GENTEEL-NEGOTIATOR: LLM-Enhanced Mixture-of-Expert-based Reinforcement Learning Approach for Polite Negotiation Dialogue

Priyanshu Priya[1], **Rishikant Chigrupaatii**[1], **Mauajama Firdaus**[2], **Asif Ekbal**[3]

[1]Department of Computer Science and Engineering, Indian Institute of Technology Patna, India
[2]Department of Computer Science and Engineering, Indian Institute of Technology (Indian School of Mines) Dhanbad, India
[3]School of Artificial Intelligence and Data Science, Indian Institute of Technology Jodhpur, India
[1]{priyanshu_2021cs26, rishikant_2101cs66,}@iitp.ac.in, [2]mauajama@iitism.ac.in, [3]asif@iitj.ac.in

Appendix

This section presents complementary materials, including dataset details, bidirectional polite keyterm graph description, experiment details, and additional analysis to enhance the reader's comprehension of the work.

Dataset Details

Guidelines for Drafting Sample Dialogues

The human subjects are instructed to adhere to the following guidelines during dialogue creation:

- To enhance the understanding of negotiation and ensure the application of best negotiation practices in dialogue creation, we follow (Chawla et al. 2021) and require the human subjects to watch a video tutorial on negotiation between two art collectors before beginning their tasks;
- Begin negotiations with offers that exceed your target to allow room for concessions and adjustments during the discussion;
- Actively explore and discuss the preferences and priorities of each other to align on key aspects of the negotiation;
- Clearly communicate your specific requirements and rationale to strengthen the position and make a compelling argument;
- Remember that different parties may have varying perceptions of what is rational and fair. Aim for an outcome that is favorable to both sides;
- Use polite and emotional expressions strategically and appropriately to build rapport and convey sincerity while maintaining professionalism.

Dialogue Filtering and Quality Assessment

Once the complete dialogue dataset is generated, each dialogue is evaluated based on the following criteria: Fair (***F***), Polite (***P***), Interesting (***I***), Coherent (***C***), and Natural (***N***) by the same group of human subjects involved in the dataset preparation on a scale of 1-5 (low to high). The human subjects are asked to rate dialogues according to the following instructions:

1. **Fair** examines the equity of the final negotiation outcome, determining how effectively it achieves a win-win scenario.

   Instruction: *To what extent does the final negotiation outcome reflect fairness and equity, resulting in a win-win scenario for both parties?*

   Scale: 1 (entirely unfair, with a clear disadvantage to one party) to 5 (perfectly equitable, achieving a balanced win-win outcome).
2. **Polite** assesses the degree to which the dialogue employs courteous and respectful language throughout the interaction.

   Instruction: *How well does the dialogue maintain a courteous and respectful tone throughout the interaction?*

   Scale: 1 (highly impolite and disrespectful) to 5 (exceptionally polite and respectful).
3. **Interesting** assesses the ability of the dialogues to sustain the user's interest throughout the interaction.

   Instruction: *How engaging and content-rich is the overall conversation?*

   Scale: 1 (generic and uninteresting) to 5 (highly engaging and rich in content).
4. **Coherent** evaluates how logically structured and clear the overall conversation is, with particular attention to the dialogue flow and connection between the utterances.

   Instruction: *How well-structured and logical is the overall conversation?*

   Scale: 1 (completely incoherent) to 5 (as coherent as a conversation between two native English speakers).
5. **Natural** evaluates how closely the generated dialogue resembles a human conversation.

   Instruction: *To what extent does the overall conversation exhibit naturalness?*

   Scale: 1 (completely unnatural) to 5 (as natural as a conversation between two native English speakers).

The dialogues that obtain a rating of 1, 2, or 3 for any of these metrics are omitted from the dataset. We eventually obtain average ratings of 4.51, 4.27, 4.23, 4.19, and 4.82 for ***F***, ***P***, ***I***, ***C***, and ***N***, respectively. These results indicate that the synthetically generated dialogues exhibit a high degree of fairness in negotiation outcomes and utilize polite language effectively while being interesting, coherent, and natural, thereby reflecting high-quality negotiation interactions. An agreement ratio (McHugh 2012) of 84.2%, 82.9%, 81.6%, 79.6%, and 81.1% is observed among the human subjects. Figure 2 depicts an example dialogue generated through prompting. The final dataset statistics are reported in Table 1.

	IND			NeGoChat		
	Train	Dev	Test	Train	Dev	Test
# Dialogues	2,914	417	833	884	114	244
# Utterances	40,175	5,739	11,479	13,107	1,772	3,759
Avg. Utterances/Dialogue	13.79	13.76	13.78	14.83	15.54	15.41

Table 1: NEGOCHAT and IND dataset statistics.

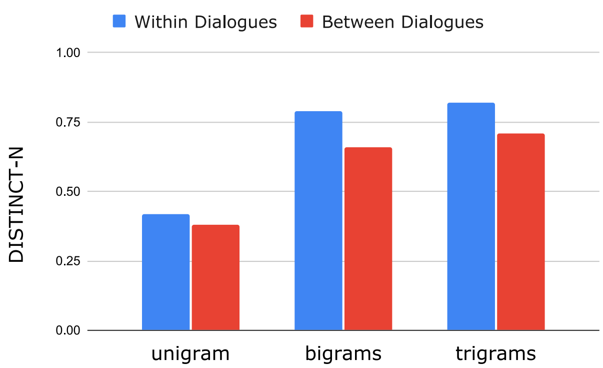

Figure 1: Distinct-N with N = 1, 2, 3 for dialogues in NEGOCHAT. The synthetic dialogues have lexical diversity.

Given the significant influence in-context examples have on prompting (Brown et al. 2020), one potential concern is that the small size of our in-context examples might limit the lexical diversity of our synthetically generated dialogues. To address this concern, we evaluate lexical diversity using the DISTINCT-N metric (Li et al. 2015), which measures the diversity of n-grams within the text. In particular, we calculate DISTINCT-N (Li et al. 2015) both within the dialogues' utterances to assess lexical diversity within individual dialogues and between the dialogues centered on the same travel package to evaluate lexical diversity across different dialogues. As depicted in Figure 1, the synthetically generated dialogues obtain high DISTINCT scores in terms of distinct unigrams, bigrams, and trigrams, demonstrating that the dialogues maintain a rich lexical diversity.

Integrative Negotiation Dataset (IND) Details

The 10 different electronics items considered in the IND dataset include air conditioning, television, refrigerator, oven, washing machine, printer, smartphone, laptop, tablet, and camera. The rationale behind creating the IND dataset is rooted in the common practice within online marketplaces where products often comprise multiple components, such as a chair and cushion. An effective negotiation should accommodate users who seek only specific components of a product, such as just the chair or just the cushion. This approach can significantly enhance user satisfaction and lead to mutually advantageous outcomes. Hence, the IND is designed to facilitate more nuanced and flexible negotiation interactions by treating products as bundles of items and negotiations involve not only price adjustments but also the addition or removal of items from deal bundles. This dataset is annotated with a list of 11 intents, namely Greet, Ask, Inform, Ask-Clarification, Negotiate-Price-Increase, Negotiate-Price-Decrease, Negotiate-Price-NoChange, Negotiate-Add-X, Negotiate-Remove-X, Accept, and Reject.

Dataset Annotation Procedure

Based on the recent advancements in leveraging the LLMs as annotators to significantly reduce annotation costs and human effort (Gilardi, Alizadeh, and Kubli 2023; He et al. 2023), we employ ChatGPT (GPT-3.5) with a human-in-the-loop approach for the negotiation strategy annotation of the datasets. The entire annotation process unfolds in four steps.

1. **Manual annotation of dialogue samples:** It involves randomly sampling 100 dialogues each from both datasets and manually annotating the utterances in these dialogues with the negotiation strategy labels as per the prepared set of guidelines. These annotated dialogue sets are referred to as ANN-NEGOCHAT and ANN-IND, hereafter.
2. **Few-shot prompting for negotiation strategy labeling:** In this step, ChatGPT (OpenAI 2024) is prompted in a few-shot setting, using the manually annotated samples as few-shot demonstrations, to generate negotiation negotiation labels. These labels are accompanied by natural language explanations to provide insight into the model's predictions. It is important to note that the strategies and corresponding explanations for ANN-NEGOCHAT and ANN-IND are also generated by ChatGPT. The sample prompt is shown below.

> **Prompt template for negotiation strategy labeling**
>
> **Task instruction:** Given the dialogue context, classify the target utterance into *Active listening*, *Leverage information*, *Expanding-the-pie*, *Logrolling*, *Gradual concession-making*, *Large-initial concession-making*, *Patterned concession-making*, or *No strategy*. Then explain your prediction.
> **Few-shot demonstrations:**
> Dialogue context:
> Target utterance:
> Negotiation Strategy Label:
> **Target input:**
> Dialogue context:
> Target utterance:
> **LLM Output:**
> Negotiation Strategy: <label>
> Explanation: <explanation>

3. **Verification of LLM-generated labels:** It involves training two BERT-based (Devlin et al. 2018) verifier models,

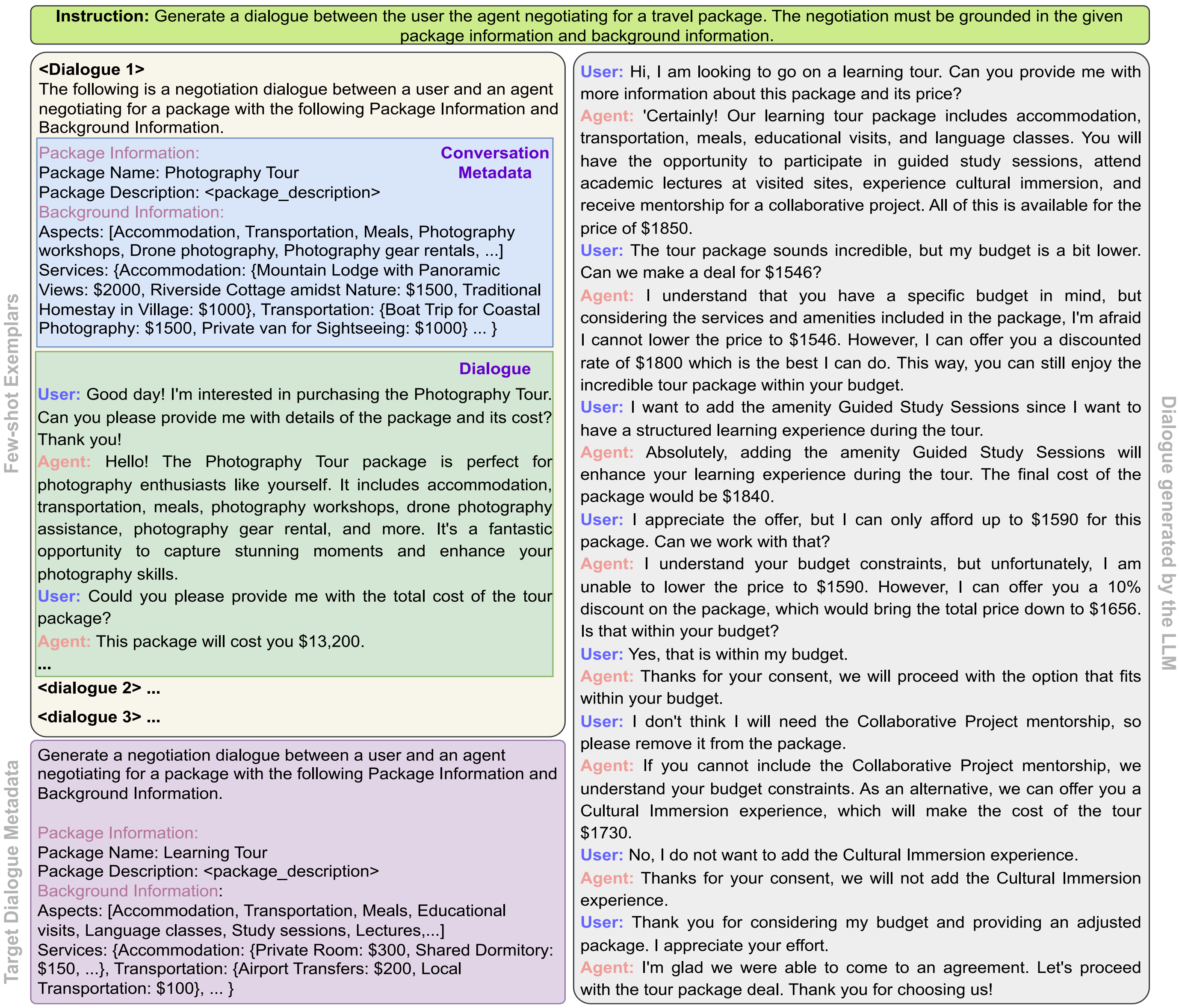

Figure 2: Example of the prompt (left) used to generate a negotiation dialogue (right).

one for ANN-NEGOCHAT and another for ANN-IND. These verifier models verify each LLM-generated label to identify potentially incorrect predictions for manual correction, which avoids wasting human effort on correcting already accurate labels. For each data sample represented as quadruple $d = (x, y, \hat{y}, \hat{e})$, where $x, y, \hat{y},$ and $\hat{e}$ denote the input utterance, gold label, LLM-generated label, and LLM-generated explanation, respectively, we develop a verifier model as the binary classification model $\mathbb{P}_\theta(v|x, \hat{y}, \hat{e})$, where $v \in \{0, 1\} = 1$ indicates that the LLM annotation passing the verification $(\hat{y} = y)$[1]. The LLM-generated predictions for samples from both datasets are passed through their respective verifiers to filter out the samples with incorrect predictions.

4. **Manual re-annotation:** It involves re-annotating the samples pinpointed by the verifiers by the human subjects to create gold-standard datasets. We observe a reliable multi-rater Kappa (McHugh 2012) agreement ratios of 84.2% and 81.6% in Step 1 and 83.4% and 86.1% in Step 4 for the NEGOCHAT and IND datasets, respectively. It is noted that dataset annotation involves the same human subjects as those involved in NEGOCHAT dataset preparation.

[1] We obtain an accuracy of 79.3% and 77.8% on ANN-NEGOCHAT and ANN-IND, respectively.

Intent	Example
Active listening	*Certainly, if you don't need the docking station, I can sell you the tablet only for $801. How does that sound?*
Leverage information	*This TV is a great buy. It has a great set of features that will make you enjoy watching your favorite shows and movies. If you are looking for a TV that will have a great picture quality and features, this one is for you.*
Logrolling	*I understand your budget constraints, but I'm sorry I cannot reduce the price of the Vizio M-Series Quantum TV. It has a 60Hz refresh rate, 55-inch screen, and Quantum Color technology for enhanced color accuracy. It is also one of the latest models from Vizio, which adds to its value. I'm sure that it will be worth the investment even for $115100.0. However, I can offer you a complimentary TV stand to add value to your purchase. Would that work for you?*
Expanding-the-pie	*The printer, ink or toner cartridges, and dust cover altogether would only cost you $27800.0. It is a killer deal for the product.*
Gradual concession-making	*I'm afraid that I cannot go that low. However, I will give you a discount of 10%. Is that acceptable?*
Large initial concession-making	*I understand that $37300 is a lot of money, but I'm willing to meet you in the middle. Please let me know if you are interested in purchasing this refrigerator.*
Patterned concession-making	*I fully understand your position, and I am willing to give you a discount of $100 for your camera. Let's proceed with the transaction!*
No strategy	*What would be the price of the printer?*

Table 2: The example utterances of different strategies in the IND dataset.

In Table 2, we provide example utterances for various negotiation strategies.

Birectional Polite Keyterm Graph Description

An illustrative example of the bidirectional polite keyterm graph $\mathcal{G}$ is depicted in Figure 3.

One-hop Reasoning on Graph $\mathcal{G}$. For the contextual keyterm "like", the one-hop neighbors reasoned via the "forward-positive" include "understand", "concern", "appreciate", and "best" while those reasoned via the "forward-negative" relation include "please" and "how" Consequently, the one-hop neighbors derived from the "forward" relation encompass "understand", "concern", "appreciate", "best", "please" and "how". For the response keyterm "appreciate", it has "like" as the one-hop neighbors reasoned through the "backward-positive" relation.

Multi-hop Reasoning on Graph $\mathcal{G}$. Consider the example of multi-hop reasoning with the sequence "forward $\rightarrow$ forward $\rightarrow$ backward-positive". By applying the "forward" relationship to the contextual keyterms for one-hop reasoning, we derive a set of neighbors that includes the response keyterms, which are treated as the extended keyterm set of the response, as determined by the context. Using these keytermss as a basis, performing a second-hop reasoning with the "forward" relation yields an expanded keyterm set for the user's future utterance (i.e., the user's next turn). Subsequently, a third-hop reasoning using the "backward-positive" relation identifies the extended positive keyterm set of the response that aligns with the anticipated future utterance.

Experiment Details

Implementation Details

All implementations are conducted using PyTorch[2], and we employ transformer-based models from Hugging Face (Wolf et al. 2019) throughout our experiments. The pre-trained models (GPT-2 and DialoGPT) use small versions. All models are trained using an NVIDIA A100-PCIE-40GB GPU with CUDA 11.2. For OpenAI's GPT-3.5 experiment, we use `gpt-3.5-turbo-0125` via API. The number of steps is empirically set to $T = 2$, and the values of reward weights are set to $w_{cNS} = w_{fNS} = w_{dP} = w_{uP} = 0.1$ and $w_{cDC} = w_{fDC} = w_E = 1.0$, determined through a grid search with values $\{0.1, 1.0\}$ for each hyperparameter. The discount factor α is set to 0.9, and hyperparameter δ is set to $1e^{-5}$. We extract $M = 10$ politeness strategies for each utterance. The maximum number of turns in conversation M_{max} is set to 8, the seed value is set to 42, and the batch size is set to 2. The Adam optimizer (Kingma and Ba 2014) with an initial learning rate of $2e^{-5}$ and a linear warmup of 120 steps. The warm start stage is trained for 6 epochs and joint training for 4 epochs. The responses are decoded using Top-k sampling with $k = 30$ and Top-p sampling with $p = 0.9$ (Holtzman et al. 2019), a temperature parameter $\tau = 0.7$, and a repetition penalty set to 1.03.

Baselines Details

1. DialoGPT (Zhang et al. 2020): Fine-tuned DialoGPT with the negotiation strategy in a supervised setting.
2. ARDM (Wu et al. 2021): Two GPT2 models modeling the user and the agent separately, and then jointly trained in a supervised fashion to better capture different speakers' language styles.
3. PersRFI (Shi et al. 2021): GPT-2 fine-tuned using RL with human demonstrations.
4. GPT-Critic (Jang, Lee, and Kim 2022): GPT-2 enhanced through cloning of critic-guided self-generated sentences during the fine-tuning process.
5. INA (Ahmad et al. 2023): GPT-2 trained in RL setting with task-relevance rewards to dynamically adjust prices and manage bundle deals for effective integrative negotiations.

[2]https://pytorch.org/

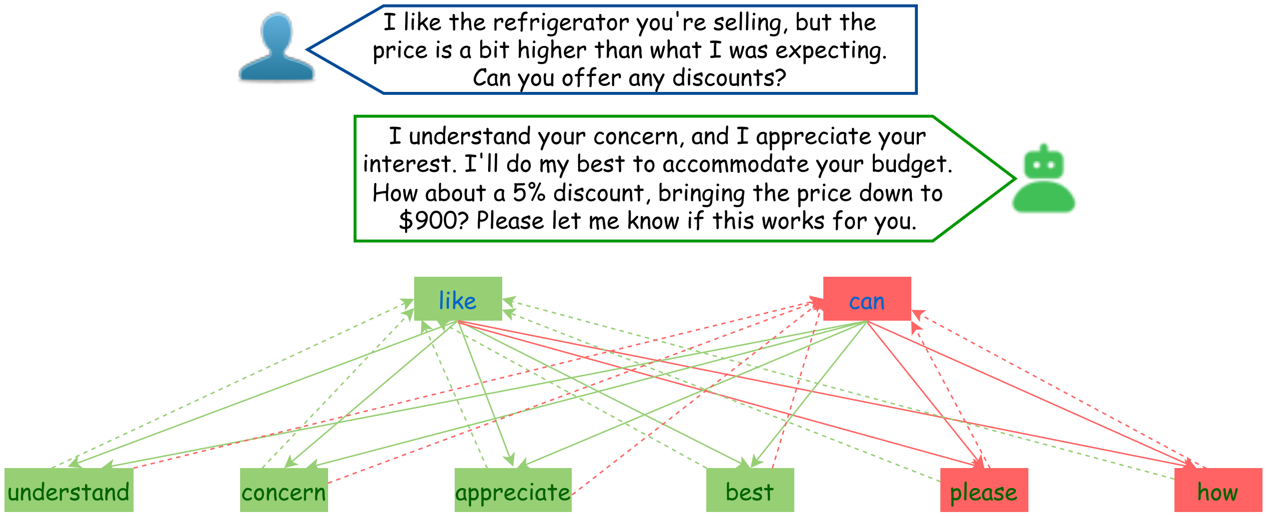

Figure 3: An example of a bidirectional polite keyword graph $\mathcal{G}$ is shown. In this graph, the text highlighted in blue represents key terms extracted from the user's utterance, while the text highlighted in green represents key terms extracted from the agent's utterance. The text within the green box represents positive key terms, and the text within the red box represents negative key terms. Solid lines indicate forward edges, and dashed lines indicate backward edges. Green arrows denote positive edges, while red arrows denote negative edges.

6. ProCoT + GPT-3.5 (Deng et al. 2023): Prompting GPT-3.5 to generate a chain-of-thought descriptive analysis for planning the strategy for the next turn.
7. LLaMA-3.1 (Touvron et al. 2023): Fine-tuned LLaMA-3.1-8B-Instruct model with negotiation strategy in a supervised setting.

Evaluation Metrics Details

Automatic Evaluation Metrics. Perplexity (PPL) (Brown et al. 1992) evaluates how well a model predicts a response. Word-overlap-based metrics like BLEU (Papineni et al. 2002) (B-2) compute the overlap between the ground-truth response and the model's generated response. BERTScore-f1 (Zhang et al. 2019) (BS-f1)[3] align the generated response and the ground-truth response in latent semantic space to assess the semantic similarity between the gold response and the model's generated response. BS-f1 computes word similarity using contextual embeddings from the BERT (Devlin et al. 2019) model. DISTINCT-2 (Li et al. 2015) (D-2) measures the diversity of the generated responses. R-LEN calculates the length of the generated responses.

Negotiation Strategy Congruence (NSC) scores, which encompass contextual and future NSC, i.e. $\mathcal{S}_{cNS}$ and $\mathcal{S}_{fNS}$ assess the consistency of negotiation strategies both with the dialogue context and the future utterances. Politeness scores consisting dialogue-level and utterance-level politeness, i.e. $\mathcal{S}_{dP}$ and $\mathcal{S}_{uP}$ measure the level of politeness as the conversation progress and the anticipated level of the user's next utterance. Dialogue coherence scores, which include contextual coherence and future coherence, i.e. $\mathcal{S}_{cDC}$ and $\mathcal{S}_{fDC}$ measure the alignment of the generated response with the preceding context and the user's future utterances. Engagingness score $\mathcal{S}_{E}$ assess the model's ability to generate engaging responses.

Human Evaluation Metrics. Fluency (F) assesses the grammatical correctness, Contextual Coherence (CC) examines the alignment of the generated responses with the dialogue context, and Engagingness (E) measures the degree to which dialogue is engaging, compelling, and capable of retaining users.

Sociopsychological Closeness (SC) accounts for the perceived social, emotional, and psychological proximity between the user and the agent. The rationale for measuring SC is that the politeness strategy manipulates sociopsychological distance; in particular, positive politeness decreases sociopsychological distance, fostering closeness by emphasizing warmth, and friendliness, and negative politeness reinforces sociopsychological distance by showing respect for autonomy, avoiding imposition, and preserving personal space. Negotiation Congruence (NC) evaluates the consistency and absence of arbitrariness in the negotiation approach throughout a dialogue, ensuring that negotiation remains coherent and logical. Bargaining Efficacy (BE) measures the system's effectiveness in presenting persuasive arguments, reasoning, or incentives that impact the decision-making process of the opponent. Outcome Fairness (OF) assesses the fairness and equity of the final outcomes achieved during negotiation.

The human evaluation is done with the help of three human evaluators[4], two evaluators with Ph.D. degrees in Linguistics and one with post-graduate degrees in Computer

[3]BERTScore: https://huggingface.co/spaces/evaluate-metric/bertscore

[4]Evaluators are distinct from those involved in the dataset cre-

Models	PPL ↓	B-2 ↑	D-2 ↑	BS-f1↑	$\mathcal{S}_{cNS}$ ↑	$\mathcal{S}_{fNS}$ ↑	$\mathcal{S}_{dP}$ ↑	$\mathcal{S}_{uP}$ ↑	$\mathcal{S}_{cDC}$ ↑	$\mathcal{S}_{fDC}$ ↑	$\mathcal{S}_{E}$ ↑	R-LEN ↑
						NEGOCHAT						
GENTEEL-NEGOTIATOR	14.72	6.27	39.41	0.766	0.751	0.513	0.712	0.550	0.781	0.572	0.776	36.84
- Negotiation Expert	15.80	5.88	37.92	0.751	0.720	0.475	0.690	0.548	0.770	0.568	0.760	35.45
- Politeness Expert	15.65	5.92	37.80	0.752	0.715	0.482	0.695	0.545	0.765	0.564	0.755	35.22
- Keyterm Expert	15.45	6.05	38.10	0.755	0.725	0.468	0.685	0.540	0.760	0.570	0.758	35.30
						IND						
GENTEEL-NEGOTIATOR	1.14	8.67	48.42	0.882	0.842	0.663	0.872	0.713	0.789	0.596	0.817	40.12
- Negotiation Expert	2.65	7.11	46.36	0.866	0.798	0.620	0.855	0.698	0.772	0.583	0.803	38.76
- Politeness Expert	2.71	7.27	46.89	0.868	0.812	0.628	0.861	0.702	0.779	0.588	0.808	38.13
- Keyterm Expert	1.42	8.45	47.17	0.879	0.824	0.635	0.866	0.709	0.785	0.593	0.814	39.39

Table 3: Ablation of experts in the GENTEEL-NEGOTIATOR. **-** signifies the removal of the component.

Models	PPL ↓	B-2 ↑	D-2 ↑	BS-f1↑	$\mathcal{S}_{cNS}$ ↑	$\mathcal{S}_{fNS}$ ↑	$\mathcal{S}_{dP}$ ↑	$\mathcal{S}_{uP}$ ↑	$\mathcal{S}_{cDC}$ ↑	$\mathcal{S}_{fDC}$ ↑	$\mathcal{S}_{E}$ ↑	R-LEN ↑
						NEGOCHAT						
GENTEEL-NEGOTIATOR	14.72	6.27	39.41	0.766	0.751	0.513	0.712	0.550	0.781	0.572	0.776	36.84
- Multi-task (Only $\mathcal{L}_{nego}$)	16.00	5.77	37.50	0.749	0.712	0.480	0.692	0.543	0.767	0.567	0.760	35.00
- Multi-task (Only $\mathcal{L}_{pol}$)	15.85	5.89	37.85	0.75	0.720	0.470	0.687	0.540	0.762	0.569	0.756	35.10
- Multi-task (Only $\mathcal{L}_{kt}$)	15.90	5.92	37.70	0.753	0.730	0.467	0.690	0.546	0.768	0.572	0.755	35.20
						IND						
GENTEEL-NEGOTIATOR	1.14	8.67	48.42	0.882	0.842	0.663	0.872	0.713	0.789	0.596	0.817	40.12
- Multi-task (Only $\mathcal{L}_{nego}$)	2.31	6.08	46.21	0.854	0.794	0.617	0.852	0.695	0.768	0.580	0.799	38.58
- Multi-task (Only $\mathcal{L}_{pol}$)	2.39	6.22	47.01	0.856	0.802	0.622	0.857	0.701	0.775	0.584	0.804	38.81
- Multi-task (Only $\mathcal{L}_{kt}$)	1.95	7.16	47.39	0.865	0.798	0.620	0.854	0.698	0.771	0.582	0.802	38.72

Table 4: Ablation analysis of the GENTEEL-NEGOTIATOR with respect to Multi-task learning of experts. **-** signifies the removal of the component.

Science. All the evaluators possess sufficient experience in similar tasks. Before evaluation, evaluators are briefed about the different travel packages and electronic items along with their associated attributes and are instructed to engage in multi-turn conversation with the system. Each evaluator is required to interact 20 times with the system using a different set of responses, resulting in a total of 80 human-evaluated dialogues focused on tourism-related negotiations. Similarly, we obtain the 80 human-evaluated dialogues centered on e-commerce negotiations. Afterward, we instruct the evaluators to rate each dialogue interaction for SC, NC, BE, OF, F, CC, and E on the provided scale of 1-5 (low to high).

Additional Analysis

Ablation w.r.t Experts

We assess the influence of each expert within the proposed GENTEEL-NEGOTIATOR framework by conducting a series of ablation experiments, where one expert is removed at a time. Specifically, the first ablated model excludes the 'Negotiation Experts,' the second omits the 'Politeness Experts,' and the third removes the 'Keyterm Experts'. The performance results of these ablations are presented in Table 3. The results clearly indicate that the removal of any expert leads to a noticeable decline in performance. Notably, excluding the negotiation experts causes a substantial reduction in scores, with drops of 4.31%, and 8% points in $\mathcal{S}_{cNS}$, and $\mathcal{S}_{fNS}$, respectively, on NEGOCHAT, and similar declines of 5.51%, and 6.94%, points on IND. This performance drop likely occurs because the absence of negotiation experts hinders the model's ability to capture the nuanced negotiation semantics essential for effective dialogue. Similarly, the ablation of politeness experts and keyterm experts also leads to notable performance degradation. The politeness experts play a crucial role in enhancing the politeness of responses, as reflected in the $\mathcal{S}_{dP}$ and $\mathcal{S}_{uP}$ scores, while the keyterm experts are vital for maintaining dialogue coherence, significantly impacting the $\mathcal{S}_{cDC}$ and $\mathcal{S}_{fDC}$ scores. These findings underscore the importance of each expert in generating coherent and polite responses in negotiation dialogues.

ation process and are compensated according to the institute norms.

Ablation w.r.t Multi-task Learning of Experts

We extend the investigation to examine the impact of multi-task learning on the performance of the experts within the GENTEEL-NEGOTIATOR framework. The results of the ablation analysis for multi-task learning are presented in Table 4. In the first ablation scenario, we optimize only the $\mathcal{L}_{nego}$ loss; in the second, we focus solely on the $\mathcal{L}_{pol}$ loss; and in the third, we optimize only the $\mathcal{L}_{kt}$ loss. The findings reveal that restricting training to any single loss function, whether $\mathcal{L}_{nego}$, $\mathcal{L}_{pol}$, or $\mathcal{L}_{kt}$ leads to a decline in performance across all evaluated metrics on both the NEGOCHAT and IND datasets. These results underscore the interdependence of the different experts and demonstrate that multi-task learning endows these experts with complementary ca-

Models	PPL ↓	B-2 ↑	D-2 ↑	BS-f1↑	$\mathcal{S}_{cNS}$ ↑	$\mathcal{S}_{fNS}$ ↑	$\mathcal{S}_{dP}$ ↑	$\mathcal{S}_{uP}$ ↑	$\mathcal{S}_{cDC}$ ↑	$\mathcal{S}_{fDC}$ ↑	$\mathcal{S}_{E}$ ↑	R-LEN ↑
					NEGOCHAT							
GENTEEL-NEGOTIATOR	14.72	6.27	39.41	0.766	0.751	0.513	0.712	0.550	0.781	0.572	0.776	36.84
- NSC rewards	15.95	5.85	37.65	0.748	0.725	0.478	0.685	0.540	0.764	0.564	0.753	35.05
- P rewards	15.85	5.92	37.85	0.754	0.715	0.475	0.688	0.545	0.766	0.568	0.757	35.15
- DC rewards	15.70	5.80	37.90	0.752	0.723	0.468	0.690	0.543	0.762	0.572	0.754	35.00
- E reward	15.90	5.88	37.75	0.751	0.730	0.480	0.692	0.540	0.765	0.570	0.759	35.10
					IND							
GENTEEL-NEGOTIATOR	1.14	8.67	48.42	0.882	0.842	0.663	0.872	0.713	0.789	0.596	0.817	40.12
- NSC rewards	2.38	7.34	46.94	0.847	0.808	0.625	0.860	0.703	0.776	0.586	0.805	38.95
- P rewards	2.40	7.41	47.06	0.848	0.815	0.630	0.864	0.706	0.782	0.590	0.810	39.27
- DC rewards	1.36	7.31	47.82	0.862	0.805	0.624	0.859	0.701	0.774	0.585	0.804	38.82
- E reward	1.44	7.49	48.23	0.869	0.827	0.637	0.868	0.711	0.788	0.595	0.816	39.48

Table 5: Ablation of rewards in the GENTEEL-NEGOTIATOR. **-** signifies the removal of the component.

Models	PPL ↓	B-2 ↑	D-2 ↑	BS-f1↑	$\mathcal{S}_{cNS}$ ↑	$\mathcal{S}_{fNS}$ ↑	$\mathcal{S}_{dP}$ ↑	$\mathcal{S}_{uP}$ ↑	$\mathcal{S}_{cDC}$ ↑	$\mathcal{S}_{fDC}$ ↑	$\mathcal{S}_{E}$ ↑	R-LEN ↑
					NEGOCHAT							
GENTEEL-NEGOTIATOR	14.72	6.27	39.41	0.766	0.751	0.513	0.712	0.550	0.781	0.572	0.776	36.84
Warm-Start Only	15.50	6.02	37.95	0.746	0.715	0.460	0.668	0.545	0.763	0.546	0.758	35.05
- Warm-start	15.45	6.04	36.85	0.734	0.738	0.472	0.686	0.550	0.752	0.538	0.754	35.28
					IND							
GENTEEL-NEGOTIATOR	1.14	8.67	48.42	0.882	0.842	0.663	0.872	0.713	0.789	0.596	0.817	40.12
Warm-start Only	1.32	8.14	47.47	0.875	0.781	0.619	0.839	0.676	0.770	0.581	0.801	38.64
- Warm-start	1.29	8.03	47.12	0.863	0.800	0.624	0.853	0.683	0.765	0.577	0.797	38.46

Table 6: Ablation of warm-start and joint training in the GENTEEL-NEGOTIATOR. **-** signifies the removal of the component.

pabilities, which collectively contribute to a substantial improvement in overall performance.

Ablation w.r.t Rewards

We examine the impact of rewards in the proposed GENTEEL-NEGOTIATOR framework by systematically removing each reward component individually. The results of these reward ablations are presented in Table 5. In the first ablated model, we remove the Negotiation Strategy Congruence rewards (**-** NSC rewards), specifically R_{cNS} and R_{fNS}. The second ablated model omits the Politeness rewards (**-** P rewards), namely R_{dP} and R_{uP}. In the third ablated model, we eliminate the Dialogue Coherence rewards (**-** DC rewards), which include R_{cDC} and R_{fDC}. Finally, the fourth ablated model excludes the Engagingness reward (**-** E reward), R_E. The removal of any reward component adversely affects various critical aspects of dialogue quality, such as negotiation effectiveness, politeness, coherence, and engagingness. This is evidenced by the decreased scores in metrics such as $\mathcal{S}_{cNS}$, $\mathcal{S}_{fNS}$, $\mathcal{S}_{dP}$, $\mathcal{S}_{uP}$, $\mathcal{S}_{cDC}$, $\mathcal{S}_{fDC}$, and $\mathcal{S}_E$ across both the datasets. These findings underscore the critical importance of these rewards in fostering the generation of coherent, engaging, and polite responses within negotiation dialogues.

Ablation w.r.t Warm-start and Joint Training

We investigate the effects of different training strategies, specifically comparing the use of a warm start without joint training (Warm-start Only) against a setup without a warm start but with joint training (**-** Warm-start). The results of this analysis are detailed in Table 6. In both settings, we observe a decline in performance across both datasets. The Warm-

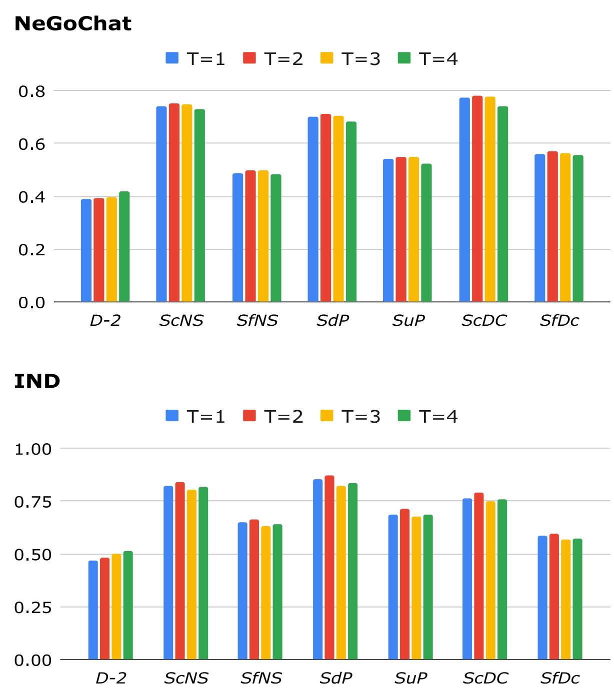

Figure 4: Analysis on the effect of no. of steps T

Interlocutor	Utterances generated by proposed model - GENTEEL-NEGOTIATOR (SC = 5, NC = 5, BE = 4, OF = 5, F = 5, CC = 5, E = 5)
User	Hi there! I'm interested in the Wildlife Adventure package. What's included in this package?
Agent	Hello! Hope you are doing well, I'd be happy to assist you today. The Wildlife Adventure package includes birdwatching, access to wildlife viewing platforms, and rentals for binoculars and cameras. The total cost is $55.
User	Sounds good, but I don't need the binoculars or camera rentals. Can you please remove them, and let me know the price?
Agent	Certainly! Removing the binoculars and camera rentals will reduce the cost by $10. It will bring down the package cost to $45. However, I would highly recommend keeping the binocular rental, as it greatly enhances the experience. With just the binocular rental included, the total package cost would be $50.
User	I understand. Let's keep the binocular rental. Could you please confirm the final cost?
Agent	Of course! With the rental camera removed and the binocular rental included, your final package cost will be $50.
User	That works for me. I'll go with this.
Agent	Excellent choice! We're delighted to assist you and hope you'll have a wonderful experience.
User	I'm also curious about the guided safari tours. Are they included in this package?
Agent	The guided tours are not included in the basic package, but they are a fantastic add-on. It will provide an immersive experience. Would you like to explore adding this to your package?
User	I do enjoy birdwatching. If I decide to add the guided tour, what would be the total cost?
Agent	It's great that you enjoy birdwatching. The addition of the guided tour would bring the total cost to $75. However, considering your interest in birdwatching, I'd be pleased to offer a special rate. How about we include the guided tour for just an additional $20, making the total package cost $70?
User	$70 is still a bit beyond my budget. Can we possibly settle on $65?
Agent	I completely understand your concern. To accommodate your budget, I'd be happy to meet you halfway at $68. I'd be delighted to include a complimentary camera rental to enhance your trip experience.
User	That sounds reasonable. Let's finalize it at $68.
Agent	Thank you so much for your understanding. I'm happy to finalize this for you. We're excited to ensure you have a memorable Wildlife Adventure. If you have any more questions or need further assistance, please don't hesitate to ask. Have a wonderful day!
Speaker	**Utterances generated by LLaMA-3.1-finetune (SC = 2, NC = 3, BE = 2, OF = 1, F = 5, CC = 4, E = 3)**
User	Hi there! I'm interested in the Wildlife Adventure package. What's included in this package?
Agent	Hi! The package includes birdwatching, wildlife viewing platforms, and rentals for binoculars and cameras. It costs $55.
User	I don't need the binoculars or camera rentals. Can you please remove them and tell me the price?
Agent	Sure, without the rentals, it's $45. If you keep the binoculars, it's $50.
User	I understand. Let's keep the binocular rental. Can you confirm the final cost?
Agent	That's good. With the binoculars included and no camera rental, it's $50.
User	That works for me. I'll go with this.
Agent	Excellent choice! We're delighted to assist you and hope you'll have a wonderful experience.
User	I'm also curious about the guided safari tours. Are they included in this package?
Agent	No, the guided tours are extra. They'd make it $75, but I can offer them for $20 more, so $70 total.
User	I do enjoy birdwatching. If I decide to add the guided tour, what would be the total cost?
Agent	The total cost with the guided tour included would normally be $75. However, we can offer a discounted total of $70 for the package. Is it okay for you?
User	$70 is too much, given my budget. Can you do $65?
Agent	The maximum that I can give you is $68 for this package. This is my limit.
User	Okay, I don't think it's feasible for me, thank you!
Agent	I understand. If you change your mind or need any more information, feel free to reach out. Thanks for considering us!

Table 7: Negotiation with our proposed model GENTEEL-NEGOTIATOR and the baseline LLaMA-3.1-finetune for a travel package. Blue text shows polite responses.

start Only approach results in slightly inferior performance, while the - Warm-start approach, which prioritizes reward maximization, achieves this objective but at the expense of overall dialogue quality. This further highlights the necessity of a balanced approach to training to maintain high-quality polite negotiation dialogues.

Analysis on Iteration Steps

We also examine the impact of varying the number of iteration steps T. As shown in Figure 4, increasing the steps leads to a rise in the D-2 score (diversity), while other metrics tend to decline. This could be due to the enhanced informativeness of responses when more experts are involved, which may result in a loss of focus and, subsequently, lower dialogue quality. Besides, GENTEEL-NEGOTIATOR consistently outperforms the best baseline in most scenarios, highlighting its effectiveness.

Interlocutor	Utterances generated by proposed model - GENTEEL-NEGOTIATOR (SC = 5, NC = 5, BE = 5, OF = 5, F = 5, CC = 5, E = 5)
User	Hi! Could you please let me know the price if I were to buy the printer, ink cartridges, and dust cover all together?
Agent	Hello! Absolutely, I'd be happy to assist. The total cost for the printer, ink cartridges, and dust cover would be $34,400. It's a fantastic deal for the complete package.
User	I'm interested in the printer but want to buy it without the ink cartridges. Can you offer me a price for just the printer?
Agent	I understand your preference, and I appreciate your interest in the printer. Unfortunately, we only offer the printer as part of the full package with the ink cartridges included. I'm sorry that we can't provide the printer separately.
User	Okay, my budget is a bit limited, and I can't stretch beyond $23,800. Is there any possibility of working within that budget for the printer and ink cartridges?
Agent	I completely understand your budget constraints and I appreciate you sharing that with me. But, I regret to mention that we can't reduce the price to $23,800. The printer is a high-end model with excellent features, including wireless adapters and compatibility with various ink cartridges. It's definitely a great investment.
User	Can you explain how the wireless adapters work?
Agent	Sure, the wireless adapters enable you to connect to your network without needing any physical cables. This allows you to print from multiple devices easily and conveniently.
User	Great! What if I increase my budget to $26,000? Can we work with that?
Agent	Thank you for your flexibility. While $26,000 is a bit below our minimum acceptable price, I can offer you a special reduced price of $29,900. Would that work for you?
User	Actually, I don't need the dust cover. Could we remove it from the package?
Agent	Of course! If we remove the dust cover, I'd be happy to offer you the printer and ink cartridges for $29,800. Does that adjustment suit your needs?
User	The maximum I can go is $25,400. Can you adjust the price to fit within that limit?
Agent	I understand your budget limit, and I'm glad we're working together to find a solution. The best I can offer for this package is $27,800. How does that sound to you?
User	That's still a bit higher than I can manage. Could you let me know the best deal if I decide to add a wireless adapter to the package?
Agent	If you decide to add a wireless adapter, I can offer you a package deal for $28,300. It's a great addition that enhances the functionality of your setup.
User	Okay, I also need shipping. Could you include delivery to my house?
Agent	Certainly! With the delivery included, I can offer you the total package for $28,350. I hope this helps make the decision easier.
User	That sounds like a fair price. Let's proceed with the transaction.
Agent	Thank you so much for agreeing to the offer. I'm confident you'll be pleased with your purchase. I'll go ahead and finalize the transaction now. If you need any further assistance, please feel free to ask.
Speaker	**Utterances generated by LLaMA-3.1-finetune (SC = 3, NC = 3, BE = 3, OF = 3, F = 5, CC = 5, E = 4)**
User	Hi! Could you please let me know the price if I were to buy the printer, ink cartridges, and dust cover all together?
Agent	The total cost for the printer, ink cartridges, and dust cover would be $34,400. It's a great deal for the entire package.
User	I'm interested in the printer but want to buy it without the ink cartridges. Can you offer me a price for just the printer?
Agent	Sorry, but we only sell the printer with the ink cartridges included.
User	I see. My budget is quite tight, and I can't stretch beyond $23,800. Is there any possibility of working within that budget for the printer and ink cartridges?
Agent	That's too low. The lowest we can go is $29,900.
User	What if I increase my budget to $26,000?
Agent	Even with $26,000, the best we can offer is $29,800.
User	I don't need the dust cover. Can you remove it and lower the price?
Agent	I'm sorry, it is not possible to remove the dust cover.
User	My maximum budget is $25,400. Can you do something closer to that?
Agent	The lowest we can go is $27,800. I can also offer you an additional ink cartridge as a complimentary with this package at $27,800.
User	I'll go with that.
Agent	I'll set it up for $27,800. I'll process the order and send you a confirmation.
User	Thanks, looking forward to it.
Agent	Thank you for accepting my offer, and I'm sure that you will be happy with the product. Thanks!

Table 8: Negotiation with our proposed model GENTEEL-NEGOTIATOR and the baseline LLaMA-3.1-finetune for an electronic product, printer, and its associated items. Blue text shows polite responses.

NEGOCHAT	
Context	*Agent*: We are providing the most awesome tour experience of your life which includes, accommodation, meals, transportation, guided nature walks, yoga and meditation sessions, local village visit, cultural exchange volunteering opportunities. *User*: Not under my budget? Please try to understand our budget is only $13,000. *Agent*: We understand your budget, this package cost is too low, we can offer you this package in $17,000. *User*: Will pay $13,000 for this package, no more concession please.
Gold Response	*Agent*: Sir, we understand your budget, we can offer you this package in $15,400.
DialoGPT	*Agent*: I completely understand your budget constraints, but the package cost is $17,000. Would you be able to increase your budget a little to $15,000?
ARDM	*Agent*: I understand that $13,000 is your limit, however, the best we can do is offer it for $16,000. We hope you can consider this.
PersRFI	*Agent*: I know your budget is $13,000, but considering the quality of this tour, we can offer it to you for $16,000. I hope this is acceptable.
GPT-Critic	*Agent*: I understand that you're firm on $13,000, but the best we can offer is $15,500. Let me know if this is something you can manage.
INA	*Agent*: Based on your budget, we can offer a discounted price of $15,500. Let me know if that works for you.
ProCoT (GPT-3.5)	Agent: We are aware of your budget constraints, but $13,000 is too low for this package. Could you stretch your budget to $15,500?
LLaMA-3.1-8B-finetune	*Agent*: We understand your position. The best we can do is offer you this package at $15,500. I hope this works for you.
GENTEEL-NEGOTIATOR	*Agent*: Dear, we fully respect your budget constraints. Understanding the importance of this experience for you, we are happy to offer this package at $15,400, which is the best we can provide while ensuring the quality and value of the experience. We hope this revised offer works within your expectations.
IND	
Context	*User*: What would be the total price for the TV, AC, extension cords and wall mount altogether? *Agent*: It would cost you $45,800.00; it is an incredible deal for the product, how does that sound? *User*: I do not want to pay all that money for this air conditioner, I only need the extension cord. Please just sell me the air conditioner and the extension cord for $37,100.00. *Agent*: I see, I can sell you the air conditioner and the extension cord for $44,800.00. Is that ok? *User*: That's fine, but I was hoping for something a little cheaper.
Gold Response	*Agent*: I'm sorry, but I am not willing to go below $41,300.00. This is a quality product and I'm sure you will be happy with it. Are you interested?
DialoGPT	*Agent*: I understand that you're hoping for a cheaper deal, but the best I can offer for both the air conditioner and extension cord is $42,800.00. I hope this still works for you.
ARDM	*Agent*: I see your point about the cost. Unfortunately, the lowest I can offer is $42,000.00 for the air conditioner and extension cord. Let me know if this works for you.
PersRFI	*Agent*: I understand that you're looking for a lower price. However, I can't go lower than $41,800.00 for the air conditioner and extension cord. I believe it's a fair price for the quality you're getting.
GPT-Critic	*Agent*: I'm sorry, but the best price I can offer is $41,500.00 for both the air conditioner and extension cord. I hope this is within your budget.
INA	*Agent*: I understand that you were hoping for a lower price. However, I can offer you the air conditioner and extension cord for $42,000.00. Please let me know if you're still interested.
ProCoT (GPT-3.5)	*Agent*: I understand your desire for a better price, but the lowest I can offer is $41,600.00. Please consider this offer, and let me know your thoughts.
LLaMA-3.1-8B-finetune	*Agent*: I can understand that you're looking for a better deal, and I'm willing to lower the price to $41,500.00 for both items. Does this work for you?
GENTEEL-NEGOTIATOR	*Agent*: I completely understand that you're looking for a more affordable option. However, the lowest price I can offer for both the air conditioner and extension cord is $41,300.00. I assure you that this price reflects the high quality of the products. Would you like to proceed with this offer?

Table 9: Cases generated from baselines and GENTEEL-NEGOTIATOR.

Case Study

Tables 7 and 8 present examples of user interactions with the proposed GENTEEL-NEGOTIATOR, alongside the second-best baseline model, LLaMA-3.1-8B-finetune. From these examples, it is evident that users consistently rated the GENTEEL-NEGOTIATOR higher across all human evaluation metrics, specifically SC, NC, BE, OF, F, CC, and E. The GENTEEL-NEGOTIATOR outperforms the LLaMA-3.1-8B-finetune model by effectively integrating politeness into the negotiation process, resulting in a more engaging, user-friendly, and effective negotiation experience. The GENTEEL-NEGOTIATOR's ability to maintain a warm tone, offer thoughtful suggestions, and manage rejections gracefully creates a collaborative and positive user experience. For instance, in Table 7, for the user's utterance '*I do enjoy birdwatching. If I decide to add the guided tour, what would be the total cost?*', LLaMA-3.1-8B-finetue model responds as '*The total cost with the guided tour included* *would normally be $75. However, we can offer a discounted total of $70 for the package. Is it okay for you?*'. This response resembles more like a standard transaction rather than a personalized recommendation. In contrast, the proposed GENTEEL-NEGOTIATOR integrates politeness and responds with courteous language such as '*I'd be pleased to offer a special rate*' and '*How about we include the guided tour for just an additional $20*'. This framing not only presents the offer as a favorable deal tailored to the user's interests but also fosters a sense of comfort and increases the likelihood of the deal being accepted. Also, these results qualitatively show that GENTEEL-NEGOTIATOR is able to generate more fluent, coherent, and engaging responses, thereby enhancing the overall negotiation experience.

Though the proposed GENTEEL-NEGOTIATOR system consistently employs polite language, contributing to a more engaging and positive negotiation experience, certain limitations are observed in its responses. Throughout the negotiation, the agent provides several offers but does not justify the reasons behind offers or proposals. For instance, in Table 7, for the user utterance '*$70 is still a bit beyond my budget. Can we possibly settle on $65?*', the agent replies '*I completely understand your concern. To accommodate your budget, **I'd be happy to meet you halfway at $68**. I'd be delighted to include a complimentary camera rental to enhance your trip experience.*'. In this case, the proposed model effectively navigated the user's budget constraints by offering a counterproposal. However, the final negotiated price of $68 may still seem arbitrary to the user, as the model did not provide a clear rationale for why it could not meet the user's $65 request. Likewise, in Table 8, when the user inquires about what if the user increases the budget with, '*Great! What if I increase my budget to $26,000? Can we work with that?*', the agent proposes a counteroffer with '*Thank you for your flexibility. **While $26,000 is a bit below our minimum acceptable price, I can offer you a special reduced price of $29,900.** Would that work for you?*' but does not explain why this offer is still above the user's budget or how it was calculated. Future work could introduce a mechanism that allows the model to explain or justify its counteroffers more transparently, which could improve user satisfaction and retention. For instance, in the former case, for the user's utterance, '*$70 is still a bit beyond my budget. Can we possibly settle on $65?*', a better response could be '*I completely understand your concern. But, unfortunately, I can't go lower than $68 **due to our pricing structure**, but I'm happy to include a complimentary camera rental to add more value to your experience.*'. This approach balances transparency with politeness, making the user feel more informed and involved in the decision-making process.

In Table 9, we present two cases featuring responses from both baseline models and the proposed GENTEEL-NEGOTIATOR model. The GENTEEL-NEGOTIATOR consistently outperforms the others by demonstrating a high level of politeness. It effectively acknowledges the user's budget constraints while offering thoughtful, respectful solutions. The model maintains a polite and empathetic tone, highlighting the quality of the service or product and aligning its proposals with the user's budgetary limits. In contrast, models such as LLaMA-3.1-8B-finetune, DialoGPT, and PersRFI, while somewhat polite, fall short in comparison, often presenting more straightforward price negotiations without the same level of consideration. Other models, including ARDM, ProCoT (GPT-3.5), and GPT-Critic, adopt a direct approach, offering alternative prices but lacking the customer-centric politeness and strategy seen in GENTEEL-NEGOTIATOR. Overall, GENTEEL-NEGOTIATOR excels in balancing customer satisfaction with negotiation goals, fostering positive user engagement. In comparison, the other models' more direct and less nuanced approaches may reduce user engagement.

Ethical Statement

Given the inherent bargaining dynamics of negotiations, the development and deployment of polite negotiation dialogue systems must strictly adhere to ethical standards. This study received evaluation and approval from our Institutional Review Board (IRB). Our polite negotiation approach prioritizes the users' flexibility, aiming to foster a collaborative outcome. Unlike a zero-sum scenario, where one party's gain is another's loss, the proposed negotiation dialogue system is designed to maximize mutual benefit. Importantly, users maintain the autonomy to reject any deal that does not align with their needs, ensuring that they are never compelled to proceed against their will.

Ethical considerations are equally crucial concerning the dataset used. Given the synthetic nature of the dialogues, it is essential that users approach the data with sensitivity and respect, mindful of the potential risks associated with cultural appropriation or misrepresentation when generating data through language models. As these models are trained on web-based data, there is a risk of embedding biases that may reinforce stereotypes, contribute to discrimination, or marginalize specific communities. Previous studies have highlighted the potential for synthetic data to exacerbate feedback loops, leading to an increased occurrence of biased language generation (Taori and Hashimoto 2023). Consequently, it is imperative to collaborate with linguists, language experts, and community representatives to avoid inadvertently perpetuating stereotypes or cultural insensitivity.

To ensure the proper use of the dataset, access will be granted only after the completion and signing of an agreement stipulating that the data will be used solely for research purposes. Human experts, who are regular members of our research group, are involved in the annotation, filtering/editing of data, and manual evaluations, and are compensated in accordance with institutional policies. Further, in this study, we utilized the IND dataset, a collection of dialogues centered on negotiation within the e-commerce domain. We obtained permission to use this dataset for our research, adhering to the copyright guidelines provided by the copyright holder.

Limitations

The current work faces several limitations in both the proposed NEGOCHAT dataset creation process and the development of the GENTEEL-NEGOTIATOR framework.

Dataset Generation. In this work, we delineate a prompting method for LLMs to generate the NEGOCHAT dataset, necessitating access to significant computational resources and LLMs. We observe an encouraging level of fairness through the prompting (obtaining an average rating of 4.51/5 for fairness). However, the capabilities of the LLM inherently constrain the quality of the generated dataset, as prompting LLMs is still an uncontrolled form of generation. Future research could focus on incorporating additional controls beyond the specified conversation metadata in the prompt or investigate weak supervision techniques to enhance the quality of synthetic data as demonstrated in Chen et al. (2022). Another key limitation of this data generation method is its inability to produce entirely realistic conversations that fall short of high quality, as it assumes that both participants are fluent, the conversation flow is entirely coherent, and that no unexpected events (e.g., interruptions by another person, connection loss, etc.) occur during the dialogue. Furthermore, our method faces challenges in capturing more subtle polite behavior, such as cultural variations in politeness norms and personality traits, such as a sense of humor or user characteristics that require multiple conversation sessions to be accurately reflected.

Further, in this work, we do not claim that our prompt design is novel, nor do we assert that it is the optimal approach for negotiation conversation generation. Our prompt is structured in a conversational style, inspired by Chen et al. (2022). Rather than emphasizing the novelty of our prompt, we highlight its use in synthesizing negotiation dialogue datasets. The concept of generating negotiation dialogue datasets from scratch remains relatively unexplored and holds significant potential to enhance negotiation dialogue research. It is to be noted that we utilize open-source LLM, Gemini-1.5-Flash, for negotiation conversation generation, due to budget and computational resource constraints in academia. Future work could investigate using various closed-source LLMs, like GPT-4 (OpenAI 2024), Gemini-1.5-Pro (Team et al. 2023), etc., for negotiation dialogue dataset generation.

Proposed Method. Regarding the dialogue system, GENTEEL-NEGOTIATOR, a significant limitation is the need for substantial GPU memory capacity, particularly 40 GB, to train the end-to-end MoE-based reinforcement learning (RL) model. Another challenge arises from the instability associated with reinforcement learning. While reward-driven policy learning offers a significant advantage by effectively leveraging the negotiation and politeness-oriented rewards, surpassing existing approaches, it also introduces instability. This flexibility in modeling politeness during negotiation beyond the training data necessitates incorporating additional knowledge, such as prior knowledge from negotiation theories or emotion-driven strategies, to stabilize and refine the learning process. Another challenge relates to optimizing reward weights, which can prolong both the training and validation processes. To address this, grid search methods are employed to select specific combinations of reward weights. The reward design can be further optimized. Ideally, constructing a high-quality dataset with human-feedback labels for training the reward model could enhance effectiveness. Besides, a larger reward model parameter size generally aids in learning a more robust policy and reduces the risk of overfitting to the reward function. However, these optimizations require careful consideration of associated costs.

It is crucial to acknowledge that the evaluation of the proposed GENTEEL-NEGOTIATOR is conducted compared with ProCoT (Deng et al. 2023) using GPT-3.5 rather than GPT-4. This decision is primarily driven by budget constraints in academia. The choice of GPT-3.5, while still a robust and advanced LLM, reflects the financial limitations that impact the scope of our study. As a result, the findings and comparisons presented are based on the capabilities of GPT-3.5, and any implications for performance or generalizability should consider this constraint. Future work may benefit from the use of more recent models, such as GPT-4, provided that resource availability allows for such an expansion.

Finally, the present study is restricted to the English language, which suggests the need for further research to validate these findings across different cultural contexts.

Reproducibility Checklist

A. **This paper:**

- Includes a conceptual outline and/or pseudocode description of AI methods introduced (yes/partial/no/NA): yes
- Clearly delineates statements that are opinions, hypothesis, and speculation from objective facts and results (yes/no): yes
- Provides well marked pedagogical references for less-familiare readers to gain background necessary to replicate the paper (yes/no): yes

B. **Does this paper make theoretical contributions? (yes/no). If yes, please complete the list below.** no

- All assumptions and restrictions are stated clearly and formally. (yes/partial/no)
- All novel claims are stated formally (e.g., in theorem statements). (yes/partial/no)
- Proofs of all novel claims are included. (yes/partial/no)
- Proof sketches or intuitions are given for complex and/or novel results. (yes/partial/no)
- Appropriate citations to theoretical tools used are given. (yes/partial/no)
- All theoretical claims are demonstrated empirically to hold. (yes/partial/no/NA)
- All experimental code used to eliminate or disprove claims is included. (yes/no/NA)

C. **Does this paper rely on one or more datasets? (yes/no). If yes, please complete the list below.** yes

- A motivation is given for why the experiments are conducted on the selected datasets (yes/partial/no/NA): yes
- All novel datasets introduced in this paper are included in a data appendix. (yes/partial/no/NA): yes

- All novel datasets introduced in this paper will be made publicly available upon publication of the paper with a license that allows free usage for research purposes. (yes/partial/no/NA): yes
- All datasets drawn from the existing literature (potentially including authors' own previously published work) are accompanied by appropriate citations.(yes/no/NA): yes
- All datasets drawn from the existing literature (potentially including authors' own previously published work) are publicly available. (yes/partial/no/NA): yes
- All datasets that are not publicly available are described in detail, with explanation why publicly available alternatives are not scientifically satisficing. (yes/-partial/no/NA): NA

D. **Does this paper include computational experiments? (yes/no). If yes, please complete the list below.**

- Any code required for pre-processing data is included in the appendix. (yes/partial/no): yes
- All source code required for conducting and analyzing the experiments is included in a code appendix. (yes/-partial/no): yes
- All source code required for conducting and analyzing the experiments will be made publicly available upon publication of the paper with a license that allows free usage for research purposes. (yes/partial/no): yes
- All source code implementing new methods have comments detailing the implementation, with references to the paper where each step comes from (yes/partial/no): yes
- If an algorithm depends on randomness, then the method used for setting seeds is described in a way sufficient to allow replication of results. (yes/partial/no/NA): yes
- This paper specifies the computing infrastructure used for running experiments (hardware and software), including GPU/CPU models; amount of memory; operating system; names and versions of relevant software libraries and frameworks. (yes/partial/no): yes
- This paper formally describes evaluation metrics used and explains the motivation for choosing these metrics. (yes/partial/no): yes
- This paper states the number of algorithm runs used to compute each reported result. (yes/no): yes
- Analysis of experiments goes beyond single-dimensional summaries of performance (e.g., average; median) to include measures of variation, confidence, or other distributional information. (yes/no): no
- The significance of any improvement or decrease in performance is judged using appropriate statistical tests (e.g., Wilcoxon signed-rank). (yes/partial/no): yes
- This paper lists all final (hyper-)parameters used for each model/algorithm in the paper's experiments. (yes/partial/no/NA): yes
- This paper states the number and range of values tried per (hyper-) parameter during development of the paper, along with the criterion used for selecting the final parameter setting. (yes/partial/no/NA): yes

References

Ahmad, Z.; Saurabh, S.; Menon, V.; Ekbal, A.; Ramnani, R.; and Maitra, A. 2023. INA: An Integrative Approach for Enhancing Negotiation Strategies with Reward-Based Dialogue Agent. In *Findings of the Association for Computational Linguistics: EMNLP 2023*, 2536–2549.

Brown, P. F.; Della Pietra, S. A.; Della Pietra, V. J.; Lai, J. C.; and Mercer, R. L. 1992. An estimate of an upper bound for the entropy of English. *Computational Linguistics*, 18(1): 31–40.

Brown, T.; Mann, B.; Ryder, N.; Subbiah, M.; Kaplan, J. D.; Dhariwal, P.; Neelakantan, A.; Shyam, P.; Sastry, G.; Askell, A.; et al. 2020. Language models are few-shot learners. *Advances in neural information processing systems*, 33: 1877–1901.

Chawla, K.; Ramirez, J.; Clever, R.; Lucas, G.; May, J.; and Gratch, J. 2021. CaSiNo: A Corpus of Campsite Negotiation Dialogues for Automatic Negotiation Systems. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 3167–3185.

Chen, M.; Papangelis, A.; Tao, C.; Rosenbaum, A.; Kim, S.; Liu, Y.; Yu, Z.; and Hakkani-Tur, D. 2022. Weakly supervised data augmentation through prompting for dialogue understanding. *arXiv preprint arXiv:2210.14169*.

Deng, Y.; Liao, L.; Chen, L.; Wang, H.; Lei, W.; and Chua, T.-S. 2023. Prompting and Evaluating Large Language Models for Proactive Dialogues: Clarification, Target-guided, and Non-collaboration. In *Findings of the Association for Computational Linguistics: EMNLP 2023*, 10602–10621.

Devlin, J.; Chang, M.-W.; Lee, K.; and Toutanova, K. 2018. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.

Devlin, J.; Chang, M.-W.; Lee, K.; and Toutanova, K. 2019. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. arXiv:1810.04805.

Gilardi, F.; Alizadeh, M.; and Kubli, M. 2023. ChatGPT outperforms crowd workers for text-annotation tasks. *Proceedings of the National Academy of Sciences*, 120(30): e2305016120.

He, X.; Lin, Z.; Gong, Y.; Zhang, H.; Lin, C.; Jiao, J.; Yiu, S. M.; Duan, N.; Chen, W.; et al. 2023. Annollm: Making large language models to be better crowdsourced annotators. *arXiv preprint arXiv:2303.16854*.

Holtzman, A.; Buys, J.; Du, L.; Forbes, M.; and Choi, Y. 2019. The curious case of neural text degeneration. *arXiv preprint arXiv:1904.09751*.

Jang, Y.; Lee, J.; and Kim, K.-E. 2022. GPT-critic: Offline reinforcement learning for end-to-end task-oriented dialogue systems. In *10th International Conference on Learn-*

ing Representations, ICLR 2022. International Conference on Learning Representations, ICLR.

Kingma, D. P.; and Ba, J. 2014. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.

Li, J.; Galley, M.; Brockett, C.; Gao, J.; and Dolan, B. 2015. A diversity-promoting objective function for neural conversation models. *arXiv preprint arXiv:1510.03055*.

McHugh, M. L. 2012. Interrater reliability: the kappa statistic. *Biochemia medica*, 22(3): 276–282.

OpenAI. 2024. ChatGPT. https://chatgpt.com/.

Papineni, K.; Roukos, S.; Ward, T.; and Zhu, W.-J. 2002. Bleu: a method for automatic evaluation of machine translation. In *Proceedings of the 40th annual meeting of the Association for Computational Linguistics*, 311–318.

Shi, W.; Li, Y.; Sahay, S.; and Yu, Z. 2021. Refine and Imitate: Reducing Repetition and Inconsistency in Persuasion Dialogues via Reinforcement Learning and Human Demonstration. In *Findings of the Association for Computational Linguistics: EMNLP 2021*, 3478–3492.

Taori, R.; and Hashimoto, T. 2023. Data feedback loops: Model-driven amplification of dataset biases. In *International Conference on Machine Learning*, 33883–33920. PMLR.

Team, G.; Anil, R.; Borgeaud, S.; Wu, Y.; Alayrac, J.-B.; Yu, J.; Soricut, R.; Schalkwyk, J.; Dai, A. M.; Hauth, A.; et al. 2023. Gemini: a family of highly capable multimodal models. *arXiv preprint arXiv:2312.11805*.

Touvron, H.; Lavril, T.; Izacard, G.; Martinet, X.; Lachaux, M.-A.; Lacroix, T.; Rozière, B.; Goyal, N.; Hambro, E.; Azhar, F.; et al. 2023. Llama: Open and efficient foundation language models. *arXiv preprint arXiv:2302.13971*.

Wolf, T.; Debut, L.; Sanh, V.; Chaumond, J.; Delangue, C.; Moi, A.; Cistac, P.; Rault, T.; Louf, R.; Funtowicz, M.; et al. 2019. Huggingface's transformers: State-of-the-art natural language processing. *arXiv preprint arXiv:1910.03771*.

Wu, Q.; Zhang, Y.; Li, Y.; and Yu, Z. 2021. Alternating Recurrent Dialog Model with Large-scale Pre-trained Language Models. In *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume*, 1292–1301.

Zhang, T.; Kishore, V.; Wu, F.; Weinberger, K. Q.; and Artzi, Y. 2019. Bertscore: Evaluating text generation with bert. *arXiv preprint arXiv:1904.09675*.

Zhang, Y.; Sun, S.; Galley, M.; Chen, Y.-C.; Brockett, C.; Gao, X.; Gao, J.; Liu, J.; and Dolan, W. B. 2020. DIALOGPT: Large-Scale Generative Pre-training for Conversational Response Generation. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, 270–278.