# 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**[1,3]

[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
{priyanshu_2021cs26, rishikant_2101cs66}@iitp.ac.in, mauajama@iitism.ac.in, asif@{iitp,iitj}.ac.in

## Abstract

Developing intelligent negotiation dialogue systems that resolve conflicts and promote equitable, inclusive, and sustainable outcomes is at the forefront of advancing automated negotiation technology for social good. Negotiation involves balancing cooperation and competition to maximize value without causing offense. Using polite language fosters mutual understanding and creates a respectful and collaborative environment essential for successful negotiations in various domains. Considering this, in this paper, we propose a polite negotiation dialogue system, GENTEEL-NEGOTIATOR for social good applications to boost the overall quality of negotiation outcomes. We focus on developing a negotiation dialogue system for two key application areas, namely tourism and e-commerce. We begin by curating a unique negotiation dialogue dataset, NEGOCHAT for tourism. We further enrich the NEGOCHAT and Integrative Negotiation Dataset (IND) for e-commerce with various negotiation strategies. These datasets are then used to develop the GENTEEL-NEGOTIATOR, leveraging the Large Language Model (LLM) and mixture-of-expert (MoE)-based reinforcement learning approach. The proposed MoE-based method employs heuristic experts dedicated to negotiation, politeness, and dialogue coherence to facilitate the learning of diverse semantics by analyzing the dialogue context. A novel reward function with negotiation strategy congruence, politeness, dialogue coherence, and engagingness rewards is designed to guide the policy's learning for generating responses. Automatic and human evaluations on NEGOCHAT and IND datasets validate the effectiveness of GENTEEL-NEGOTIATOR in generating polite responses during negotiation while maintaining conversation goals, including coherence and engagingness.

## Introduction

Negotiation dialogue systems have garnered significant interest in recent years due to their wide-ranging real-world applications (Yamaguchi, Iwasa, and Fujita 2021; Priya et al. 2024b). These systems can potentially revolutionize fields, such as e-commerce (Sree et al. 2023), conflict resolution (Kaur et al. 2022), and customer service (Stock, Petukhova, and Klakow 2022), where effective negotiation is critical. Techniques like reinforcement learning (RL) (Zhao, Xie, and Eskenazi 2019) and using large language models (LLMs) (Chen et al. 2023) have enabled the creation of systems that can simulate human-like negotiation strategies (Lewicki, Barry, and Saunders 2016), making them more adaptable and effective in complex scenarios. The inherent complexity of human negotiation (Morris and Gelfand 2004), which involves a nuanced interplay of negotiation strategies (Schoenfield 1983), and contextual understanding (Horton 2012), presents an intellectually stimulating challenge.



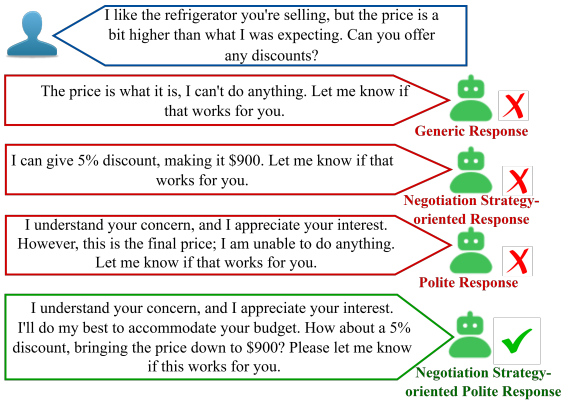


Figure 1: A conversation snippet demonstrating the use of appropriate strategy (*gradual concession-making*) and politeness during negotiation to improve the final outcome.

One critical aspect of successful negotiation in human and automated systems is the strategic use of politeness (Lee, Mason, and Malcomb 2021). Politeness (Terada, Okazoe, and Gratch 2021) plays a vital role in maintaining a cooperative atmosphere and fostering mutual respect between negotiating parties. In negotiation dialogue systems, incorporating politeness strategies (Brown and Levinson 1987) contributes to mitigating conflicts by reducing the likelihood of offending opponents, thereby maintaining or enhancing sociopsychological closeness. Figure 1 depicts a conversation snippet showcasing the relevance of negotiation strategy and politeness modeling during negotiation.

Influenced by the importance of strategies and politeness in negotiation, we introduce GENTEEL-NEGOTIATOR, a novel polite negotiation dialogue system to infuse politeness into negotiation conversations while being coherent and engaging using an LLM-enhanced Mixture-of-Expert (MoE)-

based RL approach. GENTEEL-NEGOTIATOR includes (a) an LLM for learning diverse semantics from the dialogue context; (b) three experts *viz.* negotiation, politeness, and keyterm experts to infuse politeness during negotiation while ensuring dialogue coherence; (c) an RL-based agent to strategically select experts based on an expert determination policy for generating responses. To generate engaging, coherent, polite responses with pertinent strategy, we devise a new reward function with negotiation strategy congruence, politeness, dialogue coherence, and engagingness rewards.

These automated negotiation dialogue systems are highly beneficial in high-volume sectors like e-commerce and tourism, where routine negotiations are common. These systems enhance operational efficiency and support the United Nations' Sustainable Development Goals (UNWTO 2024). To advance research in this area, we introduce NEGOCHAT, a new negotiation dialogue dataset for the tourism domain, generated through prompting the Gemini (Team et al. 2023) LLM. Further, we annotate both NEGOCHAT and the Integrative Negotiation Dataset (IND) (Ahmad et al. 2023) for e-commerce with novel negotiation strategies using ChatGPT (OpenAI 2024) accompanied by human intervention, thereby creating a comprehensive resource for analyzing negotiation behaviors in these domains.

In summary, the key contributions are outlined as follows: (i) Investigate the use of politeness by the dialogue agent on negotiation outcomes. To the best of our knowledge, this study pioneers the strategic modeling and analysis of politeness effects within negotiation conversations; (ii) Introduce NEGOCHAT, a new negotiation dialogue dataset for the tourism domain, generated using LLM with minimal manual intervention. We then automatically annotate NEGOCHAT and IND datasets with negotiation strategies using LLM prompting and human-in-the-loop techniques; (iii) Present GENTEEL-NEGOTIATOR, an LLM-enhanced MoE-based RL model to incorporate politeness into negotiation conversations while ensuring coherence and user engagement; (iv) Design a novel reward function consisting of negotiation strategy congruence, politeness, dialogue coherence, and engagingness rewards to generate engaging, coherent, and polite responses during negotiation[1].

## Related Work

Negotiation has been actively studied in diverse research areas, including game theory (Nash Jr 1950), economics (Carnevale and Pruitt 1992), and psychology (Adair, Okumura, and Brett 2001). Recently, the human-agent negotiation dialogue systems have become the center of automated negotiation literature (Fu et al. 2023). There has been a growing emphasis on employing LLMs (Fu et al. 2023; Abdelnabi et al. 2023) to develop automated negotiation agents. Moreover, Mixtures-of-Experts (MoEs) (Cai et al. 2024) that utilize a collection of $n$ "expert" sub-networks have become integral to the design of LLMs (Obando-Ceron et al. 2024).

[1]Dataset, code, and appendix are available at https://github.com/priyanshu-profile/GENTEEL-NEGOTIATOR/; https://www.iitp.ac.in/~ai-nlp-ml/resources.html#GENTEEL-NEGOTIATOR.

In recent times, the integration of politeness into dialogue systems has emerged as a critical area of research, reflecting its importance in enhancing user experience and achieving successful interactions (Silva, Semedo, and Magalhães 2022; Mishra, Priya, and Ekbal 2023a; Priya et al. 2024a). Politeness contributes to considerate and positive relationships (Golchha et al. 2019; Priya et al. 2023; Mishra, Priya, and Ekbal 2023b) and has been proven essential for effective negotiation (Terada, Okazoe, and Gratch 2021). Politeness, as a key social behavior, facilitates the development of congenial and respectful relationships, fosters mutual trust and understanding, and assists in developing rapport and cooperation during negotiation (Maaravi, Idan, and Hochman 2019).

While the existing studies have made significant strides in developing negotiation dialogue systems that model the opponent's strategy, integrating other human-like aspects, particularly politeness, into these systems remains largely unexplored. We introduce a new LLM-enhanced MoE-based RL approach that combines the advantages of LLMs, MoEs, and RL to develop a robust polite negotiation dialogue system. Further, our proposed system is designed to be scalable and adaptable across various domains.

## Dataset

We conduct experiments using two negotiation dialogue datasets, *viz.* the newly curated NEGOCHAT dataset and the Integrative Negotiation Dataset (IND) (Ahmad et al. 2023).

**NEGOCHAT Dataset Preparation.** The proposed NEGOCHAT dataset comprises dialogues between the agent and the users, specifically centered on negotiation within the tourism domain. This dataset is designed to capture the complexities involved in negotiating various components of tourism packages, including price, destinations, transportation, and additional amenities and services. By detailing these intricate interactions, NEGOCHAT provides a comprehensive resource for analyzing and enhancing dialogue systems that aim to handle real-world negotiation scenarios in the tourism sector. The dataset is developed by utilizing the extensive knowledge embedded in the LLMs to reduce dependence on expensive and limited human resources. Specifically, the dataset is generated through few-shot prompting of `Gemini-1.5-Flash` (Team et al. 2023) model, with subsequent human oversight to ensure quality control. The entire user-agent dialogue dataset creation process consists of two key stages: (a) Drafting Sample Dialogues, and (b) Generating Dialogues via Prompting.

**(a) Drafting Sample Dialogues**. To draft sample dialogues, we initially gather information about various travel packages by referring to several well-known travel websites. Based on this data, we develop 10 distinct travel packages. Each package includes the package name, a detailed description, and information on the associated aspects, amenities, services, and pricing. These details are then used to create 50 negotiation dialogues (5 dialogues for each package) between the user and the agent using a Wizard-of-Oz approach (Kelley 1984), where one human subject assumes the role of the user, and the other acts as the agent. Along with each dialogue, we maintain a brief "conversation metadata", which

| Type | Negotiation Strategy | Definition |
|---|---|---|
| Problem-solving | Active listening | Refers to comprehending the needs and perspectives of the opponent to resolve misunderstandings, build rapport, and develop mutually beneficial solutions. |
| | Leverage information | Refers to the use of credentials, past successes, and a consistent record of integrity and professionalism to establish credibility, gain the opponent's trust, and make the proposal more persuasive. |
| | Logrolling | Refers to conceding on lower-priority issues to gain concessions on higher-priority issues from the opponent to facilitate a mutually beneficial agreement through trade-offs that address both parties' key interests and preferences. |
| | Expanding-the-pie | Refers to enhancing the overall value of the deal during negotiation to transition from a zero-sum game to an integrative approach that seeks mutual benefit and collaborative value creation. |
| Concession-making | Gradual concession-making | Involves offering small, incremental concessions to show willingness to compromise, maintain negotiation leverage, assess the opponent's commitment, and encourage reciprocal concessions. |
| | Large initial concession-making | Involves making a substantial concession early in the negotiation to initiate the process, show goodwill, or create urgency, thereby encouraging the other party to make their own concessions. |
| | Patterned concession-making | Involves making concessions in a predictable and structured way, such as progressively reducing range/frequency to signal nearing limits of one party and encourage the opponent to agree before concessions become minimal. |
| | No strategy | Refers to utterances that do not employ any specific negotiation strategy. |

Table 1: The definition of different negotiation strategies. Examples are given in "Dataset Details" section of the appendix.

includes a *package information* and *background information* for two interlocutors. The *package information* represents the package name and corresponding description. The *background information* includes more fine-grained details relevant to the package, such as a list of included aspects, amenities, and services with their descriptions. These sample dialogues are created by three human subjects under the supervision of a domain specialist with business and sales background to ensure precision and relevance within specified context. These subjects possess Ph.D. degrees in Linguistics and have extensive knowledge of dialogue and negotiation concepts. 'Guidelines for Drafting Sample Dialogues' are provided in "Dataset Details" section of the appendix.

**(b) Generating Dialogues via Prompting**. The created sample dialogues are utilized to prompt `Gemini-1.5-Flash` LLM to generate synthetic dialogues in a few-shot setup. To finalize the prompt, we experimented with four manually designed prompts, including natural language instructions, few shot exemplars (three randomly selected conversation metadata and their corresponding dialogues from the aforementioned pool of sample dialogues), and the target dialogue metadata. For each prompt, we generate 20 dialogues by prompting Gemini with Top-*p* sampling (Holtzman et al. 2019) with $p = 0.95$ and temperature $\tau = 1.0$. These generated dialogues are then manually evaluated by the same three human subjects for quality in terms of negotiation efficacy on a scale of 1-3 (1-low, 2-moderate, 3-high). An inter-subject Kappa (McHugh 2012) agreement ratio of 81.6% is observed among the subjects. The prompt generating the highest number of dialogues with a score of 3 is selected as the final prompt. This prompt is then used to prompt the Gemini model for the generation of the entire dialogue dataset. 'Dialogue Filtering and Quality Assessment' and a sample dialogue are given in "Dataset Details" section of the appendix.

**Integrative Negotiation Dataset (IND).** The IND dataset (Ahmad et al. 2023) contains integrative negotiation dialogues focused on the e-commerce domain. The dialogues in IND primarily encompass negotiations on 10 distinct electronic items and their corresponding accessories.

**Dataset Annotation Scheme.** Negotiation conversations require carefully crafted strategies to assess and select from various potential actions effectively. To achieve successful negotiation and foster collaborative outcomes, the negotiators must be adept at adapting their negotiation strategies to the prevailing circumstances. Considering this, we devise a set of eight negotiation strategies arranged in a hierarchy to reflect the negotiation behavior of the involved parties. These strategies are informed by negotiation theory (Bazerman 1994) and a preliminary analysis of 40 dialogues each from the NEGOCHAT and IND datasets. The above-mentioned three human subjects independently analyzed and labeled the dialogues, discussed discrepancies, and refined the strategies accordingly. The $\kappa$ (Fleiss 1971) score in the range of $0.3 < \kappa < 0.7$ for all categories indicates moderate to fair inter-subject agreement according to (McHugh 2012). Each dialogue utterance is categorized into three types: problem-solving, concession-making, and no-strategy. Table 1 provides definitions for the strategies under these categories. Dataset annotation procedure and dataset statistics are given in the "Dataset Details" section of the appendix.

## Methodology

The overall architecture of the proposed GENTEEL-NEGOTIATOR is depicted in Figure 2.

**Dialogue Encoder.** The dialogue encoder is implemented with LLaMA-3.1-8B-Instruct (Touvron et al. 2023). For a given the dialogue context, $C = \{u_1, a_1, ..., u_{m-1}, a_{m-1}\}$ (an alternating sequence of $(m-1)$ utterances between user $(u)$ and dialogue agent $(a)$), and the target utterance $u_m$, the goal is to generate the response $a_m (= r)$. We concatenate $C$ and $u_m$ and prepend a $[CLS]$ token to form the input sequence, $X = [CLS] \oplus C \oplus u_m$, which is fed into the dialogue encoder to obtain hidden state $\mathcal{H}_X$, with the $[CLS]$ token's representation designated as $h_X$.

**Negotiation Experts.** To effectively adapt negotiation strategies and secure win-win outcomes, we integrate negotiation experts with both contextual and future strategy predictions. Each utterance in the dialogue is annotated with a negotiation strategy label, which serves as the supervised label for the negotiation strategy prediction task. We transform the contextual negotiation expert using the MLP layer:

$$\mathcal{H}_X^{ctx-nego} = \text{MLP}_{nego}(\mathcal{H}_X) \quad (1)$$

We then project the $[CLS]$ representation $h_X^{ctx-nego}$ of the

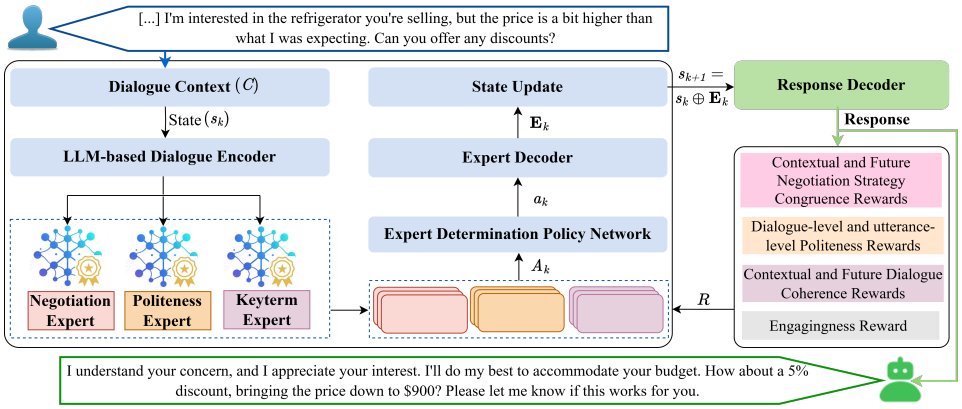


Figure 2: Architecture of the proposed polite negotiation dialogue system - GENTEEL-NEGOTIATOR.

negotiation expert to predict the negotiation strategy,

$$\mathcal{P}^{nego} = \text{softmax}(W^{nego} h_X^{ctx-nego}) \quad (2)$$

which is guided by the negotiation strategies accumulated in the $\mathcal{E}^{\text{nego}^*}$ set of the negotiator's last utterance in the dialogue context by applying cross-entropy loss:

$$\mathcal{L}^{ctx-nego} = -\frac{1}{|\mathcal{E}^{\text{nego}^*}|} \sum_{i=1}^{|\mathcal{E}^{\text{nego}^*}|} \log \mathcal{P}^{nego}(\mathcal{E}_i^{\text{nego}^*}) \quad (3)$$

For future negotiation experts, we employ the same approach to compute $\mathcal{L}^{ftr-nego}$ loss and train it to predict the negotiation strategy of the negotiator's future utterance. This enables the negotiation experts to learn diverse negotiation features through $\mathcal{L}_{nego}$ loss: $\mathcal{L}_{nego} = \mathcal{L}^{ctx-nego} + \mathcal{L}^{ftr-nego}$.

**Politeness Experts.** To monitor possible transitions in the negotiator's polite behavior, politeness experts are associated with contextual and future negotiators' politeness strategy predictions. We use linguistic features, indicative of politeness drawn from Brown and Levinson (1987), which cover two main types of politeness strategies: *positive* and *negative*. Following Danescu-Niculescu-Mizil et al. (2013), we extract $N$ detailed politeness strategies for each utterance in the dataset by pattern matching on the dependency parses of utterances. Since politeness strategies are often the linguistic markers (e.g., *hey*, *thank*, *understand*, etc.), we identify the politeness strategy of each word as positive or negative according to its linguistic feature (e.g. hedges, direct question, etc.). The high-frequency categories are used as supervised labels for politeness strategy prediction task.

We split contextual politeness experts into two types: positive and negative politeness experts, by passing them through two distinct MLP layers, which transform $\mathcal{H}_X$ into $\mathcal{H}_{X,pos}^{ctx-pol}$ and $\mathcal{H}_{X,neg}^{ctx-pol}$, respectively.

$$\mathcal{H}_{X,pos}^{ctx-pol} = \text{MLP}_{pos-pol}(\mathcal{H}_X),\ \mathcal{H}_{X,neg}^{ctx-pol} = \text{MLP}_{neg-pol}(\mathcal{H}_X) \quad (4)$$

We utilize $[CLS]$ representations $h_{X,pos}^{ctx-pol}$ and $h_{X,neg}^{ctx-pol}$ derived from positive and negative experts to identify positive politeness and negative politeness strategy, respectively:

$$\mathcal{P}^{pos-pol} = \text{softmax}(W^{pos-pol} h_{X,pos}^{ctx-pol}) \quad (5)$$

$$\mathcal{P}^{neg-pol} = \text{softmax}(W^{neg-pol} h_{X,neg}^{ctx-pol}) \quad (6)$$

which is guided by positive and negative politeness strategy accumulated in $\mathcal{E}_{\text{pos}}^{\text{pol}^*}$ and $\mathcal{E}_{\text{neg}}^{\text{pol}^*}$ sets of the negotiator's last utterance in dialogue context by applying cross-entropy loss:

$$\mathcal{L}_{pos}^{ctx-pol} = -\frac{1}{|\mathcal{E}_{\text{pos}}^{\text{pol}^*}|} \sum_{i=1}^{|\mathcal{E}_{\text{pos}}^{\text{pol}^*}|} \log \mathcal{P}^{pos-pol}(\mathcal{E}_i^{\text{pol}^*}) \quad (7)$$

$$\mathcal{L}_{neg}^{ctx-pol} = -\frac{1}{|\mathcal{E}_{\text{neg}}^{\text{pol}^*}|} \sum_{i=1}^{|\mathcal{E}_{\text{neg}}^{\text{pol}^*}|} \log \mathcal{P}^{neg-pol}(\mathcal{E}_i^{\text{pol}^*}) \quad (8)$$

It is important to acknowledge that individuals may interpret and perceive the level of politeness in an utterance differently based on their cognitive differences (Escandell-Vidal 1996; Holtgraves 2005). For future politeness experts, we employ the same approach to compute the losses, $\mathcal{L}_{pos}^{ftr-pol}$ and $\mathcal{L}_{neg}^{ftr-pol}$, and then train them to predict the politeness of the negotiator's future utterance (i.e., the subsequent utterance). This enables the politeness experts to assimilate diverse politeness-related characteristics through $\mathcal{L}_{pol}$ loss: $\mathcal{L}_{pol} = \mathcal{L}_{pos}^{ctx-pol} + \mathcal{L}_{neg}^{ctx-pol} + \mathcal{L}_{pos}^{ftr-pol} + \mathcal{L}_{neg}^{ftr-pol}$.

**Keyterm Experts.** To ensure dialogue coherence, keyterm experts are linked with keyterm predictions that help maintain coherence with both contextual and future utterances. For this, we create a bidirectional polite keyterm graph $\mathcal{G}$ (an example is given in "Bidirectional Polite Keyterm Graph" section in the appendix), which also aids in designing coherence rewards. We extract the pertinent keyterms from each utterance within the corpus and identify their politeness strategy through pattern matching on the dependency parse tree of utterances (Danescu-Niculescu-Mizil et al. 2013). The pointwise mutual information (PMI) (Church and Hanks 1990) is employed to create bidirectional edges that reflect the relationships between keyterm pairs. Specifically, the forward edge represents keyterm pairs retrieved from context and response, while the backward edge pertains to pairs drawn from future utterances and responses. Further, positive edges are created to denote keyterms associated with positive politeness strategies, while negative edges correspond to those indicative of negative politeness strategies. Each head vertex then selects tail vertices with the highest PMI scores to form connections. The vertices of $\mathcal{G}$ are utilized as supervised labels for keyterm prediction task.

Contextual keyterm experts undergo a similar transformation as politeness experts. Their $[CLS]$ representations,

$h^{ctx-kt}_{X,pos}$ and $h^{ctx-kt}_{X,neg}$ can be derived from the positive and negative keyterm experts, $\mathcal{H}^{ctx-kt}_{X,pos}$ and $\mathcal{H}^{ctx-kt}_{X,neg}$, respectively. We deduce the one-hop neighbors of contextual keyterms from 'forward-positive' and 'forward-negative' relations in $\mathcal{G}$ to improve insights for target keyterms in gold response. We employ attention mechanisms (Bahdanau, Cho, and Bengio 2014) to acquire fused embeddings $\mathcal{E}^{ctx-kt}_{pos}$ and $\mathcal{E}^{ctx-kt}_{neg}$:

$$\mathcal{E}^{ctx-kt}_{pos} = \text{Attention}(h^{ctx-kt}_{X,pos}, \mathtt{E}^{ctx-kt}_{pos}) \quad (9)$$

$$\mathcal{E}^{ctx-kt}_{neg} = \text{Attention}(h^{ctx-kt}_{X,neg}, \mathtt{E}^{ctx-kt}_{neg}) \quad (10)$$

where, $\mathtt{E}^{ctx-kt}_{pos}$ and $\mathtt{E}^{ctx-kt}_{neg}$ are embedding matrices for positive and negative neighbors, respectively, sharing parameters with dialogue encoder. We further concatenate $\mathcal{E}^{ctx-kt}_{pos}$ and $\mathcal{E}^{ctx-kt}_{neg}$ with $\mathcal{H}^{ctx-kt}_{X,pos}$ and $\mathcal{H}^{ctx-kt}_{X,neg}$, respectively, at token level and then employ MLP layers to integrate them, yielding keyterm-enhanced experts, $\mathcal{H}^{ctx-kt}_{X,pos-kt}$ and $\mathcal{H}^{ctx-kt}_{X,neg-kt}$:

$$\mathcal{H}^{ctx-kt}_{X,pos-kt}[i] = \text{MLP}(\mathcal{H}^{ctx-kt}_{X,pos}[i] \oplus \mathcal{E}^{ctx-kt}_{pos} \quad (11)$$

$$\mathcal{H}^{ctx-kt}_{X,neg-kt}[i] = \text{MLP}(\mathcal{H}^{ctx-kt}_{X,neg}[i] \oplus \mathcal{E}^{ctx-kt}_{neg} \quad (12)$$

Further, we utilize the positive and negative keyterms in gold response as supervision to optimize $\mathcal{L}^{ctx-kt}_{pos}$ and $\mathcal{L}^{ctx-kt}_{neg}$ losses using cross-entropy (analogous to politeness strategy prediction task). Similarly, employing multi-hop reasoning on $\mathcal{G}$ (illustrated in "Bidirectional Polite Keyterm Graph" section in appendix) involves sequential traversal through the graph, specifically 'forward $\rightarrow$ forward $\rightarrow$ backward-positive' and 'forward $\rightarrow$ forward $\rightarrow$ backward-negative' paths, to identify keyterms coherent with future utterance. Utilizing the positive and negative keyterms in the future utterance as prediction targets, we optimize keyterm-augmented future keyterm experts through losses, $\mathcal{L}^{ftr-kt}_{pos}$ and $\mathcal{L}^{ftr-kt}_{neg}$. This approach allows keyterm experts to acquire diverse expression-level features through $\mathcal{L}_{kt}$ loss: $\mathcal{L}_{kt} = \mathcal{L}^{ctx-kt}_{pos} + \mathcal{L}^{ctx-kt}_{neg} + \mathcal{L}^{ftr-kt}_{pos} + \mathcal{L}^{ftr-kt}_{neg}$.

**Multi-task Learning of MoE.** To preserve the fundamental semantics of the experts while ensuring their respective diversity, we compute the average of negotiation, politeness, and keyterm experts' representations to obtain $h_{X,exps}$, and align it closely with the sequence representation $h_X$ by minimizing the Mean Squared Error (MSE) loss: $\mathcal{L}_{mse} = \frac{\delta}{D_h}\sum_{i=1}^{D_h}(h_X[i] - h_{X,exps}[i])^2$, where $\delta$ and $D_h$ denote the hyperparameter and dimension of $h_X$, respectively. We then train the multi-task MoE jointly by optimizing $\mathcal{L}_{moe}$ loss: $\mathcal{L}_{moe} = \mathcal{L}_{nego} + \mathcal{L}_{pol} + \mathcal{L}_{kt} + \mathcal{L}_{mse}$.

**MoE-based Reinforcement Learning.** We adopt the conventional reinforcement learning approach (Sutton and Barto 2018) as the foundation.

**State.** We concatenate dialogue context with extracted keyterms to form initial state $s_1 = \{C, C_{kt}\} \in S$. At each step, the prompt token sequence $\mathbf{E}$ produced by the policy-determined expert (action) updates the state. We keep the observed state $s_t \in S$ at the $t^{th}$ step, denoted as $s_t = \{C, \mathbf{E}_1, \ldots, \mathbf{E}_{t-1}\}$, which is then encoded by dialogue encoder to obtain $H_{S,t}$ and $h_{S,t}$. The sequence representations of previous states are concatenated to form the current state embedding $s_t = h_{S,1} \oplus \ldots \oplus h_{S,t}$. If $t < T$ ($T$: maximum number of iterations), $s_t$ is padded with zeros to ensure consistent dimensionality. When $t > 1$, we omit the keyterms $C_{kt}$ as they have already contributed in the first iteration, and the input sequence length is limited by the LLaMA model.

**Action.** At the $t^{th}$ step, the action space $A_t$ is defined by the multi-task experts influenced by the state $s_t$. At this state, the agent determines which expert to select from $A_t$ as the chosen action $a_t$. To achieve this, we use a LLaMA-based dialogue decoder to produce the expert prompt $\mathbf{E}_t$ for $a_t$.

**Policy.** Besides employing dialogue encoder as a semantic encoding policy network, we devise an expert determination policy network using `REINFORCE` algorithm with baseline (Sutton and Barto 2018), which includes an actor-network and a value network. The actor learns a policy $\pi_\phi(a_t, s_t, A_t)$ for selecting the optimal expert action $a_t$ given the current state $s_t$ and action space $A_t$, by producing a probability distribution over the actions in $A_t$. The value network assesses the value $Q_\beta(s_t)$ of state $s_t$ to provide a baseline for `REINFORCE`. The structures of these networks are:

$$\begin{gathered} o_t = \eta\left(\left(\eta\left(s_t W_1\right) W_2\right)\right); \\ \pi_\phi\left(a_t, s_t, A_t\right) = \text{softmax}\left(A_t \odot o_t W_\phi\right); Q_\beta\left(s_t\right) = o_t W_\beta \end{gathered} \quad (13)$$

where $\eta(\cdot)$ denotes ELU activation function with a dropout layer, and $\odot$ represents Hadamard product. $A_k$ is a binarized vector used for pruning action space, and in this case, it is initialized as a full-one vector because the number of experts is relatively small.

**Rewards.** To steer policy learning, we incentivize decision-making at each step by assessing how well the response from the updated state $s_{t+1}$ contributes to effective negotiation, politeness, dialogue coherence, and user engagement.

**(1) Contextual Negotiation Strategy Congruence Reward** ensures that the generated response $r$ matches the predefined negotiation strategies in the current dialogue context. This is achieved using a RoBERTa-based negotiation strategy classifier, $G_{cNS}$ (achieve an 87.4% accuracy and an 81.5% macro-F1 score), which provides the negotiation strategy probability as the strategy score. The reward is formulated as: $R_{cNS} = G_{cNS}(C_t) - \gamma_c * G_{cNS}(r)$, where $\gamma_c \geq 1$ serves as the penalization factor.

**(2) Future Negotiation Strategy Congruence Reward** evaluates how well the generated response $r$ prepares for and aligns with potential future strategies, ensuring that the conversation remains strategic and goal-oriented to foster long-term negotiation success and adaptive dialogue. For this, we train RoBERTa-based (Liu et al. 2019) negotiation strategy classifier, $G_{fNS}$ (achieve an 82.1% accuracy and an 76.9% macro-F1 score) and design the reward as: $R_{fNS} = G_{fNS}(C_f) - \gamma_f * G_{fNS}(r)$, where $\gamma_f \geq 1$ serves as the penalization factor.

**(3) Dialogue-level Politeness Reward** aims to dynamically adjust the politeness as the conversation progresses. It is formulated as: $R_{dP} = \sum_{m=1}^{M} cos(\frac{\pi}{2} \cdot \frac{m}{M_{max}}) \cdot PD_{dP}$, where $PD_{dP} = G_P(r) - G_P(C_m)$ and $G(\cdot)$ measures the politeness level of an utterance using the state-of-the-art politeness classification model developed by Danescu-Niculescu-Mizil et al. (2013). The model is trained on two datasets containing diverse request types and achieves nearly 70% accuracy in different settings for politeness classification. Politeness scores are collected as the politeness level. We advocate for the politeness distance $PD_{dP}$ between the generated response $r$ and the contextual user's utterance $C_m$ to meet the

following criteria: (a) it should be non-negative, meaning that the generated response should at least match (equal to 0) or exceed the user's level of politeness (greater than 0); (b) it should adjust progressively with the dialogue turn $m$, reflecting that the conversation's early stages should emphasize respect and formality, while the later stages should focus on more amicable and personalized engagement. Over time, maintaining consistently respectful and appropriately polite interactions enhances sociopsychological closeness and fosters rapport and trust. Here, $M_{max}$ represents the maximum turns in conversation, and $M$ denotes the current turn.

**(4) Utterance-level Politeness Reward** assesses the feedback of the user's next utterance politeness. It is formulated as: $R_{uP} = cos(\frac{\pi}{2} \cdot \frac{M}{M_{max}}) \cdot cos(\frac{\pi}{2} \cdot PD_{uP})$, where $PD_{uP} = |G_P(r) - G_P(C_f)|$ and $PD_{uP}$ measures the relative politeness distance between the generated response $r$ and the user's future utterance $C_f$. We encourage $PD_{uP}$ to decrease as the current turn $M$ approaches maximum turn $M_{max}$ to enhance sociopsychological closeness during later stages of the conversation, thereby promoting more effective cooperation between the negotiating parties.

**(5) Contextual Dialogue Coherence Reward** ensures that the generated response $r$ remains coherent with the context $C$ by evaluating coherence at both the keyterm and sentence levels. Initially, we create a dataset that includes pairs of context and responses categorized as coherent or incoherent, where the responses of the incoherent pairs are utterances randomly sampled from the dataset. We then train a RoBERTa-based (Liu et al. 2019) classifier, $G_{cDC}$ on sentence-keyterm pairs (achieve 83.3% accuracy on NEGOCHAT and 84.7% on IND) and consider the coherence probability as the coherence score. The reward is formulated as: $R_{cDC} = G_{cDC}(C \oplus C_{kt}, r \oplus r_{kt}) \cdot e^{\frac{N_{c,kt}}{|r_{kt}|}-1}$, where $r_{kt}$ is the keyterm set of $r$ and $N_{c,kt}$ is the number of keyterms in $r_{kt}$ that are forward neighbors of contextual keyterms in $\mathcal{G}$.

**(6) Future Dialogue Coherence Reward** accounts for coherence with user's future utterance $C_f$. To this end, we create a dataset including pairs of future utterances and responses categorized as coherent or incoherent and then train another RoBERTa-based (Liu et al. 2019) classifier, $G_{fDC}$ on sentence-keyterm pairs (achieve 79.6% accuracy on NEGOCHAT and 78.5% on IND). The reward is defined as: $R_{fDC} = G_{fDC}(C_f \oplus C_{f_{kt}}, r \oplus r_{kt}) \cdot e^{\frac{N_{f,kt}}{|r_{kt}|}-1}$, where $N_{f,kt}$ is the number of keyterms in $r_{kt}$ that are the backward neighbors of $C_{f_{kt}}$ of $C_f$ in $\mathcal{G}$.

**(7) Engagingness Reward** penalizes generic and repetitive responses (*e.g., I can offer the best package at best price*) that degrades overall conversation quality (See et al. 2019). It is formulated using the Jaccard similarity between the responses, $r_m$ and $r_{m-1}$ at $m^{th}$ and $(m-1)^{th}$ turns, respectively as: $R_E = 1 - (r_{m-1} \cap r_m)/(r_{m-1} \cup r_m)$.

**Cumulative Reward** $R$ is formulated as the weighted sum of all the rewards, i.e., $R = w_{cNS} * R_{cNS} + w_{fNS} * R_{fNS} + w_{dP} * R_{dP} + w_{uP} * R_{uP} + w_{cDC} * R_{cDC} + w_{fDC} * R_{fDC} + w_{dE} * R_{dE}$.

**Optimization.** We define $T$-step iterations with the aim of agent learning being to maximize the expected cumulative reward: $J_\theta = \mathbb{E}_\pi \left[\sum_{t=1}^{T} \alpha^t r_{t+1}\right]$, where $\theta$ represents the learned parameter and $\alpha$ is the discount factor. The agent is optimized using the $\mathcal{L}_{agent}$ loss, and its policy gradient is given by: $\nabla_\theta J_\theta = \mathbb{E}_\pi [\nabla_\theta \log \pi_\phi(a_t, s_t, A_t)(G - Q_\beta(s_t))]$, where $G$ represents discounted cumulative reward from the initial state to terminal one. Subsequently, we utilize $H_{S,T+1}$, the hidden state of state $s_{T+1}$ to generate the response. The optimization of the decoder is achieved through the $\mathcal{L}_{dec}$ loss: $\mathcal{L}_{dec} = -\sum_{k=1}^{K} \log P(y_k | H_{S,T+1}, y_{<k})$.

**Warm Start.** We utilize LLaMa-3.1-8B-Instruct model for initializing the model. The initial state serves as the input for fine-tuning the model, with the warm start achieved by optimizing $\mathcal{L}_{warm} = \mathcal{L}_{moe} + \mathcal{L}_{dec}$.

**Joint Training.** The model is ultimately trained jointly by optimizing $\mathcal{L}_{joint} = \mathcal{L}_{agent} + \mathcal{L}_{dec} + \frac{1}{T+1} \sum_{t=1}^{T+1} \mathcal{L}_{moe,t}$.

## Experiments

We compare GENTEEL-NEGOTIATOR with 7 baselines: DialoGPT (Zhang et al. 2020), ARDM (Wu et al. 2021), PersRFI (Shi et al. 2021), GPT-Critic (Jang, Lee, and Kim 2022), INA (Ahmad et al. 2023), ProCoT (GPT-3.5) (Deng et al. 2023), and LLaMA-3.1-8B-finetune (Touvron et al. 2023). For automatic evaluation, we adopt Perplexity (PPL) (Brown et al. 1992), BLEU (B-2) (Papineni et al. 2002), BERTScore-f1 (BS-f1) (Zhang et al. 2019), Distinct-2 (D-2) (Li et al. 2015), and Response Length (R-LEN) to evaluate general quality of responses. To assess responses for goal accomplishment, we introduce Negotiation Strategy Congruence (NSC) scores encompassing contextual and future NSC, i.e., $\mathcal{S}_{cNS}$ and $\mathcal{S}_{fNS}$, Politeness scores consisting dialogue-level and utterance-level politeness, i.e. $\mathcal{S}_{dP}$ and $\mathcal{S}_{uP}$, Dialogue coherence scores, which include contextual and future coherence, i.e., $\mathcal{S}_{cDC}$ and $\mathcal{S}_{fDC}$, and Engagingness score $\mathcal{S}_E$. For human evaluation, we employ Fluency (F), Contextual Coherence (CC), and Engagingness (E) to assess the responses' general quality. To assess how well the generated responses achieve goals, we introduce Sociopsychological Closeness (SC), Negotiation Consistency (NC), Bargaining Efficacy (BE), and Outcome Fairness (OF) (Ahmad et al. 2023). We include 'Implementation Details', 'Baselines Details', 'Evaluation Metrics Details' in "Experiment Details" section of appendix.

## Results and Analysis

**Automatic Evaluation.** Table 2 shows that compared to baselines, proposed GENTEEL-NEGOTIATOR achieves superior dialogue quality, as indicated by its lexical (B-2) and semantic richness (BS-f1) and ability to generate more diverse (D-2) and longer responses (R-LEN) across both datasets. Specifically, on NEGOCHAT, it achieves a significant improvement of 16.3%, 4.3%, 3.1%, and 8.4% in B-2, BS-f1, D-2, and R-LEN, respectively, compared to the second-best model, LLaMA-3.1-8B-finetune. On IND, it achieves notable gains of 19.8%, 1.3%, 7.1%, and 1.6% in B-2, BS-f1, D-2, and R-LEN, respectively. GENTEEL-NEGOTIATOR obtains the highest negotiation strategy congruence ($\mathcal{S}_{cNS}$, $\mathcal{S}_{fNS}$) and politeness scores ($\mathcal{S}_{dP}$, $\mathcal{S}_{uP}$)

| 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 | | | | | | | | | | | | |
| DialoGPT | 22.02 | 3.66 | 30.21 | 0.631 | 0.565 | 0.362 | 0.570 | 0.425 | 0.650 | 0.445 | 0.625 | 23.39 |
| ARDM | 21.91 | 3.83 | 31.53 | 0.655 | 0.573 | 0.385 | 0.593 | 0.437 | 0.663 | 0.453 | 0.648 | 24.24 |
| PersRFI | 20.80 | 4.14 | 31.92 | 0.663 | 0.579 | 0.391 | 0.587 | 0.431 | 0.676 | 0.458 | 0.651 | 25.70 |
| GPT-Critic | 18.69 | 4.98 | 33.31 | 0.662 | 0.584 | 0.395 | 0.602 | 0.455 | 0.689 | 0.472 | 0.665 | 27.45 |
| INA | 17.13 | 5.92 | 34.32 | 0.671 | 0.592 | 0.402 | 0.615 | 0.467 | 0.703 | 0.496 | 0.687 | 29.02 |
| ProCoT (GPT-3.5) | 21.82 | 3.69 | 31.69 | 0.648 | 0.560 | 0.373 | 0.565 | 0.453 | 0.674 | 0.479 | 0.612 | 28.16 |
| LLaMA-3.1-8B-finetune | 16.55 | 6.24 | 38.22 | 0.734 | 0.698 | 0.499 | 0.682 | 0.499 | 0.743 | 0.532 | 0.759 | 33.98 |
| **GENTEEL-NEGOTIATOR** | **14.72** | **7.26** | **39.41** | **0.766** | **0.751** | **0.513** | **0.712** | **0.550** | **0.781** | **0.572** | **0.776** | **36.84** |
| IND | | | | | | | | | | | | |
| DialoGPT | 4.00 | 3.64 | 35.17 | 0.762 | 0.354 | 0.192 | 0.427 | 0.274 | 0.362 | 0.251 | 0.478 | 23.37 |
| ARDM | 3.89 | 3.81 | 36.50 | 0.783 | 0.375 | 0.205 | 0.448 | 0.289 | 0.383 | 0.264 | 0.497 | 25.22 |
| PersRFI | 3.78 | 4.12 | 37.89 | 0.803 | 0.396 | 0.219 | 0.471 | 0.305 | 0.407 | 0.279 | 0.518 | 25.68 |
| GPT-Critic | 3.67 | 4.96 | 39.27 | 0.823 | 0.423 | 0.239 | 0.507 | 0.332 | 0.438 | 0.302 | 0.552 | 27.43 |
| INA | 2.11 | 5.90 | 33.28 | 0.854 | 0.574 | 0.382 | 0.698 | 0.495 | 0.597 | 0.452 | 0.627 | 39.00 |
| ProCoT (GPT-3.5) | 10.78 | 3.67 | 35.66 | 0.697 | 0.523 | 0.348 | 0.637 | 0.454 | 0.544 | 0.415 | 0.569 | 28.14 |
| LLaMA-3.1-8B-finetune | 2.63 | 7.24 | 45.19 | 0.871 | 0.690 | 0.505 | 0.727 | 0.512 | 0.632 | 0.478 | 0.626 | 39.48 |
| **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** |

Table 2: Automatic evaluation results. Results are statistically significant at 5% significance level based on t-test (Welch 1947).

| Models | SC | NC | BE | OF | F | CC | E |
|---|---|---|---|---|---|---|---|
| NEGOCHAT | | | | | | | |
| INA | 2.29 | 2.85 | 2.53 | 2.74 | 2.83 | 2.79 | 2.05 |
| ProCoT (GPT-3.5) | 3.05 | 3.18 | 2.97 | 2.50 | 4.15 | 2.98 | 3.19 |
| LLaMA-3.1-8B-finetune | 3.54 | 3.07 | 3.15 | 3.01 | 3.89 | 3.26 | 3.82 |
| **GENTEEL-NOGOTIATOR** | **4.17** | **4.31** | **4.55** | **4.78** | **4.78** | **4.32** | **4.04** |
| IND | | | | | | | |
| INA | 2.12 | 2.99 | 2.64 | 2.86 | 3.14 | 3.27 | 3.29 |
| ProCoT (GPT-3.5) | 3.01 | 3.26 | 2.89 | 2.37 | 4.22 | 3.94 | 3.54 |
| LLaMA-3.1-8B-finetune | 3.65 | 3.13 | 3.32 | 3.09 | 3.93 | 3.44 | 3.85 |
| **GENTEEL-NOGOTIATOR** | **4.79** | **4.27** | **4.65** | **4.45** | **4.52** | **4.28** | **4.51** |

Table 3: Human evaluation results. Results are statistically significant at 5% significance level based on t-test (Welch 1947). All metrics are rated on a scale of 1 to 5.

while maintaining coherence ($\mathcal{S}_{cDC}$, $\mathcal{S}_{fDC}$) and engagingness ($\mathcal{S}_E$) across both the datasets, which justifies the design of the novel reward function. In particular, it yields 0.751, 0.513, 0.712, 0.550, 0.781, 0.572, and 0.776 scores for $\mathcal{S}_{cNS}, \mathcal{S}_{fNS}, \mathcal{S}_{dP}, \mathcal{S}_{uP}, \mathcal{S}_{cDC}, \mathcal{S}_{fDC}$, and $\mathcal{S}_E$, respectively on NEGOCHAT with a significant increase of 7.6%, 2.8%, 4.4%, 10.2%, 5.1%, 7.5%, and 2.2% compared to second best model. Likewise, it achieves a gain of 22.0%, 31.3%, 19.9%, 39.2%, 24.8% 24.7%, and 30.5% in $\mathcal{S}_{cNS}, \mathcal{S}_{fNS}, \mathcal{S}_{dP}, \mathcal{S}_{uP}, \mathcal{S}_{cDC}, \mathcal{S}_{fDC}$, and $\mathcal{S}_E$, respectively on IND. The LLaMA-3.1-8B-finetune model often generates less diverse and less coherent polite responses during negotiation (e.g., "*I am glad I could satisfy your needs.*" with high politeness). This may be due to the repetition of generic responses that appear in its outputs. INA often generates less polite responses, which can be justified given that it lacks politeness rewards. ProCoT (GPT-3.5) excels in natural language generation, showcasing notable strengths in dialogue coherence and engagement but performs poorly in negotiation and generating polite responses (e.g., for the user utterance, "*Can you please go down to $1250*?", the model simply responds with "*Yes, we can!*" without further negotiation or polite tone). Overall, GENTEEL-NEGOTIATOR excels across all evaluated dimensions, demonstrating the effectiveness of its advanced reward function in fostering high-quality, polite negotiation dialogues.

**Human Evaluation.** Table 3 presents human evaluation results for GENTEEL-NEGOTIATOR and baselines. We compare GENTEEL-NEGOTIATOR against INA, ProCoT (GPT-3.5), and LLaMA-3.1-8B-finetune only, as manual evaluation is expensive. It is evident that GENTEEL-NEGOTIATOR achieves better scores of 4.17, 4.31, 4.55, 4.78, 4.78, 4.32, and 4.04 for SC, NC, BE, OF, F, CC, and E, respectively, with an improvement of +0.63, +1.24, +1.40, +1.77, +0.89, +1.06, and +0.22 points for these metrics, compared to second best model, LLaMA-3.1-8B-finetune, on NEGOCHAT. A similar performance improvement is observed in the IND dataset. The superior SC, NC, BE, and OF scores indicate the effectiveness of negotiation strategy congruence and politeness rewards in facilitating polite responses during negotiation that foster sociopsychological closeness and lead to win-win outcomes. Also, the high CC and E scores reflect the pivotal role of dialogue coherence and engagingness rewards in generating coherent, and engaging responses.

**Additional Analysis.** We include more analyses - (1) Ablation w.r.t Experts, (2) Ablation w.r.t Multi-task Learning of Experts, (3) Ablation w.r.t Rewards, (4) Ablation w.r.t Warm-start and Joint Training, (5) Analysis on Iteration Steps, and (6) Case Study under "Additional Analysis" section in appendix.

## Conclusion

In this work, we introduced GENTEEL-NEGOTIATOR, a polite negotiation dialogue system to enhance the quality of negotiation outcomes in tourism and e-commerce domains. In this regard, we curated a novel NEGOCHAT negotiation dialogue dataset for tourism negotiation using LLM. GENTEEL-NEGOTIATOR leverages LLM and MoE-based RL approach employing designated negotiation, politeness, and keyword experts with well-designed negotiation strategy congruence, politeness, dialogue coherence, and engagingness rewards. Extensive experiments on NEGOCHAT and IND datsets demonstrated the promising potential of GENTEEL-NEGOTIATOR in generating polite, coherent, and engaging responses, and significantly improving the quality of negotiation outcomes.

## Acknowledgments

This research was conducted as part of the project "Conversational Agents with Negotiation and Influencing Ability" funded by Accenture Labs, Bangalore, India. Priyanshu Priya acknowledges financial support from the DST-INSPIRE Fellowship, Government of India. All authors gratefully acknowledge Google for the "Gemma Academic Program GCP Credit Award", which provided Cloud credits to support this research.

## References

Abdelnabi, S.; Gomaa, A.; Sivaprasad, S.; Schönherr, L.; and Fritz, M. 2023. Llm-deliberation: Evaluating llms with interactive multi-agent negotiation games. *arXiv preprint arXiv:2309.17234*.

Adair, W. L.; Okumura, T.; and Brett, J. M. 2001. Negotiation behavior when cultures collide: The United States and Japan. *Journal of Applied Psychology*, 86(3): 371.

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.

Bahdanau, D.; Cho, K.; and Bengio, Y. 2014. Neural machine translation by jointly learning to align and translate. *arXiv preprint arXiv:1409.0473*.

Bazerman, M. H. 1994. *Negotiating rationally*. Simon and Schuster.

Brown, P.; and Levinson, S. C. 1987. *Politeness: Some universals in language usage*. 4. Cambridge university press.

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.

Cai, W.; Jiang, J.; Wang, F.; Tang, J.; Kim, S.; and Huang, J. 2024. A survey on mixture of experts. *arXiv preprint arXiv:2407.06204*.

Carnevale, P. J.; and Pruitt, D. G. 1992. Negotiation and mediation. *Annual review of psychology*, 43(1): 531–582.

Chen, J.; Yuan, S.; Ye, R.; Majumder, B. P.; and Richardson, K. 2023. Put your money where your mouth is: Evaluating strategic planning and execution of llm agents in an auction arena. *arXiv preprint arXiv:2310.05746*.

Church, K.; and Hanks, P. 1990. Word association norms, mutual information, and lexicography. *Computational linguistics*, 16(1): 22–29.

Danescu-Niculescu-Mizil, C.; Sudhof, M.; Jurafsky, D.; Leskovec, J.; and Potts, C. 2013. A computational approach to politeness with application to social factors. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 250–259.

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.

Escandell-Vidal, V. 1996. Towards a cognitive approach to politeness. *Language sciences*, 18(3-4): 629–650.

Fleiss, J. L. 1971. Measuring nominal scale agreement among many raters. *Psychological bulletin*, 76(5): 378.

Fu, Y.; Peng, H.; Khot, T.; and Lapata, M. 2023. Improving language model negotiation with self-play and in-context learning from ai feedback. *arXiv preprint arXiv:2305.10142*.

Golchha, H.; Firdaus, M.; Ekbal, A.; and Bhattacharyya, P. 2019. Courteously Yours: Inducing courteous behavior in Customer Care responses using Reinforced Pointer Generator Network. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, 851–860.

Holtgraves, T. 2005. Social psychology, cognitive psychology, and linguistic politeness.

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

Horton, W. S. 2012. 14. Shared knowledge, mutual understanding and meaning negotiation. *Cognitive pragmatics*, 4: 375.

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

Kaur, K.; Suppiah, P. C.; Arumugam, N.; and Idham, M. 2022. Politeness and negotiation strategies in handling customers: Conflict-resolution. *International Journal of Academic Research in Business and Social Sciences*, 12(8).

Kelley, J. F. 1984. An iterative design methodology for user-friendly natural language office information applications. *ACM Transactions on Information Systems (TOIS)*, 2(1): 26–41.

Lee, A. J.; Mason, M. F.; and Malcomb, C. S. 2021. Beyond cheap talk accounts: A theory of politeness in negotiations. *Research in organizational behavior*, 41: 100154.

Lewicki, R. J.; Barry, B.; and Saunders, D. M. 2016. *Essentials of negotiation*. McGraw-Hill.

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*.

Liu, Y.; Ott, M.; Goyal, N.; Du, J.; Joshi, M.; Chen, D.; Levy, O.; Lewis, M.; Zettlemoyer, L.; and Stoyanov, V. 2019. Roberta: A robustly optimized bert pretraining approach. *arXiv preprint arXiv:1907.11692*.

Maaravi, Y.; Idan, O.; and Hochman, G. 2019. And sympathy is what we need my friend—Polite requests improve negotiation results. *Plos one*, 14(3): e0212306.

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

Mishra, K.; Priya, P.; and Ekbal, A. 2023a. Help Me Heal: A Reinforced Polite and Empathetic Mental Health and Legal

Counseling Dialogue System for Crime Victims. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, 14408–14416.

Mishra, K.; Priya, P.; and Ekbal, A. 2023b. PAL to Lend a Helping Hand: Towards Building an Emotion Adaptive Polite and Empathetic Counseling Conversational Agent. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 12254–12271.

Morris, M. W.; and Gelfand, M. J. 2004. Cultural differences and cognitive dynamics: Expanding the cognitive perspective on negotiation. *The handbook of negotiation and culture*, 45–70.

Nash Jr, J. F. 1950. The bargaining problem. *Econometrica: Journal of the econometric society*, 155–162.

Obando-Ceron, J.; Sokar, G.; Willi, T.; Lyle, C.; Farebrother, J.; Foerster, J.; Dziugaite, G. K.; Precup, D.; and Castro, P. S. 2024. Mixtures of experts unlock parameter scaling for deep rl. *arXiv preprint arXiv:2402.08609*.

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.

Priya, P.; Mishra, K.; Totala, P.; and Ekbal, A. 2023. PARTNER: A Persuasive Mental Health and Legal Counselling Dialogue System for Women and Children Crime Victims. In Elkind, E., ed., *Proceedings of the Thirty-Second International Joint Conference on Artificial Intelligence, IJCAI-23*, 6183–6191. International Joint Conferences on Artificial Intelligence Organization. AI for Good.

Priya, P.; Singh, G.; Firdaus, M.; Agrawal, J.; and Ekbal, A. 2024a. On the Way to Gentle AI Counselor: Politeness Cause Elicitation and Intensity Tagging in Code-mixed Hinglish Conversations for Social Good. In *Findings of the Association for Computational Linguistics: NAACL 2024*, 4678–4696.

Priya, P.; Yasheshbhai, D.; Joshi, R.; Ramnani, R.; Maitra, A.; Sengupta, S.; and Ekbal, A. 2024b. TRIP NEGOTIATOR: A Travel Persona-aware Reinforced Dialogue Generation Model for Personalized Integrative Negotiation in Tourism. In *Findings of the Association for Computational Linguistics: EMNLP 2024*, 16566–16595.

Schoenfield, M. K. 1983. Strategies and Techniques for Successful Negotiations. *ABAJ*, 69: 1226.

See, A.; Roller, S.; Kiela, D.; and Weston, J. 2019. What makes a good conversation? how controllable attributes affect human judgments. *arXiv preprint arXiv:1902.08654*.

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.

Silva, D.; Semedo, D.; and Magalhães, J. 2022. Polite Task-oriented Dialog Agents: To Generate or to Rewrite? In *Proceedings of the 12th Workshop on Computational Approaches to Subjectivity, Sentiment & Social Media Analysis*, 304–314.

Sree, P. D.; Kokkiligadda, M. R.; Teja, J.; and Sandeep, Y. 2023. Product negotiation in e-commerce website using chatbot. In *2023 7th International Conference on Computing Methodologies and Communication (ICCMC)*, 879–883. IEEE.

Stock, J.; Petukhova, V.; and Klakow, D. 2022. Assessment of Sales Negotiation Strategies with ISO 24617-2 Dialogue Act Annotations. In *Proceedings of the 18th Joint ACL-ISO Workshop on Interoperable Semantic Annotation within LREC2022*, 10–19.

Sutton, R. S.; and Barto, A. G. 2018. *Reinforcement learning: An introduction*.

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*.

Terada, K.; Okazoe, M.; and Gratch, J. 2021. Effect of politeness strategies in dialogue on negotiation outcomes. In *Proceedings of the 21st ACM international conference on intelligent virtual agents*, 195–202.

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*.

UNWTO. 2024. Tourism 4 SDGs: United Nations World Tourism Organization. https://www.unwto.org/tourism4sdgs.

Welch, B. L. 1947. The generalization of 'STUDENT'S'problem when several different population varlances are involved. *Biometrika*, 34(1-2): 28–35.

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.

Yamaguchi, A.; Iwasa, K.; and Fujita, K. 2021. Dialogue act-based breakdown detection in negotiation dialogues. In *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume*, 745–757.

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.

Zhao, T.; Xie, K.; and Eskenazi, M. 2019. Rethinking action spaces for reinforcement learning in end-to-end dialog agents with latent variable models. *arXiv preprint arXiv:1902.08858*.