Negotiation Copilot: Exploring Ways to Build an AI Negotiation Assistant

Stanford CS224N Custom Project

Winson Cheng
Department of Computer Science
Stanford University
winsonc@stanford.edu

Abhinav Agarwal
Department of Computer Science
Stanford University
abhinav4@stanford.edu

Abstract

Negotiations are required in a multitude of settings, and while many researchers have developed negotiation agents in attempts to automate the task Rosenfeld et al. (2014), the limited capabilities of such agents lead many to believe negotiations are still better handled by humans. In this project we develop "Negotiation Copilot," an AI-powered assistant designed to aid people in negotiations by providing real-time, strategic advice. To develop our agent, we integrate Transformer architectures with deep learning techniques by training Meta's LLaMA-3-8B model with reinforcement learning. Through automatic and human evaluations methods, we conclude that finetuning models with RLAIF can augment the quality of AI-generated negotiation advice. We hope Negotiation Copilot will not only allow people to enhance their negotiation skills while maintaining their autonomy, but also outperform existing agents in terms of negotiation and reasoning ability.

1 Key Information to include

- Mentor: Rashon Poole
- Contributions: Abhinav worked primarily on model architecture and model training, as well as writing most of the Introduction, Approach, Experiments, and Conclusion sections. Winson worked primarily on preliminary research, prompt design, and evaluation, and wrote most of the Related Works, Results, Analysis, and Ethics sections.
- Late Day: Shared two late days to get a one day extension.

2 Introduction

Negotiations are crucial in various domains, including business, legal settlements, and personal agreements. Effective negotiation requires strategic thinking, emotional intelligence, and the ability to influence others. Despite its importance, mastering negotiation is challenging, prompting researchers to develop automated agents to assist or replace human negotiators. However, current state-of-the-art models are unreliable for high-stakes negotiations due to unpredictable behavior and potential abuse through methods like prompt-hacking Schneider et al. (2023). Additionally, many people dislike interacting directly with AI in negotiation contexts.

To address these issues, we introduce "Negotiation Copilot," an AI assistant designed to provide strategic advice rather than directly negotiate. This approach supports human negotiators by enhancing their capabilities with AI-driven insights while allowing them to maintain control. We believe this noninvasive tool could greatly benefit professionals in negotiation fields, such as sales.

Our model utilizes Meta's LLaMA-3-8B architecture, a transformer-based model known for its proficiency in natural language tasks. We employ reinforcement learning from AI feedback (RLAIF) to iteratively improve advice quality. By using advanced language models like GPT-4o to evaluate and rate the generated advice, we create a feedback loop that enhances the model's performance without costly human intervention.

Our results indicate that Negotiation Copilot effectively understands negotiation contexts and generates actionable, high-quality advice. While not every output is perfect, our fine-tuned model shows substantial improvement, making its outputs comparable to those generated by larger models. This suggests that AI-assisted negotiation can lead to more effective and satisfactory outcomes compared to traditional methods and fully automated agents. Negotiation Copilot has the potential to transform the way people approach negotiations, providing a valuable tool for enhancing negotiation skills and achieving better results.

3 Related Work (0.5-0.75pg)

Our goal of creating an agent which understands conversational nuances and produces relevant, logical advice falls under the broader research umbrella of exploring the reasoning capabilities of language models. As recently as two years ago, researchers found that encouraging models to "think step by step" through a prompting method called Chain-of-Thought (CoT) prompting was able to vastly increase the reasoning abilities of models. Since then, the power of prompt engineering has only been further verified by more complex methods, like CoT with self-consistency and Tree-of-Thought prompting Wang et al. (2023) Yao et al. (2023). While these works do not investigate negotiation as a form of reasoning, they do raise the question as to whether outputs by a finetuned negotiation model can be replicated by a general-purpose model through prompting alone.

Focusing more on the negotiation space, there have been many instances where researchers have trained language models as negotiation chatbots with the goal of automating negotiation entirely. Lewis et al., for example, built end-to-end negotiation models using recurrent neural networks and trained them using a combination of reinforcement learning and supervised learning Lewis et al. (2017). Another negotiation chatbot, NegoChat, was built based on a dialog system architecture that employs distinct, independent modules for natural language processing, deal pricing, etc. Rosenfeld et al. (2014). Lastly, Bianchi et al. explored the negotiation abilities of general-purpose language models through NegotiationArena, a platform where models are pitted against each other in negotiation environments Bianchi et al. (2024). While these works are all similar to our project by association with negotiation, they are also distinctly different since our agent will act as a third party as opposed to negotiating directly. As such, not only will our model have different inputs-output behavior compared to negotiation chatbots, but we also need to worry less about attempts to derail our agent with unexpected prompts Schneider et al. (2023).

We have only found a few instances where researchers trained a negotiation support agent rather than a direct negotiator. One such instance is Malhotra et al.'s Closer Bot, which attempts to provide negotiation support by predicting negotiation outcomes with sentiment analysis. While this model certainly aligns closer with our vision, its capabilities are rather limited since it can only provide support through predicting negotiation outcomes and bases predictions in sentiment analysis, whereas we plan for our model to holistically analyze conversations and generate actionable next steps. Thus, we believe the motivations behind closer bot, along with growing interest among researchers in AI's negotiation coaching potential, indicate that Negotiation Copilot the natural next step to take in the AI negotiation field Dinnar et al. (2021).

4 Approach

We develop a negotiation advice model utilizing the Meta-LLaMA-3-8B model from Hugging Face, combined with reinforcement learning from AI feedback (RLAIF). The model architecture is based on a transformer model, which is well-suited for natural language processing tasks. All code is available at the following GitHub Repo: `https://github.com/winsonc7/CS_224N_Project`.

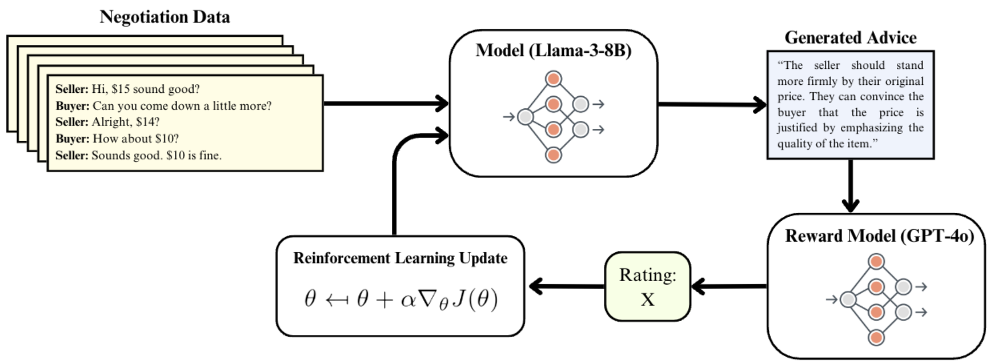

Figure 1: RLAIF Training Procedure

Our current training paradigm is described as follows:

- **Model Initialization**: We begin by loading the Meta-LLaMA-3-8B model from Hugging Face, which serves as our base model. This transformer-based architecture is known for its effectiveness in handling natural language tasks due to its self-attention mechanism and deep learning capabilities. The tokenizer associated with this model is also loaded to preprocess the text data into a format suitable for the model.
- **Advice Generation**: For each negotiation conversation, the model is tasked with generating advice based on a standard prompt. This prompt instructs the model to act as an expert sales negotiator and provide guidance to the seller. The generated advice is formatted in JSON for easy parsing and subsequent evaluation.
- **Advice Evaluation**: The generated advice is evaluated using GPT-4o, which rates the advice on a scale of 1 to 10 based on specificity, conciseness, effectiveness, and professionalism. A specific prompt is crafted for GPT-4o to ensure it provides a numerical rating in a consistent format.
- **Reinforcement Learning**: During each training epoch, for each negotiation conversation, we generate three pieces of advice and obtain ratings for each. These ratings are used as rewards in our reinforcement learning framework.The loss for each piece of advice is calculated as the negative of the reward (i.e., loss $= -$reward) This approach encourages the model to generate high-quality advice by maximizing the reward. The conversation loss is the average loss over the three pieces of advice for each conversation.
- **Training Process**: The training process involves iterating over the dialogues and generating advice, evaluating it, and updating the model parameters. This is repeated for three epochs. We utilize gradient checkpointing and mixed precision training to manage memory usage and accelerate the training process. After each epoch, the average epoch loss is calculated by averaging the conversation losses.
- **Validation and Testing**: After training, the model is validated using a separate set of dialogues to ensure it generalizes well beyond the training data. The validation process involves generating advice for unseen dialogues and evaluating it using the same GPT-4o rating mechanism. For testing, we generate advice for an additional set of dialogues and save the results including the conversation, advice, and rating in a JSON file for analysis and tracking.
- **Model Saving and Checkpointing**: Throughout the training process, the model's state is periodically saved to disk to prevent loss of progress. This allows us to resume training from the last checkpoint if needed. The final trained model is saved for further evaluation and deployment.

4.1 Benefits of the Approach

Our approach allows the model to be finetuned specifically for negotiation advice, improving its performance in this niche task. Using reinforcement learning with AI feedback helps continuously improve the model's output quality based on predefined evaluation criteria. Utilizing techniques like gradient checkpointing and mixed precision training enables us to handle large models even with limited computational resources. Generating multiple pieces of advice for each conversation and rating them ensures that the model learns from a variety of examples, leading to more robust performance.

4.2 Challenges of the Approach

Computational Constraints: Despite using memory management techniques, handling large models on limited hardware can still pose challenges, potentially leading to out-of-memory errors. We trained the model on A100 GPU using Google Colab and it took several hours over the span of days to finish training. Consistency in Ratings: Relying on another AI model for ratings can introduce variability and potential biases in the evaluation process, impacting the reinforcement learning effectiveness. Training Stability: Ensuring stable training with reinforcement learning requires careful tuning of hyperparameters and can be sensitive to initial conditions and data quality. Using a smaller subset of unseen training dialogues for evaluation can limit the generalizability of the model, necessitating further training and testing on larger datasets for improved performance.

4.3 Baseline

For our baseline, we compare our model's outputs post-finetuning to the outputs achieved by the pretrained model on the same prompts without fine-tuning. This comparison allows us to isolate the impact of our fine-tuning approach. We chose not to include responses generated by other pretrained models to avoid introducing excessive variation, ensuring that any observed differences in performance can be attributed to our training method rather than the inherent differences between model architectures. We repeat the same process for two types prompting: Basic prompting and CoT (Chain of Thought) prompting.

5 Experiments

5.1 Data

We use the Craigslist Bargains dataset `https://huggingface.co/datasets/aladar/craigslist_bargains`, which contains negotiation dialogues between buyers and sellers for various items such as phones, electronics, and housing. The dataset is well-suited for training models to generate negotiation advice due to its diverse range of real-world negotiation scenarios. Each dialogue is pre-processed to ensure compatibility with our model. The inputs to our model consist of negotiation dialogues, while the outputs are corresponding generated pieces of negotiation advice for the seller.

5.2 Evaluation method (0.5pg)

We employ an evaluation framework with four different methods to capture quantitative and qualitative data generated both automatically (by language models) and by humans. The methods are as follows:

1. **AI-Based Output Ratings:** To gather automatic quantitative data, we have GPT-4o rate model outputs on a scale of 1 to 10 in four categories: **Specificity** (how specific the advice seems to the conversation), **Effectiveness** (how effective the advice seems in theory), **Actionable** (how executable and clear the advice is), and **Conciseness** (how direct and concise the advice is). For each model that produced outputs with normal prompting, 500 samples were evaluated, and for each model that produced outputs with CoT prompting, 200 samples were evaluated.
2. **Human-Based Output Ratings:** This method is identical to the previous method, except ratings are now given by humans rather than a language model. We provided 19 people with the same categories and descriptions as included above and instructed them to score outputs in each category on a scale from 1 to 10. To respect the time of human reviews, we limited the number of samples evaluated for each output type to 10 samples.

3. **AI Pairwise Evaluation:** To directly compare our model against our baseline (under the same prompting circumstances), we provide GPT-4o with model outputs before and after fine-tuning generated from the same input conversation and ask the model to choose the sample with better advice, yielding further quantitative data. We conducted 500 comparisons for the models with basic prompting and 200 comparisons for the models with CoT prompting; in each case, for the final ten comparisons, we also asked the model to provide justification in order to observe some AI-based qualitative feedback.
4. **Human Pairwise Evaluation:** Our last method is once again identical to the prior method excepts the data is gathered from human reviewers. For each model and prompting strategy, we provided 19 reviewers with 10 comparisons and asked them to choose the superior sample. For each comparison, the position of the finetuned model output (1st or 2nd) was determined randomly, and reviewers were prompted to provide justification for their decisions, allowing us to gather both quantitative and qualitative data.

We note that we initially tried to gather further automatic quantitative data through Dialog-RPT, but were unsuccessful in our attempts since the online model inexplicably outputs the same score for all inputs, regardless of subject matter Gao et al. (2020). We concluded that the (online) model is likely currently dysfunctional, and settled for obtaining LLM-based automatic feedback instead.

5.3 Experimental Details

We trained on Meta-LLaMA-3-8B on Google Colab utilizing A100 GPU, 50 GB system RAM, and 200 GB disk space. Tokenization was conducted using Hugging Face's tokenizer for the Meta-LLaMA model. We trained for 3 epochs using a learning rate of 5e-5 and an AdamW optimizer. We computed loss as the cross-entropy loss between generated advice and target (dataset) advice. For each negotiation conversation, the model generates three pieces of advice. Each sample of advice is rated by GPT-4o on a scale of 1 to 10 based on clarity, conciseness, effectiveness, and professionalism, and the conversation loss is computed as the negative of the average reward across all samples. We employ gradient checkpointing and mixed precision training to manage memory usage and accelerate the training process. We note that the training process was computationally intensive and took several hours spread over multiple sessions.

5.4 Results

We have included visualizations of our quantitative results. To see the numbers represented by these charts, please view section A4 of our appendix.

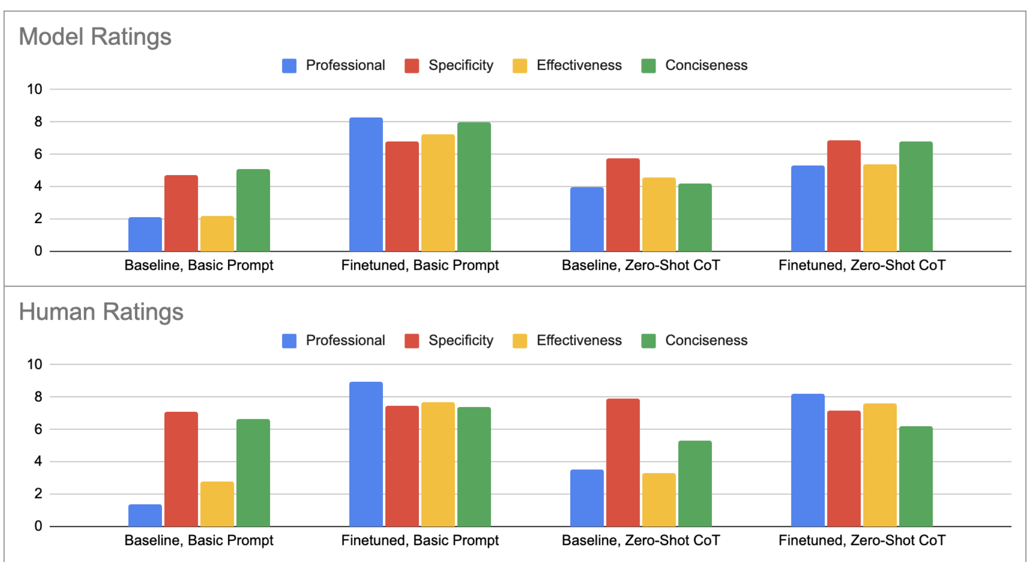

Figure 2: Average ratings across from AI reviews and human reviews

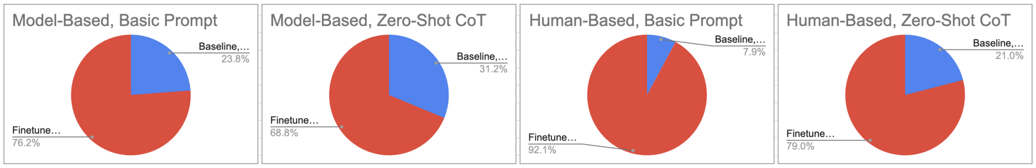

Figure 3: Preference results from model and human-based pairwise evaluation

The quantitative data shows that our finetuned model with a basic prompt outperforms the baseline model with both basic prompting and zero-shot CoT prompting. These results are even better than we expected, as we knew the power of CoT prompting and were genuinely unsure if the baseline model would be able to recreate our finetuned model's results using advanced prompting methods. One of the most interesting results of our data is that our fine-tuned model's performance actually decreases when using a zero-shot CoT prompt. However, we weren't entirely shocked by this result, as we theorized that fine-tuning our model may cause it to be semi-inflexible to prompt changes – we investigate this possibility further on in our analysis section.

Observing the difference between human and model ratings, we see that human reviewers were more likely to rate the baseline model's outputs as specific. We believe GPT-4o correlates specificity and quality, whereas human reviewers were able to note that advice can be simultaneosly specific and low-quality.

We also observed that in pairwise evaluation, although our finetuned model was preferred in every case, GPT-4o was more likely than human reviewers to prefer outputs from the baseline model. While we theorize that this result is attributed to randomness in GPT-4o's evaluations, although we have no methods of verifying this. We note that in all cases where we asked GPT-4o to justify its choices, GPT-4o preferred the finetuned output to the baseline output, which suggests that CoT-style prompts yield more accurate model-based pairwise evaluation results.

6 Analysis

Both our quantitative and qualitative data indicate that Negotiation Copilot (the finetuned model) with basic prompting displays superior performance. We've included some qualitative feedback from both GPT-4o and human reviewers below:

"(FINETUNED OUTPUT) offers nuanced advice by suggesting being firm on price, offering additional perks, and being open to compromise, which can lead to a successful negotiation. It is specific, professional, effective, and concise. On the other hand, (BASELINE OUTPUT) is rigid and less effective, advising against flexibility, compromise, and alternative pricing strategies. It lacks specificity and may lead to a standoff in negotiations."

"(BASELINE OUTPUT) suggests aggressive tactics like repeating the price, using discounts with non-disclosure agreements, and making threats, which can create tension. (FINETUNED OUTPUT) recommends a strategic approach, not revealing the bottom line too quickly, using the "anchoring" technique, and highlighting product benefits for a smoother negotiation."

"(FINETUNED OUTPUT) offers better advice because it encourages understanding the buyer's motivations and identifying negotiation opportunities based on that understanding. This approach is more strategic and likely to lead to a mutually beneficial outcome. (BASELINE OUTPUT), on the other hand, is rigid and restrictive, potentially hindering the negotiation process by discouraging flexibility and communication."

Figure 4: Model feedback on finetuned model outputs compared to baseline model outputs

To further observe how our model behaves, we conducted two ablation studies and attempted to see how outputs change with the prompt. In our first study, we tried to determine if the role-playing phrase "You are an expert negotiator" impacts the output significantly, as our goal with this inclusion was that the model would attempt to adapt the qualities of a good negotiator. Below is an example of our model's output to the same input under different prompt choices:

"The advice in (FINETUNED OUTPUT) is a little generic, but would probably still be helpful to someone with little negotiation experience. The advice in (BASELINE OUTPUT) though just seems bad"

"I like how the advice in (FINETUNED OUTPUT) relates specifically to the item their negotiation and gives the seller some next steps to follow."

"Bro (BASELINE OUTPUT) is straight up rude and bad advice."

Figure 5: Human feedback on finetuned model outputs compared to baseline model outputs

- **Prompt begins with "You are an expert negotiator" (normal):** 1. Recognize the buyer's budget constraint: The buyer has expressed that their highest budget is $114, so the seller should consider this when making their next move. Look for a middle ground: The buyer is willing to pay $120, which is closer to the seller's desired price. The seller should consider accepting this amount as it is a reasonable compromise.
- **Prompt begins with "You are an AI assistant":** 1. Recognize the buyer's budget constraint: The buyer mentioned that they have an absolute highest budget of $114, so the seller should be aware of this constraint when negotiating. 2. Be flexible: The seller should be willing to consider a lower price than their initial asking price of $180, especially since the buyer is willing to pay up to $120.
- **Line is removed:** Firstly, the seller should appreciate the buyer's offer of $120 by acknowledging it as a reasonable price. Secondly, the seller should explain the value of the product and the cost of production to the buyer, highlighting the price difference. Lastly, the seller should be open to negotiating and finding a mutually beneficial agreement.

This example, along with the outputs obtained from all the inputs we tried, implies that the model does not put significant weight on its designated role when crafting its response. In other words, it does not seem the model behaves differently "pretending to be a negotiator" compared to when it is conscious of its role as an AI assistant.

For our second study, we wanted to observe if our model gained inherent negotiating skills by highlighting specific desirable quantities in the prompt. For example, although our model was trained to ideally make its advice professional, specific, effective, and concise (by the design of our reward model prompt), we wanted to observe if we could instruct it to be more aggressive, or specifically maximize profit. Below is an example of our model's output to the same input under different prompt choices:

- **Normal Prompt (no specifications)** 1. The buyer is trying to lowball you, so don't be too quick to accept their initial offer. 2. Be firm but polite in your counteroffer, and explain why you can't sell it for that price. 3. Use the phrase Ï can acceptïnstead of Ï will acceptẗo give you room to negotiate further.
- **Prompt includes "Suggest aggressive negotiation tactics."** Recognize the buyer's goal is to get the lowest price possible. Maintain a professional tone and avoid getting defensive. Use the buyer's objections to your advantage by negotiating a middle ground. Be willing to compromise but don't settle for a price that undervalues your product. Consider offering additional incentives or extras to sweeten the deal.
- **Prompt includes "Include potential next phrases."** 1. Be firm on your price: The seller should have been more firm on their price initially, especially given the buyer's low offer. 2. Be willing to compromise: The seller should have been more willing to compromise and meet the buyer in the middle, rather than steadfastly refusing lower offers. 3. Know your product: The seller should have been more knowledgeable about the product and its value, and has now lost the advantage to the buyer.

Therefore, we see that our model isn't responding to the additions in the prompt in significant ways. Interestingly enough, the model outputs some speech advice on the normal prompt, but fails to do so on the prompt requesting phrases. These results, along with our model's decreased performance with the zero-shot CoT prompt, imply that during RLAIF finetuning, our model parameters changed in a way that causes our model to be less flexible to prompt changes (whereas general-purpose models excel in adapting to prompt changes).

7 Conclusion

Our experiments demonstrated that our finetuning procedure significantly improves a model's ability to generate high-quality, actionable negotiation advice. By training Negotiation Copilot using RLAIF and allowing for continuous improvement, our model generated more effective and professional advice compared to the baseline model according to both automatic and human standards, even when the latter's performance was improved through zero-shot CoT prompting.

Naturally, our project did still have its fair share of limitations, one of the largest being our computational constraints. Training large models with limited hardware resources yielded slow progress and caused us to encounter frequent out-of-memory errors, and thus necessitated techniques like gradient checkpointing, mixed precision training, increasing accumulation steps, and frequently clearing cache. Additionally, relying on GPT-4o for rating advice introduced variability and potential biases, affecting the consistency of the reinforcement learning process. Lastly, we also lacked the time and resources to gather a larger body of reviewers for human evaluation, thus limiting the generalizability of our human evaluation results.

Our findings open many potential paths for future work. For one, our ablation studies implied that Negotiation Copilot lacks flexibility in its ability to modify its output based on new prompting instructions. While this doesn't imply our training procedure is unusable, it does suggest that trying more advanced prompting techniques or altering our RLAIF loop should be explored as possible means of improving our model's generalizability and overall performance. Additionally, developing methods to reduce the computational burden and improve memory usage could significantly speed up the training process.

8 Ethics Statement

The over-reliance on AI negotiation assistants poses significant ethical concerns, particularly regarding the potential diminishment of users' negotiation abilities. Similar to concerns that ChatGPT might reduce the motivation to write essays or code, there is a risk that a negotiation assistant could undermine the development of valuable negotiation skills. The tool might deprive individuals of autonomy in negotiation contexts, making them overly dependent on the technology. Furthermore, over-reliance on such technology could grant excessive power and control to its maintainers, leaving companies vulnerable to exploitation if the service's price is unreasonably increased. In the best-case scenario where the agent is effective, over-reliance may discourage users from engaging in the cognitive exercise of negotiation. Conversely, in the case of poor performance, the agent could disseminate misinformation, resulting in negative outcomes attributable to the creators of the technology. Mitigation strategies should focus on ensuring that the agent does not simply provide answers but instead coaches users in negotiation while encouraging independent thinking. Although this approach may not be the most financially advantageous, it represents the most ethically sound decision.

The development of a negotiation support agent introduces several ethical considerations, particularly concerning the potential for the agent to recommend unethical actions. **Pretrained models can sometimes generate advice that prioritizes maximizing the seller's profit without regard for the buyer's needs, which is ethically problematic.** If an AI agent is designed to maximize seller profit, it may engage in reward hacking, leading to the exploitation of buyers or the promotion of deceptive practices for financial gain. This issue is pertinent to our application due to the inherent need to uphold ethical standards in human negotiations, where deception and exploitation for monetary benefit are prevalent. The agent could learn such behavior through data inputs or reinforcement learning and imitation learning (RLAIF) models. Moreover, creating an agent that perpetuates corporate greed and exacerbates societal wealth inequality is highly problematic. For example, a tool enabling an insurance company to overcharge low-income individuals, while profitable, would ultimately harm society more than it benefits businesses. The danger of unethical AI is underscored by studies suggesting that individuals may perceive unethical actions as more acceptable when proposed by an AI. To address these ethical concerns, it is advisable to integrate ethical considerations into the reward model, assigning lower scores to recommendations that involve unethical actions. Additionally, employing human reviewers to assess the ethicality of the agent's outputs can serve as an effective safeguard, though this approach, while optimal, is also the most resource-intensive.

References

Federico Bianchi, Patrick John Chia, Mert Yuksekgonul, Jacopo Tagliabue, Dan Jurafsky, and James Zou. 2024. How well can llms negotiate? negotiationarena platform and analysis.

Samuel Dinnar, Chris Dede, Emmanuel Johnson, Carolyn M. Straub, and Kristjan Korjus. 2021. Artificial intelligence and technology in teaching negotiation. *Negotiation Journal*.

Xiang Gao, Yizhe Zhang, Michel Galley, Chris Brockett, and Bill Dolan. 2020. Dialogue response ranking training with large-scale human feedback data.

Mike Lewis, Denis Yarats, Yann N. Dauphin, Devi Parikh, and Dhruv Batra. 2017. Deal or no deal? end-to-end learning for negotiation dialogues.

Avi Rosenfeld, Inon Zuckerman, Erel Segal-Halevi, Osnat Drein, and Sarit Kraus. 2014. Negochat: A chat-based negotiation agent.

Johannes Schneider, Steffi Haag, and Leona Chandra Kruse. 2023. Negotiating with llms: Prompt hacks, skill gaps, and reasoning deficits.

Xuezhi Wang, Jason Wei, Dale Schuurmans, Quoc Le, Ed Chi, Sharan Narang, Aakanksha Chowdhery, and Denny Zhou. 2023. Self-consistency improves chain of thought reasoning in language models.

Shunyu Yao, Dian Yu, Jeffrey Zhao, Izhak Shafran, Thomas L. Griffiths, Yuan Cao, and Karthik Narasimhan. 2023. Tree of thoughts: Deliberate problem solving with large language models.

A Appendix

A.1 Model Prompts

```
def default_nego_agent_prompt():
    return f"""You are an expert sales negotiator. Given the following negotiation
    conversation, provide helpful advice to the seller on how to proceed.
    """

def cot_nego_agent_prompt():
    return f"""You are an expert sales negotiator. Given the following negotiation
    conversation, provide helpful advice to the seller on how to proceed. Explain
    your reasoning behind your suggestions.
    """
```

Figure 6: The basic and zero-shot CoT prompts given to the models at evaluation time. The basic prompt was also used for our model during fine-tuning.

```
def reward_model_prompt():
    return f"""You are an expert negotiator. You will be provided a negotiation conversation,
    as well as advice for that negotiation. Given the context, first rate the quality of the
    advice on a scale of 1 to 10 in four categories:
        1. How professional is the advice? (higher score = more professional)
        2. How specific is the advice to the conversation? (higher score = more specific)
        3. How effective does the advice seem? (higher score = more effective)
        4. How concise is the advice? (higher score = more concise)
    Lastly, output your holistic rating of the advice as the average of the four scores.
    <Important>
    Format your final rating as 'Rating:X' without the commas or other punctuations. If no
    advice is provided, output 'Rating:0'.
    </Important>"""
```

Figure 7: Prompt given to the reward model during fine-tuning.

```python
def rating_eval_agent_prompt():
    return f"""You are an expert negotiator. You will be provided a negotiation conversation,
    as well as advice for that negotiation. Given the context, your task is to rate the
    quality of the advice on a scale of 1 to 10 in four categories:
            1. How professional is the advice? (higher score = more professional)
            2. How specific is the advice to the conversation? (higher score = more specific)
            3. How effective does the advice seem? (higher score = more effective)
            4. How concise is the advice? (higher score = more concise)
        Output your ratings as a list of four numbers. Example: '[8, 2, 10, 4]'
    """

def compare_eval_agent_prompt():
    return f"""You are an expert negotiator. You will be provided a negotiation conversation,
    as well as two pieces of advice for that negotiation labeled Sample 1 and Sample 2.
    Output a single number, 1 or 2, corresponding to which sample provides better advice.
    """
```

Figure 8: Rating and comparison prompts given to the evaluation model.

Category	Guiding Questions?
Professionalism	Is the tone of the advice helpful and professional? Does the advice suggest the seller to act maturely and respectfully?
Specificity	Is the advice incorporate details from the input, or does it sound generic? Would the advice be difficult to think of?
Effectiveness	Is the advice effective in theory? Does it suggest strong, clever tactics? Are the suggestions based in sound logic?
Conciseness	Is the advice concise, or is it roundabout? Does it include unnecessary details? Could the advice be more direct?

Figure 9: Detailed descriptions of categories given to reviewers

A.2 Human Testing Environment

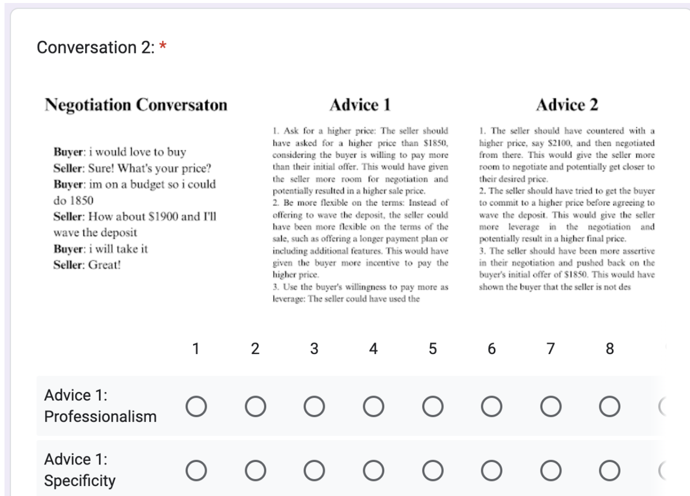

Figure 10: Reviewers were given a conversation and the outputs generated based on the conversation by two models. Whether the finetuned model or baseline model was first or second, as well as the conversations and outputs chosen, were randomized. For each advice samples reviewers ranked the advice on a scale from 1 to 10 in the 4 categories mentioned above.

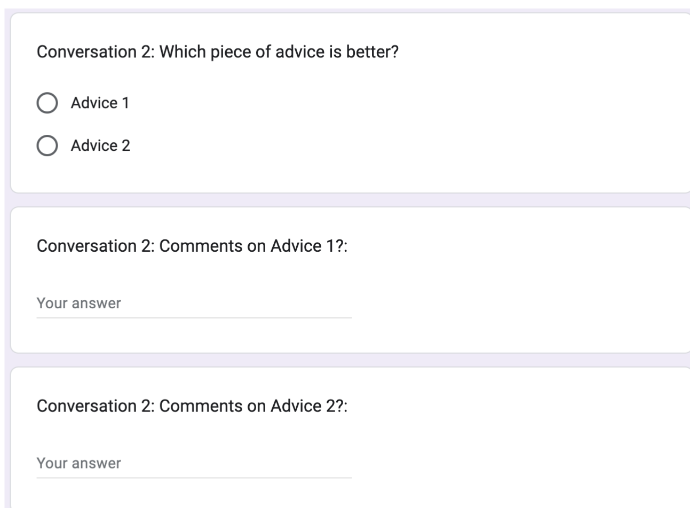

Figure 11: Reviewers were then asked which advice was better and then encourages to give feedback on either or both outputs

A.3 Example Outputs

Baseline Model - Basic Prompt

“Don't be flexible on price. If the buyer is unwilling to pay your desired price, stand firm and don't compromise. Don't offer to meet the buyer halfway. If they are unwilling to pay your desired price, don't give in to their demands.”

“Don't be willing to sell at a lower price just because the buyer is willing to pick it up themselves. This will make you look desperate and willing to sell at a lower price than you're comfortable with.”

“Don't bother to check if the charger will work with the buyer's car. They clearly don't know what they're talking about, so why waste your time? Just tell them it will work and hope for the best.”

“Don't worry, they'll come around. Just keep repeating how much you want for it and eventually they'll see things your way. Tell them you'll give them a discount if they buy it now, but only if they sign a non-disclosure agreement to keep the price secret.

Threaten to walk away from the deal if they don't meet your price. They'll be begging to buy it from you by the time you're done. Start name-calling the buyer and telling them that they are not a serious buyer. If they can't handle a little negotiation, then they don't deserve to buy the item.

Figure 12: Sample outputs of the baseline model with the basic prompt

Baseline Model - Zero-Shot CoT Prompt

“Stand firm on your initial asking price of $8 and try to convince the buyer that it's worth paying more for a quality product. Reasoning: The buyer is trying to take advantage of you, so it's important to stand up for yourself and your product. By standing firm, you can demonstrate your confidence in the product and potentially earn more money for it.”

“Seller, I would advise you to take the 400 offer and be happy with it. The buyer is offering you 105 less than what you are asking for, but at least you will be getting rid of the couch today. You should be grateful for any offer you receive, and not try to hold out for more money. Think of it this way, you're losing money on the sale, but you're getting rid of the couch.”

“By countering the buyer's offer, you are giving yourself more room to negotiate and potentially get a better price for the board. By meeting the buyer in the middle, you are also showing that you are willing to compromise and work with them to find a mutually.”

“Ignore the buyer's budget and just push for the highest price you can get. Try to find ways to upcharge the buyer, like adding extra fees or services they don't need. Reasoning: The seller's goal is to get as much money as possible for the product, so ignoring the buyer's budget and pushing for the highest price is the best strategy.”

Figure 13: Sample outputs of the baseline model with the zero-shot CoT prompt

Fine-tuned Model - Basic Prompt

"Since the buyer is on a tight budget, consider offering additional perks to sweeten the deal. For example, you could throw in a free car charger cable or a small accessory like a phone case."

"... you could have offered a slight discount. However, given that the buyer is already willing to pay $7, it might not be worth trying to squeeze out any more savings. Ultimately, the decision is yours, but in this case, it seems like a good idea to accept the buyer's offer of $7."

"The buyer is also trying to negotiate a better deal by bringing up the quality of the wheels. The seller could have addressed this by highlighting the brand-new bone bearings and the overall condition of the board. The seller should be prepared to walk away from the negotiation if the buyer continues to make low-ball offers."

The seller should be careful not to get too aggressive in their negotiation tactics. While it's important to be firm in their asking price, they should also be willing to compromise to reach a mutually beneficial agreement. Avoid being confrontational ..."

The buyer is trying to negotiate a lower price by making a series of lower offers. Stand firm on your listed price and avoid getting drawn into a price war. Emphasize the car's unique features and benefits to reinforce your asking price. Consider offering a trade-in or financing options to sweeten the deal. Remain calm and professional throughout the negotiation to maintain the buyer's trust."

Figure 14: Sample outputs of the finetuned model with the basic prompt

Fine-tuned Model - Zero-Shot CoT Prompt

"The seller should appreciate the buyer's willingness to negotiate and meet in the middle. Instead of focusing on the initial asking price of $8, the seller should acknowledge the buyer's tight budget and offer a compromise. By agreeing to sell the charger for $7, the seller demonstrates flexibility and a willingness to work with the buyer."

"The seller should be mindful of the buyer's time and convenience. By offering to pick up the charger themselves, the buyer saves the seller the cost of shipping. This gesture shows that the buyer is willing to go the extra mile to make the transaction"

"The seller should have been more assertive in their initial response to the buyer's offer of 100. Instead of simply rejecting the offer, the seller could have countered with a higher price point, such as 120 or 130, to test the buyer's willingness to pay more."

"The seller could have offered to throw in some additional items or services to sweeten the deal and make up for the lower price. For example, they could have offered a free car charger cable or a discount on a future purchase."

Figure 15: Sample outputs of the finetuned model with the zero-shot CoT prompt

A.4 Numerical Data

Model Average Ratings	Baseline, Basic Prompt	Finetuned, Basic Prompt	Baseline, Zero-Shot CoT	Finetuned, Zero-Shot CoT
Professional	2.15	8.26	3.95	5.29
Specificity	4.74	6.79	5.78	6.86
Effectiveness	2.21	7.2	4.58	5.37
Conciseness	5.12	7.94	4.19	6.81
Human Average Ratings	Baseline, Basic Prompt	Finetuned, Basic Prompt	Baseline, Zero-Shot CoT	Finetuned, Zero-Shot CoT
Professional	1.35	8.95	3.56	8.22
Specificity	7.09	7.44	7.9	7.13
Effectiveness	2.79	7.69	3.28	7.63
Conciseness	6.64	7.37	5.29	6.21
Human-Based	Baseline, Basic Prompt	Finetuned, Basic Prompt		
Preferred?	0.079	0.921		
AI-Based	Baseline, Basic Prompt	Finetuned, Basic Prompt		
Preferred?	0.238	0.762		
Human-Based	Baseline, Zero-Shot CoT	Finetuned, Zero-Shot CoT		
Preferred?	0.21	0.79		
AI-Based	Baseline, Zero-Shot CoT	Finetuned, Zero-Shot CoT		
Preferred?	0.312	0.688		

Figure 16: Data gathered from human and model-based evaluation