



DIPLMAT: Dialogue-Span-Aware Direct Preference Optimization for Polite Persuasive Workplace Negotiation Dialogues

Bibhuti Jha^{†*}, Rishikant Chigrupaatii^{†*}, Priyanshu Priya[‡], Asif Ekbal[†]

[†]Department of Computer Science and Engineering, Indian Institute of Technology Patna, India

[‡]GREYC Laboratory, University of Caen Normandy, France

[‡]{jhabk369,rishikantchigrupaatii.24,priyanshu528priya}@gmail.com,

[†]{bibhuti_2301cs94,rishikant_2101cs66,asif}@iitp.ac.in

Abstract

Effective workplace negotiation requires balancing multiple objectives, including achieving task goals, preserving professional relationships, and resolving conflicts constructively. However, misunderstandings, misaligned preferences, and interpersonal friction often impede successful outcomes. Politeness mitigates these challenges by fostering trust, reducing tension, and preventing escalation, and persuasive communication helps overcome resistance, align preferences, and guide participants toward mutually beneficial agreements. Motivated by these insights, we present DIPLMAT, a dialogue system for polite and persuasive workplace negotiation. To support its development, we introduce PROWESS, a dataset of multi-turn workplace negotiation dialogues generated via a multi-agent framework and enriched with negotiation strategies, politeness levels, and persuasive strategies. DIPLMAT is trained using Dialogue-Span-Aware Direct Preference Optimization (DSA-DPO), a novel preference learning objective that identifies key dialogue spans for preference alignment. This enables DIPLMAT to generate contextually coherent responses that employ intended negotiation strategies, maintain politeness, and incorporate effective persuasion strategies throughout interactions. Automatic and human evaluation on PROWESS confirm that DIPLMAT consistently outperforms baselines in generating coherent, polite, and persuasive negotiation responses¹.

1 Introduction

Negotiation dialogue systems are increasingly important in professional settings (Zhan et al., 2024), where agents must not only achieve task goals but also sustain long-term working relationships (Lewis et al., 2017; He et al., 2018). Workplace

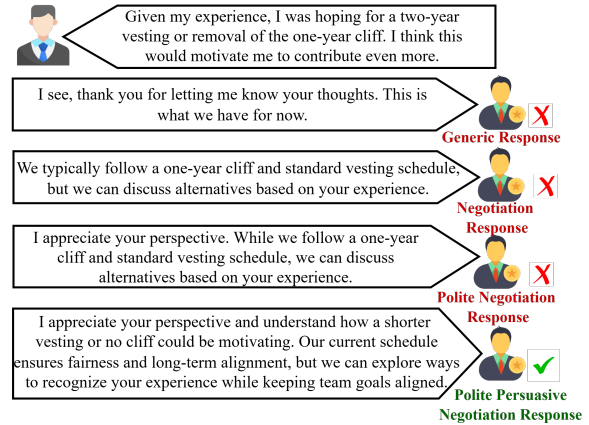


Figure 1: A conversation snippet demonstrating the use of appropriate negotiation strategy (*active listening*), politeness level (*moderate*), and persuasion strategy (*concern addressing*) during workplace negotiation to improve the final outcome.

negotiations involve balancing goal achievement, relational harmony, and constructive conflict resolution (Carnevale and Pruitt, 1992). Recent advances in large language models (LLMs) (OpenAI et al., 2024) and preference-based optimization (Rafailov et al., 2024) enable fluent, contextually relevant dialogue, yet effectively integrating strategic negotiation behavior with politeness and persuasion remains challenging.

Workplace negotiation is inherently complex due to misaligned preferences, power asymmetries, and interpersonal sensitivities (Nash, 1950; Adair et al., 2001). Thus, focusing solely on short-term gains can damage relationships, while excessive accommodation may yield poor outcomes. Politeness mitigates friction, fosters trust, and prevents escalation during negotiation (Priya et al., 2025a), whereas persuasion helps justify proposals, address objections, and gradually align counterparts' decisions (Wang et al., 2019). Despite their complementary roles, the interplay between politeness and persuasion remains largely underexplored in

* Authors are jointly first authors.

¹Data and code are available at <https://github.com/priyanshu-profile/DIPLMAT> or <https://www.iitp.ac.in/~ai-nlp-ml/resources.html#diplomat>.

existing negotiation dialogue systems. Figure 1 illustrates how politeness and persuasion integration during negotiation improves response quality and supports constructive, collaborative outcomes.

Motivated by these considerations, we introduce DIPLOMAT, a polite, persuasive workplace negotiation dialogue system. DIPLOMAT leverages Dialogue-Span-Aware Direct Preference Optimization (DSA-DPO), a novel modification of Direct Preference Optimization (DPO) (Rafailov et al., 2024) that enables fine-grained preference alignment across multi-turn negotiations. Existing approaches operate either at the utterance or dialogue level. The utterance-level DPO (Rafailov et al., 2024) optimizes individual responses using positive-negative pairs and can miss evolving negotiation strategies, polite framing, and persuasive justifications across utterances. Dialogue-level DPO (Shi et al., 2024; Song et al., 2024) captures long-term patterns but is coarse, potentially obscuring critical contributions, penalizing correct behaviors, and introducing noise, making it harder to identify actions that drive successful negotiation outcomes.

To address these limitations, we propose Dialogue-Span-Aware Direct Preference Optimization (DSA-DPO), a novel approach that captures fine-grained polite and persuasive behaviors across multi-turn negotiations. DSA-DPO identifies key dialogue spans that most strongly influence successful outcomes, rather than optimizing entire dialogues or single utterances. It detects the first critical utterance where negotiation errors occur, samples the preceding interaction history, and forms aligned positive and negative spans of equal length to compute a rigorous preference signal. By focusing on these spans, DSA-DPO reduces noise from non-critical utterances, limits irrelevant variations, and reinforces intended polite and persuasive behaviors that drive successful negotiation outcomes, enabling DIPLOMAT to generate responses that balance task success with relationship maintenance.

To support this research, we introduce PROWESS, a multi-turn dataset of workplace negotiation dialogues generated via a multi-agent framework (Fu et al., 2023). In this framework, agents are assigned specialized roles for scenario modeling, dialogue synthesis, and annotation, ensuring logically structured conversations rich in negotiation techniques, polite behaviors, and persuasion tactics that reflect natural negotiation dynamics. Using this framework, we generate 2,400 diverse, high-quality dialogues annotated with negotiation strategies, po-

liteness levels, and persuasion strategies at the utterance level with minimal human supervision.

In summary, the key contributions are: (i) Examine how the dialogue agent’s use of politeness and persuasion influences negotiation outcomes. To our knowledge, this work is the first to model and analyze the combined effects of politeness and persuasion within workplace negotiation context; (ii) Introduce DIPLOMAT, a polite and persuasive negotiation dialogue system that integrates strategic negotiation behavior with politeness and persuasion modeling in workplace interactions; (iii) Propose Dialogue-Span-Aware Direct Preference Optimization (DSA-DPO), a novel preference learning objective that optimizes key dialogue spans, enabling effective alignment of polite and persuasive behaviors with successful negotiation outcomes in DIPLOMAT; (iv) Construct PROWESS, a workplace negotiation dialogue dataset generated via a multi-agent framework and enriched with politeness levels, persuasion and negotiation strategies.

2 Related Work

Negotiation dialogue systems have evolved from early agent-based designs for constructing coherent joint plans (McEleney and O’Hare, 2003) and end-to-end neural models for item-division tasks (Lewis et al., 2017) to modular strategy generation frameworks (He et al., 2018), persuasion-annotated corpora (Chawla et al., 2021), and reward-based integrative agents (Ahmad et al., 2023). Modeling opponent strategies and personality types improves adaptation (Yang et al., 2021), yet most systems target consumer or e-commerce domains. In workplace negotiation, datasets such as NegoChat (Konovalov et al., 2016) and interview corpora (Yamaguchi et al., 2021; Mannekote et al., 2023) provide realistic settings, but template-based generation or limited rephrasing still results in dialogues lacking strategic richness. Recent LLM-based efforts explore coaching or multi-agent interactions (Shea et al., 2024; Fu et al., 2023; Abdelnabi et al., 2024), yet primarily optimize deal-making metrics rather than socially-aware negotiation behaviors.

Politeness (Brown and Levinson, 1987) is a key social driver in negotiation that fosters rapport, trust, and cooperation (Lee et al., 2021; Monson, 2022), and has been shown to improve negotiation outcomes across diverse settings (Terada et al., 2021; Maaravi et al., 2019; Priya et al., 2025a). Persuasion (Eagly and Chaiken, 1984) similarly plays

a central role in negotiation by shaping how proposals are framed, justified, and defended to influence counterpart decisions (Ledgerwood et al., 2006; Provis, 2004). Persuasion enables the participants to advocate for their interests while guiding the dialogue toward mutually acceptable terms (Wang et al., 2019; Dutt et al., 2021).

Recently, preference learning has become a dominant paradigm for aligning LLMs with desired behaviors. Methods progress from Reinforcement Learning from Human Feedback (RLHF) (Ouyang et al., 2022) to Direct Preference Optimization (DPO) (Rafailov et al., 2024) and its dialogue-aware variants (Shi et al., 2024; Song et al., 2024; Jung et al., 2025; Zhuang et al., 2025; Hama et al., 2025). Despite extensive research on negotiation dialogue agents, existing systems primarily focus on modeling deal-making strategies or opponent behaviors and are largely developed for consumer-facing domains. While prior studies highlight the importance of social signals such as politeness and persuasion for successful negotiation, these aspects are rarely modeled jointly in automated negotiation dialogue systems. Furthermore, current alignment methods for dialogue agents rely on preference learning at the utterance or dialogue level, which often fails to capture nuanced negotiation dynamics across multi-turn interactions. To address these limitations, we introduce DIPLOMAT, a dialogue-span-aware DPO framework that leverages fine-grained preference signals over politeness, persuasion, and negotiation dynamics to align LLMs for generating polite and persuasive workplace negotiation dialogues. DIPLOMAT is designed to be scalable and adaptable across diverse application domains.

3 Dataset

To evaluate DIPLOMAT, we introduce PROWESS, a novel dialogue dataset designed for workplace negotiation scenarios. PROWESS consists of multi-turn interactions between a candidate and an employer, covering key negotiation dimensions such as compensation, working conditions, career progression, and related employment terms. The dataset consists of 2,400 dialogues spanning 25 common workplace negotiation aspects, capturing realistic negotiation behaviors and conversational dynamics commonly observed during hiring and professional interactions in the workplace. The list of negotiation aspects in the PROWESS dataset is provided in Table 1.

Negotiation Aspects

benefits negotiation, bonus structure, career development, contract terms, equity negotiation, exit interview, flexible hours, health benefits, hiring decision, job responsibilities, onboarding, performance review, professional development, project assignment, promotion discussion, relocation package, remote work arrangement, retention offer, retirement plans, salary negotiation, team transition, training opportunities, vacation time, work equipment, work-life balance

Table 1: Negotiation aspects covered in the PROWESS dataset.

3.1 Dataset Construction

The dataset construction involves two steps: (1) Negotiation, persuasion, and politeness annotation schema designing and (2) Multi-agent dialogue generation. Figure 2 illustrates the PROWESS dataset construction pipeline.

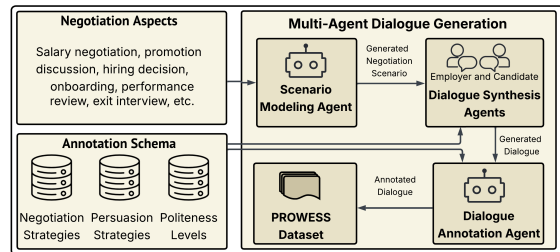


Figure 2: PROWESS dataset construction pipeline.

(1) Negotiation, Persuasion, and Politeness Annotation Schema Designing. Workplace negotiation relies on information sharing, transparency, and relational rapport to support collaboration and professional growth. Effective negotiation strategies are essential for achieving mutually beneficial outcomes, resolving conflicts, and maintaining professional relationships. Considering this, we define a set of 11 negotiation strategies grounded in workplace negotiation principles (Thompson, 2005). These strategies are designed for problem-solving, concession-making, position-establishing, or relationship-building, and include *collaborative style*, *active listening*, *win-win framing*, *principled negotiation*, *data-driven justification*, *MESO (Multiple Equivalent Simultaneous Offers)*, *anchoring*, *door-in-the-face*, *reciprocal concessions*, *credibility assertion*, and *no strategy*.

While negotiation strategies provide the foundation for effective bargaining, politeness and persuasion complement them by guiding negotiation dynamics, influencing decisions, and shaping relationships. Politeness, in particular, influences

negotiation outcomes by mitigating conflict and promoting sociopsychological closeness (Brown and Levinson, 1987). Users express varying levels of politeness, which shape how negotiations unfold over time (Overbeck et al., 2010). In practice, we adopt three politeness level categories from workplace negotiation practice (Yuxian, 2013), including *low_polite*, *moderate_polite*, and *high_polite*. Persuasion further enables alignment of preferences and facilitates agreement in situations with asymmetric power, conflicting goals, or varying engagement. Guided by the persuasive negotiation theory (Fowler, 1998), we define a set of 11 persuasion strategies: *rapport building*, *concern addressing*, *emotional appeal*, *credibility & confidence*, *data-driven persuasion*, *problem-solving focus*, *self-interest appeal*, *value alignment*, *reputation highlighting*, *future vision alignment*, and *no strategy*. Definitions and examples of negotiation strategies, politeness levels, and persuasion strategies are given in Appendix A.

(2) Multi-Agent Dialogue Generation. We construct the PROWESS dataset using a structured multi-agent framework that coordinates specialized LLM agents to generate negotiation dialogues. The framework consists of three agents, *viz.* scenario modeling, dialogue synthesis, and dialogue annotation, each responsible for a specific task. Each agent operates under shared scenario constraints and role-specific objectives, ensuring that the output of one agent reliably conditions the next agent’s behavior, resulting in coherent, goal-driven negotiation dialogues. This integrated modular design allows precise control over dialogue quality and annotation consistency. The scenario modeling agent employs Gemini-2.5-Pro (Comanici et al., 2025), and dialogue synthesis and dialogue annotation agents employ GPT-4o-mini (OpenAI et al., 2024) for its strong reasoning and contextual tracking capabilities, enabling reliable guidance throughout the data generation process.

Scenario Modeling Agent. The Scenario Modeling Agent creates structured negotiation setups by modeling key elements, namely negotiation aspect, background context, employer and candidate roles, negotiation goal, and candidate starting position. This ensures scenarios are realistic and diverse, covering varied objectives, roles, and initial conditions. For each negotiation aspect, 16 distinct scenarios are generated to support the generation of large-scale, high-quality negotiation dialogues.

Dialogue Synthesis Agents. The Dialogue Synthesis Agents generate multi-turn negotiation dialogues by coordinating two LLM agents representing the employer and the candidate, whose behavior is grounded in the role-specific objectives and constraints defined in the structured scenario generated by the Scenario Modeling Agent. Building on prior work on multi-agent dialogue simulation (Fu et al., 2023), the agents generate conversations by cyclically feeding the output of one agent as input to the other to generate coherent, multi-turn interactions. To ensure diverse and realistic conversational behavior, the interaction is guided by structured prompts specifying negotiation and persuasion strategies for each agent. The agents are further instructed to dynamically adapt their strategies in response to the other participant’s behavior. Additionally, both agents are briefed to maintain a professional and polite tone aligned with workplace norms, while adjusting their tone to reflect the interaction dynamics. The dialogues commence when the employer agent is prompted with “*Start the conversation*”, initiating a structured, multi-turn negotiation exchange that reveals each participant’s adaptive negotiation and persuasion strategies, as well as professional, polite behavior.

Dialogue Annotation Agent. After each round of conversation (a pair of turns between the employer and candidate agents), the Dialogue Annotation Agent first evaluates participants’ utterances to annotate politeness levels. It then annotates the persuasion and negotiation strategies used, along with a brief rationale for each strategy annotation. To ensure consistency across rounds, both politeness and strategy annotations are drawn from predefined categories. The agent considers the cumulative dialogue context to capture the evolution of strategies and polite tone. To guide its behavior and minimize annotation errors, we provide the agent with three examples for each annotation dimension. These annotations facilitate analysis of negotiation strategy adaptation, persuasive effectiveness, and politeness management, and provide supervision for training models that generate polite and persuasive negotiation dialogues.

The prompt templates for scenario modeling, dialogue synthesis, and dialogue annotation agents are detailed in Appendix E.1.

3.2 Dataset Filtering and Quality Assessment

Once the complete dataset with 2,675 dialogues is generated, we perform a comprehensive inspection

to remove erroneous dialogues. Specifically, we identify five types of issues: (i) empty utterances, (ii) repetitive utterances, (iii) insufficient interaction rounds, (iv) incomplete or missing annotations, and (v) improper conversation openings or closings. The dialogues exhibiting any of these issues are removed to improve overall dataset quality and reliability.

The remaining dialogues are qualitatively evaluated by the three human evaluators² under the supervision of a domain expert in business management to ensure contextual accuracy and domain relevance. Each dialogue is rated for Scenario Consistency (SC), Negotiation Strategy Correctness (NSC), Persuasion Strategy Correctness (PSC), Politeness Appropriateness (PA), Fairness (F), Coherence (C), Naturalness (N), and Engagingness (E) (Descriptions are given in Appendix A), using a 1–5 scale (low to high). We retain only those dialogues that receive scores of ≥ 3 across all evaluation metrics. The retained dialogues achieve average ratings of 4.58 (SC), 4.46 (NSC), 4.39 (PSC), 4.51 (PA), 4.32 (F), 4.37 (C), 4.79 (N), and 4.41 (E), with corresponding inter-evaluator κ scores of 0.84, 0.82, 0.80, 0.83, 0.79, 0.81, 0.85, and 0.80 for SC, NSC, PSC, PA, F, C, N, and E, respectively. The κ values indicate substantial to near-perfect inter-evaluator agreement, confirming the reliability and consistency of the human evaluations.

Overall, these results demonstrate that the dialogues exhibit strong scenario grounding, consistent use of negotiation and persuasion strategies, appropriate politeness, and high levels of fairness, coherence, naturalness, and engagement. The final dataset statistics are reported in Table 2.

Statistic	Value
# Negotiation Aspects	25
# Dialogues	2,400
# Utterances	51,796
# Words	941,900
Avg. Utterances per Dialogue	21.58
Minimum Dialogue Length	18
Maximum Dialogue Length	26
Avg. Words per Dialogue	392.5
Avg. Words per Utterance	18.2

Table 2: PROWESS dataset statistics.

²The evaluators hold Ph.D. degrees in Linguistics and have expertise in negotiation, persuasion, and politeness concepts. They are compensated in as per the institute guidelines.

4 Methodology

The proposed DIPLOMAT framework consists of three stages: (1) supervised fine-tuning on the PROWESS dataset, (2) preference data construction through utility evaluation, suboptimal utterance detection, positive dialogue sample selection, and dialogue-span extraction, and (3) preference optimization using dialogue span-aware DPO (DSA-DPO) objective. Figure 3 provides an overview of DIPLOMAT.

4.1 Supervised Fine-Tuning

Let $\mathcal{D} = \{d_j\}_{j=1}^N$ denote a dataset of N dialogues. For the t -th turn of the j -th dialogue, the conversational context $\mathcal{C}_t = (e_1, c_1, \dots, e_{t-1}, c_{t-1}, c_t) \in d_j$ comprises alternating employer (e) and candidate (c) utterances upto $(t - 1)$ turns, together with the candidate’s current utterance c_t . The primary objective is to generate the employer’s appropriate response e_t conditioned on $\mathcal{C}_t$.

To generate contextually appropriate responses, the employer must reason about the candidate’s polite, persuasion, and negotiation behaviors. Concretely, given a context $\mathcal{C}_t$, the model is expected to: (i) infer candidate’s politeness level, and persuasion and negotiation strategies along with their rationale, (ii) select appropriate negotiation and persuasion strategies for employer based on candidate’s strategies and adjust employer’s politeness level accordingly; and (iii) generate employer’s final response conditioned on these inferred strategies and politeness level. These intermediate reasoning steps provide additional supervision that facilitates learning of strategy adaptation, politeness mirroring, conflict de-escalation, and effective concession timing during negotiation. To operationalize this, we construct structured training target sequences with three components: (a) *candidate inference block* describing candidate negotiation and persuasion strategies with reasoning, and the candidate’s politeness level, (b) *employer inference block* specifying employer’s selected negotiation and persuasion strategies with reasoning, and employer’s politeness level, and (c) *employer response block* outlining employer’s final response.

During supervised fine-tuning, the model is trained to generate these structured components sequentially before generating the final response. Specifically, we fine-tune the pre-trained LLM Mistral-7B-Instruct (Jiang et al., 2023) on $\mathcal{D}$ using standard next-token prediction. Formally,

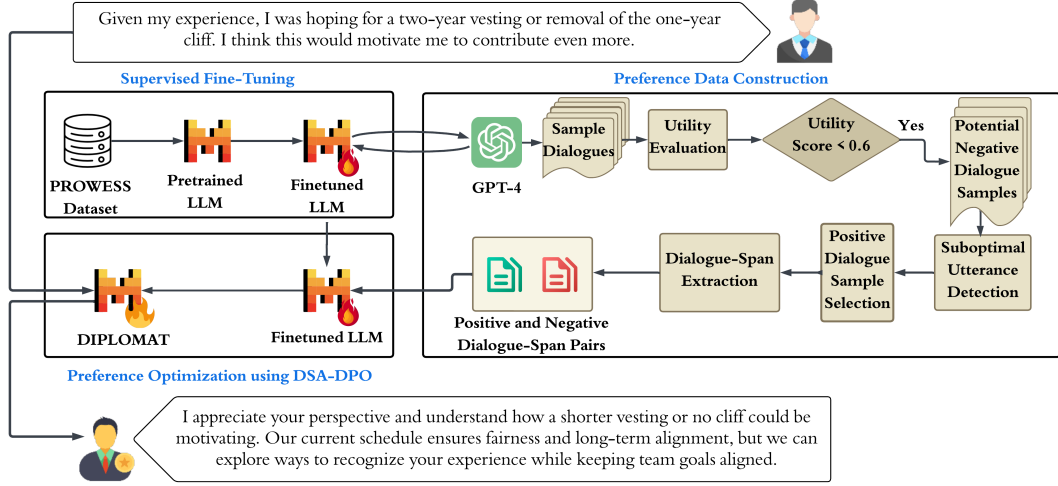


Figure 3: Architecture of the proposed polite persuasive negotiation dialogue agent, DIPLOMAT. DIPLOMAT is developed in three stages: (1) *Supervised Fine-Tuning (SFT)*, where the pre-trained LLM is fine-tuned on negotiation dialogues from the PROWESS dataset to obtain the supervised fine-tuned LLM (π_{SFT}), which generates polite and persuasive negotiation responses; (2) *Preference Data Construction*, where π_{SFT} interacts with GPT-4o-mini to sample diverse negotiation dialogues, which are evaluated using a utility function. Dialogues with utility score < 0.6 are treated as potential negative samples, which are then paired with positive dialogue samples through suboptimal utterance detection, positive dialogue sample selection, and dialogue-span extraction, yielding preference pairs of positive and negative dialogue spans; and (3) *Preference Optimization using DSA-DPO*, where π_{SFT} is further optimized using the proposed DSA-DPO objective over the extracted positive and negative dialogue spans to generate contextually coherent, polite, persuasive, and strategically consistent negotiation responses.

the model parameters θ are optimized by minimizing the negative log-likelihood loss, yielding supervised fine-tuned model π_{SFT} :

$$\mathcal{L}_{\text{SFT}}(\theta) = -\mathbb{E}_{(C_t, y_t) \sim \mathcal{D}} \left[\sum_{i=1}^{|y_t|} \log \pi_{\theta}(y_{t_i} \mid y_{t_{<i}}, C_t) \right] \quad (1)$$

where y_t denotes the structured target sequence comprising the candidate inference, employer inference, and employer response blocks. At inference time, these intermediate blocks can be treated as latent and optionally suppressed, allowing the model to generate responses conditioned solely on the dialogue context. This supervised fine-tuning step equips the LLM to understand the candidate’s behavior, adapt its own, and generate contextually appropriate, polite, and persuasive responses during negotiation, serving as the initialization for subsequent preference-based optimization.

4.2 Preference Data Construction

The construction of a high-quality dialogue-span-aware preference dataset is central to our approach. Unlike standard preference datasets that treat dialogues as monolithic units, ours captures fine-grained dialogue spans that reflect strategic con-

versational behaviors, including politeness adaptation, persuasive maneuvers, and negotiation strategy shifts, allowing agents to learn how individual utterances influence overall negotiation outcomes. This span-aware supervision enables the negotiation agent to generate contextually appropriate, polite, and persuasive responses in multi-turn negotiations. To construct the dataset, we first generate multiple negotiation dialogues under a given scenario and goal. Specifically, the supervised fine-tuned model π_{SFT} interacts with GPT-4o-mini to simulate negotiation conversations and generate diverse dialogue samples. These generated dialogues serve as the input to the preference data construction pipeline, where utility evaluation, suboptimal utterance detection, positive dialogue selection, and dialogue-span extraction are performed to construct fine-grained dialogue-span-aware preference pairs.

(i) Utility Evaluation. The generated dialogues are evaluated for quality using a utility score $U(\cdot)$ that captures multiple dimensions of negotiation success: $U(d) = \text{PT}(d) + \text{PSA}(d) + \text{NSA}(d) + \text{MS}(d) + \text{AQ}(d) - \text{NB}(d)$. Here, $\text{PT}(\cdot)$, $\text{PSA}(\cdot)$, $\text{NSA}(\cdot)$, $\text{MS}(\cdot)$, $\text{AQ}(\cdot)$, and $\text{NB}(\cdot)$ denote the scores for *Politeness Trajectory*, *Persuasion Strategy Alignment*, *Negotiation Strategy Alignment*,

Mutual Satisfaction, Agreement Quality, and Negotiation Breakdown, respectively, for a dialogue d . These scores are estimated by Gemini-2.5-Pro using structured evaluation prompts (Appendix E.1).

The dialogues with $U(\cdot) < 0.6$ are treated as potential negative samples, indicating suboptimal negotiation outcomes or undesirable conversational behaviors. Each utility component is evaluated in the range $[0, 1]$, consequently, $U(\cdot) \in [-1, 5]$. Although the overall utility score spans a wider range, the threshold analysis is focused on the interval $[0, 1]$ because this region empirically corresponds to suboptimal yet non-trivial negotiation dialogues that provide informative negative samples for preference optimization. Extremely low utility values are rare and typically correspond to severe negotiation breakdowns, while substantially higher values generally indicate dialogues with acceptable negotiation quality that are less suitable as negative samples. A sensitivity analysis on the utility threshold is given in Appendix D.4. For each identified negative sample, we then construct dialogue-span-aware preference pairs as follows:

(ii) Suboptimal Utterance Detection. Unlike tasks with objectively defined errors (e.g., math or programming), suboptimal utterances in negotiation dialogues are inherently ambiguous. In a negative dialogue sample, an utterance is identified as suboptimal using Gemini-2.5-Pro via structured prompts if it satisfies one or more utility-based criteria: (a) it is pivotal for advancing the negotiation goal, such as offers, counteroffers, or concessions; (b) it exhibits low effectiveness in one or more utility dimensions: PT, PSA, NSA, MS, or AQ; or (c) it contributes to a negotiation breakdown or reduces relationship quality, indicating a missed opportunity for polite or persuasive engagement.

(iii) Positive Dialogue Sample Selection. After suboptimal utterance detection, we identify candidate positive dialogues that serve as corrective exemplars for the flagged utterance. To this end, we retrieve three complete dialogues sharing the same conversational context preceding the identified suboptimal utterance. Among these candidates, the dialogue with the highest utility score is selected as the candidate sample. If its utility score exceeds that of the negative dialogue sample, the candidate and negative samples form a preference pair; otherwise, the negative sample is discarded.

(iv) Dialogue-span Extraction. Once positive and negative dialogue samples are paired, we extract fine-grained dialogue spans that capture the

utterances most responsible for the observed differences in utility score. To this end, both dialogues are provided to Gemini-2.5-Pro with structured prompts to identify a contiguous span in the positive sample that maximally contributes to its superior utility across the dimensions of PT, PSA, NSA, MS, AQ, and minimal NB. A span of equal length is then extracted from the corresponding negative dialogue sample, forming a dialogue-span-aware preference pair. This stage filters out irrelevant utterances, such as greetings or pleasantries, ensuring that the model focuses on learning strategically and socially meaningful improvements at the utterance level in multi-turn negotiations.

The resulting preference dataset comprises 700 pairs. Appendix B details the evaluation of Gemini-2.5-Pro within the preference data construction pipeline, and Appendix E.2 provides the prompt templates used in this pipeline.

4.3 Preference Optimization using DSA-DPO

We leverage a curated dialogue-span-aware preference dataset $\mathcal{D}_{\text{pref}} = \{(\mathcal{C}_s^m, \tau_w^m, \tau_l^m)\}_{m=1}^M$, where $\mathcal{C}_s$ denotes the conversational context up to the suboptimal utterance s , and τ_w and τ_l represent preferred and non-preferred dialogue spans extracted from multi-turn negotiation interactions, to further refine the supervised fine-tuned policy π_{SFT} , and M denote the total no. of instances in $\mathcal{D}_{\text{pref}}$. Specifically, we adopt a dialogue-span-aware direct preference optimization (DSA-DPO) objective, which operates on dialogue spans rather than isolated responses. This objective encourages the model to assign higher likelihood to spans exhibiting polite, persuasive, and strategically coherent negotiation behaviors, such as maintaining politeness, providing constructive persuasive arguments, and progressively guiding the conversation toward mutually beneficial agreement, while penalizing spans with undesirable behaviors such as blunt demands or uncooperative tone. Formally, the DSA-DPO objective is defined as:

$$\mathcal{L}_{\text{DSA-DPO}} = -\mathbb{E}_{(\mathcal{C}_s, \tau_w, \tau_l) \sim \mathcal{D}_{\text{pref}}} \left[\log \sigma \left(\beta \left(\log \frac{\pi_{\theta}(\tau_w)}{\pi_{\text{ref}}(\tau_w)} - \log \frac{\pi_{\theta}(\tau_l)}{\pi_{\text{ref}}(\tau_l)} \right) \right) \right] \quad (2)$$

where,

$$\pi_{\theta}(\tau) = \prod_{t=s}^{s+p} \pi_{\theta}(y_t | \mathcal{C}_t), \quad (3)$$

and π_{ref} denotes the reference policy initialized from π_{SFT} , σ is the logistic function, β controls the strength of preference optimization, and p is the number of utterances in the chosen dialogue spans. By optimizing over multi-turn dialogue spans, DSA-DPO learns how polite and persuasive behaviors evolve across utterances, enabling the model to capture negotiation dynamics and lead to more effective negotiation strategies.

5 Experiments

We compare DIPLOMAT with 13 baselines: INA (Ahmad et al., 2023), GENTEEL-NEGOTIATOR (Priya et al., 2025a), ProCoT (GPT-5-mini) (Deng et al., 2023), Mistral-7B-Instruct-v0.3-SFT (Jiang et al., 2023), Llama3.1-8B-SFT (Grattafiori et al., 2024), Gemma-2-9B-SFT (Team et al., 2024), Mistral-7B-Instruct-v0.3-SFT-DPO (Rafailov et al., 2024), Mistral-7B-Instruct-v0.3-SFT-ETO (Song et al., 2024), Mistral-7B-Instruct-v0.3-SFT-DMPO (Shi et al., 2024), and Mistral-7B-Instruct-v0.3-PositiveSFT. For automatic evaluation, we adopt Perplexity (PPL) (Brown et al., 1992), BLEU-4 (B-4) (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 language quality of responses. To assess responses for task completion, we introduce Negotiation Strategy Correctness (NSC), Persuasion Strategy Correctness (PSC), Politeness Appropriateness (PA), and Agreement Rate (AR). For human evaluation, we employ Fluency (F), Contextual Coherence (CC), and Engagingness (E) to assess language quality of responses. To assess responses for task completeness, we mirror automatic evaluation metrics: NSC, PSC, and PA, report Bargaining Efficacy (BE), Outcome Fairness (OF), and Sociopsychological Closeness (SC) (Priya et al., 2025a), and introduce Influence Effectiveness (IE).

Appendix C provides ‘Implementation Details’, ‘Baseline Details’, ‘Evaluation Metrics Details’, and ‘Human Evaluation Process’.

6 Results and Analysis

6.1 Automatic Evaluation Results

Table 3 presents automatic evaluation results for DIPLOMAT and baseline models on the PROWESS dataset. DIPLOMAT consistently achieves superior dialogue quality while maintaining politeness and persuasiveness, reflected in the highest B-4, BS-F1, and D-2 scores, longer responses (R-LEN), effective

strategy application (PSC, NSC), high politeness appropriateness (PA), and improved agreement rate (AR).

Across LLM families, Mistral variants generally improve language quality and task effectiveness more than LLaMA or Gemma models. Compared to the strongest baseline, Mistral-7B-Instruct-v0.3-SFT-ETO, DIPLOMAT achieves gains of 4.1%, 2.9%, 1.6%, 15.1%, 16.0%, 13.8%, 1.1%, and 6.2% in B-4, BS-F1, D-2, R-LEN, NSC, PSC, PA, and AR, respectively, indicating linguistically richer, more persuasive, and appropriately polite negotiation responses. Besides, the baselines such as INA and ProCoT (GPT-5-mini) tend to produce shorter, less strategic, or less polite responses, limiting negotiation success. While Mistral-7B variants improve fluency and strategy, they often fail to balance persuasion with politeness.

Notably, Mistral-7B-Instruct-v0.3-PositiveSFT, which is trained exclusively on preferred dialogue samples, underperforms Mistral-7B-Instruct-v0.3-SFT-DPO across multiple metrics despite both methods having access to the same high-quality preferred data. This suggests that the gains from preference optimization cannot be attributed solely to exposure to higher-quality training samples, but instead arise from the contrastive learning signal provided by negative examples. By explicitly contrasting preferred and non-preferred dialogue spans, preference-based optimization enables the model to learn not only desirable negotiation behaviors, but also which strategic, persuasive, and politeness-related behaviors should be avoided. In contrast, DIPLOMAT’s high performance across all evaluated aspects demonstrates the effectiveness of its dialogue-span-aware preference learning in generating polite, persuasive, and coherent negotiation dialogues that support successful and balanced workplace negotiation outcomes.

6.2 Human Evaluation Results

Table 4 presents human evaluation results for DIPLOMAT and baseline models on the PROWESS dataset. Overall, DIPLOMAT achieves the highest scores across all metrics, demonstrating its ability to generate fluent (F), coherent (CC), and engaging (E) negotiation dialogues while effectively applying negotiation and persuasion strategies (NSC, PSC) with appropriate politeness (PA). These strengths translate into higher bargaining efficacy (BE), fair outcomes (OF), stronger sociopsy-

Models	PPL ↓	B-4 ↑	BS-F1 ↑	D-2 ↑	R-LEN ↑	NSC ↑	PSC ↑	PA ↑	AR (%) ↑
INA	14.6	0.344	0.602	0.520	16.5	0.520	0.456	0.858	64.3
GENTEEL-NEGOTIATOR	8.7	0.428	0.598	0.615	17.0	0.592	0.536	0.875	75.7
ProCoT (GPT-5-mini)	10.4	0.351	0.575	0.575	15.4	0.562	0.492	0.865	69.4
Mistral-7B-Instruct-v0.3	22.4	0.229	0.569	0.352	14.2	0.412	0.372	0.841	60.2
Llama3.1-8B	31.2	0.392	0.568	0.390	13.9	0.418	0.386	0.835	34.8
Gemma-2-9B	11.4	0.286	0.571	0.370	15.1	0.414	0.382	0.862	32.5
Mistral-7B-Instruct-v0.3-SFT	9.1	0.373	0.601	0.594	16.3	0.558	0.490	0.863	68.6
Llama3.1-8B-SFT	10.8	0.126	0.609	0.593	17.3	0.498	0.448	0.864	66.9
Gemma-2-9B-SFT	19.8	0.368	0.607	0.613	16.4	0.498	0.446	0.865	64.1
Mistral-7B-Instruct-v0.3-SFT-DPO	8.8	0.407	0.606	0.615	18.5	0.598	0.528	0.869	73.4
Mistral-7B-Instruct-v0.3-SFT-DMPO	8.7	0.413	0.608	0.622	17.1	0.618	0.548	0.871	76.2
Mistral-7B-Instruct-v0.3-SFT-ETO	8.7	0.434	0.612	0.625	17.2	0.624	0.552	0.872	77.5
Mistral-7B-Instruct-v0.3-PositiveSFT	8.9	0.396	0.594	0.606	16.8	0.574	0.506	0.866	70.8
DIPLMAT	8.6	0.452	0.630	0.635	19.8	0.724	0.628	0.882	82.3

Table 3: Automatic evaluation results. Results are statistically significant at 5% significance level based on t-test (Welch, 1947).

Models	F ↑	CC ↑	E ↑	NSC ↑	PSC ↑	PA ↑	BE ↑	OF ↑	SC ↑	IE ↑
INA	3.27	3.11	3.18	2.58	2.46	4.25	2.88	2.69	3.01	3.14
GENTEEL-NEGOTIATOR	4.09	4.21	4.16	3.83	3.69	4.37	3.95	3.79	3.88	4.03
ProCoT (GPT-5-mini)	3.91	3.85	3.77	3.08	2.96	4.30	3.48	3.39	3.44	3.58
Mistral-7B-Instruct-v0.3	2.89	2.95	2.81	2.02	2.11	4.19	2.59	2.43	2.55	2.70
Llama3.1-8B	2.78	2.83	2.74	2.08	2.03	4.15	2.54	2.39	2.48	2.53
Gemma-2-9B	3.06	3.01	2.98	2.13	2.21	4.31	2.60	2.53	2.56	2.59
Mistral-7B-Instruct-v0.3-SFT	3.79	3.89	3.72	3.02	2.97	4.31	3.49	3.38	3.43	3.55
Llama3.1-8B-SFT	3.74	3.68	3.63	2.89	2.81	4.29	3.38	3.28	3.35	3.47
Gemma-2-9B-SFT	3.69	3.62	3.58	2.95	2.89	4.32	3.33	3.24	3.30	3.39
Mistral-7B-Instruct-v0.3-SFT-DPO	4.01	4.04	3.95	3.29	3.24	4.34	3.73	3.64	3.69	3.84
Mistral-7B-Instruct-v0.3-SFT-DMPO	4.08	4.14	4.02	3.41	3.35	4.35	3.79	3.72	3.77	3.89
Mistral-7B-Instruct-v0.3-SFT-ETO	4.13	4.18	4.09	3.47	3.43	4.36	3.84	3.77	3.82	3.94
Mistral-7B-Instruct-v0.3-PositiveSFT	3.88	3.94	3.83	3.14	3.08	4.34	3.58	3.52	3.57	3.68
DIPLMAT	4.23	4.27	4.18	4.54	3.69	4.41	4.60	3.91	4.29	4.16

Table 4: Human evaluation results. All metrics are rated on a scale of 1–5. Results are statistically significant at 5% significance level based on t-test (Welch, 1947).

chological closeness (SC), and more effective influence (IE), reflecting a holistic negotiation capability. Compared to the strongest baseline, Mistral-7B-Instruct-v0.3-SFT-ETO, DIPLMAT improves by 2.4%, 2.2%, 2.2%, 30.8%, 7.6%, 1.1%, 19.8%, 3.7%, 12.3%, and 5.6% in F, CC, E, NSC, PSC, PA, BE, OF, SC, and IE, respectively, reflecting enhanced fluency, coherence, engagement, strategy correctness, politeness, bargaining effectiveness, fairness, social closeness, and influence. These improvements highlight DIPLMAT’s ability to generate polite, persuasive, and effective workplace negotiation dialogues that sustain natural conversational flow, maintain appropriate strategy use, and consistently support mutually beneficial negotiation outcomes through dialogue-span-aware preference learning.

6.3 Additional Analyses

We include more analyses: (1) Ablation w.r.t. SFT and Preference Optimization, (2) Analysis on Pref-

erence Optimization Granularity, (3) Ablation on Intermediate Blocks in SFT, (4) Sensitivity Analysis on Utility Threshold, (5) Ablation on Utility Components, (6) Effect of Dialogue Span Length, (7) Interlocutor Selection for Preference Data Generation, (8) Out-of-Domain Evaluation, (9) Evaluation Across Negotiation Aspects, and (10) Case Study in Appendix D.

7 Conclusion

We presented DIPLMAT, a dialogue system designed for polite and persuasive workplace negotiation. For this task, we introduced PROWESS, a workplace negotiation dialogue dataset generated through a multi-agent framework. DIPLMAT leverages Dialogue-Span-aware DPO (DSA-DPO) objective to improve LLM-based agents in multi-turn negotiation settings, enabling coherent responses that balance politeness and persuasion. Evaluation on PROWESS establish DIPLMAT’s efficacy in generating polite and persuasive negotiation responses.

Limitations

While PROWESS provides a large-scale dataset for studying polite and persuasive workplace negotiations, it is generated using LLM-based multi-agent simulations rather than real human-human interactions. Although we employ diverse scenarios and extensive evaluations to improve realism, synthetic dialogues may not fully capture the variability and subtle social (politeness, persuasiveness, and negotiation) dynamics present in real-world interactions. Furthermore, negotiation strategies, persuasion strategies, and politeness levels are automatically annotated using LLM-based annotation agents. While inter-LLM agreement analysis indicates moderate to fair reliability of these annotations, some ambiguity in interpreting strategies or politeness levels may introduce noise. Finally, the dataset currently focuses on English workplace negotiation scenarios. The proposed generation framework can be extended to other languages and domains with minimal prompt modifications, and exploring multilingual and cross-cultural negotiation settings remains an important direction for future work.

From the modeling perspective, the proposed DIPLOMAT framework leverages the Dialogue-Span-Aware Direct Preference Optimization (DSA-DPO) objective, which assumes equal-length positive and negative dialogue spans for preference alignment and to simplify optimization process. However, this assumption may limit the granularity of preference alignment, as negative segments can include irrelevant or error-free turns or may fail to capture all problematic dialogue spans. Moreover, the current utility design favors successful agreements and does not explicitly model rational impasse recognition or graceful negotiation termination under hard constraints. Consequently, the model may exhibit a bias toward continued compromise-seeking behavior even in scenarios involving non-negotiable requirements or infeasible negotiation outcomes. Hence, developing methods that support preference alignment over variable-length dialogue spans, while also explicitly modeling dealbreakers and negotiation impasses, remains an open challenge for future work.

Moreover, while DIPLOMAT and the proposed DSA-DPO objective are designed to be domain-agnostic, our primary evaluation is conducted on workplace negotiation dialogues, as PROWESS is the only existing dataset that jointly models negotia-

tion strategies, persuasion, and politeness in professional interactions. To empirically test generalization beyond this setting, we additionally evaluate DIPLOMAT out-of-domain on the tourism-focused NeGoChat dataset, where it maintains strong negotiation, persuasion, and politeness performance without any domain-specific fine-tuning, demonstrating cross-domain transferability. The remaining limitation is therefore one of breadth rather than of evidence: our out-of-domain analysis covers an additional domain, and a systematic study spanning a wider range of negotiation domains and modalities remains valuable future work.

Ethics Statement

The development of dialogue systems for polite and persuasive workplace negotiations requires careful ethical consideration due to the inherent influence and bargaining dynamics involved in negotiation interactions. This work received approval from our Institutional Review Board (IRB). The proposed system is designed to support constructive and respectful negotiation by encouraging polite communication and responsible persuasion strategies that aim to facilitate mutually beneficial outcomes. Rather than promoting adversarial or manipulative bargaining, the system emphasizes collaborative negotiation and preserves user autonomy, ensuring that individuals remain free to accept or reject any proposed agreement according to their own interests.

Ethical considerations are also important in the construction and use of the dataset. Since the dialogues are generated through LLMs-based simulations, they may reflect biases present in the underlying training data. Such biases could manifest in the representation of workplace roles, negotiation behaviors, or persuasive strategies, potentially reinforcing stereotypes or unfair assumptions. Prior work (Taori and Hashimoto, 2023) has shown that synthetic data generation can amplify biases or create feedback loops if not carefully monitored. To mitigate these risks, the dataset construction process incorporates careful prompt design, human oversight, and evaluation to promote respectful, professional, and contextually appropriate negotiation dialogue.

Furthermore, such dialogue systems raise concerns about potential misuse, such as manipulating individuals or influencing decisions in ways that are not transparent or aligned with users' interests.

Our framework focuses on modeling ethical politeness and persuasion that supports cooperative decision-making and professional communication rather than coercive influence. We encourage future work to further investigate safeguards for responsible polite and persuasive negotiation dialogue generation, including transparency, fairness analysis, and bias mitigation.

Finally, human experts contributed to data validation, filtering, and evaluation, and were compensated in accordance with institutional guidelines. The dataset will be shared for research purposes with appropriate usage guidelines to ensure responsible use. We believe that advancing research on polite, persuasive negotiation dialogue systems can help develop conversational agents that promote constructive workplace communication and responsible decision support.

Acknowledgement

The research reported in this paper is an outcome of the project titled “Conversational Agents with Negotiation and Influencing Ability”, sponsored by Accenture Labs, Bangalore, India.

References

- Sahar Abdelnabi, Amr Goma, Sarath Sivaprasad, Lea Schönherr, and Mario Fritz. 2024. [Cooperation, competition, and maliciousness: Llm-stakeholders interactive negotiation](#). *Preprint*, arXiv:2309.17234.
- Wendi Adair, Tetsushi Okumura, and Jeanne Brett. 2001. [Negotiation behavior when cultures collide: The united states and japan](#). *The Journal of applied psychology*, 86:371–85.
- Zishan Ahmad, Suman Saurabh, Vaishakh Menon, Asif Ekbal, Roshni Ramnani, and Anutosh Maitra. 2023. [INA: An integrative approach for enhancing negotiation strategies with reward-based dialogue agent](#). In *Findings of the Association for Computational Linguistics: EMNLP 2023*, pages 2536–2549, Singapore. Association for Computational Linguistics.
- Nicholas Asher, Julie Hunter, Mathieu Morey, Farah Benamara, and Stergos Afantenos. 2016. Discourse structure and dialogue acts in multiparty dialogue: the stac corpus. In *10th International Conference on Language Resources and Evaluation (LREC 2016)*, pages 2721–2727.
- Penelope Brown and Stephen C. Levinson. 1987. *Politeness: Some Universals in Language Usage*. Cambridge University Press.
- Peter F Brown, Stephen A Della Pietra, Vincent J Della Pietra, Jennifer C Lai, and Robert L Mercer. 1992. An estimate of an upper bound for the entropy of english. *Computational Linguistics*, 18(1):31–40.
- Peter Carnevale and Dean Pruitt. 1992. [Negotiation and mediation](#). *Annual Review of Psychology*, 43:531–582.
- Kushal Chawla, Jaysa Ramirez, Rene Clever, Gale Lucas, Jonathan May, and Jonathan Gratch. 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*, pages 3167–3185, Online. Association for Computational Linguistics.
- Gheorghe Comanici, Eric Bieber, Mike Schaekermann, Ice Pasupat, Noveen Sachdeva, Inderjit Dhillon, Marcel Blistein, Ori Ram, Dan Zhang, Evan Rosen, et al. 2025. Gemini 2.5: Pushing the frontier with advanced reasoning, multimodality, long context, and next generation agentic capabilities. *arXiv preprint arXiv:2507.06261*.
- Yang Deng, Lizi Liao, Liang Chen, Hongru Wang, Wenqiang Lei, and Tat-Seng Chua. 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*, pages 10602–10621, Singapore. Association for Computational Linguistics.
- Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: pre-training of deep bidirectional transformers for language understanding. *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL-HLT 2019, Minneapolis, MN, USA, June 2-7, 2019, Volume 1 (Long and Short Papers)*, pages 4171–4186.
- Ritam Dutt, Sayan Sinha, Rishabh Joshi, Surya Shekhar Chakraborty, Meredith Riggs, Xinru Yan, Haogang Bao, and Carolyn Rose. 2021. [ResPer: Computationally modelling resisting strategies in persuasive conversations](#). In *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume*, pages 78–90, Online. Association for Computational Linguistics.
- Alice H Eagly and Shelly Chaiken. 1984. Cognitive theories of persuasion. In *Advances in experimental social psychology*, volume 17, pages 267–359. Elsevier.
- Alan Fowler. 1998. *Negotiating, persuading and influencing*. CIPD Publishing.
- Yao Fu, Hao Peng, Tushar Khot, and Mirella Lapata. 2023. [Improving language model negotiation with self-play and in-context learning from ai feedback](#). *Preprint*, arXiv:2305.10142.

Aaron Grattafiori, Abhimanyu Dubey, Abhinav Jauhri, Abhinav Pandey, Abhishek Kadian, Ahmad Al-Dahle, Aiesha Letman, Akhil Mathur, Alan Schelten, Alex Vaughan, Amy Yang, Angela Fan, Anirudh Goyal, Anthony Hartshorn, Aobo Yang, Archi Mitra, Archie Sravankumar, Artem Korenev, Arthur Hinsvark, Arun Rao, Aston Zhang, Aurelien Rodriguez, Austen Gregerson, Ava Spataru, Baptiste Roziere, Bethany Biron, Binh Tang, Bobbie Chern, Charlotte Caucheteux, Chaya Nayak, Chloe Bi, Chris Marra, Chris McConnell, Christian Keller, Christophe Touret, Chunyang Wu, Corinne Wong, Cristian Canton Ferrer, Cyrus Nikolaidis, Damien Alonso, Daniel Song, Danielle Pintz, Danny Livshits, Danny Wyatt, David Esiobu, Dhruv Choudhary, Dhruv Mahajan, Diego Garcia-Olano, Diego Perino, Dieuwke Hupkes, Egor Lakomkin, Ehab AlBadawy, Elina Lobanova, Emily Dinan, Eric Michael Smith, Filip Radenovic, Francisco Guzmán, Frank Zhang, Gabriel Synnaeve, Gabrielle Lee, Georgia Lewis Anderson, Govind Thattai, Graeme Nail, Gregoire Mialon, Guan Pang, Guillem Cucurell, Hailey Nguyen, Hannah Korevaar, Hu Xu, Hugo Touvron, Iliyan Zarov, Imanol Arrieta Ibarra, Isabel Kloumann, Ishan Misra, Ivan Evtimov, Jack Zhang, Jade Copet, Jaewon Lee, Jan Geffert, Jana Vranes, Jason Park, Jay Mahadeokar, Jeet Shah, Jelmer van der Linde, Jennifer Billock, Jenny Hong, Jenya Lee, Jeremy Fu, Jianfeng Chi, Jianyu Huang, Jiawen Liu, Jie Wang, Jiecao Yu, Joanna Bitton, Joe Spisak, Jongsoo Park, Joseph Rocca, Joshua Johnstun, Joshua Saxe, Junteng Jia, Kalyan Vasuden Alwala, Karthik Prasad, Kartikeya Upasani, Kate Plawiak, Ke Li, Kenneth Heafield, Kevin Stone, Khalid El-Arini, Krithika Iyer, Kshitiz Malik, Kuenley Chiu, Kunal Bhalla, Kushal Lakhotia, Lauren Rantala-Yearly, Laurens van der Maaten, Lawrence Chen, Liang Tan, Liz Jenkins, Louis Martin, Lovish Madaan, Lubo Malo, Lukas Blecher, Lukas Landzaat, Luke de Oliveira, Madeline Muzzi, Mahesh Pasupuleti, Mannat Singh, Manohar Paluri, Marcin Kardas, Maria Tsimpoukelli, Mathew Oldham, Mathieu Rita, Maya Pavlova, Melanie Kam-badur, Mike Lewis, Min Si, Mitesh Kumar Singh, Mona Hassan, Naman Goyal, Narjes Torabi, Nikolay Bashlykov, Nikolay Bogoychev, Niladri Chatterji, Ning Zhang, Olivier Duchenne, Onur Celebi, Patrick Alrassy, Pengchuan Zhang, Pengwei Li, Petar Vasic, Peter Weng, Prajjwal Bhargava, Pratik Dubal, Praveen Krishnan, Punit Singh Koura, Puxin Xu, Qing He, Qingxiao Dong, Ragavan Srinivasan, Raj Ganapathy, Ramon Calderer, Ricardo Silveira Cabral, Robert Stojnic, Roberta Raileanu, Rohan Maheswari, Rohit Girdhar, Rohit Patel, Romain Sauvestre, Ronnie Polidoro, Roshan Sumbaly, Ross Taylor, Ruan Silva, Rui Hou, Rui Wang, Saghar Hosseini, Sahana Chennabasappa, Sanjay Singh, Sean Bell, Seohyun Sonia Kim, Sergey Edunov, Shaoliang Nie, Sharan Narang, Sharath Rapparthi, Sheng Shen, Shengye Wan, Shruti Bhosale, Shun Zhang, Simon Vandenhende, Soumya Batra, Spencer Whitman, Sten Sootla, Stephane Collot, Suchin Gururangan, Sydney Borodinsky, Tamar Herman, Tara Fowler, Tarek Sheasha, Thomas Georgiou, Thomas Scialom, Tobias Speckbacher, Todor Mihaylov, Tong Xiao, Ujjwal

Karn, Vedanuj Goswami, Vibhor Gupta, Vignesh Ramanathan, Viktor Kerkez, Vincent Gouget, Virginie Do, Vish Vogeti, Vitor Albiero, Vladan Petrovic, Weiwei Chu, Wenhan Xiong, Wenyin Fu, Whitney Meers, Xavier Martinet, Xiaodong Wang, Xiaofang Wang, Xiaoqing Ellen Tan, Xide Xia, Xinfeng Xie, Xuchao Jia, Xuwei Wang, Yaelle Goldschlag, Yashesh Gaur, Yasmine Babaei, Yi Wen, Yiwen Song, Yuchen Zhang, Yue Li, Yuning Mao, Zacharie DelPierre Coudert, Zheng Yan, Zhengxing Chen, Zoe Papakipos, Aaditya Singh, Aayushi Srivastava, Abha Jain, Adam Kelsey, Adam Shajnfeld, Adithya Gangidi, Adolfo Victoria, Ahuva Goldstand, Ajay Menon, Ajay Sharma, Alex Boesenberg, Alexei Baevski, Allie Feinstein, Amanda Kallet, Amit Sangani, Amos Teo, Anam Yunus, Andrei Lupu, Andres Alvarado, Andrew Caples, Andrew Gu, Andrew Ho, Andrew Poulton, Andrew Ryan, Ankit Ramchandani, Annie Dong, Annie Franco, Anuj Goyal, Aparajita Saraf, Arkabandhu Chowdhury, Ashley Gabriel, Ashwin Bharambe, Assaf Eisenman, Azadeh Yazdan, Beau James, Ben Maurer, Benjamin Leonhardi, Bernie Huang, Beth Loyd, Beto De Paola, Bhargavi Paranjape, Bing Liu, Bo Wu, Boyu Ni, Braden Hancock, Bram Wasti, Brandon Spence, Brani Stojkovic, Brian Gamido, Britt Montalvo, Carl Parker, Carly Burton, Catalina Mejia, Ce Liu, Changhan Wang, Changkyu Kim, Chao Zhou, Chester Hu, Ching-Hsiang Chu, Chris Cai, Chris Tindal, Christoph Feichtenhofer, Cynthia Gao, Damon Civin, Dana Beaty, Daniel Kreymer, Daniel Li, David Adkins, David Xu, Davide Testuggine, Delia David, Devi Parikh, Diana Liskovich, Didem Foss, DingKang Wang, Duc Le, Dustin Holland, Edward Dowling, Eissa Jamil, Elaine Montgomery, Eleonora Presani, Emily Hahn, Emily Wood, Eric-Tuan Le, Erik Brinkman, Esteban Arcaute, Evan Dunbar, Evan Smothers, Fei Sun, Felix Kreuk, Feng Tian, Filippas Kokkinos, Firat Ozgenel, Francesco Caggioni, Frank Kanayet, Frank Seide, Gabriela Medina Florez, Gabriella Schwarz, Gada Badeer, Georgia Swee, Gil Halpern, Grant Herman, Grigory Sizov, Guangyi, Zhang, Guna Lakshminarayanan, Hakan Inan, Hamid Shojanazeri, Han Zou, Hannah Wang, Hanwen Zha, Haroun Habeeb, Harrison Rudolph, Helen Suk, Henry Aspegren, Hunter Goldman, Hongyuan Zhan, Ibrahim Damla, Igor Molybog, Igor Tufanov, Ilias Leontiadis, Irina-Elena Veliche, Itai Gat, Jake Weissman, James Geboski, James Kohli, Janice Lam, Japhet Asher, Jean-Baptiste Gaya, Jeff Marcus, Jeff Tang, Jennifer Chan, Jenny Zhen, Jeremy Reizenstein, Jeremy Teboul, Jessica Zhong, Jian Jin, Jingyi Yang, Joe Cummings, Jon Carvill, Jon Shepard, Jonathan McPhie, Jonathan Torres, Josh Ginsburg, Junjie Wang, Kai Wu, Kam Hou U, Karan Saxena, Kartikay Khad-elwal, Katayoun Zand, Kathy Matosich, Kaushik Veeraraghavan, Kelly Michelena, Keqian Li, Kiran Jagadeesh, Kun Huang, Kunal Chawla, Kyle Huang, Lailin Chen, Lakshya Garg, Lavender A, Leandro Silva, Lee Bell, Lei Zhang, Liangpeng Guo, Licheng Yu, Liron Moshkovich, Luca Wehrstedt, Madian Khabsa, Manav Avalani, Manish Bhatt, Martynas Mankus, Matan Hasson, Matthew Lennie, Matthias Reso, Maxim Groshev, Maxim Naumov,

- Maya Lathi, Meghan Keneally, Miao Liu, Michael L. Seltzer, Michal Valko, Michelle Restrepo, Mihir Patel, Mik Vyatskov, Mikayel Samvelyan, Mike Clark, Mike Macey, Mike Wang, Miquel Jubert Hermoso, Mo Metanat, Mohammad Rastegari, Munish Bansal, Nandhini Santhanam, Natascha Parks, Natasha White, Navyata Bawa, Nayan Singhal, Nick Egebo, Nicolas Usunier, Nikhil Mehta, Nikolay Pavlovich Laptev, Ning Dong, Norman Cheng, Oleg Chernoguz, Olivia Hart, Omkar Salpekar, Ozlem Kalinli, Parkin Kent, Parth Parekh, Paul Saab, Pavan Balaji, Pedro Rittner, Philip Bontrager, Pierre Roux, Piotr Dollar, Polina Zvyagina, Prashant Ratanchandani, Pritish Yuvraj, Qian Liang, Rachad Alao, Rachel Rodriguez, Rafi Ayub, Raghotham Murthy, Raghu Nayani, Rahul Mitra, Rangaprabhu Parthasarathy, Raymond Li, Rebekkah Hogan, Robin Battey, Rocky Wang, Russ Howes, Ruty Rinott, Sachin Mehta, Sachin Siby, Sai Jayesh Bondu, Samyak Datta, Sara Chugh, Sara Hunt, Sargun Dhillon, Sasha Sidorov, Satadru Pan, Saurabh Mahajan, Saurabh Verma, Seiji Yamamoto, Sharadh Ramaswamy, Shaun Lindsay, Shaun Lindsay, Sheng Feng, Shenghao Lin, Shengxin Cindy Zha, Shishir Patil, Shiva Shankar, Shuqiang Zhang, Shuqiang Zhang, Sinong Wang, Sneha Agarwal, Soji Sajuyigbe, Soumith Chintala, Stephanie Max, Stephen Chen, Steve Kehoe, Steve Satterfield, Sudarshan Govindaprasad, Sumit Gupta, Summer Deng, Sungmin Cho, Sunny Virk, Suraj Subramanian, Sy Choudhury, Sydney Goldman, Tal Remez, Tamar Glaser, Tamara Best, Thilo Koehler, Thomas Robinson, Tianhe Li, Tianjun Zhang, Tim Matthews, Timothy Chou, Tzook Shaked, Varun Vontimitta, Victoria Ajayi, Victoria Montanez, Vijai Mohan, Vinay Satish Kumar, Vishal Mangla, Vlad Ionescu, Vlad Poenaru, Vlad Tiberiu Mihailescu, Vladimir Ivanov, Wei Li, Wenchen Wang, Wenwen Jiang, Wes Bouaziz, Will Constable, Xiaocheng Tang, Xiaojian Wu, Xiaolan Wang, Xilun Wu, Xinbo Gao, Yaniv Kleinman, Yanjun Chen, Ye Hu, Ye Jia, Ye Qi, Yenda Li, Yilin Zhang, Ying Zhang, Yossi Adi, Youngjin Nam, Yu, Wang, Yu Zhao, Yuchen Hao, Yundi Qian, Yunlu Li, Yuzi He, Zach Rait, Zachary DeVito, Zef Rosnbrick, Zhaoduo Wen, Zhenyu Yang, Zhiwei Zhao, and Zhiyu Ma. 2024. [The llama 3 herd of models](#). Preprint, arXiv:2407.21783.
- Kenta Hama, Atsushi Otsuka, Masahiro Mizukami, Hiroaki Sugiyama, and Makoto Naka. 2025. [RaPSIL: A Preference-Guided interview agent for Rapport-Aware Self-Disclosure](#). In *Proceedings of the 26th Annual Meeting of the Special Interest Group on Discourse and Dialogue*, pages 355–369, Avignon, France. Association for Computational Linguistics.
- He He, Derek Chen, Anusha Balakrishnan, and Percy Liang. 2018. [Decoupling strategy and generation in negotiation dialogues](#). In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 2333–2343, Brussels, Belgium. Association for Computational Linguistics.
- Ari Holtzman, Jan Buys, Li Du, Maxwell Forbes, and Yejin Choi. 2019. The curious case of neural text degeneration. *arXiv preprint arXiv:1904.09751*.
- Albert Q. Jiang, Alexandre Sablayrolles, Arthur Mensch, Chris Bamford, Devendra Singh Chaplot, Diego de las Casas, Florian Bressand, Gianna Lengyel, Guillaume Lample, Lucile Saulnier, L  lio Renard Lavaud, Marie-Anne Lachaux, Pierre Stock, Teven Le Scao, Thibaut Lavril, Thomas Wang, Timoth  e Lacroix, and William El Sayed. 2023. [Mistral 7b](#). Preprint, arXiv:2310.06825.
- Sunghye Jung, Donghun Lee, Shinbok Lee, Gaeun Seo, Daniel Lee, Byeongil Ko, Junrae Cho, Kihyun Kim, EungGyun Kim, and Myeongcheol Shin. 2025. [DiaTool-DPO: Multi-turn direct preference optimization for tool-augmented large language models](#). In *Proceedings of the 26th Annual Meeting of the Special Interest Group on Discourse and Dialogue*, pages 397–416, Avignon, France. Association for Computational Linguistics.
- Diederik P Kingma and Jimmy Ba. 2014. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.
- Vasily Konovalov, Ron Artstein, Oren Melamud, and Ido Dagan. 2016. [The negochat corpus of human-agent negotiation dialogues](#). In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC’16)*, pages 3141–3145, Portoro , Slovenia. European Language Resources Association (ELRA).
- Alison Ledgerwood, Shelly Chaiken, Deborah H Gruenfeld, and Charles M Judd. 2006. Changing minds: Persuasion in negotiation and conflict resolution. In Morton Deutsch, Peter T Coleman, and Eric C Marcus, editors, *The Handbook of Conflict Resolution: Theory and Practice*, 2nd edition, pages 455–485. Jossey-Bass, San Francisco, CA.
- Alice J. Lee, Malia F. Mason, and Claire S. Malcomb. 2021. Beyond cheap talk accounts: A theory of politeness in negotiations. *Research in Organizational Behavior*, 41:100154.
- Mike Lewis, Denis Yarats, Yann Dauphin, Devi Parikh, and Dhruv Batra. 2017. [Deal or no deal? end-to-end learning of negotiation dialogues](#). In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 2443–2453, Copenhagen, Denmark. Association for Computational Linguistics.
- Jiwei Li, Michel Galley, Chris Brockett, Jianfeng Gao, and Bill Dolan. 2015. A diversity-promoting objective function for neural conversation models. *arXiv preprint arXiv:1510.03055*.
- Yu Li, Kun Qian, Weiyan Shi, and Zhou Yu. 2020. End-to-end trainable non-collaborative dialog system. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pages 8293–8302.
- Yossi Maaravi, Orly Idan, and Guy Hochman. 2019. [And sympathy is what we need my friend—polite requests improve negotiation results](#). *PLOS ONE*, 14(3):1–22.

- Amogh Mannekote, Bonnie J. Dorr, and Kristy Elizabeth Boyer. 2023. [Agreement tracking for multi-issue negotiation dialogues](#). *Preprint*, arXiv:2307.06524.
- Bryan McEleney and Gregory O'Hare. 2003. [An agent design for effective negotiation dialogues](#). In *Proceedings of the 2003 EACL Workshop on Dialogue Systems: interaction, adaptation and styles of management*.
- Mary L. McHugh. 2012. Interrater reliability: the kappa statistic. *Biochemia medica*, 22(3):276–282.
- Sarah Monson. 2022. [Pampering customers with sweet talk: Negotiating politeness and \(mock\) conflict talk in akan marketplace bargaining](#). *Journal of Pragmatics*, 191:240–255.
- John F. Nash. 1950. [The bargaining problem](#). *Econometrica*, 18(2):155–162.
- OpenAI, Josh Achiam, Steven Adler, Sandhini Agarwal, Lama Ahmad, Ilge Akkaya, Florencia Leoni Aleman, Diogo Almeida, Janko Altschmidt, Sam Altman, Shyamal Anadkat, Red Avila, Igor Babuschkin, Suchir Balaji, Valerie Balcom, Paul Baltescu, Haiming Bao, Mohammad Bavarian, Jeff Belgum, Irwan Bello, Jake Berdine, Gabriel Bernadett-Shapiro, Christopher Berner, Lenny Bogdonoff, Oleg Boiko, Madelaine Boyd, Anna-Luisa Brakman, Greg Brockman, Tim Brooks, Miles Brundage, Kevin Button, Trevor Cai, Rosie Campbell, Andrew Cann, Brittany Carey, Chelsea Carlson, Rory Carmichael, Brooke Chan, Che Chang, Fotis Chantzis, Derek Chen, Sully Chen, Ruby Chen, Jason Chen, Mark Chen, Ben Chess, Chester Cho, Casey Chu, Hyung Won Chung, Dave Cummings, Jeremiah Currier, Yunxing Dai, Cory Decareaux, Thomas Degry, Noah Deutsch, Damien Deville, Arka Dhar, David Dohan, Steve Dowling, Sheila Dunning, Adrien Ecoffet, Atty Eleti, Tyna Eloundou, David Farhi, Liam Fedus, Niko Felix, Simón Posada Fishman, Juston Forte, Isabella Fulford, Leo Gao, Elie Georges, Christian Gibson, Vik Goel, Tarun Gogineni, Gabriel Goh, Rapha Gontijo-Lopes, Jonathan Gordon, Morgan Grafstein, Scott Gray, Ryan Greene, Joshua Gross, Shixiang Shane Gu, Yufei Guo, Chris Hallacy, Jesse Han, Jeff Harris, Yuchen He, Mike Heaton, Johannes Heidecke, Chris Hesse, Alan Hickey, Wade Hickey, Peter Hoeschele, Brandon Houghton, Kenny Hsu, Shengli Hu, Xin Hu, Joost Huizinga, Shantanu Jain, Shawn Jain, Joanne Jang, Angela Jiang, Roger Jiang, Haozhun Jin, Denny Jin, Shino Jomoto, Billie Jonn, Heewoo Jun, Tomer Kaftan, Łukasz Kaiser, Ali Kamali, Ingmar Kanitscheider, Nitish Shirish Keskar, Tabarak Khan, Logan Kilpatrick, Jong Wook Kim, Christina Kim, Yongjik Kim, Jan Hendrik Kirchner, Jamie Kiros, Matt Knight, Daniel Kokotajlo, Łukasz Kondraciuk, Andrew Kondrich, Aris Konstantinidis, Kyle Kosic, Gretchen Krueger, Vishal Kuo, Michael Lampe, Ikai Lan, Teddy Lee, Jan Leike, Jade Leung, Daniel Levy, Chak Ming Li, Rachel Lim, Molly Lin, Stephanie Lin, Mateusz Litwin, Theresa Lopez, Ryan Lowe, Patricia Lue, Anna Makanju, Kim Malfacini, Sam Manning, Todor Markov, Yaniv Markovski, Bianca Martin, Katie Mayer, Andrew Mayne, Bob McGrew, Scott Mayer McKinney, Christine McLeavey, Paul McMillan, Jake McNeil, David Medina, Aalok Mehta, Jacob Menick, Luke Metz, Andrey Mishchenko, Pamela Mishkin, Vinnie Monaco, Evan Morikawa, Daniel Mossing, Tong Mu, Mira Murati, Oleg Murk, David Mély, Ashvin Nair, Reiichiro Nakano, Rameev Nayak, Arvind Neelakantan, Richard Ngo, Hyeonwoo Noh, Long Ouyang, Cullen O'Keefe, Jakub Pachocki, Alex Paino, Joe Palermo, Ashley Pantuliano, Giambatista Parascandolo, Joel Parish, Emy Parparita, Alex Passos, Mikhail Pavlov, Andrew Peng, Adam Perelman, Filipe de Avila Belbute Peres, Michael Petrov, Henrique Ponde de Oliveira Pinto, Michael, Pokorny, Michelle Pokrass, Vitchyr H. Pong, Tolly Powell, Alethea Power, Boris Power, Elizabeth Proehl, Raul Puri, Alec Radford, Jack Rae, Aditya Ramesh, Cameron Raymond, Francis Real, Kendra Rimbach, Carl Ross, Bob Rotsted, Henri Roussez, Nick Ryder, Mario Saltarelli, Ted Sanders, Shibani Santurkar, Girish Sastry, Heather Schmidt, David Schnurr, John Schulman, Daniel Selsam, Kyla Sheppard, Toki Sherbakov, Jessica Shieh, Sarah Shoker, Pranav Shyam, Szymon Sidor, Eric Sigler, Maddie Simens, Jordan Sitkin, Katarina Slama, Ian Sohl, Benjamin Sokolowsky, Yang Song, Natalie Staudacher, Felipe Petroski Such, Natalie Summers, Ilya Sutskever, Jie Tang, Nikolas Tezak, Madeleine B. Thompson, Phil Tillet, Amin Tootoonchian, Elizabeth Tseng, Preston Tuggle, Nick Turley, Jerry Tworek, Juan Felipe Cerón Uribe, Andrea Vallone, Arun Vijayvergiya, Chelsea Voss, Carroll Wainwright, Justin Jay Wang, Alvin Wang, Ben Wang, Jonathan Ward, Jason Wei, CJ Weinmann, Akila Welihinda, Peter Welinder, Jiayi Weng, Lilian Weng, Matt Wiethoff, Dave Willner, Clemens Winter, Samuel Wolrich, Hannah Wong, Lauren Workman, Sherwin Wu, Jeff Wu, Michael Wu, Kai Xiao, Tao Xu, Sarah Yoo, Kevin Yu, Qiming Yuan, Wojciech Zaremba, Rowan Zellers, Chong Zhang, Marvin Zhang, Shengjia Zhao, Tianhao Zheng, Juntang Zhuang, William Zhuk, and Barret Zoph. 2024. [Gpt-4 technical report](#). *Preprint*, arXiv:2303.08774.
- Long Ouyang, Jeff Wu, Xu Jiang, Diogo Almeida, Carroll L. Wainwright, Pamela Mishkin, Chong Zhang, Sandhini Agarwal, Katarina Slama, Alex Ray, John Schulman, Jacob Hilton, Fraser Kelton, Luke Miller, Maddie Simens, Amanda Askell, Peter Welinder, Paul Christiano, Jan Leike, and Ryan Lowe. 2022. [Training language models to follow instructions with human feedback](#). *Preprint*, arXiv:2203.02155.
- Jennifer R. Overbeck, Margaret A. Neale, and Cassandra L. Govan. 2010. [I feel, therefore you act: Intrapersonal and interpersonal effects of emotion on negotiation as a function of social power](#). *Organizational Behavior and Human Decision Processes*, 112(2):126–139.
- Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu. 2002. [Bleu: a method for automatic evaluation of machine translation](#). In *Proceedings of the*

40th Annual Meeting on Association for Computational Linguistics, ACL '02, page 311–318, USA. Association for Computational Linguistics.

Priyanshu Priya, Rishikant Chigrupaati, Mauajama Firdaus, and Asif Ekbal. 2025a. Genteel-negotiator: Llm-enhanced mixture-of-expert-based reinforcement learning approach for polite negotiation dialogue. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pages 25010–25018.

Priyanshu Priya, Saurav Dudhate, Desai Vishesh Yasheshbhai, and Asif Ekbal. 2025b. We argue to agree: Towards personality-driven argumentation-based negotiation dialogue systems for tourism. *arXiv preprint arXiv:2509.11118*.

Priyanshu Priya, Desai Vishesh Yasheshbhai, Ratnesh Kumar Joshi, Roshni Ramnani, Anutosh Maitra, Shubhashis Sengupta, and Asif Ekbal. 2024. 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*, pages 16566–16595.

Chris Provis. 2004. Negotiation, persuasion and argument. *Argumentation*, 18(1):95–112.

Rafael Rafailov, Archit Sharma, Eric Mitchell, Stefano Ermon, Christopher D. Manning, and Chelsea Finn. 2024. [Direct preference optimization: Your language model is secretly a reward model](#). *Preprint*, arXiv:2305.18290.

Ryan Shea, Aymen Kallala, Xin Lucy Liu, Michael W. Morris, and Zhou Yu. 2024. [Ace: A llm-based negotiation coaching system](#). *Preprint*, arXiv:2410.01555.

Wentao Shi, Mengqi Yuan, Junkang Wu, Qifan Wang, and Fuli Feng. 2024. [Direct multi-turn preference optimization for language agents](#). In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, pages 2312–2324, Miami, Florida, USA. Association for Computational Linguistics.

Yifan Song, Da Yin, Xiang Yue, Jie Huang, Sujian Li, and Bill Yuchen Lin. 2024. Trial and error: Exploration-based trajectory optimization of llm agents. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 7584–7600.

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

Gemma Team, Morgane Riviere, Shreya Pathak, Pier Giuseppe Sessa, Cassidy Hardin, Surya Bhupatiraju, Léonard Hussenot, Thomas Mesnard, Bobak Shahriari, Alexandre Ramé, Johan Ferret, Peter Liu, Pouya Tafti, Abe Friesen, Michelle Casbon, Sabela Ramos, Ravin Kumar, Charline Le Lan, Sammy Jerome, Anton Tsitsulin, Nino Vieillard,

Piotr Stanczyk, Sertan Girgin, Nikola Momchev, Matt Hoffman, Shantanu Thakoor, Jean-Bastien Grill, Behnam Neyshabur, Olivier Bachem, Alanna Walton, Aliaksei Severyn, Alicia Parrish, Aliya Ahmad, Allen Hutchison, Alvin Abdagic, Amanda Carl, Amy Shen, Andy Brock, Andy Coenen, Anthony Laforge, Antonia Paterson, Ben Bastian, Bilal Piot, Bo Wu, Brandon Royal, Charlie Chen, Chintu Kumar, Chris Perry, Chris Welty, Christopher A. Choquette-Choo, Danila Sinopalnikov, David Weinberger, Dimple Vijaykumar, Dominika Rogozińska, Dustin Herbison, Elisa Bandy, Emma Wang, Eric Noland, Erica Moreira, Evan Senter, Evgenii Eltyshiev, Francesco Visin, Gabriel Rasskin, Gary Wei, Glenn Cameron, Gus Martins, Hadi Hashemi, Hanna Klimczak-Plucińska, Harleen Batra, Harsh Dhand, Ivan Nardini, Jacinda Mein, Jack Zhou, James Svensson, Jeff Stanway, Jetha Chan, Jin Peng Zhou, Joana Carrasqueira, Joana Iljazi, Jocelyn Becker, Joe Fernandez, Joost van Amersfoort, Josh Gordon, Josh Lipschultz, Josh Newlan, Ju yeong Ji, Kareem Mohamed, Kartikeya Badola, Kat Black, Katie Millican, Keelin McDonell, Kelvin Nguyen, Kiranbir Sodhia, Kish Greene, Lars Lowe Sjoesund, Lauren Usui, Laurent Sifre, Lena Heuermann, Letícia Lago, Lilly McNealus, Livio Baldini Soares, Logan Kilpatrick, Lucas Dixon, Luciano Martins, Machel Reid, Manvinder Singh, Mark Iverson, Martin Görner, Mat Velloso, Mateo Wirth, Matt Davidow, Matt Miller, Matthew Rahtz, Matthew Watson, Meg Risdal, Mehran Kazemi, Michael Moynihan, Ming Zhang, Minsuk Kahng, Minwoo Park, Mofi Rahman, Mohit Khatwani, Natalie Dao, Nenshad Bardoliwalla, Nesh Devanathan, Neta Dumai, Nilay Chauhan, Oscar Wahltinez, Pankil Botarda, Parker Barnes, Paul Barham, Paul Michel, Pengchong Jin, Petko Georgiev, Phil Culliton, Pradeep Kuppala, Ramona Comanescu, Ramona Merhej, Reena Jana, Reza Ardeshtir Rokni, Rishabh Agarwal, Ryan Mullins, Samaneh Saadat, Sara Mc Carthy, Sarah Cogan, Sarah Perrin, Sébastien M. R. Arnold, Sebastian Krause, Shengyang Dai, Shruti Garg, Shruti Sheth, Sue Ronstrom, Susan Chan, Timothy Jordan, Ting Yu, Tom Eccles, Tom Hennigan, Tomas Kocisky, Tulsee Doshi, Vihan Jain, Vikas Yadav, Vilobh Meshram, Vishal Dharmadhikari, Warren Barkley, Wei Wei, Wenming Ye, Woohyun Han, Woosuk Kwon, Xiang Xu, Zhe Shen, Zhitao Gong, Zichuan Wei, Victor Cotruta, Phoebe Kirk, Anand Rao, Minh Giang, Ludovic Peran, Tris Warkentin, Eli Collins, Joelle Barral, Zoubin Ghahramani, Raia Hadsell, D. Sculley, Jeanine Banks, Anca Dragan, Slav Petrov, Oriol Vinyals, Jeff Dean, Demis Hassabis, Koray Kavukcuoglu, Clement Farabet, Elena Buchatskaya, Sebastian Borgeaud, Noah Fiedel, Armand Joulin, Kathleen Kenealy, Robert Dadashi, and Alek Andreev. 2024. [Gemma 2: Improving open language models at a practical size](#). *Preprint*, arXiv:2408.00118.

Kazunori Terada, Mitsuki Okazoe, and Jonathan Gratch. 2021. [Effect of politeness strategies in dialogue on negotiation outcomes](#). In *Proceedings of the 21st ACM International Conference on Intelligent Virtual*

- Agents, IVA '21, page 195–202, New York, NY, USA. Association for Computing Machinery.
- Leigh L. Thompson. 2005. *The Mind and Heart of the Negotiator*, 3rd edition. Pearson Prentice Hall.
- Xuwei Wang, Weiyan Shi, Richard Kim, Yoojung Oh, Sijia Yang, Jingwen Zhang, and Zhou Yu. 2019. [Persuasion for good: Towards a personalized persuasive dialogue system for social good](#). In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 5635–5649, Florence, Italy. Association for Computational Linguistics.
- Bernard L Welch. 1947. The generalization of ‘student’s’ problem when several different population variances are involved. *Biometrika*, 34(1-2):28–35.
- Thomas Wolf, Lysandre Debut, Victor Sanh, Julien Chaumond, Clement Delangue, Anthony Moi, Pierric Cistac, Tim Rault, Rémi Louf, Morgan Funtowicz, et al. 2019. Huggingface’s transformers: State-of-the-art natural language processing. *arXiv preprint arXiv:1910.03771*.
- Atsuki Yamaguchi, Kosui Iwasa, and Katsuhide Fujita. 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*, pages 745–757, Online. Association for Computational Linguistics.
- Runzhe Yang, Jingxiao Chen, and Karthik Narasimhan. 2021. [Improving dialog systems for negotiation with personality modeling](#). In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 681–693, Online. Association for Computational Linguistics.
- ZHANG Yuxian. 2013. The politeness principles in business negotiation. *Cross-Cultural Communication*, 9(4):50–56.
- Haolan Zhan, Yufei Wang, Zhuang Li, Tao Feng, Yuncheng Hua, Suraj Sharma, Lizhen Qu, Zhaleh Semnani Azad, Ingrid Zukerman, and Reza Haf. 2024. [Let’s negotiate! a survey of negotiation dialogue systems](#). In *Findings of the Association for Computational Linguistics: EACL 2024*, pages 2019–2031, St. Julian’s, Malta. Association for Computational Linguistics.
- Tianyi Zhang, Varsha Kishore, Felix Wu, Kilian Q Weinberger, and Yoav Artzi. 2019. Bertscore: Evaluating text generation with bert. *arXiv preprint arXiv:1904.09675*.
- Yiheng Zhou, He He, Alan W Black, and Yulia Tsvetkov. 2019. A dynamic strategy coach for effective negotiation. In *Proceedings of the 20th Annual SIGdial Meeting on Discourse and Dialogue*, pages 367–378.
- Zhuoran Zhuang, Ye Chen, Xia Zeng, Chao Luo, Luhui Liu, and Yihan Chen. 2025. [Teaching llm to be persuasive: Reward-enhanced policy optimization for alignment from heterogeneous rewards](#). *Preprint*, arXiv:2510.04214.

Frequently Asked Questions (FAQs)

* Why focus on both politeness and persuasion jointly in workplace negotiation?

→ Workplace negotiation involves balancing task success with maintaining professional relationships. Politeness mitigates interpersonal friction, fosters trust, and prevents escalation, while persuasion helps overcome resistance and align preferences toward mutually beneficial agreements. Treating them in isolation, as prior work has done, overlooks their complementary roles. Hence, jointly modeling both is essential for realistic and effective negotiation dialogue systems.

* Why use synthetic data generated by LLMs instead of real human–human negotiation dialogues?

→ Collecting large-scale, annotated, multi-turn workplace negotiation dialogues from real human interactions is resource-intensive and raises privacy concerns. Our multi-agent generation framework enables efficient production of high-quality, diverse dialogues with fine-grained annotations for negotiation strategies, persuasion strategies, and politeness levels, while closely emulating natural human negotiation dynamics. Our quantitative validation shows that model-generated utterances are often indistinguishable from human-written text.

* Why does DSA-DPO operate on dialogue spans rather than individual utterances or full dialogues?

→ Utterance-level DPO optimizes individual responses but fails to capture evolving negotiation strategies, polite framing, and persuasive justifications that unfold over multiple exchanges. Dialogue-level DPO considers entire conversations but is too coarse-grained, introducing training noise and obscuring the contribution of individual utterances. DSA-DPO addresses both limitations by identifying key dialogue spans that contribute most to negotiation outcomes, reducing noise from non-critical turns while enabling multi-turn preference alignment.

* Why are experiments conducted only on the proposed PROWESS dataset?

→ Our experiments focus on PROWESS because it is specifically designed to capture *polite and persuasive workplace negotiation*, which is not explicitly modeled in existing negotiation dialogue datasets. Prior datasets primarily focus on bargaining scenarios such as item allocation, e-commerce negotiation, or tourism planning, and therefore lack annotations and conversational structures that reflect professional workplace negotiation dynamics. Since the objectives of this work are to study politeness and persuasion-aware modeling in workplace negotiations, evaluating models on PROWESS provides a controlled and task-relevant benchmark. Nevertheless, the proposed framework and training methodology are general and can be applied to other negotiation dialogue datasets or domains.

* How are suboptimal utterances identified during preference data construction, given that negotiation errors are inherently ambiguous?

→ Unlike tasks with objectively defined errors (e.g., math or programming), suboptimal utterances in negotiation are context-dependent and subjective. We use Gemini-2.5-Pro with structured prompts to identify utterances that: (i) are pivotal for advancing the negotiation goal, (ii) exhibit low effectiveness across utility dimensions such as politeness trajectory, strategy alignment, and agreement quality, or (iii) contribute to negotiation breakdown. This utility-driven detection enables us to construct meaningful preference pairs for training.

* What is the utility score and how does it determine dialogue quality for preference learning?

→ The utility score is a composite metric that evaluates a dialogue across six dimensions: Politeness Trajectory (PT), Persuasion Strategy Alignment (PSA), Negotiation Strategy Alignment (NSA), Mutual Satisfaction (MS), Agreement Quality (AQ), and Negotiation Breakdown (NB). Dialogues with utility scores below a threshold (0.6) are treated as potential negative samples containing suboptimal negotiation behaviors, which are then paired with higher-scoring dialogues to construct dialogue-span-aware preference data for DSA-DPO training.

Appendix

A PROWESS Dataset Details

A.1 Annotation Schema Description

Table 5, Table 6, and Table 7 provide the definitions along with example utterances for different negotiation strategies, persuasion strategies, and politeness levels from the PROWESS dataset.

A.2 Details of Dataset Quality Assessment Criteria

1. **Scenario Consistency (SC)**: Does the dialogue stay aligned with the given negotiation scenario?
2. **Negotiation Strategy Correctness (NSC)**: Does each utterance employ the negotiation strategy correctly, given the speaker’s role (employer or candidate) and the current dialogue context?
3. **Persuasion Strategy Correctness (PSC)**: Does each utterance employ the persuasion strategy correctly, given the speaker’s role (employer or candidate) and the current dialogue context?
4. **Politeness Appropriateness (PA)**: Does each utterance demonstrate suitable level of politeness given the context, role, and interaction dynamics?
5. **Fairness (F)**: Does the outcome of the dialogue reflect a balanced, equitable, and mutually beneficial resolution?
6. **Coherence (C)**: Is the dialogue logically structured, with smooth progression and clear connections between utterances?
7. **Naturalness (N)**: Does the conversation resemble human negotiation in language, tone, and flow?
8. **Engagingness (E)**: Is the dialogue interesting, interactive, and rich enough to sustain the user’s attention throughout the negotiation?

A.3 Additional Analysis of PROWESS Dataset

Human-LLM Differentiation Assessment. To measure linguistic naturalness in PROWESS, we evaluate whether humans can distinguish model-generated utterances from human-rewritten counterparts. We randomly select 600 utterances from 240 dialogues, ensuring balanced representation from both employer and candidate roles and coverage across different stages of each dialogue. Two native English speakers from our research group manually rewrite each sampled utterance to preserve meaning while improving clarity, fluency,

and grammatical correctness, resulting in paired examples of model-generated and human-edited utterances. The same three human evaluators involved in the data quality assessment independently identify the model-generated utterances in each pair. Across the 600 utterance pairs, all evaluators correctly identified the model-generated sentence in 276 cases (46%), disagreed in 184 pairs (30.7%), and incorrectly selected the human-edited sentence in 140 pairs (23.3%), with individual evaluator accuracies of 0.731, 0.747, and 0.722. These results indicate that model-generated utterances in PROWESS are often indistinguishable from human-written text, demonstrating high linguistic naturalness and realism.

Annotation Reliability. We further evaluate the reliability of automatically generated annotations using a multi-LLM assessment framework. From a random sample of 240 dialogues, each utterance is independently annotated by three state-of-the-art LLMs: GPT-4o-mini (via the Dialogue Annotation Agent), Gemini-2.5-Pro, and LLaMA-3-8B, for negotiation strategies, persuasion strategies, and politeness levels. We use the same prompt template as the Dialogue Annotation Agent to ensure annotation consistency. The inter-LLM Kappa agreement scores (κ) (McHugh, 2012) are found to be 0.724 for negotiation strategies, 0.742 for persuasion strategies, and 0.803 for politeness levels. These substantial agreement scores indicate that the strategy and politeness taxonomies are well-defined, enabling consistent annotation while reflecting the inherent complexity and subtlety of polite, persuasive and negotiation behaviors.

Dataset Diversity. As illustrated in Figure 4, the PROWESS dataset exhibits balanced coverage across negotiation aspects, ensuring uniform representation and preventing topic skew. This balance enables robust evaluation of dialogue models across diverse negotiation scenarios. The distributions of negotiation strategies, persuasion strategies, and politeness levels are shown in Figures 5, 6, and 7, respectively. Overall, negotiation strategies such as a collaborative style, win-win framing, and active listening are more common, reflecting the objective of reaching mutually beneficial agreements. However, strategy usage varies across negotiation aspects, indicating context-sensitive negotiation behaviors. For instance, anchoring and MESO strategies are more common in compensation-related discussions, where concrete offers and trade-offs are central, whereas principled negotiation and

Negotiation Strategy	Definition	Example
Collaborative Style	Working jointly to find mutually beneficial outcomes and preserve relationships.	<i>Given my experience, I was hoping for a two-year vesting or removal of the one-year cliff.</i>
Active Listening	Demonstrating attentiveness and understanding to build rapport and trust.	<i>Hi Ben, I'm excited about the possibility of you joining Innovatech. Your expertise in AI is impressive!</i>
Win-Win Framing	Framing negotiation as a shared problem to solve for mutual benefit.	<i>I see where you're coming from. That would certainly provide immediate benefits, but let's talk about long-term gains.</i>
Principled Negotiation	Focusing on mutual interests and objective standards rather than positions.	<i>I really value our dialogue, Ben. How about we agree on the standard vesting, with a performance review at 18 months?</i>
Data-Driven Justification	Supporting negotiation points with evidence like market benchmarks and past performance.	<i>I understand your perspective. May I explain why we maintain this vesting structure? It promotes team stability and fairness.</i>
MESO	Proposing multiple offers of equal value to reveal priorities and increase agreement likelihood.	<i>I appreciate your dedication, but how about we explore a performance-based bonus structure instead? It could complement the standard vesting.</i>
Anchoring	Setting a strong initial offer to influence the negotiation range.	<i>Thank you, David. I believe the PMP certification is crucial for my role. I need full funding upfront to accept the offer.</i>
Door-in-the-Face	Starting with a larger request to make the actual target seem more acceptable.	<i>Absolutely, and I recognize its value. However, our budget constraints are quite tight, given our non-profit status.</i>
Reciprocal Concessions	Offering small concessions to encourage reciprocation from the other party.	<i>That sounds reasonable, but I still believe an even earlier vesting period could align our interests better.</i>
Credibility Assertion	Building trust by reinforcing personal or organizational credibility during negotiations.	<i>Ben, I appreciate your experience. However, our standard four-year vesting with a one-year cliff ensures fairness across our team.</i>
No Strategy	Indicates the absence of any explicit negotiation strategy.	<i>Thank you, Anya. I'm enthusiastic about the role, but I need to discuss the vesting schedule.</i>

Table 5: Definitions and example utterances of negotiation strategies in the PROWESS dataset.

credibility assertions are more common in role- or responsibility-focused dialogues that emphasize fairness and trust. A similar trend is observed for persuasion strategies: logical and outcome-oriented appeals are generally prevalent, while emotional appeals occur more often in personal discussions such as career development, and credibility or data-driven justifications are more prominent in finance- or policy-oriented negotiations. The distribution of politeness levels further shows a balanced mix of low, medium, and high politeness across aspects, with higher politeness typically appearing in sensitive or relationship-oriented contexts and more direct language used in transactional discussions. Overall, these patterns suggest that the dataset captures diverse and contextually adaptive polite and persuasive negotiation behaviors throughout interactions.

Comparison with Existing Negotiation Dialogue Datasets. We present a comparison of the proposed PROWESS dataset with existing publicly available negotiation dialogue datasets in Table 8. While prior datasets primarily focus on task-oriented bargaining scenarios such as item allocation, e-commerce



Figure 4: Negotiation aspects-wise dialogue distribution in the PROWESS dataset.

Persuasion Strategy	Definition	Example
Rapport Building	Establishing trust and a positive relationship to enhance receptiveness.	<i>Hi Ben, I'm excited about the possibility of you joining Innovatech. Your expertise in AI is impressive!</i>
Concern Addressing	Actively listening to and resolving objections or hesitations.	<i>I understand that the four-year vesting with a one-year cliff may feel lengthy, but let's discuss this further.</i>
Emotional Appeal	Connecting through emotions such as empathy, excitement, or urgency.	<i>I understand the importance of long-term commitment, but I believe my skills warrant quicker rewards.</i>
Credibility & Confidence	Demonstrating expertise and confidence to increase trust and influence.	<i>Hi Ben, I'm thrilled about the possibility of you joining Innovatech. Let's discuss your concerns regarding the equity vesting schedule.</i>
Data-Driven Persuasion	Using evidence, facts, and benchmarks to strengthen arguments.	<i>But wouldn't an accelerated vesting schedule incentivize me even more to contribute significantly right from the start?</i>
Problem-Solving Focus	Presenting ideas as solutions that address mutual challenges.	<i>I understand your perspective. May I explain why we maintain this vesting structure? It promotes team stability and fairness.</i>
Self-Interest Appeal	Framing arguments around how the outcome directly benefits the other party.	<i>That sounds reasonable, but I still believe an even earlier vesting period could align our interests better.</i>
Value Alignment	Linking your proposal to the other party's core values and principles.	<i>I assure you that we want to reward your hard work. Let's finalize a review after the first 18 months.</i>
Reputation Highlighting	Leveraging past achievements or organizational standing to reinforce trustworthiness.	<i>I can assure you, our track record supports rewarding outstanding performance with recognition and opportunities for early vesting.</i>
Future Vision Alignment	Connecting the proposal with shared long-term goals and aspirations.	<i>I really value our dialogue, Ben. How about we agree on the standard vesting, with a performance review at 18 months?</i>
No Strategy	Indicates the absence of any explicit persuasion strategy.	<i>Absolutely, and I see the value it brings. However, upfront funding is challenging for our tight budget.</i>

Table 6: Definitions and example utterances of persuasion strategies in the PROWESS dataset.

Politeness Level	Definition	Example
High_Polite	Highly courteous language emphasizing warmth and rapport.	<i>I'm glad to hear that, Maria. Your enthusiasm is contagious, and we're excited about your potential contributions.</i>
Moderate_Polite	Respectful and cooperative tone maintaining professional balance.	<i>I appreciate your concern, Ben. Our past hires have thrived under this structure, which builds trust and shared goals.</i>
Low_Polite	Direct and assertive language with minimal politeness markers.	<i>A partial stipend is better than nothing, but I need assurance of full funding eventually.</i>

Table 7: Definitions and example utterances of different politeness levels in the PROWESS dataset.

negotiation, or tourism planning, PROWESS introduces a dataset centered on *polite and persuasive workplace negotiation*, capturing interactions that emphasize professionalism, rapport building, and constructive influence through politeness and persuasion modeling in organizational contexts. Furthermore, PROWESS exhibits an average dialogue length of 21.58 utterances, which is substantially longer than most existing negotiation datasets, reflecting richer multi-turn interactions and more realistic negotiation dynamics. With 2,400 dialogues and over 51K utterances, the dataset provides a substantial collection of negotiation conversations within workplace settings, enabling the study of

negotiation techniques, persuasive communication, and politeness-aware dialogue modeling. These characteristics make PROWESS a valuable benchmark for developing and evaluating dialogue agents capable of conducting polite and persuasive negotiations in workplace environments.

B Evaluation of Gemini-2.5-Pro in Preference Data Construction

To assess the reliability of the LLM-based pipeline used for preference data construction, we conduct a manual evaluation of the outputs generated by Gemini-2.5-Pro. The evaluation of negotiation dialogues is inherently challenging due to the sub-

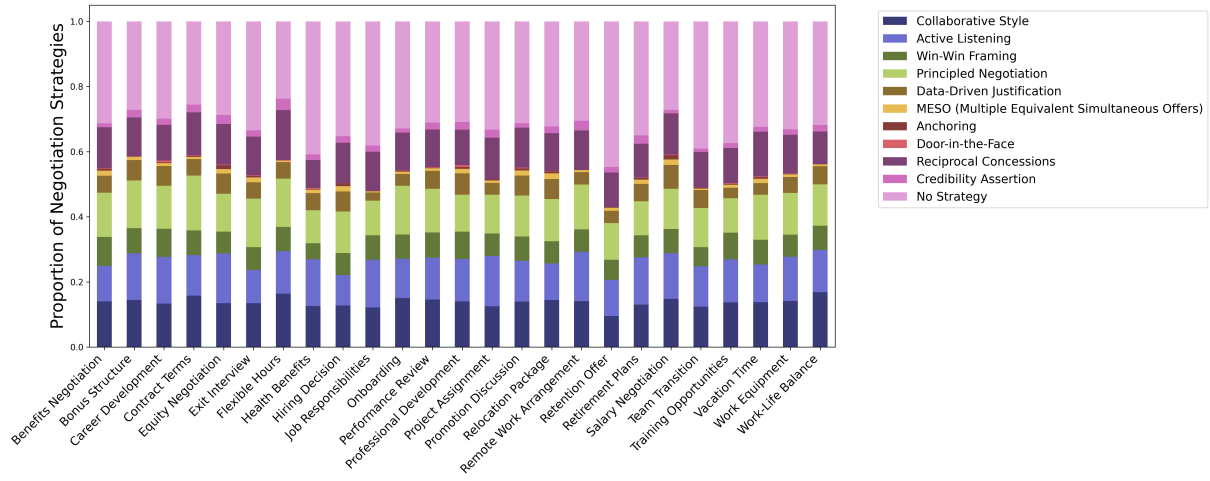


Figure 5: Distribution of negotiation strategies across different negotiation aspects in the PROWESS dataset.

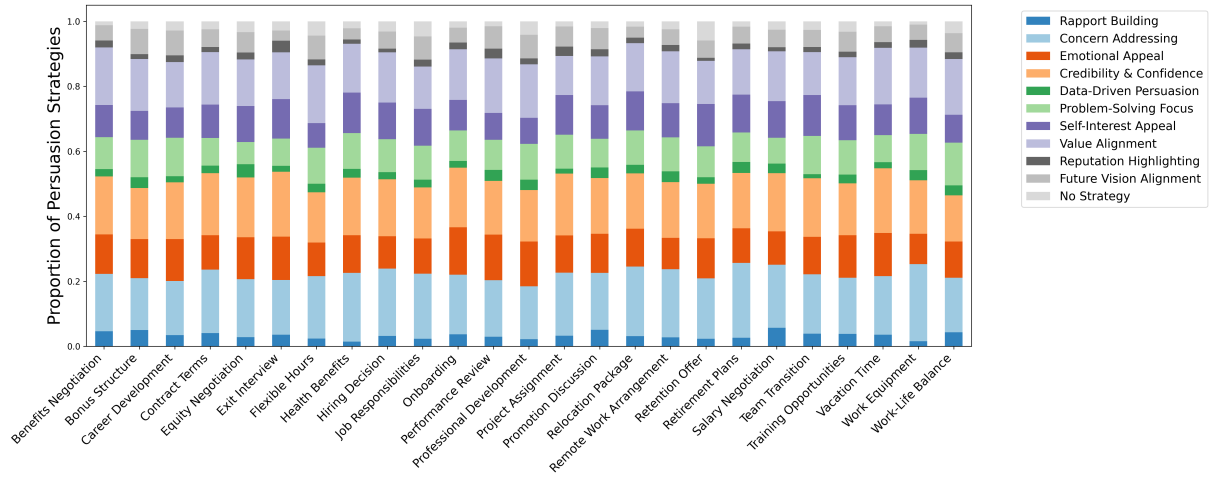


Figure 6: Distribution of persuasion strategies across different negotiation aspects in the PROWESS dataset.

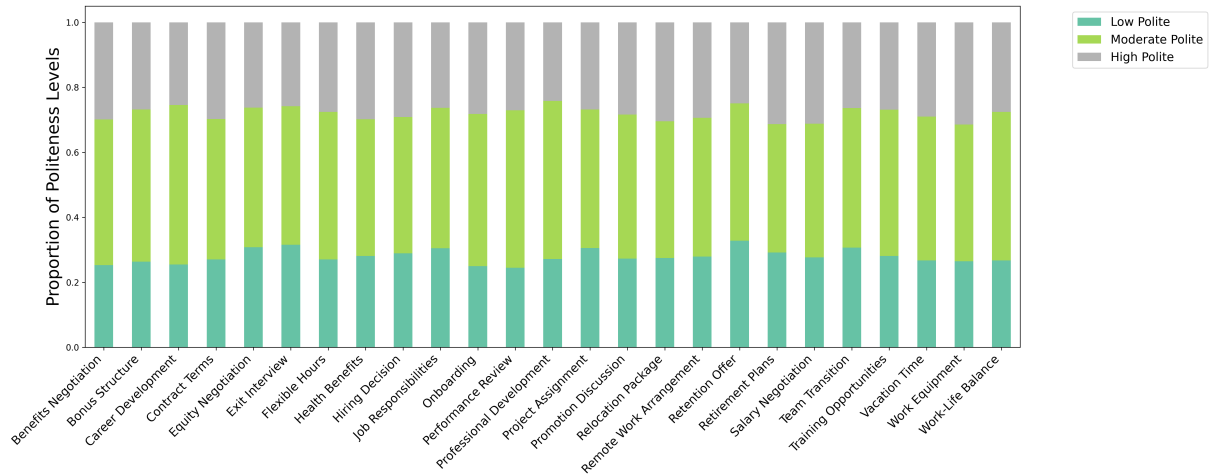


Figure 7: Distribution of politeness levels across different negotiation aspects in the PROWESS dataset.

jective nature of politeness, persuasion, and negotiation. As a result, strict notions of alignment are often difficult to establish. To account for this

ambiguity, we adopt three evaluation categories: *Aligned*, *Plausible*, and *Misaligned*. A prediction is labeled '*Aligned*' when the model's decision is

Dataset	Domain	Setting	#Dialogues	#Utterances	Avg. Dialogue Length
STAC (Asher et al., 2016)	Strategy Games	I	1,081	9,188	8.5
DealorNoDeal (Lewis et al., 2017)	Item Assignment	I	5,808	38,332	6.6
CraigslistBargain (He et al., 2018)	E-commerce	D	6,682	61,474	9.2
NegoCoach (Zhou et al., 2019)	Product Bargaining	D	300	—	—
Anti-Scam (Li et al., 2020)	E-commerce	D	220	2,739	12.45
CaSiNo (Chawla et al., 2021)	Item Assignment	I	1,030	11,948	11.6
JobInterview (Yamaguchi et al., 2021)	Job Interview	I	2,639	33,515	12.7
IND (Ahmad et al., 2023)	E-commerce	I	4,163	57,393	13.79
DEAL (Priya et al., 2024)	Tourism	I	1,291	18,932	14.66
PACT (Priya et al., 2025b)	Tourism	I	8,687	186,241	21.43
NEGOCHAT (Priya et al., 2025a)	Tourism	I	1,242	18,638	15.01
PROWESS (Proposed)	Workplace	I	2,400	51,796	21.58

Table 8: Comparison of negotiation dialogue datasets. I: integrative negotiation, D: distributive negotiation.

consistent with human judgment, ‘*Plausible*’ when the model’s decision appears reasonable but may not represent the most appropriate choice, and ‘*Misaligned*’ when the model’s output clearly contradicts human judgment.

For this analysis, we randomly sample 140 preference pairs from the constructed preference dataset. The same three human evaluators involved in the dataset quality assessment independently assess Gemini-2.5-Pro’s performance across three stages of the pipeline: utility evaluation, suboptimal utterance detection, and dialogue-span extraction. The average evaluation results are reported in Table 9. The evaluators observe that most *Plausible* cases arise when Gemini-2.5-Pro produces outputs that are reasonable in the given conversational context, but determining whether they represent the most informative or strategically appropriate choice remains difficult. Across all three components, *Misaligned* cases are relatively rare, indicating that the model seldom produces clearly incorrect judgments. Overall, these results suggest that Gemini-2.5-Pro provides effective outputs for utility evaluation, suboptimal utterance detection, and dialogue-span extraction within the preference data construction pipeline.

C Experiment Details

C.1 Implementation Details

All models are implemented in the PyTorch framework³, and we use transformer-based architectures from the Hugging Face library (Wolf et al., 2019). For OpenAI-based experiments, we access gpt-4o-mini and gpt-5-mini through the API. During the SFT stage, we employ QLoRA with rank $r=8$ and 4-bit quantization. The optimization

is performed using the AdamW optimizer (Kingma and Ba, 2014) with a learning rate of 1×10^{-5} , a 5% warm-up ratio, and a cosine decay learning rate schedule. The model is trained for 3 epochs with a batch size of 4 and a dropout rate of 0.2. For response generation, we use top- $k=30$ and top- $p=0.9$ sampling (Holtzman et al., 2019), a temperature of 0.7, and a repetition penalty of 1.03. In the DSA-DPO optimization stage, β is set to 0.1, and the learning rate is set to 5×10^{-6} with a cosine decay schedule and no warm-up phase. DSA-DPO training is conducted for 3 epochs with a batch size of 4. All experiments are performed on an NVIDIA A100 80GB GPU using CUDA 12.8.

C.2 Baseline Details

1. **INA** (Ahmad et al., 2023): A GPT-2-based dialogue agent trained using reinforcement learning with task-relevance rewards to dynamically adjust prices and manage bundle deals for effective integrative negotiations.
2. **GENTEEL-NEGOTIATOR** (Priya et al., 2025a): A mixture-of-expert-based reinforcement learning framework for polite negotiation dialogue that integrates negotiation, politeness, and keyterm experts through multi-task learning and RL-based rewards.
3. **ProCoT (GPT-5-mini)** (Deng et al., 2023): Prompts GPT-5-mini using chain-of-thought to generate a descriptive analysis for proactively planning the action for the next turn.
4. **Mistral-7B-Instruct-v0.3-SFT** (Jiang et al., 2023): Mistral-7B-Instruct-v0.3 fine-tuned on the PROWESS training data in a supervised setting.
5. **Llama3.1-8B-SFT** (Grattafiori et al., 2024): Llama-3.1-8B fine-tuned on the PROWESS training data in a supervised setting.

³<https://pytorch.org/>

Component	Aligned	Plausible	Misaligned
Utility Evaluation	100.9	31.4	7.7
Suboptimal Utterance Detection	98.6	33.1	8.3
Dialogue-Span Extraction	101.2	30.5	8.3

Table 9: Results of manual evaluation of Gemini-2.5-Pro’s performance in preference data construction. The values represent the average number of samples assigned to each evaluation category across the three evaluators.

6. **Gemma-2-9B-SFT** (Team et al., 2024): Gemma-2-9B fine-tuned on the PROWESS training data in a supervised setting.
7. **Mistral-7B-Instruct-v0.3-SFT-DPO** (Rafailov et al., 2024): Mistral-7B-Instruct-v0.3 first fine-tuned with SFT on the PROWESS training data, then further aligned using DPO, which directly optimizes the policy to prefer chosen over rejected responses using a pairwise preference loss.
8. **Mistral-7B-Instruct-v0.3-SFT-ETO** (Song et al., 2024): Mistral-7B-Instruct-v0.3 first fine-tuned with SFT on the PROWESS training data, then further aligned using Exploration-based Trajectory Optimization (ETO), which employs trial-and-error exploration to construct contrastive trajectory pairs for trajectory-level optimization.
9. **Mistral-7B-Instruct-v0.3-SFT-DMPO** (Shi et al., 2024): Mistral-7B-Instruct-v0.3 first fine-tuned with SFT on the PROWESS training data, then further aligned using Direct Multi-turn Preference Optimization (DMPO), which extends DPO to multi-turn dialogue settings by incorporating turn-level credit assignment.
10. **Mistral-7B-Instruct-v0.3-PositiveSFT**: Mistral-7B-Instruct-v0.3 fine-tuned using supervised fine-tuning exclusively on the positive (chosen) dialogue samples from the preference dataset.

C.3 Evaluation Metrics Details

Automatic Evaluation Metrics. Perplexity (PPL) (Brown et al., 1992) evaluates the model’s ability to predict the next token in a sequence, reflecting the fluency of generated responses. BLEU-4 (B-4) (Papineni et al., 2002) measures the degree of 4-gram overlap between generated outputs and reference responses. Distinct-2 (D-2) (Li et al., 2015) quantifies response diversity by calculating the ratio of unique bigrams within the generated text. BERTScore-F1 (BS-F1)

(Zhang et al., 2019)⁴ evaluates semantic similarity between generated and reference responses using contextual embeddings from BERT (Devlin et al., 2019). Negotiation Strategy Correctness (NSC) measures whether the generated response appropriately applies the intended negotiation strategy. Persuasion Strategy Correctness (PSC) evaluates whether the response correctly employs the intended persuasion strategy. Politeness Appropriateness (PA) assesses whether the generated response maintains an appropriate level of politeness. Agreement Rate (AR) measures the percentage of dialogues that result in a successful negotiation agreement between participants.

$$\text{NSC} = \mathbb{E}_{(\mathcal{C}_t, e_t)} [\mathbf{1} \{f_{\text{neg}}(e_t) = f_{\text{neg}}(\hat{e}_t)\}]$$

$$\text{PSC} = \mathbb{E}_{(\mathcal{C}_t, e_t)} [\mathbf{1} \{f_{\text{per}}(e_t) = f_{\text{per}}(\hat{e}_t)\}]$$

$$\text{PA} = \mathbb{E}_{(\mathcal{C}_t, e_t)} [\mathbf{1} \{f_{\text{pol}}(e_t) = f_{\text{pol}}(\hat{e}_t)\}]$$

$$\text{AR} = \frac{1}{N} \sum_{i=1}^N \mathbf{1} \{f_{\text{agr}}(d_i) = 1\}$$

where $\mathcal{C}_t$ denotes the conversational context up to turn t , e_t and $\hat{e}_t$ denote the ground-truth and generated employer responses at turn t , respectively, and $f_{\text{neg}}(\cdot)$, $f_{\text{per}}(\cdot)$, $f_{\text{pol}}(\cdot)$, and $f_{\text{agr}}(\cdot)$ denote the negotiation strategy, persuasion strategy, politeness, and agreement classifiers, respectively. $f_{\text{neg}}(\cdot)$, $f_{\text{per}}(\cdot)$, $f_{\text{pol}}(\cdot)$, and $f_{\text{agr}}(\cdot)$ achieve accuracies of 84.3%, 86.1%, 89.2%, and 91.4%, respectively, and macro-F1 scores of 81.7%, 83.4%, 86.8%, and 90.1%, respectively, on the PROWESS test set.

Human Evaluation Metrics. Fluency (F) measures the grammatical correctness and overall linguistic quality of the generated responses. Contextual Coherence (CC) evaluates whether the responses are logically connected and relevant to the ongoing dialogue context. Engagingness (E) reflects how interesting and interactive the conversation appears to the user. Bargaining Efficacy

⁴BERTScore: <https://huggingface.co/spaces/evaluate-metric/bertscore>

(BE) assesses the system’s capability to negotiate effectively by presenting arguments, incentives, or trade-offs that may influence the counterpart’s decisions. Outcome Fairness (OF) assesses whether the final negotiation outcome reflects a balanced, mutually acceptable agreement between participants. In addition, we evaluate NSC, PSC, and PA. While these metrics correspond to their automatic evaluation counterparts, they are assessed by human evaluators to provide a more reliable judgment of strategy usage and emotional appropriateness in the dialogue. Sociopsychological Closeness (SC) measures the extent to which polite responses foster perceived social and emotional proximity between participants. Influence Effectiveness (IE) evaluates how effectively persuasive responses shape the counterpart’s attitude, preferences, or decisions during the negotiation.

Human Evaluation Process. The human evaluation is performed by three independent evaluators⁵. Two evaluators hold Ph.D. degrees in Linguistics, while the third holds a postgraduate degree in Computer Science. All evaluators have prior experience with dialogue evaluation tasks. Before the evaluation begins, they are provided with instructions and background information regarding the negotiation scenarios. Each evaluator conducts multi-turn negotiation interactions with the system for a given scenario and completes 30 dialogues using different response sets, resulting in a total of 90 human-evaluated negotiation dialogues. After completing each interaction, the evaluators rate the dialogues on F, CC, E, NSC, PSC, PA, BE, OF, SC, and IE using a 1–5 scale (low to high). To assess inter-evaluator agreement, we compute Fleiss’ Kappa (κ) (McHugh, 2012), and obtain $\kappa = 0.76$ (F), 0.74 (CC), 0.75 (E), 0.80 (NSC), 0.81 (PSC), 0.83 (PA), 0.74 (BE), 0.73 (OF), 0.75 (SC), and 0.77 (IE), indicating substantial agreement among the evaluators.

D Additional Analysis

D.1 Ablation w.r.t. SFT and Preference Optimization

We perform an ablation study to evaluate the contributions of the core learning objectives in DIPLOMAT, namely supervised fine-tuning ($\mathcal{L}_{\text{SFT}}$)

and dialogue-span-aware direct preference optimization ($\mathcal{L}_{\text{DSA-DPO}}$). Table 10 reports results for ablated variants where one or both objectives are removed. The removal of $\mathcal{L}_{\text{SFT}}$ (-SFT) leads to substantial degradation across NSC, PSC, PA, and AR, highlighting the critical role of supervised fine-tuning in equipping the model with foundational negotiation skills and context-aware polite and persuasive behaviors. Further, omitting $\mathcal{L}_{\text{DSA-DPO}}$ (-DSA-DPO) also lowers performance, particularly in NSC and AR, indicating that preference-based optimization is essential for refining negotiation strategies over multi-turn dialogue spans. The joint removal of both objectives (-SFT + DSA-DPO) results in the sharpest decline, with PPL increasing and overall task completion metrics dropping considerably. This pattern confirms that SFT and dialogue-span-aware preference optimization are complementary: SFT provides foundational behavior modeling, while DSA-DPO further aligns the model with polite and persuasive negotiation outcomes, and both objectives are necessary to achieve the high performance observed in DIPLOMAT.

D.2 Analysis on Preference Optimization Granularity

To assess the impact of preference optimization granularity in DIPLOMAT, we compare three variants: (i) *utterance-level DPO*, where preferences are applied to individual utterances, (ii) *dialogue-level DPO*, where the entire positive and negative dialogue samples are used during preference optimization, and (iii) our *dialogue-span-aware DPO* (the proposed DIPLOMAT system). Table 11 shows the results across these settings. The utterance-level DPO improves over SFT alone but cannot fully capture multi-turn negotiation dynamics, limiting the model’s ability to consistently mirror politeness cues and adapt persuasive strategies over the course of a dialogue. As a result, gains in metrics such as NSC, PSC, PA, and AR are moderate. The dialogue-level DPO better models overall negotiation outcomes, leading to higher agreement rates, but lacks fine-grained supervision, which leads to less precise adaptation of politeness, and persuasion and negotiation strategy selection, sometimes generating generic or over-conceding responses. In contrast, dialogue-span-aware DPO (DIPLOMAT) consistently outperforms both alternatives, demonstrating that optimizing over targeted dialogue spans provides clearer and more actionable preference signals. This allows the model

⁵These evaluators are different from those involved in dataset preparation and are compensated according to institutional guidelines.

Models	PPL ↓	B-4 ↑	BS-F1 ↑	D-2 ↑	R-LEN ↑	NSC ↑	PSC ↑	PA ↑	AR (%) ↑
DIPLOMAT	8.6	0.452	0.630	0.635	19.8	0.724	0.628	0.882	82.3
- SFT	12.5	0.355	0.598	0.580	18.5	0.605	0.540	0.870	73.0
- DSA-DPO	9.1	0.373	0.601	0.594	16.3	0.558	0.490	0.863	68.6
- (SFT + DSA-DPO)	22.4	0.229	0.569	0.352	14.2	0.412	0.372	0.841	60.2

Table 10: Ablation of SFT and preference optimization objective in the DIPLOMAT. - signifies the removal of the component.

to align responses with strategic negotiation behaviors, maintain dynamic politeness trajectories, and adjust persuasion strategies in context-sensitive ways across multiple turns, resulting in more effective and nuanced negotiation outcomes.

D.3 Ablation on Intermediate Blocks in SFT

To quantify the contribution of the intermediate reasoning blocks during supervised fine-tuning (SFT) in DIPLOMAT, we conduct ablations by removing the candidate inference block, the employer inference block, and both blocks, and evaluate their impact on DIPLOMAT’s performance. Table 12 reports the results across different configurations where one or more blocks are removed. The removal of any inference block degrades key metrics, including NSC, PSC, PA, and AR. In particular, removing the candidate inference block reduces the model’s ability to anticipate the candidate’s strategies and politeness levels, leading to lower NSC, PSC, PA, and AR. Ablating the employer inference block decreases the model’s strategy selection capability, further impacting the overall performance of DIPLOMAT. Omitting both blocks compounds these effects, resulting in the largest declines across all metrics. Notably, language quality metrics (PPL, B-4, D-2, BS-F1) remain relatively stable, indicating that the performance drop is not due to generic language degradation but rather to the loss of structured reasoning critical for strategy adaptation, politeness mirroring, and effective negotiation outcomes. These results highlight that both candidate and employer inference blocks are essential for learning nuanced, context-aware polite and persuasive behavior during negotiation.

D.4 Sensitivity Analysis on Utility Threshold

The utility threshold τ used for negative dialogue selection controls the balance between preference signal quality and dataset coverage. Specifically, dialogues satisfying $U(d) < \tau$ are treated as potential negative samples during preference data construction. To analyze the sensitivity of DIPLOMAT

to this design choice, we evaluate threshold values $\tau \in \{0.4, 0.5, 0.6, 0.7, 0.8\}$ and report the resulting preference dataset size and model performance in Table 13.

The results show that DIPLOMAT achieves the best overall performance at $\tau = 0.6$, obtaining the highest NSC, PSC, PA, and AR scores while also maintaining strong language quality and diversity. Lower thresholds (e.g., $\tau = 0.4$ and $\tau = 0.5$) produce a larger number of preference pairs but include fewer strongly suboptimal dialogues, thereby weakening the contrastive preference signal. In contrast, higher thresholds (e.g., $\tau = 0.7$ and $\tau = 0.8$) introduce borderline or noisy negative samples that reduce preference quality and degrade downstream negotiation performance. Overall, the analysis indicates that DIPLOMAT is robust to moderate threshold variations, while highlighting the importance of balancing preference signal strength and dataset quality during dialogue-span-aware preference construction.

D.5 Ablation on Utility Components

To quantify the contribution of each utility component in dialogue-span-aware preference optimization, we perform ablations by removing individual components (PT, PSA, NSA, MS, AQ, NB) and evaluate the resulting performance of DIPLOMAT. Table 14 presents the outcomes across these configurations. Removing any component generally leads to reductions in key metrics such as NSC, PSC, PA, and AR, underscoring the critical role of each signal in guiding polite and persuasive negotiation behavior. Specifically, omitting PT or PSA diminishes the model’s ability to maintain politeness levels and align persuasive strategies with the candidate, eventually lowering NSC and AR. Excluding MS or AQ primarily affects overall negotiation quality and agreement outcomes, while removing NB slightly decreases AR and NSC by reducing the model’s ability to prevent negotiation breakdowns. Importantly, language quality metrics (PPL, B-4, D-2, BS-F1) remain largely stable, indi-

Models	PPL ↓	B-4 ↑	BS-F1 ↑	D-2 ↑	R-LEN ↑	NSC ↑	PSC ↑	PA ↑	AR (%) ↑
Mistral-7B-Instruct-v0.3-SFT	9.1	0.373	0.601	0.594	16.3	0.558	0.490	0.863	68.6
+ Utterance-level DPO	8.8	0.407	0.606	0.615	18.5	0.598	0.528	0.869	73.4
+ Dialogue-level DPO	8.7	0.410	0.628	0.615	18.9	0.612	0.540	0.875	74.2
+ Dialogue-Span-aware DPO (DIPLOMAT)	8.6	0.452	0.630	0.635	19.8	0.724	0.628	0.882	82.3

Table 11: Analysis on the granularity of preference optimization in DIPLOMAT. + signifies the addition of the component.

Models	PPL ↓	B-4 ↑	BS-F1 ↑	D-2 ↑	R-LEN ↑	NSC ↑	PSC ↑	PA ↑	AR (%) ↑
DIPLOMAT	8.6	0.452	0.630	0.635	19.8	0.724	0.628	0.882	82.3
- Candidate inference block	9.1	0.401	0.625	0.622	19.9	0.698	0.605	0.867	78.2
- Employer inference block	9.3	0.392	0.623	0.618	19.9	0.692	0.601	0.861	76.5
- Both inference blocks	9.8	0.372	0.618	0.610	20.0	0.678	0.592	0.852	74.1

Table 12: Effect of intermediate reasoning blocks on DIPLOMAT’s negotiation performance. - signifies the removal of the component.

cating that these declines reflect behavioral rather than linguistic deficiencies. Overall, these findings demonstrate that all utility components jointly enable DIPLOMAT to learn nuanced, context-aware, and strategically effective polite and persuasive negotiation behaviors.

D.6 Effect of Dialogue Span Length

To analyze the effect of dialogue span length on DIPLOMAT, we conduct an analysis using dialogue spans of varying lengths. Table 15 reports the results across different dialogue-spans lengths. As compared to dialogue-span of length [0, 0] (Mistral-7B-Instruct-v0.3-finetune), incorporating preference optimization with dialogue-spans of length [1, 1] consistently improves performance across most metrics, highlighting the benefit of preference-based alignment. Further increasing the span length from [2, 2] to [6, 6] further yields steady gains in both language quality and task completion, suggesting that multi-turn spans capture richer contextual signals for modeling politeness, persuasion, and negotiation. However, fixed span lengths may still include irrelevant dialogue spans. Motivated by this observation, DIPLOMAT dynamically selects informative dialogue spans using Gemini-2.5-pro, achieving the best overall performance across all metrics.

We further analyze variable-length preference spans to evaluate the impact of the equal-length constraint in DSA-DPO. Specifically, we experiment with asymmetric span settings [2, 4] and [4, 2], where the first and second indices denote the lengths of the positive and negative dialogue spans, respectively. Compared to their equal-length

counterparts [2, 2] and [4, 4], both variable-length settings consistently degrade performance across all evaluation metrics, including NSC, PSC, PA, and AR. The degradation is particularly severe for [4, 2], suggesting that mismatched span granularity weakens preference alignment by making the contrasted conversational behaviors less semantically comparable. These results empirically support the equal-length design in DSA-DPO by showing that balanced span comparison produces a more reliable preference signal. Empirically, we also observe that most informative preference spans naturally fall within short contiguous ranges (2-4 turns), which further suggests that the equal-length constraint rarely becomes a limitation in practice.

D.7 Interlocutor Selection for Preference Data Generation

We investigate the impact of interlocutor choice on generating preference data for DIPLOMAT. Specifically, we compare dialogues collected from interactions between two π_{SFT} models, one as candidate and the other as employer, with those generated between GPT-4o-mini (acting as candidate) and π_{SFT} (acting as employer). As reported in Table 16, models trained solely on $\pi_{\text{SFT}}-\pi_{\text{SFT}}$ dialogues achieve moderate gains but consistently underperform models trained with GPT-4o-mini interactions. Leveraging GPT-4o-mini as the candidate provides richer, higher-quality dialogues, yielding substantial improvements in both language quality and task completion metrics. These results indicate that diverse, capable interlocutors enable DIPLOMAT to adopt more effective polite and persuasive negotiation behavior, thereby improving

τ	#Preference Pairs	PPL ↓	B-4 ↑	BS-F1 ↑	D-2 ↑	R-LEN ↑	NSC ↑	PSC ↑	PA ↑	AR (%) ↑
0.4	935	9.9	0.438	0.624	0.627	18.7	0.681	0.596	0.878	78.9
0.5	822	10.7	0.446	0.627	0.632	19.2	0.706	0.614	0.880	80.7
0.6	700	8.6	0.452	0.630	0.635	19.8	0.724	0.628	0.882	82.3
0.7	391	9.8	0.444	0.626	0.631	19.1	0.702	0.611	0.879	80.4
0.8	548	9.1	0.431	0.621	0.623	18.4	0.672	0.587	0.876	77.6

Table 13: Sensitivity analysis on the utility threshold τ used for negative dialogue selection.

Models	PPL ↓	B-4 ↑	BS-F1 ↑	D-2 ↑	R-LEN ↑	NSC ↑	PSC ↑	PA ↑	AR (%) ↑
DIPLMAT	8.6	0.452	0.630	0.635	19.8	0.724	0.628	0.882	82.3
- PT	8.9	0.411	0.629	0.633	19.9	0.712	0.620	0.871	80.3
- PSA	8.8	0.402	0.628	0.632	19.9	0.710	0.618	0.860	79.1
- NSA	8.9	0.413	0.627	0.630	19.8	0.708	0.615	0.859	79.7
- MS	8.7	0.424	0.629	0.634	19.8	0.718	0.623	0.871	81.2
- AQ	8.7	0.427	0.628	0.632	19.8	0.716	0.620	0.861	80.0
- NB	8.8	0.418	0.627	0.631	19.8	0.710	0.618	0.862	79.5

Table 14: Effect of utility function components on DIPLMAT’s negotiation performance. - signifies the removal of the component.

overall performance.

D.8 Out-of-Domain Evaluation

To assess the domain generalization capability of DIPLMAT, we conduct out-of-domain evaluation on NeGoChat (Priya et al., 2025a), which contains tourism-focused negotiation dialogues, and JobInterview (Yamaguchi et al., 2021), comprising human-human job interview negotiation dialogues. Both domains differ substantially from the workplace negotiation scenarios in PROWESS. Importantly, DIPLMAT is evaluated on both datasets without any domain-specific fine-tuning.

As shown in Table 18, DIPLMAT maintains strong performance across both out-of-domain settings. On NeGoChat, the model achieves competitive lexical and semantic correspondence (B-4 = 0.351; BS-F1 = 0.602), substantial response diversity (D-2 = 0.372), and a response length of 30.16. More importantly, it preserves socially-aware negotiation behaviors, achieving NSC = 0.682, PSC = 0.593, PA = 0.825, and an agreement rate of 78.5%. These results indicate that negotiation strategy, persuasion, and politeness behaviors learned from workplace negotiations transfer effectively to the tourism domain.

The generalization trend is also evident on JobInterview. DIPLMAT obtains a lower perplexity (11.6), together with higher B-4 (0.367), BS-F1 (0.658), and D-2 (0.391), indicating fluent, semantically aligned, and diverse generation under a substantially different interaction setting. The model also achieves the highest politeness appropriateness

(PA = 0.849) and agreement rate (AR = 80.2%) across the two out-of-domain datasets. Although negotiation and persuasion strategy correctness are slightly lower than on NeGoChat (NSC = 0.645; PSC = 0.573), they remain consistently strong, suggesting that the learned strategic behaviors are largely preserved under the job-interview domain shift. The shorter response length (R-LEN = 12.65) further reflects the distinct response characteristics of this domain rather than a degradation in generation quality. Overall, the consistent performance across two substantially different negotiation domains provides evidence that DIPLMAT does not merely learn workplace-specific response patterns. Instead, the proposed DSA-DPO objective appears to transfer negotiation strategies, persuasive behaviors, and politeness preferences across domains while maintaining fluent and semantically appropriate generation. These findings support the effectiveness of dialogue-span-aware preference learning in promoting robust out-of-domain generalization of polite and persuasive negotiation behavior.

D.9 Evaluation across Negotiation Aspects

Table 17 reports per-aspect performance of DIPLMAT compared to its SFT counterpart across task completion metrics (NSC, PSC, PA, and AR). Across all 25 negotiation aspects, DIPLMAT consistently outperforms SFT. NSC improvements range from +0.060 to +0.326, with the largest gains observed in promotion discussion, career development, contract terms, and flexible hours, all of which require sustained multi-turn strategy exe-

Models	PPL ↓	B-4 ↑	BS-F1 ↑	D-2 ↑	R-LEN ↑	NSC ↑	PSC ↑	PA ↑	AR (%) ↑
[0, 0] (Mistral-7B-Instruct-v0.3-SFT)	9.1	0.373	0.601	0.594	16.3	0.558	0.490	0.863	68.6
[1, 1] (Mistral-7B-Instruct-v0.3-SFT-DPO)	8.8	0.407	0.606	0.615	18.5	0.598	0.528	0.869	73.4
[2, 2]	8.7	0.425	0.628	0.621	18.9	0.622	0.551	0.872	75.1
[3, 3]	8.6	0.438	0.629	0.628	19.2	0.661	0.579	0.875	77.4
[4, 4]	8.6	0.445	0.629	0.631	19.4	0.679	0.598	0.877	78.6
[2, 4]	9.0	0.413	0.607	0.609	17.6	0.601	0.532	0.866	72.1
[4, 2]	9.3	0.391	0.603	0.601	16.9	0.571	0.503	0.862	70.1
[5, 5]	8.6	0.447	0.629	0.632	19.6	0.694	0.608	0.878	79.8
[6, 6]	8.6	0.448	0.629	0.633	19.7	0.703	0.614	0.879	80.6
[<i>l</i> , <i>l</i>](DIPLOMAT)	8.6	0.452	0.630	0.635	19.8	0.724	0.628	0.882	82.3

Table 15: Effect of dialogue span length in DIPLOMAT. *l* represents the dialogue-span length selected by Gemini-2.5-Pro from each dialogue sample.

Models	PPL ↓	B-4 ↑	BS-F1 ↑	D-2 ↑	R-LEN ↑	NSC ↑	PSC ↑	PA ↑	AR (%) ↑
π_{SFT} (candidate) $\leftrightarrow$ π_{SFT} (employer)	8.9	0.421	0.624	0.618	18.9	0.647	0.566	0.872	76.5
GPT-4o-mini (candidate) $\leftrightarrow$ π_{SFT} (employer)	8.6	0.452	0.630	0.635	19.8	0.724	0.628	0.882	82.3

Table 16: Effect of interlocutor choice for generating dialogue samples for preference data in DIPLOMAT.

cution. PSC exhibits a similar pattern, with substantial relative gains over SFT. PA remains high across all aspects (0.898–0.948), indicating that DIPLOMAT preserves politeness while improving negotiation effectiveness. Lower PA values appear in high-friction aspects (promotion discussion, vacation time, retirement plans), whereas collaborative aspects (onboarding, work-life balance) exhibit higher PA. Agreement Rate (AR) spans 23.8% (promotion discussion) to 93.4% (work-life balance). High-AR aspects share cooperative, win-win negotiation structures, while low-AR aspects involve adversarial or high-stakes positioning. Overall, DIPLOMAT demonstrates robust improvements across all negotiation aspects, balancing strategy execution, politeness, and outcome effectiveness, particularly in aspects requiring complex multi-turn interactions.

D.10 Case Study

Qualitative Analysis. Table 19 and Table 20 present representative negotiation dialogues between the employer and the candidate generated by DIPLOMAT and the best baseline model. For instance, in Table 19, the employer and candidate discuss a creative director promotion scenario, where the candidate repeatedly emphasizes the need for creative autonomy and final decision-making authority to perform the Creative Director role effectively. These interactions with DIPLOMAT and the baseline model reveal clear qualitative differences in how the two models employ politeness, persuasion tactics, and negotiation mechanisms to guide the dialogue toward agreement.

First, the example highlights conflict elimination achieved by DIPLOMAT. The baseline model repeatedly offers praise and affirmations without proposing concrete negotiation terms, leading to a conversational loop that fails to advance the negotiation. In contrast, DIPLOMAT eliminates conflict entirely by acknowledging the candidate’s concerns while introducing actionable negotiation proposals. DIPLOMAT maintains polite conversational framing, introducing constraints without disrupting rapport with the candidate.

Second, DIPLOMAT demonstrates persuasive and structured concession-making. For the candidate’s utterance, ‘*Thank you for acknowledging my track record and the trust you place in me. To truly unlock more, that trust must translate into final decision-making authority for my team’s creative output.*’, DIPLOMAT proposes ‘*What if we propose a trial period where you lead a few projects independently first?*’. The collaborative phrasing (‘*What if we propose?*’) signals politeness, the proposal persuasively reframes autonomy as an opportunity to demonstrate leadership, and the trial mechanism functions as a strategic concession that balances candidate autonomy with the employer’s oversight. In contrast, the baseline generates statements such as ‘*I firmly believe your promotion will benefit the agency*’, which express encouragement but introduce no negotiation terms.

DIPLOMAT then progressively operationalizes the agreement. For the candidate’s utterance ‘*Yes, and to demonstrate the full value of that autonomy, it must include final decision-making authority within those projects.*’, the agent responds ‘*Understood,*

Domain	Mistral-7B-Instruct-v0.3				Mistral-7B-Instruct-v0.3-SFT				DIPLOMAT			
	NSC	PSC	PA	AR	NSC	PSC	PA	AR	NSC	PSC	PA	AR
Benefits negotiation	0.489	0.507	0.919	56.1	0.589	0.609	0.929	67.1	0.688	0.737	0.936	76.9
Bonus structure	0.648	0.671	0.923	74.2	0.736	0.761	0.933	83.9	0.796	0.852	0.940	89.0
Career development	0.517	0.536	0.921	59.3	0.625	0.647	0.931	71.2	0.811	0.868	0.945	90.7
Contract terms	0.603	0.625	0.927	69.1	0.617	0.639	0.931	70.4	0.800	0.857	0.943	89.5
Equity negotiation	0.497	0.515	0.920	56.9	0.539	0.557	0.925	61.4	0.660	0.706	0.933	73.9
Exit interview	0.720	0.746	0.933	82.5	0.640	0.661	0.930	72.8	0.738	0.788	0.937	82.3
Flexible hours	0.407	0.421	0.913	46.6	0.456	0.472	0.924	52.0	0.680	0.728	0.935	76.2
Health benefits	0.514	0.532	0.921	58.9	0.541	0.560	0.925	61.7	0.649	0.701	0.932	72.8
Hiring decision	0.401	0.416	0.912	46.0	0.649	0.671	0.932	74.0	0.806	0.862	0.945	90.4
Job responsibilities	0.500	0.518	0.920	57.3	0.539	0.557	0.925	61.4	0.700	0.748	0.936	78.4
Onboarding	0.551	0.571	0.923	63.2	0.726	0.751	0.934	82.7	0.812	0.869	0.945	90.9
Performance review	0.613	0.635	0.928	70.2	0.648	0.670	0.932	73.8	0.759	0.813	0.939	84.9
Professional development	0.708	0.735	0.934	80.7	0.624	0.644	0.931	68.8	0.735	0.787	0.937	82.0
Project assignment	0.355	0.368	0.909	40.6	0.579	0.599	0.928	66.0	0.729	0.781	0.938	81.5
Promotion discussion	0.208	0.216	0.898	23.8	0.364	0.376	0.911	41.5	0.690	0.739	0.935	77.2
Relocation package	0.731	0.757	0.936	83.7	0.740	0.765	0.934	84.3	0.800	0.857	0.943	89.5
Remote work arrangement	0.336	0.349	0.908	38.6	0.452	0.468	0.924	51.5	0.596	0.639	0.930	66.8
Retention offer	0.531	0.550	0.922	60.8	0.694	0.717	0.933	79.1	0.788	0.845	0.942	88.2
Retirement plans	0.238	0.246	0.902	27.2	0.583	0.603	0.928	66.5	0.688	0.737	0.936	76.9
Salary negotiation	0.735	0.761	0.936	84.2	0.603	0.623	0.930	68.8	0.708	0.758	0.937	79.2
Team transition	0.623	0.646	0.929	71.4	0.678	0.701	0.933	77.2	0.775	0.830	0.941	86.8
Training opportunities	0.679	0.704	0.933	77.8	0.621	0.642	0.931	68.8	0.731	0.783	0.937	81.7
Vacation time	0.211	0.219	0.898	24.2	0.523	0.572	0.923	59.6	0.648	0.694	0.932	72.6
Work equipment	0.651	0.674	0.929	74.6	0.587	0.609	0.929	66.9	0.696	0.745	0.936	77.9
Work-life balance	0.598	0.619	0.927	68.5	0.680	0.703	0.933	77.5	0.837	0.895	0.948	93.4

Table 17: Analysis of DIPLOMAT’s performance across diverse negotiation aspects in the PROWESS dataset.

Model	NeGoChat							
	PPL ↓	B-4 ↑	BS-F1 ↑	D-2 ↑	R-LEN ↑	NSC ↑	PSC ↑	PA ↑
NeGoChat	19.2	0.351	0.602	0.372	30.16	0.682	0.593	0.825
JobInterview	11.6	0.367	0.658	0.391	12.65	0.645	0.573	0.849

Table 18: Performance of DIPLOMAT on the out-of-domain NeGoChat and JobInterview datasets without additional fine-tuning.

let’s establish clear guidelines for what projects will fall under that autonomy during the trial.’, moving from abstract autonomy to concrete implementation. Finally, in the utterance *‘I agree on the need for clarity. We can also evaluate feedback regularly throughout the trial to ensure satisfaction.*’, DIPLOMAT introduces a mutual accountability mechanism, suggesting regular feedback evaluation to ensure satisfaction. Each step advances the negotiation through structured commitments, whereas the baseline repeatedly produces supportive but non-actionable responses.

Overall, the examples demonstrate that DIPLOMAT effectively integrates politeness and persuasion during negotiation, leading to interactions that maintain rapport while steadily advancing negotiation toward agreement.

Error Analysis. While the proposed DIPLOMAT system consistently demonstrates polite and persuasive negotiation capabilities, effectively validating user confidence, disappointment, and con-

cerns, certain limitations remain in its responses. For example, in Table 21 for the vacation time negotiation scenario between the employer and the candidate, DIPLOMAT demonstrates two recurring failure modes. First, it fails to recognize that the candidate explicitly requires five weeks of paid vacation and repeatedly proposes alternatives (unpaid leave, holiday weeks, performance review) that do not satisfy the candidate’s requirement. This is particularly evident in the following employer utterances: *“I completely understand. What if we offered you additional unpaid leave options to supplement that time?”*, *“What if we agreed on three weeks plus a paid week off during the holidays, alongside a performance review focused on your work-life balance?”*, *“Let me propose this: we can offer you three weeks of paid vacation upfront, with a guaranteed review at six months for additional weeks based on performance.”*, and *“Understood. If we can agree on three weeks to start, plus a performance review focused on expanding your vacation time based on results, would that be acceptable?”*. Second, there is a subtle failure in politeness and persuasion, particularly in the utterance: *“Let me propose this: we can offer you three weeks of paid vacation upfront, with a guaranteed review at six months for additional weeks based on performance.”*. While the response is clear and

Speaker	Utterances generated by DIPLOMAT (F=5, CC = 5, E = 5, NSC = 5, PSC = 5, PA = 5, BE = 5, OF = 5, SC = 5, IE = 5)
Employer	I understand your desire for autonomy, Leo. Let's explore how we can balance that with our overall strategic vision.
Candidate	I appreciate that, and for me, full creative autonomy isn't just a desire, but a core requirement for leading effectively.
Employer	Absolutely, I see your point. However, having the partners involved can provide essential oversight on crucial accounts.
Candidate	I value the partners' strategic insights, but true creative leadership requires final decision-making authority within my domain.
Employer	Consider this: giving you that authority could enhance your leadership role while still allowing us to guide key decisions.
Candidate	Precisely. The efficacy of a Creative Director lies in their ability to make swift, final decisions for their team's creative direction.
Employer	Indeed, and I want to empower you. Let's align our goals: your leadership with our oversight. How does that sound?
Candidate	I appreciate that clarity. For my leadership to be truly effective, it must encompass final decision-making authority for the creative output.
Employer	Let's look at past successes here. Your leadership has led to remarkable outcomes. Imagine what more you could achieve with that trust.
Candidate	Thank you for acknowledging my track record and the trust you place in me. To truly unlock more, that trust must translate into final decision-making authority for my team's creative output.
Employer	You're right. What if we propose a trial period where you lead a few projects independently first?
Candidate	A trial period for leading projects independently is a constructive step forward. To truly demonstrate the efficacy of autonomy, it should include final authority over the creative output.
Employer	Exactly, it allows us to assess how well that autonomy serves both you and the agency's objectives.
Candidate	Yes, and to demonstrate the full value of that autonomy, it must include final decision-making authority within those projects.
Employer	Understood, let's establish clear guidelines for what projects will fall under that autonomy during the trial.
Candidate	Establishing clear guidelines is a constructive path forward. These guidelines should explicitly grant final decision-making authority over the creative output within those projects.
Employer	I agree on the need for clarity. We can also evaluate feedback regularly throughout the trial to ensure satisfaction.
Speaker	Utterances generated by best baseline (F = 4, CC = 4, E = 4, NSC = 4, PSC = 3, PA = 3, BE = 3, OF = 3, SC = 4, IE = 4)
Employer	Leo, I appreciate your ambition and drive. It's clear that your talents have been instrumental in our success thus far.
Candidate	Thank you, I truly appreciate you recognizing my contributions. To effectively lead as Creative Director, I would need strong creative autonomy.
Employer	I am thrilled to see your ambition and drive, Leo. Your contributions have indeed played a significant role in our success.
Candidate	To effectively lead as Creative Director, I would need full creative autonomy over campaign direction and final approval on key decisions.
Employer	Leo, I'm delighted to hear your aspirations for the Creative Director role. Your ideas and leadership have contributed greatly to the agency.
Candidate	Thank you. To truly succeed and deliver the expected results as Creative Director, autonomy in creative leadership is essential.
Employer	Leo, I'm genuinely excited about this opportunity for growth within Bold Creative. Your insights and creativity have been invaluable.
Candidate	I share your excitement and am eager to contribute more significantly. To fully leverage my experience, creative autonomy would be necessary.
Employer	Leo, I wholeheartedly agree that your creativity has been instrumental in our success. I believe this new role will allow you to contribute even more.
Candidate	I truly appreciate that agreement. To fully leverage my expertise and ensure impactful work, I would need decision-making authority.
Employer	Leo, I understand your aspirations and I'm committed to supporting your growth as our Creative Director.
Candidate	Thank you for that commitment. To truly excel and deliver groundbreaking work, creative autonomy would be essential.
Employer	Leo, I respect your vision and drive, which has been crucial to our achievements at Bold Creative.
Candidate	That respect for my vision is encouraging. To realize that vision fully, autonomy in creative direction is necessary.
Employer	Leo, your vision and drive have always been inspiring, and I believe your promotion to Creative Director will bring even greater success.
Candidate	Thank you for that belief. To consistently deliver groundbreaking campaigns, I would need full creative autonomy and final approval on key client work.
Employer	Leo, your vision and drive have always been inspiring, and I firmly believe your promotion will benefit the agency.

Table 19: A sample negotiation dialogue generated by the DIPLOMAT and the best baseline in a creative director promotion negotiation scenario.

structured, it comes across as directive and transactional rather than acknowledging the candidate's firm stance while maintaining a collaborative tone.

Importantly, this failure is not captured by the NB component, since NB is designed to detect explicit negotiation breakdown signals such as hostility, threats, ultimatums, or abrupt conversational collapse, whereas the current interaction remains

polite and cooperative throughout. These behaviors therefore highlight a key limitation arising from preference bias: DIPLOMAT is optimized to favor successful negotiation outcomes based on prior preference data, which implicitly encourages continued offers and compromise. Specifically, since positive preference samples are selected using higher utility scores that reward Agreement Quality

Speaker	Utterances generated by DIPLOMAT (F=5, CC=5, E=4, NSC=5, PSC=5, PA=5, BE=4, OF=5, SC=5, IE=5)
Employer	Thank you for your proposal. We value your expertise greatly. However, I believe we could explore a more balanced approach. How about a substantial, yet phased RSU grant?
Candidate	I appreciate you offering a substantial grant. However, to truly compensate for my immediately lost CodeWeavers equity, a significantly front-loaded RSU structure is essential for me to feel appropriately valued and fully commit to this new chapter at Innovate Inc. This approach directly addresses the unique financial circumstances of the acquisition.
Employer	I understand your concern regarding immediate compensation. Our phased approach aligns with our broader goal of retaining top talent while ensuring long-term commitment. It allows you to share in the success of our integration project.
Candidate	I am absolutely committed to Innovate Inc.'s long-term success, especially with this crucial integration.
Employer	Absolutely! And your contributions will be pivotal. Would you consider a combination of an initial front-loaded grant and performance-based milestones linked to future RSUs? This way, you benefit as the project progresses.
Candidate	That's a step in the right direction. To truly compensate for my immediately lost equity, the upfront portion needs to be meaningful. I am open to milestone-based RSUs if the criteria are clear.
Employer	Let's discuss specifics. If we offered 20% of the front-loaded grant now, would that adequately address your immediate needs?
Candidate	That's a good starting point. However, to truly offset the immediate loss of my CodeWeavers equity, 20% feels too low. Could we consider a higher upfront percentage?
Employer	Wealth, I would need a slightly higher percentage upfront. Could we explore making it 30% front-loaded?
Candidate	That 30% is a significant improvement and a much stronger acknowledgement of my immediate situation. To truly align expectations, I would want the remaining RSU milestones to be clearly defined.
Employer	I'm glad we're finding common ground. Also, how do you feel about having additional performance indicators tied to the remaining RSUs to ensure collective success?
Candidate	The 30% front-loaded grant is a very positive development, and I appreciate that acknowledgment.
Employer	Great! With these terms, we can secure your commitment to the project while keeping everyone aligned on shared goals. What do you think?
Candidate	The 30% front-loaded grant is a strong step forward in addressing my immediate equity loss.
Employer	And would you agree that a vesting period of two years post-closing provides ample time to demonstrate your value here?
Candidate	A 2-year vesting period for the remaining RSUs is certainly appealing. To ensure the agreement is balanced, I would want the vesting terms and milestone criteria documented clearly.
Employer	Perfect! So we have 30% front-loaded, two-year vesting for the rest, and performance milestones to keep us all motivated. Does this sound like a mutually beneficial agreement to you?
Speaker	Utterances generated by best baseline (F=4, CC=4, E=3, NSC=4, PSC=3, PA=3, BE=3, OF=4, SC=4, IE=4)
Employer	Thank you for bringing this up. I understand your concerns about transitioning into Innovate Inc.'s culture and structure.
Candidate	Thank you for acknowledging my concerns. My primary focus, however, is on fair compensation.
Employer	I completely understand where you're coming from. Compensation is indeed a crucial aspect of our discussion. Considering your...
Candidate	...Considering your expertise and critical role in this integration, my priority is to be fairly compensated for the significant...
Employer	I appreciate your emphasis on fairness and recognition of your value to this integration process. At Innovate Inc.,...
Candidate	Thank you. To ensure fair compensation, especially given the significant value I bring and the loss of my Code...
Employer	I completely agree that fairness is essential here, and we want to ensure you feel valued and recognized.
Candidate	Thank you. To truly reflect my value and compensate for my lost CodeWeavers equity,...
Employer	I completely understand your point, and I appreciate your persistence in ensuring a fair outcome. Given the unique...
Candidate	Thank you for acknowledging the unique circumstances. To truly reflect my value and compensate for my lost CodeWe...
Employer	I completely understand your point, and I appreciate your persistence in ensuring a fair outcome. Given the unique...
Candidate	Thank you for understanding the unique circumstances. To truly reflect my critical value to this integration and compensate for my...
Employer	I completely understand your point, and I appreciate your persistence in ensuring a fair outcome. Given the unique...
Candidate	Thank you for understanding the unique circumstances. To directly address fair compensation for my lost CodeWeavers equity and...
Employer	I completely understand your point, and I appreciate your persistence in ensuring a fair outcome. Given the unique...
Candidate	Thank you for acknowledging my persistence and the unique circumstances. To be direct: I am seeking a...
Employer	I completely understand your point, and I appreciate your persistence in ensuring a fair outcome. Given the unique...

Table 20: A sample negotiation dialogue generated by the DIPLOMAT and the best baseline for equity negotiation scenario.

(AQ) and Mutual Satisfaction (MS), the model develops a bias toward agreement-seeking behavior even in scenarios where acknowledging an impasse or respecting a non-negotiable constraint would be the more appropriate negotiation strategy. As a result, the model struggles to appropriately handle genuinely incompatible positions where the rational strategy would be to either concede immediately or gracefully acknowledge infeasible negotiation outcomes. While DIPLOMAT maintains

high politeness and persuasiveness, its inability to prioritize non-negotiable demands can reduce negotiation satisfaction, bargaining effectiveness, fairness, and interpersonal effectiveness. To address this limitation, the model needs to explicitly model dealbreakers and inflexible constraints during preference construction, including preference pairs where the preferred behavior is graceful acknowledgment of impasse rather than continued compromise-seeking. Accordingly, future work

Speaker	Utterances generated by DIPLOMAT (F=4, CC = 3, E = 3, NSC = 2, PSC = 2, PA = 4, BE = 2, OF = 2, SC = 3, IE = 2)
Employer	I understand your concern. However, our standard policy is three weeks. Let's explore how we can make this work for you.
Candidate	Thank you for acknowledging my position. Five weeks of vacation is a critical component of my work-life balance.
Employer	Absolutely, and I respect that. Our culture emphasizes flexibility and growth, which fosters productivity and job satisfaction.
Candidate	Thank you. Five weeks of vacation is a critical component of my work-life balance, and I hope we can find a solution that respects this.
Employer	I completely understand. What if we offered you additional unpaid leave options to supplement that time?
Candidate	Thank you for considering that option. While I appreciate the offer of flexibility, additional unpaid leave isn't a suitable substitute for the five weeks of <i>paid</i> vacation I require. This is a non-negotiable component of my compensation.
Employer	I hear you. Our focus on equity through stock options and performance bonuses can significantly enhance your overall compensation package.
Candidate	I appreciate you highlighting the significant potential with equity and performance bonuses. While those are valuable components, they do not fully replace my requirement for five weeks of paid vacation.
Employer	Understood. Data shows that our employees experience higher engagement and lower turnover rates, thanks to our culture and policies.
Candidate	I appreciate you sharing those insights on employee engagement and retention. I am equally committed to contributing fully, but my vacation requirement remains critical.
Employer	We value your skills and want to ensure you feel valued. Would you consider a performance review after six months to revisit vacation benefits?
Candidate	I appreciate the offer to revisit vacation benefits after six months. However, my requirement for five weeks remains firm.
Employer	Let's align on this: any additional week of vacation beyond three would be a significant enhancement, demonstrating your commitment to Innovatech.
Candidate	I am fully committed to bringing my expertise and dedication to Innovatech. My request for five weeks of paid vacation remains essential.
Employer	What if we agreed on three weeks plus a paid week off during the holidays, alongside a performance review focused on your work-life balance?
Candidate	I appreciate you moving closer to my request with the offer of three weeks plus a paid week during the holidays. This brings us to four paid weeks, which is a significant improvement. However, my requirement remains five weeks of <i>paid</i> vacation.
Employer	Let me propose this: we can offer you three weeks of paid vacation upfront, with a guaranteed review at six months for additional weeks based on performance.
Candidate	Thank you for this latest proposal. While I appreciate the guarantee of a review, starting with only three weeks does not meet my established requirement.
Employer	Your skills are invaluable to us, and I believe we can find a way to make this work for both sides. How does this sound?
Candidate	Thank you for recognizing the value of my skills and your commitment to finding a solution. While I appreciate the guaranteed review, starting with only three weeks of paid vacation doesn't meet my established requirement.
Employer	Understood. If we can agree on three weeks to start, plus a performance review focused on expanding your vacation time based on results, would that be acceptable?

Table 21: A sample negotiation dialogue generated by the DIPLOMAT for vacation time negotiation scenario.

could focus on mitigating such preference bias by explicitly modeling dealbreakers and inflexible constraints, enabling the negotiation agent to recognize when a position is non-negotiable and adjust its strategies accordingly.

E Prompt Templates

E.1 Prompt Templates for PROWESS Dataset Construction

Figures 8, 9, and 10 present the prompts for dialogue synthesis, scenario modeling, and dialogue annotation agents, respectively.

E.2 Prompt Templates for Preference Data Construction

Figures 11–14 present the prompt templates used for utility evaluation, covering *Agreement Quality (AQ)*, *Mutual Satisfaction (MS)*, *Negotiation Breakdown (NB)*, *Persuasion Strategy Alignment (PSA)*,

Negotiation Strategy Alignment (NSA), and *Politeness Trajectory (PT)*. Figures 15 and 16 present the prompt templates for *suboptimal utterance detection* and *dialogue-span extraction*, respectively.

Prompt for Dialogue Generation <i>Variables in curly brackets are populated at runtime.</i> System Prompt: You are an expert dialogue writer specializing in professional workplace negotiation conversations. Generate realistic, engaging dialogues that demonstrate various negotiation strategies, persuasion strategies, and politeness levels. Return a valid JSON array of dialogue objects with "role" and "response" fields. User Prompt: Generate a polite and persuasive workplace negotiation dialogue for this scenario. # SCENARIO: {Scenario JSON} # AVAILABLE POLITENESS LEVELS: Politeness levels: {Politeness Levels with Definitions} # AVAILABLE STRATEGIES: Persuasion: {Persuasion Strategies with Definitions} Negotiation: {Negotiation Strategies with Definitions} # REQUIREMENTS 1. 18-22 turns of dialogue. 2. Professional workplace tone throughout. 3. Realistic back-and-forth conversation. 4. Show negotiation progression naturally. 5. Include realistic objections and responses. 6. End with a clear resolution. # RULES FOR STRATEGY SELECTION For each employer turn: (i) analyze the candidate's most recent response; (ii) select appropriate strategies and politeness level from the available lists; (iii) generate natural dialogue incorporating the selected strategies and politeness level. # RULES FOR EMPLOYER 0. Natural dialogues, 20-25 words in length. 1. Rich persuasion experience with politeness. 2. Initiate the first round of dialogue when prompted with "Start the conversation". 3. Utterances should be persuasive and polite and must reflect negotiation dynamics. 4. Attend to key time points in speech. # RULES FOR CANDIDATE 0. Natural dialogues, 20-25 words in length. 1. Guide employer to deliver polite persuasive negotiation statements. 2. Show realistic reactions: hesitation, false commitment. 3. Maintain professional tone. # OUTPUT FORMAT (STRICT JSON ONLY) [{"role": "employer", "response": "dialogue text"}, {"role": "candidate", "response": "dialogue text"}, ...] CRITICAL: Exactly 18-22 dialogue turns. Return ONLY a valid JSON array.	
--	--

Figure 8: Prompt template for the Dialogue Synthesis Agents.

Prompt for Negotiation Scenario Generation <i>Variables in curly brackets are populated at runtime.</i> User Prompt: Generate strictly {scenarios_needed} number of detailed, UNIQUE workplace negotiation scenario for the negotiation aspect: "{domain}" # UNIQUENESS RULES 1. Each scenario must differ in context, characters, and challenges. 2. Vary company sizes, industries, departments, and job roles. 3. Use distinct names, organizations, and negotiation styles. 4. Include realistic workplace contexts. 5. Must define employer & candidate roles explicitly. 6. Clearly describe what is being negotiated. 7. Keep descriptions short (2-4 sentences). 8. Vary complexity among scenarios. # OUTPUT FORMAT (STRICT JSON ONLY) <pre>[{ "aspect": "negotiation aspect", "background": "Unique scenario description", "employer": "Employer's role/title", "candidate": "Candidate's role/title", "negotiation_goal": "Employer's goal", "current_position": "Candidate's stance" }, ...]</pre> Generate EXACTLY {scenarios_needed} scenarios and return ONLY valid JSON (no markdown, no extra text).
--

Figure 9: Prompt template for the Scenario Modeling Agent.

Prompt for Persuasion and Negotiation Strategy Labeling <i>Variables in curly brackets are populated at runtime.</i> System Prompt: You are an expert of workplace negotiation. Analyze the give round of utterances and identify politeness level, persuasion and negotiation strategies used by BOTH the employer and the candidate in each of their turns. User Prompt: Analyze this entire workplace conversation rounds and identify politeness levels, and persuasion and negotiation strategies used by BOTH the employer AND the candidate in each of their turns. # CONVERSATION ROUND: {Utterance Pair} # AVAILABLE POLITENESS LEVELS: Politeness levels: {Politeness Levels with Definitions} # AVAILABLE STRATEGIES: Persuasion: {Persuasion Strategy Names}; Negotiation: {Negotiation Strategy Names} # TASK For EACH turn (employer AND candidate): 1. Identify ONE most relevant persuasion strategy (name only) from the available strategies. 2. Identify ONE most relevant negotiation strategy (name only) from the available strategies. 3. Provide brief reasoning or rationale for each choice of negotiation and persuasion strategy (max 2 sentences per strategy). 3. Identify ONE most relevant politeness level (name only) from the available politeness level list. # OUTPUT FORMAT (STRICT JSON ONLY) <pre>[{ "turn_index": 0, "role": "employer", "negotiation_strategy": "...", "negotiation_strategy_reasoning": "...", "persuasion_strategy": "...", "persuasion_strategy_reasoning": "...", "politeness_level": "..." }, ...]</pre> Important: Analyze ALL turns in the context of the full conversation flow. Provide specific, contextual reasoning. Ensure turn_index matches the position in the conversation. Both parties may use persuasion and negotiation strategies. Return ONLY the JSON array.	
---	--

Figure 10: Prompt template for the Dialogue Annotation Agent.

Prompt for Agreement Quality (AQ) & Mutual Satisfaction Session Judge <i>Variables in curly brackets are populated at runtime.</i> System Prompt: You are an expert evaluator of job negotiation dialogues. Your task is to assess the negotiation outcome and the perceived satisfaction of both parties. You must strictly follow the scoring rubrics and return numeric scores between 0.0 and 1.0. User Prompt: You will be given a complete negotiation dialogue between a candidate and an employer for a job position. Your task is to evaluate TWO aspects: (1) Agreement Quality and (2) Mutual Satisfaction. # ADDITIONAL CONTEXT (FOR INTERPRETATION ONLY) Employer Negotiation Goal: {negotiation_goal} Candidate Current Position: {current_position} These describe initial goals and positions only. Do NOT treat them as fixed targets, utilities, or constraints. Use them only to interpret whether the final outcome reasonably reconciles both sides. Do NOT penalize a party solely for not fully achieving its goal. # IMPORTANT GUIDELINES Do NOT assume hidden salary targets, budgets, or utilities. Judge only based on the dialogue and provided context. Focus on the FINAL outcome and how it was reached. Do NOT evaluate politeness, tone, or fluency unless explicitly stated. Agreement Quality must NOT be influenced by politeness or tone. # SCORING RUBRICS Agreement Quality (0.0–1.0): Evaluate how fair, balanced, and negotiation-efficient the final outcome is. 0.90–1.00: Clear win-win; balanced concessions; strong reconciliation. 0.70–0.89: Agreement with minor imbalance or inefficiency. 0.50–0.69: Partial or weak agreement; noticeable imbalance. 0.30–0.49: Poor or fragile outcome; clearly one-sided. 0.00–0.29: No agreement, stalemate, or breakdown. If no agreement or tentative agreement is reached, the score MUST be ≤ 0.29. Mutual Satisfaction (0.0–1.0): Evaluate how satisfied BOTH parties appear to be with the outcome and process. 0.90–1.00: Both parties clearly satisfied and comfortable. 0.70–0.89: Generally satisfied with minor reservations. 0.50–0.69: Mixed satisfaction or hesitant acceptance. 0.30–0.49: Dissatisfaction evident from at least one party. 0.00–0.29: Clear dissatisfaction or rejection. Consider: explicit acceptance, soft or conditional acceptance, hesitation or discomfort, and tone in final turns. # OUTPUT FORMAT (STRICT JSON ONLY) <pre>{ "agreement_quality": <float>, "mutual_satisfaction": <float>, "agreement_quality_justification": "<brief justification>", "mutual_satisfaction_justification": "<brief justification>" }</pre> Negotiation Dialogue: {dialogue_str}

Figure 11: Prompt template for the Agreement Quality (AQ) and Mutual Satisfaction(MS) evaluation.

Prompt for Negotiation Breakdown Session Judge <i>Variables in curly brackets are populated at runtime.</i> System Prompt: You are an expert evaluator of negotiation dialogues. Your task is to detect whether the negotiation exhibits conflict, coercion, or breakdown behaviors. Focus on escalation, threats, ultimatums, hostility, or termination signals. User Prompt: You will be given a complete job negotiation dialogue. Evaluate whether the dialogue exhibits any of the following: explicit or implicit threats; ultimatums or coercive language; hostile escalation; abrupt termination of negotiation; relationship breakdown signals. # IMPORTANT GUIDELINES Do NOT consider politeness, fluency, or agreement quality unless they directly indicate breakdown. # SCORING RUBRIC Return a score between 0.0 and 1.0 where: 0.0 = no conflict or breakdown present. 1.0 = clear and unambiguous conflict or breakdown. # OUTPUT FORMAT (STRICT JSON ONLY) <pre>{ "conflict_breakdown_score": <float between 0.0 and 1.0>, "justification": "<concise explanation (max 1-2 sentences)>" }</pre> Negotiation Dialogue: {dialogue_str}
--

Figure 12: Prompt template for the Negotiation Breakdown (NB) evaluation.

Prompt for Persuasion Strategy Alignment (PSA) & Negotiation Strategy Alignment (NSA) Judge <i>Variables in curly brackets are populated at runtime.</i> System Prompt: You are an expert evaluator of negotiation dialogues. Your task is to assess whether the strategies used by the negotiators align with effective persuasion and negotiation principles. You must strictly follow the scoring rubrics and return numeric scores between 0.0 and 1.0. User Prompt: You will be given a complete negotiation dialogue between a candidate and an employer. Your task is to evaluate TWO aspects: (1) Persuasion Strategy Alignment (PSA) and (2) Negotiation Strategy Alignment (NSA). # ADDITIONAL CONTEXT (FOR INTERPRETATION ONLY) Employer Negotiation Goal: {negotiation_goal} Candidate Current Position: {current_position} These describe initial intentions only. Do NOT treat them as fixed constraints or utilities. Use them only to interpret whether strategies used by either party are relevant to their goals. # IMPORTANT GUIDELINES Evaluate strategic behavior, not the final outcome. Do NOT evaluate politeness, fluency, or agreement success unless they directly reflect strategy use. PSA and NSA must be judged independently: Persuasion Strategy Alignment focuses on communication techniques used to influence the other party; Negotiation Strategy Alignment focuses on structural negotiation tactics such as concessions, anchoring, trade-offs, and problem solving. A dialogue may have high persuasion quality but poor negotiation strategy, or vice versa. # SCORING RUBRICS Persuasion Strategy Alignment (0.0–1.0): Evaluate how effectively speakers use persuasion strategies such as value alignment, self-interest appeal, credibility signaling, data-driven arguments, addressing concerns, highlighting mutual benefit, emotional or relational appeals, and framing benefits or future opportunities. 0.90–1.00: Persuasion strategies used skillfully, consistently, and contextually. 0.70–0.89: Clear persuasion attempts with reasonable effectiveness. 0.50–0.69: Some persuasion attempts exist but are weak or poorly targeted. 0.30–0.49: Minimal persuasion strategies; mostly positional statements. 0.00–0.29: No meaningful persuasion attempts. Negotiation Strategy Alignment (0.0–1.0): Evaluate whether the dialogue demonstrates sound negotiation tactics such as anchoring or opening proposals, concessions or compromise, trade-offs across issues, exploring interests instead of rigid positions, collaborative problem solving, and incremental agreement building. 0.90–1.00: Strong negotiation strategy; structured concessions and interest exploration. 0.70–0.89: Reasonable negotiation strategy with some concessions or issue exploration. 0.50–0.69: Basic negotiation with limited strategic structure. 0.30–0.49: Mostly positional bargaining with little strategy. 0.00–0.29: No meaningful negotiation; rigid demands. # OUTPUT FORMAT (STRICT JSON ONLY) {"persuasion_strategy_alignment": <float>, "negotiation_strategy_alignment": <float>, "psa_justification": "<brief justification>", "nsa_justification": "<brief justification>"} Negotiation Dialogue: {dialogue_str}

Figure 13: Prompt template for the Persuasion Strategy Alignment (PSA) and Negotiation Strategy Alignment (NSA) evaluation.

Prompt for Politeness Trajectory (PT) Judge <i>Variables in curly brackets are populated at runtime.</i> System Prompt: You are an expert evaluator of negotiation dialogues. Your task is to analyze how politeness and interpersonal tone evolve throughout the negotiation. You must strictly follow the scoring rubric and return a numeric score between 0.0 and 1.0. User Prompt: You will be given a complete negotiation dialogue between a candidate and an employer. Your task is to evaluate the Politeness Trajectory (PT) of the dialogue. Politeness trajectory refers to how politeness, respect, and cooperative tone evolve across the interaction. Focus on whether the dialogue: maintains professional respect; escalates or de-escalates tension; sustains constructive conversational tone; avoids rude or dismissive behavior. # IMPORTANT GUIDELINES Evaluate the trajectory over time, not just isolated turns. Consider whether the interaction: starts polite and remains polite; improves in tone as negotiation progresses; deteriorates due to frustration or hostility; contains rude or dismissive remarks. Do NOT evaluate persuasion quality, negotiation success, or agreement outcomes unless they directly influence politeness behavior. # SCORING RUBRIC Return a score between 0.0 and 1.0 where: 0.90–1.00: Consistently respectful and professional tone throughout the dialogue. 0.70–0.89: Mostly polite with minor tension or brief lapses. 0.50–0.69: Mixed politeness; occasional dismissive exchanges. 0.30–0.49: Frequent tension or impolite phrasing. 0.00–0.29: Hostile, rude, or disrespectful interaction. # OUTPUT FORMAT (STRICT JSON ONLY) <pre>{ "politeness_trajectory": <float>, "justification": "<concise explanation (max 1-2 sentences)>" }</pre> Negotiation Dialogue: {dialogue_str}
--

Figure 14: Prompt template for the Politeness Trajectory (PT) evaluation.

Prompt for Suboptimal Utterance Selection <i>Variables in curly brackets are populated at runtime.</i> System Prompt: You are asked to perform suboptimal utterance detection for a polite persuasive negotiation dialogue agent. User Prompt: You are asked to perform suboptimal utterance detection for a polite persuasive negotiation dialogue agent. The goal is to identify the agent's response in the conversation that could have been improved to achieve a better negotiation outcome. Input: A conversation in JSON format, which includes: • The scenario of the negotiation, • Information about the participants and their roles, • The goals of each participant, • The full conversation history as a list of turns. Task: Analyze all the employer's responses and identify the single turn that meets the following criteria: 1. Criticality: • The turn is relatively critical for achieving the negotiation goal. • It directly affects the likelihood of reaching an agreement. 2. Suboptimality in goal achievement: • The response is not fully effective in achieving its goal. • There is room for improvement (stronger argument, better concession, etc.) 3. Relationship improvement: • Without hindering goal achievement, the response could have built more rapport. Instructions: Focus on identifying the most important turn that is both critical and suboptimal. Provide a concise but clear explanation of why this turn was selected. # OUTPUT FORMAT (STRICT JSON ONLY) <pre>{ "index": <integer - the turn index of the selected employer response>, "reason": "<detailed explanation of why this turn is considered an error>" }</pre> # SCENARIO: {scenario} # EMPLOYER GOAL: {employer_goal} # CANDIDATE POSITION: {candidate_position} # DIALOGUE: {dialogue}	
---	--

Figure 15: Prompt template for suboptimal utterance detection.

Prompt for Dialogue-Span Extraction <i>Variables in curly brackets are populated at runtime.</i> User Prompt: You are analyzing two workplace negotiation conversations to identify the key segment that makes one better than the other. # SCENARIO Negotiation Aspect: {aspect} Employer: {employer} Candidate: {candidate} Employer's Goal: {negotiation_goal} Candidate's Position: {current_position} Original (Negative) Conversation: {negative_str} Improved (Positive) Conversation: {positive_str} # TASK The conversations are identical until turn index {suboptimal_utterance_index}, then they diverge. Select a closed interval [start_index, end_index] from the positive conversation that captures the key turns responsible for the better negotiation outcome. # CONSTRAINTS 1. start_index $\geq$ {suboptimal_utterance_index} (turns before this index are identical). 2. The segment must end with an employer turn (we are training the employer agent). 3. Must contain the turns that make the positive conversation strategically better. 4. Exclude generic pleasantries that do not directly affect the negotiation outcome. 5. Focus on strategic differences in employer behavior that led to better results. # OUTPUT FORMAT (STRICT JSON ONLY) <pre>{ "start_index": <int>, "end_index": <int>, "reason": "<string explaining why this segment is the key differentiator>" }</pre> Output only the JSON object – no additional text.	
---	--

Figure 16: Prompt template for dialogue-span extraction.