط تمويل متعددة، يمكن له أن يختار ما يناسبه منها.  
**المشتري:** أعتقد أنني مهتم باللون الأزرق. هل يمكنني تجربة القيادة؟  
**البايع:** بالتأكيد، سأجهز لك السيارة للتجربة. هل يمكن أن أعرف اسمها لأدونه في السجل؟  
**المشتري:** اسمي أحمد. وشكرًا لك على المساعدة.  
**البايع:** لا شكر على واجب، سأخبرهم أن أحمد في انتظار تجربة القيادة.

---

### English Prompt:

**Seller:** Hello, how can I assist you today?

**Buyer:** I am interested in buying a new car, but I have some questions.

**Seller:** Of course, I am here to answer all [her] inquiries. (Inconsistency: should be "your" inquiries to match the buyer's gender and the direct conversation context.)

**Buyer:** Is the car available in other colors?

**Seller:** Yes, it is available in blue, red, and black.

**Buyer:** And what are the financing options available to me?

**Seller:** We have multiple financing plans, [he] can choose what suits [him] from them. (Inconsistency: should be "you" to match the direct conversation with the buyer.)

**Buyer:** I think I am interested in the blue one. Can I test drive it?

**Seller:** Certainly, I will prepare the car for the test drive. May I know [his] name to note it in the record? (Inconsistency: should be "your" name to directly address the buyer.)

**Buyer:** My name is Ahmed. And thank you for your help.

**Seller:** No need to thank, I will inform them that Ahmed is waiting for the test drive. (The pronoun

**Arabic Prompt:**

**المشتري:** أريد أن أشتري هذه المجموعة من الأجهزة الإلكترونية، لكنني أحتاج فقط إلى نصف عدد الأجهزة. هل يمكننا تقسيم المجموعة؟

**البائع:** بالطبع، يمكننا ذلك. لدينا عشرة أجهزة لابتوب بإجمالي قيمة 5000 دولار، يمكننا تقسيمها إلى مجموعتين بقيمة 2500 دولار لكل مجموعة.

**المشتري:** جيد، ولكن أريد أيضًا تقسيم الطابعة ثلاثية الأبعاد التي تبلغ قيمتها 2000 دولار إلى نصفين، حتى أدفع 1000 دولار فقط.

**البائع:** حسنًا، سأقوم بتقسيم الطابعة إلى نصفين وأبيعك نصفها بألف دولار.

**المشتري:** ممتاز، هذا يناسبني.

**English Prompt:**

**Buyer:** I want to buy this set of electronic devices, but I only need half the number of devices. Can we split the set?

**Seller:** Of course, we can do that. We have ten laptops with a total value of 5000 dollars, we can split them into two groups, each valued at 2500 dollars.

**Buyer:** Good, but I also want to split the 3D printer, which is worth 2000 dollars, in half, so I only pay 1000 dollars.

**Seller:** Okay, I will split the printer in half and sell you one half for a thousand dollars.

Illustration of Chain of Thought (COT) Limitations in GPT-3.5 along with its English Translation

### Arabic Dialogue:

Prompt: مشتري يريد التفاوض على سعر سيارة 50,000 ريال

**Buyer:** السلام عليكم، أنا مهتم في السيارة التي نشرتها للبيع على الإنترنت. هل لا زالت متوفرة؟

**Seller:** وعليكم السلام، نعم السيارة لا زالت متوفرة. هل ترغب في رؤيتها؟

**Buyer:** نعم بالتأكيد، ولكن قبل ذلك أود معرفة المزيد عن السيارة. هل يمكنك توضيح مواصفاتها وحالتها؟

**Seller:** بالطبع، السيارة بحالة جيدة جداً وموديلها من عام 2015. وهي سيارة أوتوماتيكية، ماشية 100,000 كيلومتر فقط. كما أنها مجهزة بنظام تكييف وسماعات ستريو.

**Buyer:** أرى، هذه مواصفات جيدة. وماذا عن السعر؟

**Seller:** أنا أطلب 50,000 ريال.

**Buyer:** هذا السعر يبدو قليلاً مرتفعاً. هل يمكنك النزول إلى 45,000 ريال؟

**Seller:** أعتذر، لكن السعر الذي طلبته هو السعر الأدنى الذي أستطيع بيع السيارة به.

**Buyer:** ولكن السيارة قد قطعت 100,000 كيلومتر، وهذا يعني أنها قد استخدمت بشكل كبير. أعتقد أن 45,000 ريال هو سعر أكثر ملاءمة.

**Seller:** أنا أفهم ما تقوله، ولكن السيارة بحالة جيدة جداً وتم العناية بها بشكل جيد. لكن لأنك جاد في شرائها، أقبل بتخفيض السعر إلى 48,000 ريال.

**Buyer:** أقدر تعاونك واستعدادك للتفاوض. ماذا لو قمنا بتقسيم الفرق، أقترح السعر 46,500 ريال؟

**Seller:** حسناً، هذا مقبول. يمكننا الإتفاق على 46,500 ريال.

**Buyer:** رائع، أتطلع لرؤية السيارة وإتمام الصفقة. شكراً لك.

### English Dialogue:

Prompt: A buyer is interested in negotiating the price of a car listed at 50,000 Riyals.

**Buyer:** Hello. I'm interested in the car you've advertised for sale online. Is it still available?

**Seller:** Hello. Yes, the car is still available. Would you like to come and see it?

**Buyer:** Definitely, but before that, I'd like to know more about it. Could you tell me about its specifications and condition?

**Seller:** Certainly. The car is in excellent condition. It's a 2015 model, automatic, and has only 100,000 kilometers on it. It's equipped with air conditioning and stereo speakers.

**Buyer:** Sounds like good specs. What's your asking price?

**Seller:** I'm asking for 50,000 Riyals.

**Buyer:** That seems a bit high. Would you consider lowering the price to 45,000 Riyals?

**Seller:** I'm sorry, but the price I've set is the minimum I can sell the car for.

**Buyer:** Considering the car has 100,000 kilometers on it, which indicates significant use, I believe 45,000 Riyals would be a more appropriate price.

**Seller:** I understand your point, but the car is in very good condition and has been well-maintained. However, since you are earnest about buying, I'm willing to reduce the price to 48,000 Riyals.

**Buyer:** I appreciate your cooperation and readiness to negotiate. How about we meet halfway? I suggest 46,500 Riyals.



### Arabic Dialogue:

**Prompt:** 11 مليون درهم إماراتي هو السعر الأولي للسيارة هو 11 مليون درهم إماراتي الفاخرة. السعر الأولي للسيارة هو 11 مليون درهم إماراتي.

**Buyer:** مساء الخير، أنا مهتم بشراء البوغاتي فيرون التي أمامنا.

**Seller:** مساء النور، هذا خيار رائع. هذه السيارة تبلغ قيمتها 11 مليون درهم إماراتي.

**Buyer:** أعرف، ولكنني أعتقد أن السعر مبالغ فيه قليلاً. هل يمكننا التفاوض عليه؟

**Seller:** بالطبع، نحن نرحب بالتفاوض. كم تتوقع أن تدفع؟

**Buyer:** ما رأيك في 9 مليون درهم؟

**Seller:** هذا السعر أقل بكثير من قيمة السيارة. ولكن، لأنك جاد في الشراء، أستطيع أن أقدم لك خصم 10% على السعر الأصلي. هذا يعني أن السعر الجديد سيكون 9.9 مليون درهم.

**Buyer:** أقدر ذلك، ولكنني مازلت أعتقد أن السعر مرتفع. ماذا لو قمت بدفع 50% الآن و 50% على مدى السنة القادمة. هل يمكنك خصم أكبر في هذه الحالة؟

**Seller:** هذا اقتراح جيد. إذا كنت ستدفع نصف القيمة الآن، يمكنني أن أقدم لك خصم 15% على القيمة الإجمالية. هذا يعني أنك ستدفع 4.675 مليون درهم الآن، ونفس المبلغ في المستقبل.

**Buyer:** يبدو هذا عرضاً عادلاً. سأقبل العرض. شكراً لتعاونك.

**Seller:** شكراً لك. ستكون سعيداً بهذه السيارة الرائعة.

---

### English Dialogue:

**Prompt:** The scenario takes place in the luxury car market in Dubai, where negotiations are underway for the price of a luxurious Bugatti Veyron. The initial price of the car is 11 million United Arab Emirates Dirhams.

**Buyer:** Good evening, I am interested in purchasing the Bugatti Veyron in front of us.

**Seller:** Good evening, that's an excellent choice. This car is valued at 11 million United Arab Emirates Dirhams.

**Buyer:** I know, but I think the price is a bit exaggerated. Can we negotiate it?

**Seller:** Of course, we welcome negotiations. How much do you expect to pay?

**Buyer:** How about 9 million Dirhams?

**Seller:** That price is much lower than the car's value. However, since you are serious about buying, I can offer you a 10% discount on the original price. This means the new price will be 9.9 million Dirhams.

**Buyer:** I appreciate that, but I still think the price is high. What if I pay 50% now and 50% over the next year. Could you offer a bigger discount in this case?

**Seller:** That's a good proposal. If you're going to pay half the value now, I can offer you a 15% discount on the total value. This means you'll pay 4.675 million Dirhams now, and the same amount in the future.

**Buyer:** That seems like a fair offer. I accept. Thank you for your cooperation.

**Seller:** Thank you. You will be happy with this magnificent car.

### Arabic Dialogue:

#### Prompt:

في هذه اللعبة التفاوض، يتم تقديم مبدئيًا لمشاركين، أمانى وسمير، مع العناصر التالية: كتاب واحد، 3 قبعات، وكرة واحدة. القيم لهذه العناصر بالنسبة لكل مشارك هي: يتم التقسيم حسب القيمة المعنوية للاغراض لكل شخص وفق مقياس من 1 الى 5 بحيث يكون 1 الاقل اهمية و 5 الاكثر اهمية. يتم تقسيم الاهمية كالتالي

أمانى: الكتاب = 1، القبعة = 3، الكرة = 0

سمير: الكتاب = 2، القبعة = 2، الكرة = 2.

**Seller:** "نظرًا لأنني لا أرى قيمة كبيرة في الكرة، أنا على استعداد لمبادلتها معك بقبعة واحدة"

**Buyer:** "أنفهم عرضك، لكن الكتاب له نفس القيمة بالنسبة لي ولك. أقترح أن أبادلك الكرة بقبعتين، مع الحفاظ على الكتاب"

**Seller:** "أرغب في الكتاب أيضًا، إذا كنت ستأخذ قبعتين. ما رأيك في الكتاب وقبعتين مقابل الكرة؟"

**Buyer:** "الكتاب له قيمة كبيرة بالنسبة لي. ماذا عن الكتاب مقابل الكرة مباشرة، وتأخذ أنت قبعتين؟"

**Seller:** "موافقة، هذا يبدو عادلاً. أنا أخذ الكتاب وقبعتين، وأنت تأخذ الكرة وقبعة واحدة"

**Buyer:** "إتفاننا. لدينا الآن توزيع نهائي يناسب كلانا"

---

### English Dialogue:

#### Prompt:

In this negotiation game, initially, two participants, Amani and Samir, are presented with the following items: one book, 3 hats, and one ball. The values of these items for each participant are: The division is based on the sentimental value of the items for each person, on a scale from 1 to 5, where 1 is the least important and 5 is the most important. The importance is divided as follows:

Amani: Book = 1, Hat = 3, Ball = 0.

Samir: Book = 2, Hat = 2, Ball = 2.

**Seller:** "Since I don't see much value in the ball, I am willing to exchange it with you for one hat."

**Buyer:** "I understand your offer, but the book has the same value for both of us. I propose that I exchange the ball for two hats, while keeping the book."

**Seller:** "I am also interested in the book, if you are taking two hats. What do you think about the book and two hats in exchange for the ball?"

**Buyer:** "The book has a great value to me. How about the book directly in exchange for the ball, and you take two hats?"

**Seller:** "Agreed, that seems fair. I'll take the book and two hats, and you take the ball and one hat."

**Arabic Dialogue:**

**Prompt:** مديرة قسم تقنية المعلومات (أنثى) في شركة كبرى تتفاوض مع ممثل شركة برمجيات (ذكر) حول ترخيص برمجيات للشركة بقيمة 3 ملايين ريال. المديرية تستخدم تكتيك التفاوض على القيمة المضافة.

**Buyer:** "نحن مهتمون ببرمجياتكم، لكننا نريد التأكد من أننا نحصل على أفضل قيمة ممكنة. ماذا يمكنكم تقديمه بالإضافة إلى الترخيص؟"

**Seller:** "يا سيدتي، بالإضافة إلى الترخيص، نحن نقدم دعمًا فنيًا شاملاً وتحديثات منتظمة. يمكننا أيضًا تقديم خصم بنسبة 5%."

**Buyer:** "الدعم الفني والتحديثات مهمة، لكننا نبحث عن شيء يعزز كفاءتنا. هل يمكنكم تضمين تدريب مخصص لفريقنا وتحليلات متقدمة لبياناتنا؟"

**Seller:** "إنها إضافات قيمة. يمكننا تضمين جلسات تدريب مخصصة وتحليلات بيانات متقدمة. مع هذه الإضافات، هل يمكننا الاتفاق على الصفقة؟"

**Buyer:** "نعم، هذه الشروط تلبي احتياجاتنا. دعونا نتابع لتوقيع العقد."

**Seller:** "ممتاز، ساعد العقد ونتابع لإتمام الاتفاقية."

-----  
**English Dialogue:**

**Prompt:** An IT department manager (female) at a major company is negotiating with a representative of a software company (male) about a software licensing deal worth 3 million Riyals. The manager, a female, uses a negotiation tactic focusing on added value.

**Buyer (Female):** "We are interested in your software, but we want to ensure that we're getting the best possible value. What can you offer in addition to the licensing?"

**Seller (Male):** "Madam, in addition to the license, we offer comprehensive technical support and regular updates. We can also offer a 5% discount."

**Buyer (Female):** "Technical support and updates are important, but we're looking for something that enhances our efficiency. Can you include customized training for our team and advanced analytics for our data?"

**Seller (Male):** "Those are valuable additions. We can include customized training sessions and advanced data analytics. With these additions, can we agree on the deal?"

**Buyer (Female):** "Yes, these terms meet our needs. Let's proceed to sign the contract."

**Seller (Male):** "Excellent, I'll prepare the contract and we can proceed to finalize the agreement."

Example of Pronoun Consistency in Arabic Negotiation Dialogue with English Translation

### Arabic Dialogue:

**Prompt:** 4 ، و 500 دولار) ، قيمة كل آلة موسيقية (قيمة كل آلة 150 دولار) ، 2 آلات موسيقية (قيمة كل آلة 500 دولار) ، و 4 ساعات حائط (قيمة كل ساعة 200 دولار). يتم التقسيم حسب القيمة المعنوية للاغراض لكل شخص. يتم تقسيم الاهمية كالتالي

البائع (هو): يقدر الآلات الموسيقية والساعات

المشترية (هي): مهتمة بالآلات الرياضية وتريد آلة موسيقية

**Seller:** "أرغب في الاحتفاظ بآلة واحدة وثلاث ساعات. هل تأخذين الآلة الأخرى ومجموعتين رياضيتين مقابلهم؟"

**Buyer:** "أريد الآلتين ومجموعتين. ماذا عن الآلتين مقابل ثلاث مجموعات رياضية؟"

**Seller:** "أستطيع التخلي عن آلة واحدة فقط. هل يمكن أن نضيف مجموعة رياضية ثالثة؟"

**Buyer:** "أوافق، لكن أريد إضافة ساعة للصفقة"

**Seller:** "إذا، آلة واحدة وساعة مقابل أربع مجموعات رياضية؟"

**Buyer:** "هذا يعادل 700 دولار مقابل 600 دولار. أضيف مجموعة خامسة ونتفق"

**Seller:** "أوافق على ذلك، صفقة مقبولة"

---

### English Dialogue:

**Prompt:** Negotiating the division of 6 sets of sports equipment (each set valued at \$150), 2 musical instruments (each valued at \$500), and 4 wall clocks (each valued at \$200). The division is based on the sentimental value of the items for each person. The importance is divided as follows:

**Seller (Male):** Values the musical instruments and the clocks.

**Buyer (Female):** Interested in the sports equipment and wants a musical instrument.

**Seller:** "I would like to keep one instrument and three clocks. Would you take the other instrument and two sports sets in exchange?"

**Buyer:** "I want both instruments and two sets. How about the two instruments in exchange for three sports sets?"

**Seller:** "I can only part with one instrument. Can we add a third sports set?"

**Buyer:** "I agree, but I want to add a clock to the deal."

**Seller:** "So, one instrument and a clock in exchange for four sports sets?"

**Buyer:** "That equals \$700 against \$600. I'll add a fifth set and we have a deal."

### Arabic Dialogue:

**Prompt:** بائع (ذكر) في شركة تكنولوجيا يتفاوض مع مشتريّة (أنثى) تمثّل شركة كبرى حول ترقيّة نظام تكنولوجيا بقيمة 1.2 مليون ريال. المشتريّة تستخدم تكتيك التفاوض على خصم متدرج بناءً على الحجم والخدمات المضافة.

**Buyer:** "نحن نخطط لترقيّة نظامنا التكنولوجي، وتبلغ تكلفة المشروع 1.2 مليون ريال. هل يمكنكم تقديم خصم يعتمد على حجم المشروع؟"

**Seller:** "بالنظر إلى حجم المشروع، يمكننا تقديم خصم 10٪، مما يخفض السعر إلى 1.08 مليون ريال."

**Buyer:** "أقدر العرض، لكننا نبحث أيضًا عن خدمات إضافية. ماذا عن تضمين دعم فني موسع وتدريب الموظفين مقابل خصم إضافي؟"

**Seller:** "إذا كنتم تبحثون عن دعم فني موسع وتدريب، يمكننا زيادة الخصم إلى 15٪ مع تضمين هذه الخدمات، مما يجعل السعر النهائي 1.02 مليون ريال."

**Buyer:** "هذا عرض جيد. ولكن، إذا قمتم بتضمين تحديثات البرمجيات لمدة عامين، سأوافق على الفور."

**Seller:** "إنها صفقة كبيرة، لكننا على استعداد لتضمين تحديثات البرمجيات لمدة عامين والحفاظ على السعر عند 1.02 مليون ريال لضمان شراكة طويلة الأمد."

**Buyer:** "ممتاز، هذا يلبي جميع متطلباتنا. دعونا نتابع لتوقيع العقد."

### English Dialogue:

**Prompt:** A male seller in a technology company is negotiating with a female buyer representing a major company about upgrading a technological system valued at 1.2 million Riyals. The buyer uses a negotiation tactic focusing on graduated discounts based on project size and added services.

**Buyer (Female):** "We are planning to upgrade our technological system, and the project's cost is 1.2 million Riyals. Can you offer a discount based on the project's size?"

**Seller (Male):** "Considering the project's size, we can offer a 10% discount, reducing the price to 1.08 million Riyals."

**Buyer (Female):** "I appreciate the offer, but we are also looking for additional services. How about including extended technical support and employee training for an additional discount?"

**Seller (Male):** "If you are looking for extended technical support and training, we can increase the discount to 15% with these services included, making the final price 1.02 million Riyals."

**Buyer (Female):** "That's a good offer. However, if you include software updates for two years, I'll agree immediately."

**Seller (Male):** "It's a significant deal, but we are willing to include two years of software updates and maintain the price at 1.02 million Riyals to ensure a long-term partnership."

**Buyer (Female):** "Excellent, this meets all our requirements. Let's proceed to sign the contract."

### Arabic Dialogue:

#### Prompt:

المفاوضة على تقسيم 2 أجهزة تلفاز، 3 ساعات ذكية، و4 كتب تاريخية. يتم التقسيم حسب القيمة المعنوية للاغراض لكل شخص وفق مقياس من 1 الى 5 بحيث يكون 1 الاقل اهمية و 5 الاكثر اهمية. . يتم تقسيم الاهمية كالتالي

البائعة (هي): الأجهزة = 3، الساعات = 4، الكتب = 1

المشتري (هو): الأجهزة = 1، الساعات = 2، الكتب = 3

**Seller:** "أنا بحاجة إلى جهاز تلفاز وساعة ذكية. هل تأخذ الجهاز الآخر وساعة مقابل كتابين؟"

**Buyer:** "أريد كلا الأجهزة. ماذا عن الأجهزة مقابل كتاب وساعتين؟"

**Seller:** "لا أستطيع التخلي عن الساعات كلها. هل يمكن أن نضيف شيئاً آخر؟"

**Buyer:** "حسناً، الأجهزة مقابل كتابين وساعة"

**Seller:** "هذا عرض جيد، نتفق عليه"

---

### English Dialogue:

#### Prompt:

Negotiating the division of 2 televisions, 3 smartwatches, and 4 historical books. The division is based on the sentimental value of the items for each person, rated on a scale from 1 to 5, where 1 is the least important and 5 is the most important. The importance is divided as follows:

Seller (Female): Televisions = 3, Watches = 4, Books = 1.

Buyer (Male): Televisions = 1, Watches = 2, Books = 3.

**Seller:** "I need one television and one smartwatch. Would you take the other television and a watch in exchange for two books?"

**Buyer:** "I want both televisions. How about the televisions in exchange for one book and two watches?"

**Seller:** "I can't give away all the watches. Can we add something else?"

**Buyer:** "Okay, the televisions in exchange for two books and one watch."

**Seller:** "That's a good offer, we agree on it."

	Arabic Text	English Translation
<b>Prompt</b>	<p>انشئ حوار تفاوضي طويلاً ومعقد بين بائع ومشتري استناداً إلى السيناريو التالي: المشتري في مكتبة خلال الساعات الأخيرة من تخفيض اليوم واحد يقدم خصماً بنسبة 30٪ على جميع الكتب. المشتري صميل دائم ويحاول التفاوض على خصم إضافي أو صفقات خاصة على شراء بالجملة لمجموعة متنوعة من الكتب</p>	<p>Create a lengthy and complex negotiation dialogue between a seller and a buyer based on the following scenario: The buyer is in a bookstore during the final hours of a one-day sale and offers a 30% discount on all books. The buyer is a loyal customer and is attempting to negotiate an additional discount or special deals for bulk purchases of a variety of books.</p>
<b>Base Model</b>	<p>المشتري: مرحباً، أنا أبحث عن سيارة مستعملة. البائع: نعم، لدينا بعض السيارات المستعملة المتاحة للبيع. ما هو نوع السيارة التي تبحث عنها؟ المشتري: أود سيارة صغيرة بأربعة أبواب ومحرك قوي. البائع: بالطبع، لدينا عدة خيارات تناسب متطلباتك. هل تفضل سيارة من فئة السيدان أم سيارة رياضية الشكل؟ المشتري: أفضل سيارة من فئة السيدان. البائع: لدينا سيارات من عدة ماركات شهيرة مثل تويوتا وهوندا ونيسان. المشتري: أفضل تويوتا. هل لديك سيارة كورولا مستعملة؟ البائع: نعم، لدينا بعض سيارات كورولا مستعملة. المشتري: نعم، ذلك يناسبني. أود أن أشتريها بسعر 10000 دولار. هل يمكن أن تقل هذا العرض؟ البائع: السعر المطلوب قليل بعض الشيء. هل يمكن أن تزيد العرض إلى 12000 دولار؟ المشتري: 12000 دولار هو سعر مرتفع بعض الشيء بالنسبة لي. هل يمكن أن تقبل 11000 دولار؟ البائع: لا يمكنني أن أقبل 11000 دولار. هل يمكن أن تزيد العرض قليلاً؟ المشتري: حسناً، يمكنني أن أقبل 11500 دولار. البائع: حسناً، يمكنني أن أقبل 11500 دولار.</p>	<p>Buyer: "Hello, I'm looking for a used car." Seller: "Yes, we have some used cars available for sale. What type of car are you looking for?" Buyer: "I'd like a small car with four doors and a powerful engine." Seller: "Certainly, we have several options that meet your requirements. Do you prefer a sedan or a sporty-looking car?" Buyer: "I prefer a sedan." Seller: "We have cars from several well-known brands like Toyota, Honda, and Nissan." Buyer: "I prefer Toyota. Do you have a used Corolla?" Seller: "Yes, we have some used Corolla cars." Buyer: "Great, that suits me. I'd like to buy it for \$10,000. Can you accept this offer?" Seller: "The price you're offering is a bit low. Can you increase the offer to \$12,000?" Buyer: "\$12,000 is a bit high for me. Can you accept \$11,000?" Seller: "I can't accept \$11,000. Can you slightly increase the offer?" Buyer: "Alright, I can accept \$11,500." Seller: "Alright, I can accept \$11,500"</p>
<b>Fine Tuned Model</b>	<p>البائع: "مرحباً، يمكنني مساعدتك في شيء ما؟" المشتري: "أنا أبحث عن جهاز كمبيوتر محمول جديد. هل لديك أي شيء في المخزون؟" البائع: "نعم، لدينا العديد من الخيارات المتاحة. ما هو ميزانيتك؟" المشتري: "أنا أفكر في قضاء حوالي 4000 ريال." البائع: "هذا المبلغ منخفض للغاية. لكن يمكنني تقديم جهاز بمواصفات جيدة بـ 4500 ريال." المشتري: "هذا أكثر مما أستطيع تحمله. ماذا عن 4200 ريال؟" البائع: "أعترض، ولكن هذا العرض لا يمكن قبوله. لكن يمكنني تقديم خصم بنسبة 5٪ إذا اشتريت اليوم." المشتري: "هذا سيجعل السعر 4275 ريال. هل يمكنك جعله 4250 ريال وأقوم بالشراء اليوم؟" البائع: "حسناً، لدينا اتفاق. السعر سيكون 4250 ريال إذا أجريت الشراء اليوم."</p>	<p>Seller: "Hello, can I assist you with something?" Buyer: "I'm looking for a new laptop. Do you have anything in stock?" Seller: "Yes, we have several options available. What's your budget?" Buyer: "I'm thinking of spending around 4000 Riyals." Seller: "That amount is quite low. However, I can offer a laptop with good specifications for 4500 Riyals." Buyer: "That's more than I can afford. How about 4200 Riyals?" Seller: "I'm sorry, but that offer is not acceptable. However, I can give you a 5% discount if you purchase today." Buyer: "That would make the price 4275 Riyals. Can you make it 4250 Riyals, and I'll make the purchase today?" Seller: "Alright, we have a deal. The price will be 4250 Riyals if you make the purchase today."</p>

Comparative Dialogue Outputs from Original and Fine-Tuned Models - This figure presents side-by-side dialogue examples illustrating the adherence to the negotiation scenario and the contextual relevance achieved by the fine-tuned model versus the base GPT-3.5 model.

	Arabic Text	English Translation
<b>Prompt</b>	أنشئ حوار تفاوضي طويل ومعقد بين بائع ومشتري استنادًا إلى السيناريو التالي: مشتري يرغب في بدلة مصممة حسب الطلب، والخياط يقدم في البداية عرضًا بقيمة 4000 ريال سعودي.	Create a long and complex negotiation dialogue between a seller and a buyer based on the following scenario: The buyer wants a custom-tailored suit, and the tailor initially offers a price of 4000 Saudi Riyals.
<b>Base Model</b>	<p>المشتري: أنا أرغب في بدلة رسمية بقماس عالي الجودة ولون أسود. أود أن تكون مصممة بشكل متقن وتحتوي على تفاصيل خاصة مثل أزرار من الذهب الخالص.</p> <p>البائع: لدينا مجموعة واسعة من الأقمشة العالية الجودة والألوان المختلفة. نقتح عليك استخدام الصوف الإيطالي الناعم للقماس ونفضل البدلة بالكامل حسب تفضيلاتك. بالنسبة للأزرار، يمكننا استخدام أزرار من الذهب الخالص، ولكن هذا سيؤثر على سعر البدلة. هل توافق على ذلك؟</p> <p>المشتري: نعم، أنا مستعد لدفع سعر أعلى من أجل الجودة والتفاصيل الخاصة ما هو السعر المقترح؟</p> <p>البائع: سعر البدلة المصممة حسب الطلب مع الأزرار من الذهب الخالص سيكون 1000 دولار. ولكن إذا كنت ترغب في خفض السعر قليلاً، يمكننا استخدام أزرار من البلاستيك مطلية بالذهب بدلاً من الذهب الخالص مقابل 800 دولار. ما رأيك في ذلك؟</p>	<p>Buyer: I want a formal suit made from high-quality fabric in black color. I'd like it to be finely crafted with special details like pure gold buttons.</p> <p>Seller: We have a wide range of high-quality fabrics and different colors. We suggest using soft Italian wool for the fabric and tailoring the suit entirely to your preferences. As for the buttons, we can use pure gold buttons, but this will affect the price of the suit. Do you agree to that?</p> <p>Buyer: Yes, I'm willing to pay a higher price for quality and special details. What is the proposed price?</p> <p>Seller: The price for the custom-tailored suit with pure gold buttons will be 1000 dollars. However, if you'd like to lower the price slightly, we can use plastic buttons coated with gold instead of pure gold, for 800 dollars. What do you think?</p>
<b>Fine Tuned Model</b>	<p>البائع: لدينا العديد من الخيارات للبدل الرجالية المصممة حسب الطلب. كمية العمل والنسيج المستخدم يحددان السعر.</p> <p>المشتري: هل يمكنني الحصول على فكرة عن الأسعار؟</p> <p>البائع: بدءاً من 5000 ريال سعودي وحتى 10000 ريال سعودي.</p> <p>المشتري: أعتقد أن هذا مرتفع جداً. هل يمكننا التفاوض؟</p> <p>البائع: يمكننا النقاش، ولكن السعر يعتمد على التفاصيل التي ترغب فيها.</p> <p>المشتري: أبحث عن بدلة بسيطة بنسيج جيد. ما هو أفضل سعر يمكنك تقديمه؟</p> <p>البائع: أقل سعر يمكنني تقديمه هو 4000 ريال سعودي.</p> <p>المشتري: هل يمكنك أن تقدم لي خصمًا إذا اشتريت اثنين؟</p> <p>البائع: إذا اشتريت اثنين، يمكنني تقديمهما بـ 7500 ريال سعودي.</p>	<p>Seller: We have many options for custom-made men's suits. The amount of work and the fabric used determine the price.</p> <p>Buyer: Can I get an idea about the prices?</p> <p>Seller: Starting from 5000 Saudi Riyals up to 10000 Saudi Riyals.</p> <p>Buyer: I think that's too high. Can we negotiate?</p> <p>Seller: We can discuss, but the price depends on the details you want.</p> <p>Buyer: I'm looking for a simple suit with good fabric. What's the best price you can offer?</p> <p>Seller: The lowest price I can offer is 4000 Saudi Riyals.</p> <p>Buyer: Can you give me a discount if I buy two?</p> <p>Seller: If you buy two, I can offer them for 7500 Saudi Riyals.</p>

Example Dialogues Comparison - This figure presents a side-by-side comparison of dialogue outputs from the base and fine-tuned models, showcasing the adherence to the negotiation scenario and currency specifications