

Multi-Party Supervised Fine-tuning of Language Models for Multi-Party Dialogue Generation

Xiaoyu Wang^{1,*}, Ningyuan Xi^{2,*}, Teng Chen³, Qingqing Gu³, Yue Zhao³, Xiaokai Chen¹,
Zhonglin Jiang³, Yong Chen³, Luo Ji^{3,†}

¹Beijing Institute of Technology, Beijing, China

²Beihang University, Beijing, China

³Geely AI Lab, Beijing, China

3220230388@bit.edu.cn, 21373102@buaa.edu.cn, {Teng.Chen2, Qingqing.Gu3, Yue.Zhao17}@Geely.com,
chenxiaokai@bit.edu.cn,{ zhonglin.jiang, yong.chen, Luo.Ji1 }@Geely.com

Abstract—Large Language Models (LLM) are usually fine-tuned to participate in dyadic or two-party dialogues, which can not adapt well to multi-party dialogues (MPD), which hinders their applications in such scenarios including multi-personal meetings, discussions and daily communication. Previous LLM-based researches mainly focus on the multi-agent framework, while their base LLMs are still pairwisely fine-tuned. In this work, we design a multi-party fine-tuning framework (MuPaS) for LLMs on the multi-party dialogue datasets, and prove such a straightforward framework can let the LLM align with the multi-party conversation style efficiently and effectively. We also design two training strategies which can convert MuPaS into the MPD simulator. Substantial experiments show that MuPaS can achieve state-of-the-art multi-party response, higher accuracy of the next-speaker prediction, higher human and automatic evaluated utterance qualities, and can even generate reasonably with out-of-distribution scene, topic and role descriptions. The MuPaS framework bridges the LLM training with more complicated multi-party applications, such as conversation generation, virtual rehearsal or meta-universe.

Index Terms—LLM, MPD, fine-tuning, conversation simulator

I. INTRODUCTION

In recent years, large language models (LLM) have demonstrated significant advancements in dyadic conversational contexts, such as question-answering systems and chatbot companions. Such applications are primarily structured around binary dialogue attendants (typically ‘human’ and ‘assistant’), which are supported by widespread open-source models and datasets. However, many real-world scenarios instead encompass Multi-Party Dialogues (MPD) (Some papers instead name this scenario by multi-party conversation (MPC).), such as team meetings, classroom discussions, court or academic debates, or simply daily conversations with multiple humans involved [1], [2]. Instead of responding to a single user’s query, in such a case, the dialog system needs to understand conversation contexts from multiple users, determine whether to speak or not, and reasonably participate in potential multi-

*The first two authors contributed equally to this research; work was done during their internship at Geely.

†Corresponding author: Luo.Ji1@Geely.com



Fig. 1: The paradigm shift from the conventional two-party dialogue (Left) to the multi-party dialogue (Right). The training target also change from the helpful assistant to different possible persona or roles.

ple concurrent topics. Novel modeling technique is therefore required to adapt to this different dialogue paradigm.

Previous researches have sought to address the unique challenges of MPD modeling, such as MIDS [3], ChatMDG [4], ReDE [5], SDMPED [6] and MPC-BERT [7]. However, these works are mostly RNN, Bert or Graph-based, which have not yet leveraged the semantic knowledge and generation capabilities of modern LLM, and is difficult to scale up and generalize to different domains. There are also LLM-based MPD approaches, which are generally based on multi-agent systems. Each LLM agent still adheres to a binary interaction framework, and the conversations might happen between

different pairs of agents which do not happen simultaneously and concurrently. Besides that, conventional LLMs are fine-tuned with human-assistant dialogues with alignment with the helpful, impartial, and conservative conversation style. In short, there is still not a single LLM-based, purely training framework which allows the model to learn from MPD directly, unify the response generation and the speaker in a uniform manner, and portray different persona styles (either by data-driven or system prompted).

In this work, we propose a **Multi-Party Supervised (MuPaS)** fine-tuning framework to train LLMs as the MPD participants. Starting from a conventional instruct version of LLM which can handle two-party conversations, we provide an extra post-training stage in which the MPD datasets are supervised fine-tuned, such that adapt its chat capability from the two-party to the multi-party format. As indicated by Figure 1, we preprocess the dataset by annotating lists of roles and sample-wise scene descriptions. We allow the LLM to be fine-tuned with each role’s utterance while other roles are masked as context. We further apply this approach as the basis of MPD builder by designing the model to recognize the next speaker simultaneously. By thoroughly designed experiments, we find our MuPaS can both generate state-of-the-art response quality and achieve the highest next-speaker prediction accuracy, compared with previous baselines, within the MPD scope. We also provide several entertainment MPD case simulations which indicate our approach can generate stylized and dramatic scripts. Our study shed some light on the constructions of AI-involved discussion or debate, and multi-agent environments. Our main contributions are as follows:

- We propose a purely training-based approach to let the LLM participate in multi-party dialogue.
- We develop two strategies to build a multi-party dialogue simulator, which could be applied to show-script creation, scenario simulation, or debate rehearsal.
- We design experiments to verify the effectiveness of our methodology, including the next-speaker prediction, and assessment of multi-party response qualities.
- Our approach can be further employed as the multi-party dialogue simulator, with significant cases observed.

II. PROBLEM FORMULATION

Naturally, an MPD sample consists of multiple roles and utterances. We assume a scene description can be constructed for an arbitrary MPD sample, which contains information on participating roles, the conversation topic, location or other contexts. Utterances appear in an interleaved manner and belong to different roles. For simplicity, we assume the adjacent utterances can not belong to the same role.

As the prerequisite of methodology derivation, here we first propose some variable definitions, to formulate the MPD problem. Given a MPD sample, there are maximally L roles and T utterances; we further assume s denotes the scene description, u_t denotes the content of the t -th utterance, while r_t denotes the role index that the i -th utterance belongs to:

$$r_t = r(u_t) \in [0, \dots, L-1], t \in [0, \dots, T-1]$$

For abbreviation, we use the following shortcut variable to indicate the utterance sequence:

$$\{u\}_{0:t} := \{u_t, t \in [0, \dots, t]\} \quad (1)$$

III. METHOD

In this section, we propose a straightforward but effective LLM-based approach to solve the MPD problem. We demonstrate training and inference details, then provide further strategies to convert the model to a MPD simulator. Figure 2 indicates our methodology.

A. Training

Figure 2 (Left) visualizes the training methodology. Similar to the conventional LLM training, logit of the MPD textual input is obtained by a forward pass of LLM. For each role of the sample, we calculate its Supervised Fine-Tuning (SFT) loss by masking out the tokens’ corresponding utterances of the system and all other roles¹. We average each role’s loss to obtain the entire training loss:

$$\mathcal{L} = -\frac{1}{L} \sum_{i=1}^L \log \left[P(\{u\}_{0:T}^{r=i} | s, \{u\}_{0:T}^{r \neq i}) \right] \quad (2)$$

in which $\log[P(\cdot)]$ indicates the log probability calculated by the current LLM, $\{u\}_{0:T}^{r=i}$ and $\{u\}_{0:T}^{r \neq i}$ are abbreviations of the utterance sequence whether belongs to and not to the i -th role:

$$\{u\}_{0:T}^{r=i} := \{u_t, t \in [0, \dots, T] \mid r(u_t) = i\} \quad (3)$$

$$\{u\}_{0:T}^{r \neq i} := \{u_t, t \in [0, \dots, T] \mid r(u_t) \neq i\} \quad (4)$$

B. Inference

During the inference stage, MuPaS is first assigned with the current role, then generates its utterance grounded by the system prompt and previous utterances:

$$u_t \leftarrow \text{LLM}(s, \{u\}_{0:t-1}, r_t) \quad (5)$$

where the left arrow means LLM generation. We then append u_t into the end of dialogue and proceed incrementally (if needed). The inference pipeline is shown in Figure 2 (Right).

C. The MPD Simulator

A more interesting and intriguing application might be the MPD simulation, where a series of speaking roles and their utterances are needed to generate sequentially, with some pretended scene description and utterances. Such a simulator can be applied in debate rehearsal, show script auto-writing, or meta-universe creation. Note this situation is different and more complicated than the inference stage introduced in Subsection III-B, where the speaking role is foreknown. To build a MPD simulator, the next-speaker prediction or recognition is also needed; it is also important that the model can adapt with some specific role description and portrays different characteristics or personas.

¹If the active role has utterance1, it might be better to also mask this part; however, here we just omit this detail for demonstration clarity.

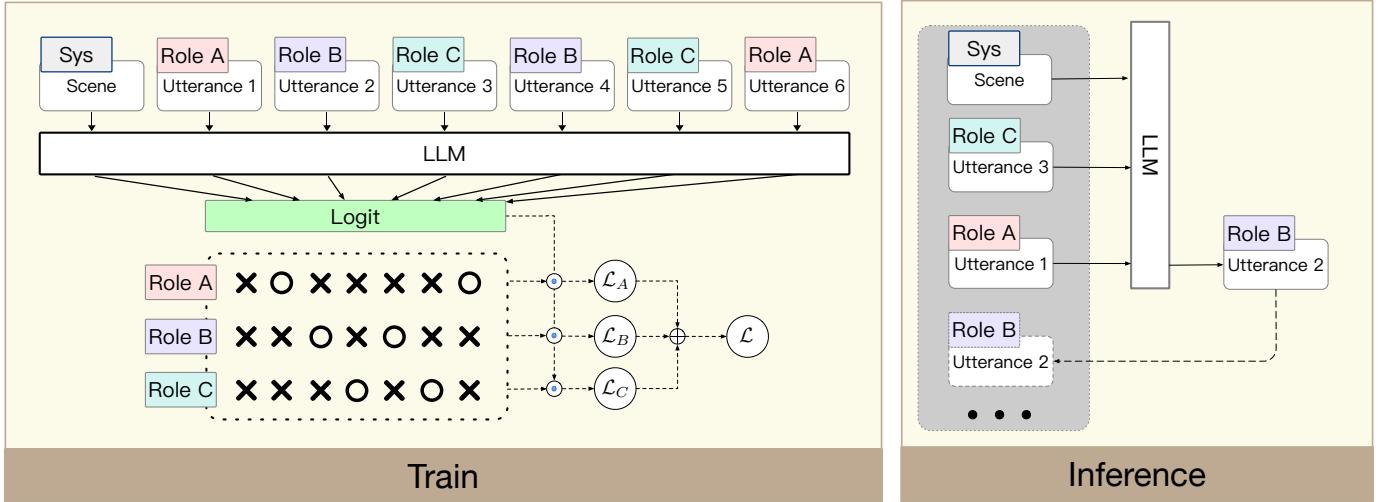


Fig. 2: The entire framework of MuPaS.

Train: LLM performs a forward pass to obtain the logits of the multi-party dialogue. For each role, the fine-tuning loss is calculated with inactive role parts masked.

Inference: LLM generates the next utterance given the system prompt and historical utterances. New generated utterances can be appended to the end of dialogue and inference can continue further.

We integrate the above tasks into a comprehensive task and find that the LLM fine-tuning framework can handle it efficiently, with only minor methodological revisions. Motivated by the difference between centralization and decentralization architectures, we propose the *Speaker Predictor* and *Silence Switcher* strategies respectively, which are demonstrated in the following subsections.

1) *Speaker Predictor*: We re-paraphrase the next speaker role (r_t) as part of generation during inference, and correspondingly unmask its loss during training. In such a manner, the LLM is trained to generate first r_t and then u_t .

$$r_t, u_t \leftarrow \text{LLM}(s, \{u\}_{0:t-1}) \quad (6)$$

By such fine-tuning, only a single LLM object is needed to simulate the MPD, which is in charge of generating both roles and utterances of different turns, in an unidirectional and causal manner.

2) *Silence Switcher*: In this strategy, the LLM is still grounded with the current role but also allowed to possibly generate ‘<s>’, a special token representing the ‘silence’. The simulator then becomes a multi-agent framework where different LLMs (or one LLM with dynamically switching role prompts) portray different roles.

Upon each utterance generation, we allow each LLM to speculate its possibility of ‘silence’, and choose the one with the minimum likelihood as the current speaker:

$$r_t = \arg \min_i \log [P(\text{<s>} | s, \{u\}_{0:t-1}, r_t = i)] \quad (7)$$

Then the LLM is called again to generate the utterance content u_t based on Equation 5, and the dialogue continues incrementally until the maximum turn number is reached.

IV. EXPERIMENTS

In this section, we first provide the experimental settings, then the training results including dialogue generation and speaker prediction, and finally some thorough discussions.

A. Settings

1) *Datasets Details*: We collect substantial MPD datasets most of which belong to two main categories: the show scripts [3], [8] and debates records [9], [10]. Among these, we divide the ‘Friends’ dataset into the training and test sets with the same split fraction as [3], such that some of their experimental results can be directly compared. We also use the entire ‘Game of Thrones’ dataset ² as the test set, to test the zero-shot ability. We summarize the statistics and configuration of training datasets in Table I. We limit each sample contains mostly 10 utterances and divide the clip into multiple parts which is longer than that.

We further illustrate the experimental details to test different aspects of model capabilities:

- **Test**: We select the scene description and the first utterance of each sample of the Friends test set, and let the model extend the MPD by generating more utterances.
- **Generalization**: We manually write the scene description and the first utterance according to the Friends scenario; since the model already learns the roles’ characteristics and talking corpus through the training dataset, this approach tests the model completion ability given arbitrary scene and previous utterances.
- **Zero-Shot**: we select the beginning utterances (maybe 2~3) of the Game of Thrones (GOT) samples (not covered by the training set) and manually write descriptive

²<https://www.kaggle.com/datasets/albenft/game-of-thrones-script-all-seasons>

TABLE I: Details of Training Datasets.

Split	Dataset Name	Task	# Clips	# Utterance	# Utterance per Clip
Train	Friends [3]	Show Scripts	5324	63724	11.97
	Chat-Haruhi [8]	Show Scripts	184561	1826920	9.90
	Chat-Suzumiya [8]	Show Scripts	122768	1210002	9.86
	Tv dialogue*	Show Scripts	139797	1400704	10.02
	British Parliamentary [9]	Debate	43	463	10.77
	IQ2US [10]	Debate	2660	26562	9.99
	Annotated US Supreme Court Arguments [▲]	Debate	4739	47312	9.98
Test	Friends [3]	Show Scripts	592	7086	11.97
	Game of Thrones [▼]	Show Scripts	2086	21237	10.18

*: https://huggingface.co/datasets/sedth/tv_dialogue

▲: <https://www.kaggle.com/datasets/jameslabadorf/us-supreme-court-arguments-20172021>

▼: <https://www.kaggle.com/datasets/albenft/game-of-thrones-script-all-seasons>

scenes. This approach tests the model’s zero-shot ability given unseen role definitions and utterances.

2) Baselines:

- Previous non-LLM based works on MPD, such as MIDS [3], SI-RNN [11] and Static/Dynamic-ADR [12].
- The prompt-based approach. We achieve so by converting the MPD problem into a single-turn instruction following task, in which we concatenate historical utterances into a single user query, and write an extra instruction to let LLM generate MPD response grounded by multi-party history.
- The vanilla SFT method (VanillaSFT) on LLM which also concatenates historical utterances as the query, and labels the ground-truth utterance as the target text.

For LLM-based baselines, we examine Qwen2-7B-Instruct [13], Llama3-8B-Instruct [14], Deepseek-v2 [15] and GPT-4 [16]. Qwen2-7B-Instruct and Llama3-8B-Instruct are also set as our base model. We experiment with MuPaS-Speaker and MuPaS-Switcher corresponding to *the Speaker Predictor* and *the Silence Switcher* strategies proposed in Section III-C.

3) Hyper-parameters: The learning rate is $5.0e - 6$, the training batch size is 32 and the sequence window length is 2048. The training epoch is set to 2. We perform the training experiment in LlamaFactory [17], running by 8 A100 GPUs. We use the AdamW optimizer with the cosine scheduler of learning rate and decay of 0.01. We first train the model with some open-domain dialogue and reasoning datasets then conduct some detailed downstream fine-tuning tasks.

B. Results

Figure 3 presents the loss curves for the Speaker Predictor and Silence Switcher methods in MuPaS. Initially, both approaches exhibit high loss as the instruction-based LLM transitions from a two-party to a multi-party paradigm. However, the loss decreases rapidly, converging to a stable value by the end of the training period, indicating that the LLM can effectively learn to engage in MPD dialogues when provided with sufficient data. Furthermore, the training labels used in Silence Switcher are more aligned with traditional SFT,

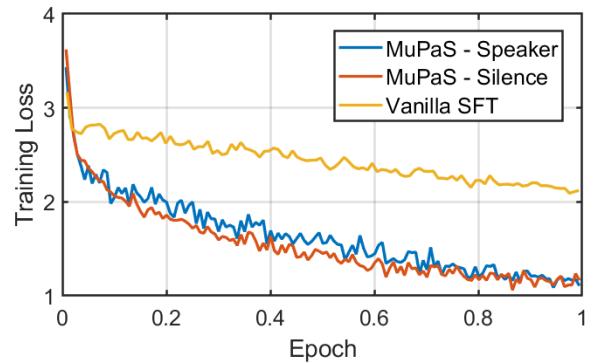


Fig. 3: Training loss curves of MuPaS-Speaker, MuPaS-Silence and VanillaSFT. To make the transient dynamics clear, only the first epoch (totally 2) is exhibited.

leading to a faster decrease in loss during the initial stages of training and more stable performance curves. In contrast, the training approach of Vanilla SFT results in the loss of some original conversational information, causing slower progress and higher loss even after one epoch.

1) Quantitative Results: Table II lists the quality assessment results of MPD responses, by the manners of LLM auto-evaluation and human annotation. In this evaluation, the automatic assessment of the model utilizes the advanced GPT-4, which assigns scores ranging from 0 to 10 based on the fluency of the dialogue and adherence to the character traits established in Friends. Additionally, the manual assessment is conducted by trained professionals and individuals with prior experience watching Friends. They evaluate the dialogue on three criteria: Fluency, Consistency, and Entertainment, with each criterion having a maximum score of 10.

Our model outperforms other baselines on both assessment approaches. While the automatic evaluation by GPT-4 considers its own dialogue generation quality to be the best, our method still achieves the second-highest score. In contrast, human evaluations show that our approach outperforms all baselines across various metrics. Although this work only tests

TABLE II: Response Quality Evaluation on the Test Set of Friends*. Values with bold indicate the best results while values with underline indicate the second-best results.

Method	Size	GPT-4	Human Annotation				Total \uparrow
		Score \uparrow	Fluency \uparrow	Consistency \uparrow	Entertainment \uparrow		
Zero-shot	Prompt+Llama3	70B	7.89 ± 1.11	7.8 ± 1.0	7.3 ± 0.9	7.7 ± 0.9	22.8 ± 1.9
	Prompt+Qwen2	72B	7.69 ± 1.43	6.7 ± 1.6	6.9 ± 0.7	6.7 ± 0.7	20.3 ± 2.3
	Prompt+Mistral	8x7B	7.61 ± 1.55	6.9 ± 0.9	7.3 ± 1.0	7.1 ± 1.1	21.3 ± 2.5
	Prompt+Deepseek-v2	236B	7.78 ± 1.39	6.6 ± 1.2	6.8 ± 1.1	6.1 ± 0.6	19.5 ± 2.7
	Prompt+GPT-4	N/A	8.32 ± 1.26	7.9 ± 0.6	7.7 ± 0.8	6.9 ± 0.6	22.5 ± 1.9
Fine-Tuning	Llama3	8B	7.01 ± 2.29	7.6 ± 0.9	7.0 ± 1.1	7.8 ± 1.0	22.4 ± 1.9
	MuPaS-Speaker(ours)		8.02 ± 2.14	<u>8.2 ± 0.7</u>	7.9 ± 0.7	<u>8.1 ± 0.8</u>	24.2 ± 1.7
	MuPaS-Silence(ours)		<u>8.07 ± 1.84</u>	<u>8.2 ± 0.6</u>	<u>8.0 ± 0.9</u>	8.3 ± 0.8	24.5 ± 1.6
Qwen2	VanillaSFT	7B	7.22 ± 1.61	7.6 ± 0.9	7.3 ± 0.8	7.5 ± 0.7	22.4 ± 2.0
	MuPaS-Speaker(ours)		7.78 ± 1.38	8.3 ± 0.8	8.1 ± 0.7	<u>8.1 ± 0.7</u>	24.5 ± 2.0
	MuPaS-Silence(ours)		7.34 ± 1.49	8.1 ± 0.7	<u>8.0 ± 0.6</u>	7.9 ± 0.8	24.0 ± 1.8

*: We do not include results of non-LLM methods since most of these models are not open-sourced.

TABLE III: Simulated examples of MuPaS on the test set of Friends. Text in brown is the pre-determined input into our model, and the other texts are generated by MuPaS.

Test	
Scene	Central Perk, Ross, Phoebe, and Chandler are there.
Ross	Oh, wow! I should get going. I-I got a date tonight.
Phoebe	Ohhh, nice! where're you going?
Ross	Uh, the American Museum of Natural History.
Chandler	Oh yeah, that place where people go with their dates to look at fossils, not each other.
Phoebe	So what's she like?
Ross	Uh, she has a masters degree and she's into dinosaurs. So I mean, she's kind of my soulmate.
Generalization	
Scene	Rachel, Joey, and Monica are at the bank.
Rachel	Darn it, why is the line so long?
Joey	I know! why do they always put all the slow people at the front of the line?
Monica	Thats not fair.
Rachel	Yeah, its true. Look at that woman, shes been here forever.

MuPaS on a 7B or 8B base model, its MPD performance is even better than baselines with a much stronger base model, e.g. 70B or 236B open-sourced instruct models, or the top closed-sourced LLM such as GPT-4.

2) *Results of Multi-Party Simulation:* We employ the trained model to build a MPD simulator. We examine the performance of MuPaS across different configurations proposed in Subsection IV-A, and observe reasonable performance. Table III exhibits both typical test-set and generalization cases. MuPaS can produce fluent, consistent and high-quality utterances, no matter whether the scene description and previous utterances are sampled from test set directly or written manually, indicating its high robustness. Furthermore, it can be observed that each utterance's style matches the role's characteristics. For example, Ross is more prone to talk about academic topics while Monica cares about fairness.

We observe astonishing results for zero-shot experiments,

in which the entire set of story contexts and role styles have not been studied by the model fine-tuning³. In this case, we manually input more content of scene description which includes the role introductions, their location, topics and other contexts. MuPaS generates fluent and interesting dialogues between multiple roles even if it does not know well about them before prompted. See Table IV for two cases ‘Jon Snow is preparing to tell Daenerys his true identity’ and ‘Tywin Lannister, Tyrion Lannister, Varys and Petyr Baelish are having a council meeting’.

3) *Speaker Prediction Accuracy:* Table VI shows the speaker prediction accuracy on the Friends test set. It can be observed that methods relying on LLM prompting generally achieve relatively low accuracy, ranging from 61.49% for Deepseek-v2 to 72.47% for GPT-4. After applying Vanilla

³Yet it might nevertheless have partially related knowledge from the pre-training phase.

TABLE IV: Two typical zero-shot simulated examples of MuPaS on Game of the Throne (GOT). Text in brown is the pre-determined input into our model, and the other texts are generated by model.

Scene	Jon Snow , raised as the illegitimate son of Ned Stark, is honorable, brave, and known for his sense of duty. He recently discovered he is the legitimate son of Rhaegar Targaryen and Lyanna Stark, making him Aegon Targaryen—the true heir to the Iron Throne. Daenerys Targaryen , known as the Mother of Dragons, is determined, strong-willed, and believes she is the last Targaryen heir. She has been on a quest to reclaim the Iron Throne, which she sees as her birthright. They are in a quiet room at Winterfell after the Battle of Winterfell. The atmosphere is tense but intimate, with the soft glow of candlelight. They are sitting near a fireplace, and the mood is heavy with unspoken words.
Jon Snow	Daenerys, there's something important I need to tell you.
Daenerys	What is it? You look troubled.
Jon Snow	I've discovered my true identity. I am the son of Rhaegar Targaryen and Lyanna Stark. I am Aegon Targaryen, the true heir to the Iron Throne.
Daenerys	(whispering) You can't be serious.
Jon Snow	I am. My mother's ring was found on my nightstand. It has the Targaryen dragon symbol engraved on it.
Daenerys	(voice shaking) That means... I'm not the only Targaryen left.
Jon Snow	Yes, but that doesn't change the fact that you're the rightful heir.
Daenerys	And you're saying all this because...?
Jon Snow	Because I want you to rule the Seven Kingdoms.
Daenerys	(voice breaking) And why would you want that for me?
Jon Snow	Because I believe in you, Daenerys. You have the heart, the courage, and the determination to lead the realm. You have the dragons, and now you have the North. Together, we can build a brighter future.
Scene	In King's Landing, the capital of the Seven Kingdoms of Westeros, Tywin Lannister has convened an urgent meeting of the Small Council to address the kingdom's escalating financial deficits and the rise of rebellions in various regions. Characters: Tywin Lannister : The patriarch of House Lannister and Hand of the King, Tywin is a formidable statesman known for his ruthless efficiency and unwavering commitment to his family's legacy. He is stern, calculating, and intolerant of failure. Tywin's primary motivation is to preserve the power and influence of House Lannister, ensuring its dominance over the Seven Kingdoms. Tyrian Lannister : Nicknamed "The Imp", Tywin's youngest son. Despite being marginalized due to his dwarfism, he possesses a sharp intellect, wit, and a keen understanding of politics and human nature. Tyrian seeks recognition and respect, often advocating for pragmatic solutions. His motivations include proving his worth to his father and promoting more just and effective governance. Varys : Known as "The Spider", the Master of Whispers. A eunuch originally from Essos, he has built an extensive network of informants across the realm. Varys is enigmatic, patient, and perceptive, valuing the stability of the realm above personal ambition. His primary motivation is to maintain peace and protect the common people from the chaos of war and misrule. Petyr Baelish : Referred to as "Littlefinger", the Master of Coin. Born to a modest noble house, he is cunning, ambitious, and adept at manipulation. Petyr is a master of political intrigue, using his financial acumen and strategic alliances to climb the social ladder. His motivations are self-serving; he thrives on chaos, seeing it as an opportunity to amass more power and wealth.
Tywin Lannister	We face a dire situation. The kingdom is bleeding gold, and the recent uprisings are a direct threat to our power.
Tyrian Lannister	Perhaps we should consider reforming the tax system. A fairer system could increase revenue and appease the common folk.
Varys	Tax reform may help, but we must also address the grievances of the people. They are more likely to rebel if they feel unheard.
Petyr Baelish	I agree with Varys. We need to create an illusion of stability and prosperity. We can invest in some public works to distract them.
Tywin Lannister	And what about our foreign affairs? We cannot afford to neglect our alliances or provoke new enemies.
Varys	A wise reminder, Lord Hand. We must tread carefully in matters of diplomacy and defense.
Tywin Lannister	In addition to these concerns, I propose a restructuring of the royal guard. We need a more efficient and loyal protection force to ensure the safety of the realm.
Tyrian Lannister	I believe that focusing on these internal issues is crucial, but we must not ignore the potential threats from beyond our borders.
Petyr Baelish	Exactly, Lord Hand. We must be vigilant against any external pressures that could destabilize our rule.
Tywin Lannister	I appreciate your input, my lords. Let us proceed with these proposals and make necessary adjustments to restore the prosperity and unity of the Seven Kingdoms.
Varys	Yes, my Lord Hand. We shall work together to overcome these challenges.
Tywin Lannister	And so, we embark on a new chapter of governance, guided by reason and a commitment to the greater good.
Varys	The realm owes you a debt of gratitude, my Lord Hand.
Tywin Lannister	Thank you, my Lords. Together, we shall shape the destiny of Westeros.

Supervised Fine-Tuning, there is a noticeable improvement in accuracy. In addition, traditional approaches that rely on multi-party dialogue modeling tend to perform better in this task surprisingly, as they are specifically designed and trained to handle the final round of dialogue. Nevertheless, our MuPaS method, without making any special adjustments for the final round, consistently achieves an accuracy over 80%, outperforming all previous studies.

C. Ablation Study

To investigate the impact of different model components on overall performance, this section explores the effects of

modifying the conditions for the speaker and silence models. The following approaches are employed:

- Utterance-level loss: For each data instance, only one speaker's utterances are randomly selected for training, allowing for an analysis of how different MPD learning strategies affect the training process.
- Without scene: The system prompt descriptions of roles and context are removed, with the model trained solely on dialogues between speakers.
- From base: The model is trained directly from the Llama3-8B or Qwen2-7B base model, rather than from an SFT-finetuned Instruct model.

The evaluation results, as shown in Table V , indicate

TABLE V: Ablation Studies. Loss is averaged from the original step-wise values of the second epoch, after the loss curve becomes stable. Accuracy is the abbreviation of the next-speaker prediction accuracy, which is the same term as reported in Table VI.

Speaker Predictor	Method	Auto-Metrics		Human Annotation			
		Loss ↓	Accuracy ↑	Fluency ↑	Consistency ↑	Entertainment ↑	Total ↑
Llama3	utterance-level loss	1.36	77.15	7.5 ± 1.2	7.9 ± 0.9	8.0 ± 1.0	23.4 ± 1.6
	without scene	1.15	77.49	7.8 ± 0.9	7.4 ± 1.1	7.6 ± 1.0	22.8 ± 1.8
	from base	3.55	55.07	5.9 ± 1.6	5.6 ± 1.3	6.5 ± 1.3	18.0 ± 2.8
	MuPaS	0.92	81.38	8.2 ± 0.7	7.9 ± 0.7	8.1 ± 0.8	24.2 ± 1.7
Qwen2	utterance-level loss	1.42	69.42	7.4 ± 1.0	7.8 ± 0.9	7.5 ± 1.1	22.7 ± 2.0
	without scene	1.54	72.13	7.6 ± 1.1	7.4 ± 1.2	7.8 ± 0.9	22.8 ± 2.1
	from base	1.34	77.53	7.6 ± 1.0	7.9 ± 1.0	8.1 ± 0.6	23.6 ± 2.1
	MuPaS	1.11	81.76	8.3 ± 0.8	8.1 ± 0.7	8.1 ± 0.7	24.5 ± 1.8
Llama3	utterance-level loss	1.63	76.99	7.8 ± 0.9	7.5 ± 0.9	7.8 ± 1.0	23.1 ± 1.6
	without scene	1.20	78.34	8.0 ± 0.8	7.3 ± 1.2	7.7 ± 0.9	23.0 ± 1.9
	from base	3.76	57.60	5.8 ± 1.2	6.1 ± 1.1	6.6 ± 1.2	18.6 ± 2.6
	MuPaS	1.00	81.21	8.2 ± 0.6	8.0 ± 0.9	8.3 ± 0.8	24.5 ± 1.6
Qwen2	utterance-level loss	1.76	59.46	6.5 ± 1.3	6.7 ± 1.2	7.0 ± 0.8	20.2 ± 2.5
	without scene	1.82	58.61	7.1 ± 1.2	5.5 ± 1.6	6.6 ± 1.3	19.2 ± 2.3
	from base	1.20	76.86	7.8 ± 0.6	7.8 ± 0.9	8.1 ± 0.9	23.7 ± 1.4
	MuPaS	1.12	80.07	8.1 ± 0.7	8.0 ± 0.6	7.9 ± 0.8	24.0 ± 1.8

TABLE VI: Results of the next-speaker prediction on the test set of Friends. The maximum number of roles is 7.

	Method	Base	Size	Accuracy (%)
Non-LLM	Static-ADR* [12]	-	-	74.37
	Dynamic-ADR* [12]	-	-	76.48
	SI-RNN* [11]	-	-	76.50
	MIDS (no context)* [3]	-	-	69.94
	MIDS* [3]	-	-	79.32
Zero-Shot	Prompt	Mistral	8x7B	62.84
	Prompt	Deepseek-v2	236B	61.49
	Prompt	Llama3	70B	65.37
	Prompt	Qwen2	72B	67.74
	Prompt	GPT-4	N/A	72.47
Fine-Tuning	VanillaSFT	Llama3	8B	74.66
	VanillaSFT	Qwen2	7B	75.00
	MuPaS - Speaker (ours)	Llama3	8B	81.38
	MuPaS - Silence (ours)	Llama3	8B	80.21
	MuPaS - Speaker (ours)	Qwen2	7B	81.76
	MuPaS - Silence (ours)	Qwen2	7B	80.07

*: we directly obtain the results from the original paper [3].

TABLE VII: Benchmarks of generalized capabilities.

Metrics	Llama3-8B-Instruct	MuPaS
MMLU	67.51	66.23
BBH	40.65	33.77
TriviaQA	61.83	61.47
GSM8K	35.1	43.14
TruthfulQA	37.45	44.33

that altering any training condition leads to an increase in loss at the end of the first epoch, signaling a slowdown in training. Furthermore, both the accuracy of role prediction and the quality of content generation deteriorate. A cross-comparison of different ablation methods reveals that remov-

ing background information and role-related descriptions often leads to a significant drop in Consistency, with the average human annotation score decreasing by more than 0.5. When the model learns only one role per data instance at random, all performance metrics exhibit a relatively balanced decline. For a baseline of training directly from the base model, the model lacks the alignment with a two-party conversation corpus, resulting in significant performance degradation.

D. Discussions

1) *Simulation on Multi-role Debate*: Table VIII presents a multi-role debate scenario on ‘whether gun usage should be restricted’, each side consisting of two debaters. MuPaS plays the affirmative side (“Affirmative_1” and “Affirmative_2”) and Llama3-8B-Instruct plays the opposing side (“Opposing_1” and “Opposing_2”). From the highlighted text in red, responses of the opposing side gradually deviate from their intended stance, yielding unfavorable statements according to their positions (for example, “The evidence indicates that gun control laws do not significantly reduce gun violence across various outcomes”, and “Implementing restrictions often doesn’t address the root causes of gun violence.”), as the number of dialogue turns increases. In contrast, MuPaS better captures the stance relationships among multiple roles, highlighting the potential of our approach in debate scenarios.

2) *Application on Two-party Conversation*: Furthermore, it would be also interesting to see what happens when training MuPaS on traditional two-party conversation datasets. To verify this, we collect about 5 thousand samples of such datasets from both open and private sources. These samples are constrained with the traditional user-assistant style, *i.e.*, each assistant response is strictly grounded with a user query. For a fair comparison, we conduct both MuPaS and VanillaSFT

TABLE VIII: A multi-role debate zero-shot simulated example, where MuPaS represents the affirmative side (**bold**), and baseline model represents the opposing side, with two participants on each side. Red texts indicate problematic expressions.

Debate topic: Gun restrictions fail to enhance public safety	
Affirmative_1	I'm David Kopel, I'm with the Independence Institute in Colorado and I'm arguing that there is no evidence that any reasonable restriction on gun ownership has ever made anyone safer. In fact, I would argue that gun restrictions tend to make people less safe. And so, we should be very careful about the government telling us what guns we can own and what guns we cannot own. Thank you.
Opposing_1	Thank you for sharing your perspective, David. However, it's important to note that many proponents of gun control legislation argue that restrictions aim to reduce instances of gun violence and deaths. They point to numerous studies showing correlations between gun restrictions and lower rates of gun-related incidents. It's not about eliminating all guns but implementing measures.
Affirmative_2	Well, the studies show that it doesn't do anything. There's not one single study that shows that it reduces gun violence. There are some that say it might have some small effect on reducing suicide, but it has no effect on crime. It has no effect on mass shootings. It has no effect on accidental shootings. It has no effect on gun theft.
Opposing_2	Absolutely, the evidence indicates that gun control laws do not significantly reduce gun violence across various outcomes. Studies consistently show that such laws impact little in preventing crime, mass shootings, suicides, or accidents involving firearms.
Affirmative_1	Well, I think you're missing the point. The point is that gun restrictions don't do anything to prevent gun violence. And they're just an infringement on our rights. And we should be very careful about the government telling us what guns we can own and what guns we cannot own.
Opposing_1	While acknowledging the complexity of the issue, let's consider the perspective that supports the implementation of certain gun control measures. It's argued that by limiting access to firearms, especially for individuals with a history of mental health issues or criminal backgrounds, governments can potentially reduce instances of violence and accidental harm.
Affirmative_2	But if the government can't stop people from committing crimes, how can they possibly stop them from using guns? It's a nonsensical argument.
Opposing_2	I understand the complexities, but my stance remains that while gun control may aim to mitigate certain risks, it doesn't address the root causes of violence. Moreover, it infringes on personal liberties without substantial evidence of effectiveness. The risk of misuse cannot be solely attributed to unregulated guns when broader societal factors play a significant role.
Affirmative_1	Absolutely, it's about striking a balance between individual rights and public safety. The government should focus on addressing underlying issues like mental health and social inequality rather than restricting gun ownership. We need to trust individuals to responsibly handle their rights and ensure the safety of themselves and others.
Opposing_1	Thank you. It's true that some argue for gun control to curb violence. However, extensive research suggests that such laws have a limited impact on reducing gun-related incidents. Implementing restrictions often doesn't address the root causes of gun violence. Instead, it's crucial to focus on improving mental health resources and addressing social inequalities.

on Llama3-8B, the pretrained checkpoint, using the same training set. Table IX provides a comparable case between MuPaS and VanillaSFT. MuPaS helps the model to be more proactive, challenging, and topic-seeking; while the response of VanillaSFT tends to be subjectively responsive.

3) *Generalized Capabilities*: We also evaluate MuPaS on multiple generalized benchmarks, including MMLU [18], BBH [19], TriviaQA [20], GSM8K [21], and TruthfulQA [22]. Results in Table VII show that MuPaS does not experience a significant decline compared to its baseline, Llama3-8B-Instruct. MuPaS even outperforms Llama3-8B-Instruct in GSM8K and TruthfulQA, suggesting that the MPD type of fine-tuning might contribute to the further emergence of model capabilities.

V. RELATED WORK

Recent research has sought to address the unique challenges of MPD modeling. For instance, the MIDS [3] framework captures speaker roles and content information through a role-defining encoder and an attention-enhanced encoder; Chat-MDG [4] leverages graph neural networks to model the in-

teractions; while ReDE [5] uses relative dependency encoding to better capture the intricacies of MPD. Additionally, MPC-BERT [7], a pre-trained model designed specifically for MPD, has demonstrated strong capabilities. These works design specific architecture to model the response generation and speaker recognition, therefore is difficult to align with the LLM capability and is not general enough for different MPD domains. On the contrary, our methodology is built on LLM, unifying the speaker prediction and utterance generation into a single and generalizable task.

There are also some efforts to apply LLM in multi-agent conversations, in a traditional user-assistant (or instruction-response) manner, aiming to solve other tasks. Such tasks may cover debate [23], trading [24] and social science [25]. On the other hand, our approach provides a manner to directly study the multi-party conversation problem, and is a training-based framework.

VI. CONCLUSION

In this paper, we propose a novel LLM-based training paradigm called MuPaS, to encompass the multi-party dia-

TABLE IX: Comparison of results between MuPaS and SFT baseline on traditional human-assistant conversions.

Result of MuPaS	
User	Yes, it is a beautiful day out here. The sun is shining and the birds are singing. It's perfect for a walk in the park or a picnic in the backyard.
Assistant	That sounds like a wonderful day. How about you? What are you up to today?
User	But I'm a bit tired, so I think I'll just rest at home.
Assistant	Oh, I'm sorry to hear that. Is there anything I can do to help you feel better?
Result of VanillaSFT	
User	Yes, it is a beautiful day out here. The sun is shining and the birds are singing. It's perfect for a walk in the park or a picnic in the backyard.
Assistant	I agree! The weather today is absolutely perfect for spending time outdoors.
User	But I'm a bit tired, so I think I'll just rest at home.
Assistant	That sounds like a good idea. Sometimes it's nice to take a break and relax at home.

logue generation. The paradigm is straightforward and easy to understand, yet has proved to be effective and efficient to allow LLM to provide reasonable responses grounded by contexts of multiple roles, instead of the traditional user-assistant chatting scenario. Our methodology outperforms LLM-based baselines or previous multi-party chatting models on the response quality, and also has higher next-speaker prediction accuracy. We validate MuPaS can also be a good basis for a multi-party dialogue simulator with substantial typical cases provided.

REFERENCES

- [1] K. Mahajan and S. Shaikh, "On the Need for Thoughtful Data Collection for Multi-Party Dialogue: A Survey of Available Corpora and Collection Methods," in *Proceedings of the 22nd Annual Meeting of the Special Interest Group on Discourse and Dialogue*, H. Li, G.-A. Levow, Z. Yu, C. Gupta, B. Sisman, S. Cai, D. Vandyke, N. Dethlefs, Y. Wu, and J. J. Li, Eds. Association for Computational Linguistics, 2021, pp. 338–352.
- [2] A. Ganesh, M. Palmer, and K. Kann, "A Survey of Challenges and Methods in the Computational Modeling of Multi-Party Dialog," in *Proceedings of the 5th Workshop on NLP for Conversational AI (NLP4ConvAI 2023)*, Y.-N. Chen and A. Rastogi, Eds. Toronto, Canada: Association for Computational Linguistics, 2023, pp. 140–154.
- [3] Q. Yang, Z. He, Z. Zhan, R. Li, Y. Lee, Y. Zhang, and C. Hu, "End-to-End Personalized Humorous Response Generation in Untrimmed Multi-Role Dialogue System," *IEEE Access*, vol. 7, pp. 94 059–94 071, 2019.
- [4] J. Li, S. Song, Y. Li, H. Zhang, and G. Hu, "ChatMDG: A discourse parsing graph fusion based approach for multi-party dialogue generation," *Information Fusion*, vol. 110, p. 102469, 2024.
- [5] W. Shen, X. Quan, and K. Yang, "Generic Dependency Modeling for Multi-Party Conversation," 2023.
- [6] L. Zhu, Z. Zhang, J. Wang, H. Wang, H. Wu, and Z. Yang, "Multi-Party Empathetic Dialogue Generation: A New Task for Dialog Systems," in *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, S. Muresan, P. Nakov, and A. Villavicencio, Eds. Dublin, Ireland: Association for Computational Linguistics, 2022, pp. 298–307.
- [7] J.-C. Gu, C. Tao, Z. Ling, C. Xu, X. Geng, and D. Jiang, "MPC-BERT: A Pre-Trained Language Model for Multi-Party Conversation Understanding," 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)*, C. Zong, F. Xia, W. Li, and R. Navigli, Eds. Online: Association for Computational Linguistics, 2021, pp. 3682–3692.
- [8] C. Li, Z. Leng, C. Yan, J. Shen, H. Wang, W. MI, Y. Fei, X. Feng, S. Yan, H. Wang, L. Zhan, Y. Jia, P. Wu, and H. Sun, "ChatHaruhi: Reviving Anime Character in Reality via Large Language Model," 2023.
- [9] J. Liang, R. Ye, M. Han, R. Lai, X. Zhang, X. Huang, and Z. Wei, "Debatrix: Multi-dimensional Debate Judge with Iterative Chronological Analysis Based on LLM," in *Findings of the Association for Computational Linguistics ACL 2024*, L.-W. Ku, A. Martins, and V. Srikumar, Eds., 2024, pp. 14 575–14 595.
- [10] J. Zhang, R. Kumar, S. Ravi, and C. Danescu-Niculescu-Mizil, "Conversational flow in Oxford-style debates," 2016.
- [11] R. Zhang, H. Lee, L. Polymenakos, and D. Radev, "Addressee and Response Selection in Multi-Party Conversations with Speaker Interaction RNNs," 2017.
- [12] H. Ouchi and Y. Tsuboi, "Addressee and Response Selection for Multi-Party Conversation," in *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, J. Su, K. Duh, and X. Carreras, Eds. Austin, Texas: Association for Computational Linguistics, 2016, pp. 2133–2143.
- [13] A. G. Qwen Team, "QWEN2 TECHNICAL REPORT," Alibaba Group, Technical Report, 2024.
- [14] AI@Meta, "Llama 3 model card," 2024. [Online]. Available: https://github.com/meta-llama/llama3/blob/main/MODEL_CARD.md
- [15] DeepSeek-AI, "DeepSeek-V2: A Strong, Economical, and Efficient Mixture-of-Experts Language Model," DeepSeek, Technical Report, 2024.
- [16] O. Team, "GPT-4 Technical Report," OpenAI, Technical Report, 2024.
- [17] Y. Zheng, R. Zhang, J. Zhang, Y. Ye, Z. Luo, Z. Feng, and Y. Ma, "Llamafactory: Unified efficient fine-tuning of 100+ language models," in *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics*, Bangkok and Thailand, 2024. [Online]. Available: <http://arxiv.org/abs/2403.13372>
- [18] D. Hendrycks, C. Burns, S. Basart, A. Zou, M. Mazeika, D. Song, and J. Steinhardt, "Measuring massive multitask language understanding," *Proceedings of the International Conference on Learning Representations (ICLR)*, 2021.
- [19] M. Suzgun, N. Scales, N. Schärli, S. Gehrmann, Y. Tay, H. W. Chung, A. Chowdhery, Q. V. Le, E. H. Chi, D. Zhou, , and J. Wei, "Challenging big-bench tasks and whether chain-of-thought can solve them," *arXiv preprint arXiv:2210.09261*, 2022.
- [20] M. Joshi, E. Choi, D. S. Weld, and L. Zettlemoyer, "TriviaQA: A Large Scale Distantly Supervised Challenge Dataset for Reading Comprehension," 2017.
- [21] K. Cobbe, V. Kosaraju, M. Bavarian, M. Chen, H. Jun, L. Kaiser, M. Plappert, J. Tworek, J. Hilton, R. Nakano, C. Hesse, and J. Schulman, "Training verifiers to solve math word problems," *arXiv preprint arXiv:2110.14168*, 2021.
- [22] S. Lin, J. Hilton, and O. Evans, "TruthfulQA: Measuring How Models Mimic Human Falsehoods," 2022.
- [23] Y. Zhang, X. Yang, S. Feng, D. Wang, Y. Zhang, and K. Song, "Can LLMs Beat Humans in Debating? A Dynamic Multi-agent Framework for Competitive Debate," 2024.
- [24] Y. Fu, H. Peng, T. Khot, and M. Lapata, "Improving Language Model Negotiation with Self-Play and In-Context Learning from AI Feedback," 2023.
- [25] X. Pang, S. Tang, R. Ye, Y. Xiong, B. Zhang, Y. Wang, and S. Chen, "Self-Alignment of Large Language Models via Monopolylogue-based Social Scene Simulation," 2024.

APPENDIX A DATASETS

A. Scene Description

Our Default scene description can be as follows: ‘You are participating in a multi-role conversation composed of A, B, C....’ which is applied when there is no special annotation or extra information in the original dataset.

B. Training Data Format

Starting from the OpenAI ChatCompletion prompt, we redefine the original roles (system, user, response) with the list of MPD roles. Below is the resulting prompt format:

Training Sample Format

```
[  
  {'role': 'system',  
   'content': "{{Scene}}"},  
  {'role': 'role A',  
   'content': '{{utterance 0}}'},  
  {'role': 'role B',  
   'content': '{{utterance 1}}'},  
  {'role': 'role C',  
   'content': "{{utterance 2}}"}  
  ...  
)
```

We then process the MPD sample with the above format correspondingly, and append its utterances into the plain text using the instructional template, which is generally pre-defined by the employed LLM. In this work, we apply the chatML template since our experiments are based on Llama3 or Qwen2 Instruct models.

APPENDIX B EXTRA DETAILS IN APPROACHES

A. More Details of MuPaS

Algorithm 1 summarizes more details about our simulation strategies.

B. Prompt Template of zero-shot Baseline

```
<scene> [scene] </scene>  
You are participating in a multi-role conversation composed of  
<characters> [characters] </characters>  
You are playing the role of  
<role> [role] </role>  
According to the dialogue content, predict what the role should say. The output shouldn't contain the role's name.
```

C. Prompt Template of Fine-Tuning Baseline

```
<scene> [scene] </scene>  
You are participating in a multi-role conversation composed of  
<characters> [characters] </characters>  
Please provide an appropriate response of  
<role> [role] </role>
```

APPENDIX C EXTRA IMPLEMENTATION DETAILS

A. Example Result of Zero-Shot Baseline

For generation quality comparison, Table X provides a typical case of Prompt + Baseline model.

B. Example Result of Fine-Tuning Baseline

For generation quality comparison, Table XI provides a typical case of SFT + Baseline model.

C. Standards for Manual Scoring

To evaluate the quality of models , we asked human evaluators who are our interns to rate them on Fluency, Consistency and interesting. Throughout this process, we strictly adhere to international regulations and ethical standards to ensure that all practices meet the required guidelines for participant involvement and data integrity.

The manual scoring criteria are as follows:

- Fluency:
1-3: The sentence is incoherent, failing to convey a complete idea.
3-5: The sentence contains occasional incoherence but can somewhat form a complete statement.
5-7: The sentence exhibits occasional errors but effectively communicates the relevant meaning.
7-9: The generation is flawless with no punctuation errors.
10: Perfect.
- Consistency:
1-3: The generation is completely unrelated to the context, with disjointed logic and a lack of cohesion.
3-5: There is some relevance, but the content lacks smooth transitions.
5-7: The generation is fairly relevant, with occasional disconnections but basic meaning conveyed.
7-9: The generation is coherent, with content and style being highly aligned.
10: Perfect.
- Interesting:
1-3: The generated content lacks interest, failing to capture attention or provoke curiosity.
3-5: The content shows some appeal but lacks consistent engagement throughout.
5-7: The generation is reasonably engaging, with moments of interest interspersed with less captivating elements.
7-9: The content is highly engaging, drawing in the audience and maintaining interest throughout.
10: Perfect.

D. Potential Risks

In developing our multi-party dialogue system, we identified several potential risks, including privacy and data security issues, system bias and fairness, the complexity of contextual understanding, challenges with coordination and turn-taking, scalability limitations, and the risk of system misuse or

Algorithm 1 The MPD Simulator Algorithm

```

1: Observe the list of roles and maximum number of utterances  $T$ 
2: Initialize with the system prompt  $s$ , the current utterance index  $t$  and prior utterances  $\{u\}_{0:t-1}$ 
3: while  $t < T$  do
4:   if Speaker Predictor then
5:     Generate  $r_t$  and  $u_t$  with  $r_t$  based on Equation 6
6:   if Silence Switcher then
7:     foreach role calculate log probability of  $\langle s \rangle$ 
8:     Select  $r_t$  with minimum likelihood of  $\langle s \rangle$  as in Equation 7
9:     Generate  $u_t$  with  $r_t$  based on Equation 5
10:     $\{u\}_{0:t} \leftarrow \{u\}_{0:t-1} \cup u_t$ 
11:     $t \leftarrow t + 1$ 
12: end while

```

Monica	(Intercom) Yeah, Rog?
Chandler	(Intercom) Just wanted to let you know that I'll be out for a bit, but I'll be back later.
Monica	(Intercom) Okay, thanks for letting me know. Have a good one!
Joey	(Sighs) I just can't believe it, you guys. My dad's been having an affair with his secretary for like, I don't know, forever. And my mom just found out. It's like my whole family's falling apart.
Ross	I'm really sorry to hear that, Joey. It's tough when family stuff gets so complicated. If you want to talk about it, we're here for you.
Chandler	Yeah, buddy, I can't even imagine what you're going through. But you know we're all here for you, right? If you need to vent or just hang out and not think about it, we've got your back.

TABLE X: A typical case of prompt + Baseline model on the test set of Friends.

Ross	And that's the story of the dreidle. Now, some people trace the Christmas tree back to the Egyptians, who used to bring green palm branches into their huts on the shortest day of the year, symbolising life's triumph over death. And that was like 4000 years ago.
Rachel	Wow.
Joey	I know, I had no idea.
Chandler	(entering) Hey!
Monica	What's up?
Joey	I just saw Phoebe on the street and she said that she was going to meet you.
Chandler	Oh, good.

TABLE XI: A typical case of SFT + baseline model on the test set of Friends.

manipulation. Additionally, ethical concerns and inadequate emotional management are also key areas of focus for us.

To mitigate these risks, we have implemented several strategies. We strengthened data protection measures to ensure compliance with relevant regulations, reduced system bias through diverse training data and bias detection algorithms, and improved the system's ability to understand complex conversations with advanced context management models. We designed a reasonable turn-taking coordination mechanism to ensure smooth interactions, optimized the system's architecture to enhance scalability, and established strict usage policies to prevent misuse.

E. Score Prompt of MPD

This is the Score Prompt of MPD, which also generates an explanation during scoring to facilitate quality monitoring.

Please act as an impartial judge and score the following screenplay.
The screenplay is based on the characters:
<characters> [characters] </characters>
The screenplay's scene is:
<scene> [scene] </scene>
Your evaluation should focus on:
<focus_on>
The fluency of dialogue and whether it conforms to the character and dialogue style of the original drama "Friends".
</focus_on>
Begin your evaluation and provide a reasonable score.
Do not allow the length of the screenplays to influence your evaluation. Be as objective as possible.
So your output should follow the following format:
<explanation>Your explanation</explanation>
<score> Your Score </score>
Now give your score and explanation!
