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

Anonymous ACL submission

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.

1 Introduction

In recent years, large language model (LLM) has 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 the Multi-Party Dialogues (MPD) [1], such as team meetings, classroom discussions, court or academic debates, or simply daily conversions with multiple humans involved (Mahajan and Shaikh, 2021; Ganesh et al., 2023). Instead of responding to a single user's query, in such case the dialog system needs to understand conversation contexts from multiple users, determine whether to speak or not, and reasonably participate in potential multiple concurrent topics. Novel modeling technique is therefore required to adapted with this different dialogue paradigm.

Figure 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.

Previous researches have sought to address the unique challenges of MPD modeling, such as MIDS (Yang et al., 2019), ChatMDG (Li et al., 2024), ReDE (Shen et al., 2023), SDMPED (Zhu et al., 2022) and MPC-BERT (Gu et al., 2021). 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-

[1] Some papers instead name this scenario by multi-party conversation (MPC).

based MPD approaches, which are generally based on multi-agent systems. Each LLM agent is still adhere to a binary interaction framework, and the conversations might happen between different pairs of agents which are not happened simultaneously and concurrently. Besides that, conventional LLMs are fine-tuned with human-assistant dialogues with the alignment with the helpful, impartial, and conservative conversation style. In short, there is still not a LLM-based training framework which allows the model 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 **Mu**lti-**Pa**rty **S**upervised (**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 pre-process 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 interesting 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 code and model are open-sourced on `https://anonymous.4open.science/r/MuPas-5F53`. Our main contributions are as follows:

- We propose an LLM-based approach to learn to participate in multi-party dialogue, which to the best of our knowledge, is the first time to introduce LLM in this area.
- 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.
- We examine the capability of our approach as the multi-party dialogue simulator, with substantial case performance observed.

2 Problem Formulation

Naturally, an MPD sample is consisting of multiple roles and utterances. We further 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[2]. Utterances appear in an interleaved manner and belong to different roles. For simplicity, we just 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:

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

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

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

3 Method

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

3.1 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

[2]If there is no MPD scene description in the original dataset, we will automatically generate a scene description based on the role list, *e.g.*, 'A, B, C are participating a conversation.'. See more details in the Appendix.

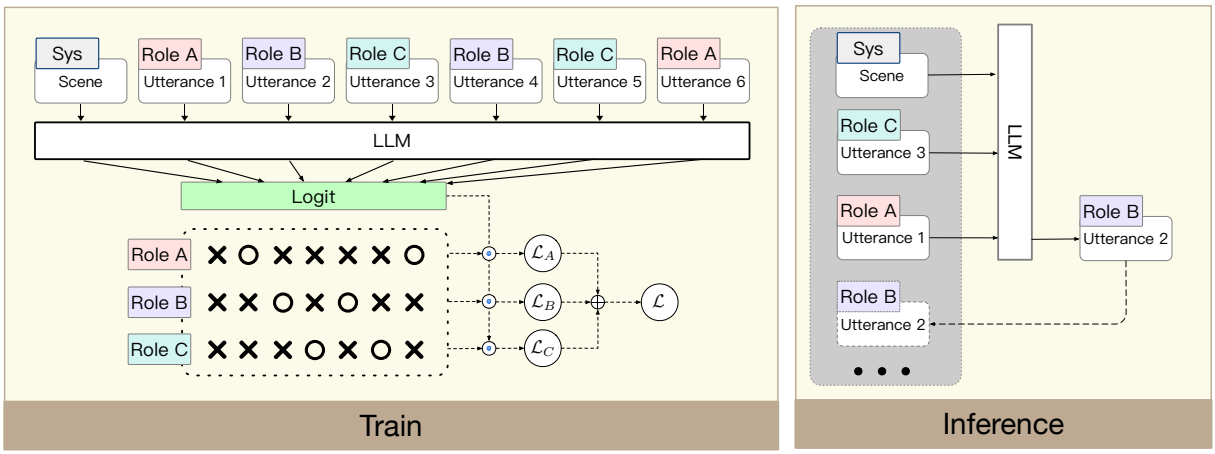

Figure 2: The entire framework of MuPaS.
Train: LLM performs a forward pass to obtain the logits of the multi-party dialogue. For each role, the fine-tuning loss is calculated with inactive role parts masked.
Inference: LLM generates the next utterance given the system prompt and historical utterances. New generated utterances can be appended to the end of dialogue and inference can continue further.

its Supervised Fine-Tuning (SFT) loss by masking out the tokens' corresponding utterances of the system and all other roles [3]. We average each role's loss to obtain the entire training loss:

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

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

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

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

3.2 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:

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

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

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

3.3 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 3.2, where the speaking role is foreknown. To build a MPD Simulator, the next-speaker prediction or recognition is also needed; and it is also important that the model can adapt with some specific role description and portrays different characteristics or personas.

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.

3.3.1 Speaker Predictor

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

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

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.

3.3.2 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:

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

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.

We further summarize and compare details of the above two strategies by Algorithm 1 in Appendix B.1.

4 Experiments

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

4.1 Settings

We collect substantial MPD datasets most of which belong to two main categories: the show scripts and debates records. One can refer to Appendix A.1 for completed details. Among these, we divide the 'Friends' dataset into the training and test test with the same split fraction as (Yang et al., 2019), such that some of their experimental results can be directly compared. We also use the entire 'Game of Thrones' dataset as the test test, to test the zero-shot ability. 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 test, and let the model to 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 learn the roles' characteristic and talking corpus through the training dataset, this approach test 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 scenes. This approach test the model's zero-shot ability given unseen role definitions and utterances.

4.2 Baselines

- Previous non-LLM based works on MPD, such as MIDS (Yang et al., 2019), SI-RNN (Zhang et al., 2017) and Static/Dynamic-ADR (Ouchi and Tsuboi, 2016).
- The prompt-based approach. We achieve so by converting the MPD problem in to 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 to 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 Instruct (Qwen Team, 2024), Llama3 Instruct (AI@Meta, 2024), Deepseek-v2 (DeepSeek-AI, 2024) and GPT4 (Team, 2024). Qwen2-7B-Instruct is 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 3.3.

4.3 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,

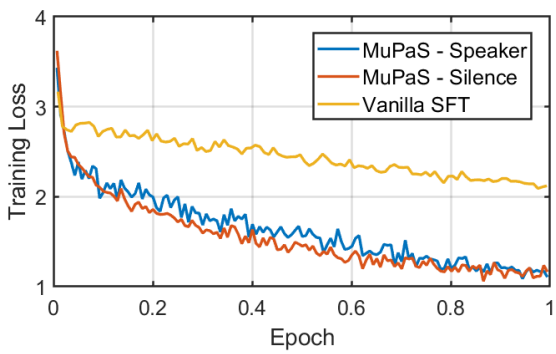

Figure 3: Training loss curves of MuPaS-Speaker, MuPaS-Silence and VanillaSFT. All training results are all based on Qwen2 7B Instruct. To make the transient dynamics clear, only the first epoch (totally 2) is exhibited.

the training labels used in Silence Switcher are more aligned with traditional SFT, 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.

4.3.1 Quantitative Results

Table 1 list 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. The prompt used for this evaluation is provided in the Appendix. Additionally, the manual assessment is conducted by trained professionals or individuals with prior experience watching Friends. They evaluate the dialogue on three criteria: Fluency, Consistency, and Interestingness, with each criterion having a maximum score of 10. Detailed scoring criteria are also included in the Appendix.

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 we only test MuPaS on a 7B base model, its MPD performance is even better than baselines with a much stronger base model, *e.g.* 72B or 236B open-sourced instruct models, or the top closed-sourced LLM such as GPT4.

4.3.2 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 4.1, and observe reasonable performance. Table 2 exhibits both typical test-set and generalization cases. MuPaS can produce fluent, consistent and high-quality utterances, no matter that 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 matching the role's characteristics. For example, Ross is more prone to talk about academic topic while Monica cares about the fair.

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-tuning [4]. In this case, we manually input more contents of scene description which including the role introductions, their location, topics and other contexts ('Jon Snow are preparing to tell Daenerys his true identity'). MuPaS generates fluent and interesting dialogues between 'Jon Snow' and 'Daenerys' even it does not know well about them before prompted. Table 7 provides another case which include more than two roles ('Tywin Lannister, Tyrion Lannister, Varys and Petyr Baelish are having a council meeting') in the Appendix C.2.

4.3.3 Speaker Prediction Accuracy

Table 2 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 50.51% for Qwen2 7B to 72.47% for GPT-4. The accuracy is also positively correlated with model size and conversational capabilities. After applying a simple 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, 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, achieves an accuracy of 81.76%, outperforming all previous studies.

[4] But it might nevertheless have partially related knowledge from the pre-training phase.

	Method	Base	GPT4 Score↑	Human Annotation Fluency↑	Consistency↑	Interesting↑	Total↑
Zero-Shot	Prompt + Qwen2	72B	7.69	6.7	6.9	6.7	20.3
	Prompt + Deepseek-v2	236B	7.78	6.6	6.8	6.1	19.5
	Prompt + GPT4	N/A	**8.32**	7.9	7.7	6.9	22.5
Fine-Tuning	VanillaSFT + Qwen2	7B	7.22	7.6	7.3	7.5	22.4
	MuPaS-Speaker (ours)	7B	7.78	**8.3**	**8.1**	**8.1**	**24.5**
	MuPaS-Silence (ours)	7B	7.34	8.1	8.0	7.9	24.0

Table 1: Response Quality Evaluation on the Test Set of Friends*. Values with bold indicates the best results while values with underline indicates the second best results.

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

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.

Table 2: Simulated examples of MuPaS on the test test of Friends. Text in brown is the pre-determined input into our model, and the other texts are generated by model.

4.4 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:

- One role/sample: 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 Qwen2 7B base model, rather than from an SFT-finetuned Instruct model.

The evaluation results, as shown in Table 5 , indicate 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 training directly from the base model is less affected, while the other two groups experience a more significant impact. Finally, the speaker predictor and silence predictor exhibit different sensitivities to the experimental conditions. The decline in generation quality metrics for the speaker predictor is relatively minor. In particular, the removal of descriptions related to context and roles results in the silence predictor losing substantial amounts of critical information, leading to decreases in accuracy and consistency of 25.22% and 31.25%, respectively, which significantly affects its overall performance.

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.

Table 3: A typical zero-shot simulated example 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.

5 Related Work

In this section, we review previous works and the recent progress of LLM on multi-party dialogues.

5.1 Modeling on Multi-Party Dialogue

Recent research has sought to address the unique challenges of MPD modeling. For instance, the MIDS (Yang et al., 2019) framework captures speaker roles and content information through a role-defining encoder and an attention-enhanced encoder. Other models, such as ChatMDG (Li et al., 2024), leverage graph neural networks to model the interactions, while ReDE (Shen et al., 2023) uses relative dependency encoding to better capture the intricacies of MPD. Additionally, MPC-BERT (Gu et al., 2021), a pre-trained model designed specifically for MPD, has demonstrated strong capabilities in tasks such as addressee recognition and response selection by incorporating structural modeling of speaker roles. 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 which takes advantage of its dyadic chatting capability and can unify the speaker prediction and utterance generation into a single and generalized task.

5.2 Multiple LLM-Agent Conversations

There are also some efforts to apply LLM on multi-agent conversations. Such approach employs LLM in a traditional user-assistant (or instruction-response) manner and aim to solve other tasks. For example, LLM-debate (Yung-Sung Chuang, 2023) and Agent4Detate (Zhang et al., 2024) let the agents to participate the debate and study the Society of Mind (Minsky, 1988) on the LLM basis. Trade (Fu et al., 2023) model uses multiple agents to simulate distinct roles in game-theoretic scenarios. MATRIX (Pang et al., 2024) simulate the multi-role interaction process and acknowledge

	Method	Size	Accuracy (%)
non-LLM	Static-ADR* (Ouchi and Tsuboi, 2016)	-	74.37
	Dynamic-ADR* (Ouchi and Tsuboi, 2016)	-	76.48
	SI-RNN* (Zhang et al., 2017)	-	76.50
	MIDS (no context)* (Yang et al., 2019)	-	69.94
	MIDS* (Yang et al., 2019)	-	79.32
Zero-Shot	Prompt + Qwen2	7B	50.51
	Prompt + Qwen2	72B	67.74
	Prompt + Llama3	70B	65.37
	Prompt + GPT4	N/A	72.47
Fine-Tuning	VanillaSFT + Llama3	8B	74.66
	VanillaSFT + Qwen2	7B	75.00
	MuPaS - Speaker (ours)	7B	**81.76**
	MuPaS - Silence (ours)	7B	78.38

Table 4: Result of the next-speaker prediction on the test set of Friends. The maximum number of roles is 7. $*$: we directly obtain the results from the original paper (Yang et al., 2019).

	Method	Auto-Metrics		Human Annotation			
		Loss↓	Accuracy↑	Fluency↑	Consistency↑	Interesting↑	Total↑
Speaker Predictor	one role/sample	1.42	69.42	7.4	7.8	7.5	22.7
	without scene	1.54	72.13	7.6	7.4	7.8	22.8
	from base	1.34	77.53	7.6	7.9	**8.1**	23.6
	MuPaS	**1.11**	**81.76**	**8.3**	**8.1**	**8.1**	**24.5**
Silence Switcher	one role/sample	1.76	59.46	6.5	6.7	7.0	20.2
	without scene	1.82	58.61	7.1	5.5	6.6	19.2
	from base	1.20	76.86	7.8	7.8	**8.1**	23.7
	MuPaS	**1.16**	**78.38**	**8.1**	**8.0**	7.9	**24.0**

Table 5: 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 4.

the social consequence. Similar efforts have be proposed to solve the math problem (Liu et al., 2023) or software development (Qian et al., 2024). On the other hand, our approach provide a manner to directly study the multi-party conversation problem, and is a training-based framework.

6 Conclusion

In this paper, we propose a novel LLM-based training paradigm called MuPaS, to encompass the multi-party dialogue generation. The paradigm is straightforward and easy to understand, but is proved to be effective and efficient to allow LLM provide reasonable response grounded by context 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 of multi-party dialogue simulator with substantial typical cases provided.

7 Limitation

MuPaS is trained with general MPD datasets, and currently do not cover multimodal or multi-thread topics. The MPD scenario can be classified into scripted (such as show, movie scripts) and unscripted (such as daily and open-domain conversations), while we provide a general training-based and data-driven solution, while do not study these scenario differences.

Our work propose an academic solution to generate MPD while it does have the possibility that our MuPaS can be used to create fake and fraud stories. Such application should be prohibited.

References

AI@Meta. 2024. Llama 3 model card.

DeepSeek-AI. 2024. DeepSeek-V2: A Strong, Economical, and Efficient Mixture-of-Experts Language Model. Technical report, DeepSeek.

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.

Ananya Ganesh, Martha Palmer, and Katharina Kann. 2023. 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)*, pages 140–154, Toronto, Canada. Association for Computational Linguistics.

Jia-Chen Gu, Chongyang Tao, Zhenhua Ling, Can Xu, Xiubo Geng, and Daxin Jiang. 2021. 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)*, pages 3682–3692, Online. Association for Computational Linguistics.

Cheng Li, Ziang Leng, Chenxi Yan, Junyi Shen, Hao Wang, Weishi MI, Yaying Fei, Xiaoyang Feng, Song Yan, HaoSheng Wang, Linkang Zhan, Yaokai Jia, Pingyu Wu, and Haozhen Sun. 2023. ChatHaruhi: Reviving Anime Character in Reality via Large Language Model. *Preprint*, arXiv:2308.09597.

Jingyang Li, Shengli Song, Yixin Li, Hanxiao Zhang, and Guangneng Hu. 2024. ChatMDG: A discourse parsing graph fusion based approach for multi-party dialogue generation. *Information Fusion*, 110:102469.

Jingcong Liang, Rong Ye, Meng Han, Ruofei Lai, Xinyu Zhang, Xuanjing Huang, and Zhongyu Wei. 2024. Debatrix: Multi-dimensional Debate Judge with Iterative Chronological Analysis Based on LLM. In *Findings of the Association for Computational Linguistics ACL 2024*, pages 14575–14595, Bangkok, Thailand and virtual meeting. Association for Computational Linguistics.

Zijun Liu, Yanzhe Zhang, Peng Li, Yang Liu, and Diyi Yang. 2023. Dynamic LLM-Agent Network: An LLM-agent Collaboration Framework with Agent Team Optimization. *Preprint*, arXiv:2310.02170.

Khyati Mahajan and Samira Shaikh. 2021. 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*, pages 338–352, Singapore and Online. Association for Computational Linguistics.

M. Minsky. 1988. *Society of mind*, 1st edition.

Hiroki Ouchi and Yuta Tsuboi. 2016. Addressee and Response Selection for Multi-Party Conversation. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pages 2133–2143, Austin, Texas. Association for Computational Linguistics.

Xianghe Pang, Shuo Tang, Rui Ye, Yuxin Xiong, Bolun Zhang, Yanfeng Wang, and Siheng Chen. 2024. Self-Alignment of Large Language Models via Monopolylogue-based Social Scene Simulation. *Preprint*, arXiv:2402.05699.

Chen Qian, Yufan Dang, Jiahao Li, Wei Liu, Zihao Xie, Yifei Wang, Weize Chen, Cheng Yang, Xin Cong, Xiaoyin Che, Zhiyuan Liu, and Maosong Sun. 2024. Experiential Co-Learning of Software-Developing Agents. *Preprint*, arXiv:2312.17025.

Alibaba Group Qwen Team. 2024. QWEN2 TECHNICAL REPORT. Technical report, Alibaba Group.

Weizhou Shen, Xiaojun Quan, and Ke Yang. 2023. Generic Dependency Modeling for Multi-Party Conversation. *Preprint*, arXiv:2302.10680.

OpenAI Team. 2024. GPT-4 Technical Report. Technical report, OpenAI.

Qichuan Yang, Zhiqiang He, Zhiqiang Zhan, Rang Li, Yanwei Lee, Yang Zhang, and Changjian Hu. 2019. End-to-End Personalized Humorous Response Generation in Untrimmed Multi-Role Dialogue System. *IEEE Access*, 7:94059–94071.

Hongyin Luo Yoon Kim James Glass Pengcheng He Yung-Sung Chuang, Yujia Xie. 2023. Dola: Decoding by contrasting layers improves factuality in large language models. *Preprint*, arXiv:2309.03883.

Justine Zhang, Ravi Kumar, Sujith Ravi, and Cristian Danescu-Niculescu-Mizil. 2016. Conversational flow in Oxford-style debates. *Preprint*, arXiv:1604.03114.

Rui Zhang, Honglak Lee, Lazaros Polymenakos, and Dragomir Radev. 2017. Addressee and Response Selection in Multi-Party Conversations with Speaker Interaction RNNs. *Preprint*, arXiv:1709.04005.

Yiqun Zhang, Xiaocui Yang, Shi Feng, Daling Wang, Yifei Zhang, and Kaisong Song. 2024. Can LLMs Beat Humans in Debating? A Dynamic Multi-agent Framework for Competitive Debate. *Preprint*, arXiv:2408.04472.

Yaowei Zheng, Richong Zhang, Junhao Zhang, Yanhan Ye, Zheyan Luo, Zhangchi Feng, and Yongqiang Ma. 2024. 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.

Ling.Yu Zhu, Zhengkun Zhang, Jun Wang, Hongbin Wang, Haiying Wu, and Zhenglu Yang. 2022. 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)*, pages 298–307, Dublin, Ireland. Association for Computational Linguistics.

A Datasets

A.1 Datasets Details

We collect substantial MPD datasets which can be classified into two main categories: the show scripts and debates records. To align with the experimental settings of (Yang et al., 2019), we also divide the 'Friends' dataset into the training and test test with the same split fraction. We also use the entire 'Game of Thrones' dataset as the test test. All other datasets are part of training set. We summarize the statistics and configuration of training datasets in Table 6. We limit each sample contains mostly 10 utterances and divide the clip into multiple parts which is longer than that.

A.2 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.

A.3 Training Data Format

Starting from the OpenAI ChatCompletion prompt, we re-define 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 Qwen2 Instruct models.

Split	Dataset Name	Task	# of Clips	# of Utterance	# of Utterance per Clip
Train	Friends (Yang et al., 2019)	Show Scripts	5324	63724	11.97
	Chat-Haruhi(Li et al., 2023)	Show Scripts	184561	1826920	9.90
	Chat-Suzumiya(Li et al., 2023)	Show Scripts	122768	1210002	9.86
	Tv dialogue*	Show Scripts	139797	1400704	10.02
	British Parliamentary(Liang et al., 2024)	Debate	43	463	10.77
	IQ2US(Zhang et al., 2016)	Debate	2660	26562	9.99
	Annotated US Supreme Court Arguments▲	Debate	4739	47312	9.98
Test	Friends (Yang et al., 2019)	Show Scripts	592	7086	11.97
	Game of Thrones▼	Show Scripts	2086	21237	10.18

Table 6: Details of Training Datasets.
∗: https://huggingface.co/datasets/sedthh/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

B Extra Details in Approaches

B.1 More details of MuPaS

Algorithm 1 summarizes more details about our simulation strategies.

B.2 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.
"""

B.3 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>
"""

C Experimental Details

C.1 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 (Zheng et al., 2024), 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.

C.2 Extra Results of MuPaS

Table 7 provides a multi-role case about the scene of 'Tywin Lannister, Tyrion Lannister, Varys and Petyr Baelish are having a council meeting'. In this case, generated utterances are still solid and reasonable, and one can find that the entire meeting is completed. This result indicate that our method can be potentially applied on the virtual rehearsal of meeting, debate or similar discussions.

Furthermore, we also find that simply training MuPaS on traditional two-party conversation datasets can help the model to be more proactive, challenging and topic-seeking, by learning the response of user grounded by the assistant. See Table 8 for a comparison.

C.3 Example Result of Zero-Shot Baseline

A typical case of Prompt + Qwen2-7B-Instruct:

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!

Algorithm 1 The MPD Simulator Algorithm

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

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.

C.4 Example Result of Fine-Tuning Baseline

A typical case of SFT + Qwen2-7B-Instruct:

Ross: and thats 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 lifes triumph over death. And that was like 4000 years ago.

Rachel: Wow.

Joey: I know, I had no idea.

Chandler: (entering) Hey!

Monica: Whats up?

Joey: I just saw Phoebe on the street and she said that she was going to meet you.

Chandler: Oh, good.

C.5 Score Prompt of MPD

"""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! """

C.6 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.

Scene	Background: 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 for generations to come. Tyrion Lannister: Nicknamed "The Imp," Tyrion is 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. Tyrion 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," Varys is the Master of Whisperers. 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," Petyr is 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. Setting: The meeting is held in the council chamber of the Red Keep, a room with high ceilings and narrow windows overlooking the bustling city below. Rich tapestries depicting historical battles adorn the stone walls, and a large wooden table dominates the center of the room. The atmosphere is heavy with tension and unspoken agendas. Candles flicker atop iron sconces, casting elongated shadows that mirror the looming challenges facing the kingdom. Each member of the council is acutely aware that their decisions today will have profound consequences for the future of Westeros.
Tywin Lannister	We face a dire situation. The kingdom is bleeding gold, and the recent uprisings are a direct threat to our power.
Tyrion Lannister	Perhaps we should consider reforming the tax system. A fairer system could increase revenue and appease the common folk.
Varys	Tax reform may indeed 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. Perhaps 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.
Tyrion 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.
Tyrion 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.
Tyrion Lannister	Thank you, my Lords. Together, we shall shape the destiny of Westeros.

Table 7: Another zero-shot simulated example of MuPaS on Game of Thrones (GOT) with more than two roles.

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.

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 SFT baseline	
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.

Table 8: Comparison of results between MuPaS and SFT baseline on traditional human-assistant conversions.

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.

C.7 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 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. Furthermore, we introduced emotion analysis and conflict management mechanisms to handle complex emotional scenarios, ensuring that the system performs reliably in multi-party dialogue settings.