Selective Prompting Tuning for Personalized Conversations with LLMs

Qiushi Huang[1,2 *], **Xubo Liu**[1], **Tom Ko**[3], **Bo Wu**[4], **Wenwu Wang**[1], **Yu Zhang**[2†] , **Lilian Tang**[1 †]
[1]University of Surrey, [2]Southern University of Science and Technology,
[3]ByteDance AI Lab, [4]MIT-IBM Watson AI Lab
{qiushi.huang, xubo.liu, w.wang, h.tang}@surrey.ac.uk,
{tomkocse, yu.zhang.ust}@gmail.com, bo.wu@ibm.com

Abstract

In conversational AI, personalizing dialogues with persona profiles and contextual understanding is essential. Despite large language models' (LLMs) improved response coherence, effective persona integration remains a challenge. In this work, we first study two common approaches for personalizing LLMs: textual prompting and direct fine-tuning. We observed that textual prompting often struggles to yield responses that are similar to the ground truths in datasets, while direct fine-tuning tends to produce repetitive or overly generic replies. To alleviate those issues, we propose **S**elective **P**rompt **T**uning (SPT), which softly prompts LLMs for personalized conversations in a selective way. Concretely, SPT initializes a set of soft prompts and uses a trainable dense retriever to adaptively select suitable soft prompts for LLMs according to different input contexts, where the prompt retriever is dynamically updated through feedback from the LLMs. Additionally, we propose context-prompt contrastive learning and prompt fusion learning to encourage the SPT to enhance the diversity of personalized conversations. Experiments on the CONVAI2 dataset demonstrate that SPT significantly enhances response diversity by up to 90%, along with improvements in other critical performance indicators. Those results highlight the efficacy of SPT in fostering engaging and personalized dialogue generation. The SPT model code is publicly available for further exploration. [1]

1 Introduction

Personalization in dialogue systems enhances user interaction by creating a coherent and customized experience. It involves adapting conversations to individual preferences, backgrounds, and real-time context, ensuring each dialogue feels personally relevant. This tailored approach fosters a deeper connection between users and technology, making interactions more intuitive and engaging. By understanding and anticipating user needs, personalized dialogues can offer more than just relevant responses; they provide a seamless, conversational experience that mirrors human interaction, enriching the overall quality of digital communication.

PersonaChat (Zhang et al., 2018) has become a pivotal dataset for personalization research in conversational AI, offering persona profiles that detail an interlocutor's preferences and background in four to five sentences. These profiles guide conversational agents in creating dialogues that are both engaging and consistent with the persona's characteristics and prior conversational context. This area has seen diverse approaches for enhancing personalization, such as attention mechanisms (Huang et al., 2023b), reinforcement learning with multiple rewards (Song et al., 2021; Liu et al., 2020), and persona profile enrichment through stories (Huang et al., 2023a), demonstrating the breadth of innovation in making interactions more personalized and meaningful.

Recently, the advent of large language models (LLMs) (Zhang et al., 2022; Touvron et al., 2023) has opened new avenues for dialogue generation, offering the potential for creating conversations that align with human preferences. However, fully leveraging LLMs to achieve the level of personalization showed in PersonaChat is a promising yet underexplored area. Currently, LLMs are primarily guided by direct textual prompts or through parameter-efficient fine-tuning like prompt tuning (Lester et al., 2021) that only tunes a few virtual tokens instead of whole LLMs for specific tasks.

However, designing personalized conversational agents with LLMs faces two main challenges. The primary issue lies in diverse settings in conversations, which encompass a wide array of dialogues, each characterized by unique persona profiles and

[*] Work done as an intern at ByteDance AI Lab.
[†] Corresponding authors.
[1] https://github.com/hqsiswiliam/SPT

varying lengths of conversation. This diversity necessitates an understanding of the distinct conversational settings within the data. Through textual prompting, it is hard to guide the model to generate desired responses aligned with the target texts. Simply fine-tuning LLMs through prompt tuning without careful conversational setting analysis risks producing responses that lack specificity and depth, resulting in a generic and bland generation.

Secondly, another equally critical challenge arises from the limitations inherent to the datasets used for persona-based dialogue generation. Typically small and lacking in diversity, these datasets can restrict the model's exposure to a wide range of conversational scenarios. When LLMs (e.g., Llama2-7B (Touvron et al., 2023)) are tuned through trainable soft prompts on PersonaChat, they risk overfitting to specific persona profiles. This overfitting manifests in the model's responses, which become repetitive and overly aligned with the persona, often at the cost of dynamic and contextually appropriate interactions. Although this might lead to improvements in metrics such as F1 or BLEU scores, it detracts from the overall diversity and engagingness of the dialogues, undermining the model's ability to emulate authentic human conversation.

To handle those two challenges when designing personalized conversations with LLMs, we propose a Selective Prompt Tuning (SPT) model. Specifically, to tackle the first challenge, it is crucial to identify inherent data patterns without explicit annotations. To achieve this, it is intuitive to utilize a group of multiple soft prompts to handle different conversational settings when tuning the model in a parameter-efficient way. However, as previously mentioned, the annotations for the dialogue settings are missing and even hard to discover and annotate. If we naively concurrently tune all prompts without clear distinctions, this would yield only marginal differences compared with tuning one soft prompt. Therefore, to build effective multiple prompts to discover the inherent data pattern inside the personalized dialogue, the proposed SPT model utilizes a dense retriever to adaptively select a proper soft prompt from the soft prompt group based on the given input context. To distinguish the effectiveness of soft prompts, we utilize the loss from LLMs as feedback to guide the update of the dense retriever without explicit annotations. Based on this, the proposed SPT model could discover patterns intrinsically associated with different dialogues. In this way, the retriever and soft prompt group evolve together, benefiting from continuous interactions that enrich their capability to discriminate and generate diverse, contextually relevant responses.

To address the second challenge that LLM may overfit small-scale datasets such as PersonaChat, the proposed SPT method integrates two complementary mechanisms: context-prompt contrastive learning and prompt fusion learning. The context-prompt contrastive learning mechanism ensures diversity by encouraging the use of different soft prompts for varied dialogue contexts, preventing repetitive responses. Concurrently, prompt fusion learning aggregates all prompt predictions during back-propagation, optimizing towards a unified output. This dual strategy not only preserves response diversity across contexts but also enhances overall model performance, demonstrating their cooperative effectiveness in tackling overfitting while maintaining the performance.

By integrating the above two parts into the SPT method, experiments on the CONVAI2 dataset (Dinan et al., 2019) with LLMs (i.e., Llama2 (Touvron et al., 2023) and OPT (Zhang et al., 2022)) not only demonstrate marked improvements in response diversity and engagingness but also indicate enhancements in other key performance metrics. Quantitatively, the proposed SPT model consistently outperforms baselines across models with various sizes. Moreover, SPT offers profound insights into different dialogue scenarios, particularly in the model's strategic prompt selection. Comprehensive ablation studies highlight the adaptability of different prompts to specific dialogue contexts.

Overall, our contributions can be summarized as follows.

- We present the novel SPT method by integrating a trainable dense retriever with dynamic soft prompt selection to improve dialogue personalization and enhance both the diversity and engagingness.

- In the proposed SPT method, we introduce the context-prompt contrastive mechanism and prompt fusion learning within a unified framework to foster prompt diversity and adaptability.

- Extensive experiments show the effectiveness of the proposed SPT method.

2 Related Work

2.1 Personalized Dialogue Generation

The CONVAI2 dataset, curated from the PersonaChat dataset (Zhang et al., 2018), features a persona profile with four to five sentences for each interlocutor (Dinan et al., 2019). This dataset has been established as a benchmark for personalized dialogue generation. Building upon this dataset, a variety of studies have explored different methods. For example, Wolf et al. (2019) extend the GPT2 model (Radford et al., 2019) with fine-tuning techniques specific to persona-based conversations. Differently, Song et al. (2021) employed a tripartite BERT architecture (Devlin et al., 2019), optimized through reinforcement learning, to craft responses. Similarly, Liu et al. (2020) introduces a transmitter-receiver model by applying reinforcement learning with custom rewards to refine the dialogue generation process. Cao et al. (2022) leverage model-agnostic data augmentation techniques to enrich the training dataset with pseudo-samples using models like GPT2 and BERT. Huang et al. (2023b) develop an adaptive attention mechanism that coherently integrates persona and context information. Huang et al. (2023a) propose a LAPDOG method to incorporate an external story corpus to enhance persona profiles for richer response generation. In contrast to those methods, the proposed SPT framework decomposes the task with multiple soft prompts without necessitating additional annotations or reliance on external corpora, which enables the generation of diverse and engaging responses while maintaining the integrity of the conversational context.

2.2 Language Models and Personalization

Language models (LMs) estimate text sequence probabilities, with recent models expanding from millions (Radford et al., 2019; Zhang et al., 2022) to billions of parameters (Brown et al., 2020; Zhang et al., 2022), and training corpora now including vast web texts and instructional data (Ouyang et al., 2022; Touvron et al., 2023). Such advancements have notably improved the performance of LMs on various NLP tasks, especially in generating high-quality text for conversational applications. While those LMs are adept at providing user-centric responses, personalization remains a challenge. The prevalent strategy involves appending manually crafted hard prompts to LMs, which is overly simplistic and can result in the 'lost in the middle' problem (Liu et al., 2023). This occurs when the output of the LM is generically correct but lacks personalized context, struggling to reconcile broad training data with specific user prompts. To counteract this, the proposed SPT method enables the LLM to adapt its responses to varying personalized contexts more effectively. As a result, SPT fosters the generation of dialogue responses that are not only consistent but also highly personalized, addressing the core challenge of maintaining context relevance in user interactions.

3 Methodology

In this section, we introduce the proposed SPT method.

3.1 Problem Settings

In persona-based dialogue sessions, a context is represented as $C = \{P, U\}$, where $P = \{p_1, \ldots, p_e\}$ denotes the persona comprising e sentences (e.g., $4 \leq e \leq 5$) to provide background information for a machine interlocutor m and $U = \{u_{h,1}, u_{m,1}, \ldots, u_{h,n}\}$ denotes the dialogue context initiated by the human h to capture the exchange between human h and machine m. The goal is to generate a machine's response $r = u_{m,n}$ that aligns with its persona P and the context U.

3.2 Architecture

Figure 1 illustrates the SPT framework, consisting of a soft prompt group, a dense retriever, and a frozen LLM. Within this framework, the dense retriever selects an appropriate soft prompt from the soft prompt group by determining the closest match to the given context C. The chosen prompt is then merged with C to guide the LLM to produce compelling responses. The SPT framework restricts the soft prompt group and dense retriever to be trainable, while maintaining the LLM in a frozen state, which could significantly reduce the memory footprint and optimize resource utilization during training.

Soft Prompt Group The soft prompt group, denoted by $SP = \{sp_1, ..., sp_K\}$, consists of K soft prompts with random initialization. Each prompt features $L \times D$ virtual tokens, where D denotes the hidden dimension of the LLM and L denotes the length of prompts. These prompts are fine-tuned during training while the LLM remains frozen.

Soft Prompt Selection The soft prompt selection is done by a trainable retriever, $Ret(\cdot, \cdot)$,

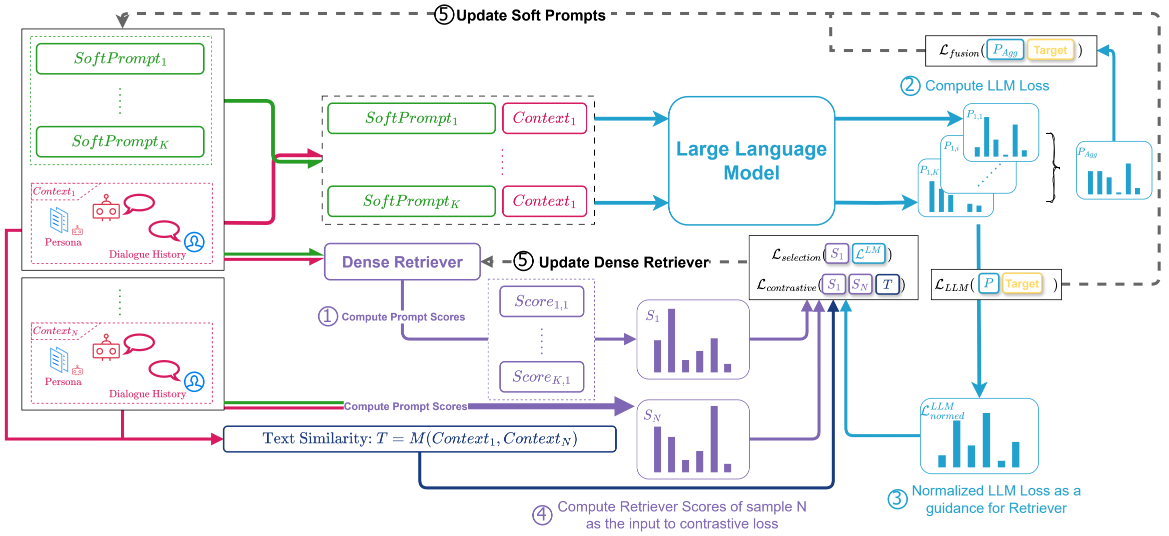


Figure 1: Selective Prompt Tuning (SPT) process for personalized dialogue generation with large language models (LLMs). The process starts by computing similarity scores for K soft prompts given the context, followed by LLM loss computation. The prompts are then fed into the LLM along with the context to generate multiple LLM losses which are normalized. A dense retriever computes another set of scores for a different context to inform the contrastive loss. The four computed losses guide the updates to the soft prompts and the retriever to enhance response diversity and relevance.

which calculates the similarity score $s_{C,sp} = \{s_{C,1}, ..., s_{C,K}\}$ between the context embedding emb_C from the LLM and each candidate sp_i in the soft prompt group SP. It ranks all the soft prompts based on the computed similarity score $\{s_{C,i}\}_{i=1}^{K}$ to identify the most suitable prompt for the context.

LLMs The LLMs deployed here are the decoder-only causal language model with frozen weights and initialized from pre-trained models.

3.3 Computing Similarity between Soft Prompts and Context

To reduce computational overhead, the dense retriever Ret utilizes two linear layers, i.e., lin_C and lin_{sp}, for computing the similarity scores $\{s_{C,i}\}$. Those similarity scores are calculated using the context embedding $emb_C \in \mathbb{R}^{M\times D}$ obtained by the LLM's word embedding layer LLM_{emb} and the soft prompt representation in $\mathbb{R}^{L\times D}$. The similarity score is computed as

$$\begin{aligned}
emb_C &= \text{LLM}_{emb}(C), \\
v_C &= \text{lin}_C(emb_C), \\
v_{sp,i} &= \text{lin}_{sp}(sp_i), \\
\bar{v}_C &= \text{Avg}_{\text{dim}=0}(v_C), \\
\bar{v}_{sp,i} &= \text{Avg}_{\text{dim}=0}(v_{sp,i}), \\
s^{raw}_{c,i} &= \frac{\bar{v}_C \cdot \bar{v}_{sp,i}}{\|\bar{v}_C\|_2 \cdot \|\bar{v}_{sp,i}\|_2}, \\
s_{C,i} &= \text{Softplus}(s^{raw}_{C,i}),
\end{aligned} \tag{1}$$

where $\text{Avg}_{\text{dim}=0}(\cdot)$ denote the averaging operation along the length dimension to address the sequence length discrepancy between emb_C and sp_i, $\text{Softplus}(\cdot)$ denotes the softplus activation function to ensure that $s_{C,i}$ remains in the range $[0, 1]$ and enhance the numerical stability during training, and $s_{C,i}$ represents the normalized similarity score between the context C and the soft prompt sp_i.

3.4 Learning Prompt Selection

Navigating the lack of explicit annotations in complex dialogue scenarios poses a challenge in accurately guiding the retriever to assess the similarity between the context and each soft prompt. A naive method, which independently fine-tunes the entire soft prompt group and then selects candidates based on the similarity score during decoding, might lead to sub-optimal performance, akin to tuning a single soft prompt. To address this, we leverage context-driven losses from soft prompts, refining similarity score computations and enabling informed retriever decisions during training, as introduced in the next two subsections.

3.4.1 Soft Prompt Loss

For simplicity, consider the case with a single context. Given a context c_n from persona and dialogue history and its corresponding ground truth response $target_n$, we calculate the negative log-likelihood loss for each soft prompt as

$$\begin{aligned} pred_{i,n} &= \text{LLM}(\text{concat}(sp_i, c_n)), \\ \mathcal{L}_i^{LLM} &= \text{NLL}(pred_{i,n}, target_n), \end{aligned} \tag{2}$$

where concat$(\cdot,\cdot)$ denotes the concatenation operation, $\text{LLM}(\cdot)$ denotes the LLM's forward operation, which takes a text sequence as the input and returns the predicted token probability distribution as the output, and NLL$(\cdot,\cdot)$ denotes the negative log-likelihood loss. This process generates K losses $\mathcal{L}^{LLM} = \{\mathcal{L}_1^{LLM}, ..., \mathcal{L}_K^{LLM}\}$ to measure the predictive ability of each soft prompt.

3.4.2 Prompt Selection Loss

In the absence of explicit annotations for conversational settings, updating the retriever to identify the most effective soft prompt for a given context is challenging. However, by using soft prompts in LLMs with the same context, the loss from different prompts can serve as a guide to determine which soft prompt is most suitable. Based on this consideration, we use the soft prompt loss (i.e., $\mathcal{L}^{LLM}$ defined in Eq. (2)) to gauge each candidate sp_i in the soft prompt group SP within c_n. Aligning the LLM's performance evaluation with the retriever's similarity scores is achieved by using the KL divergence between the negative language model loss (as guidance) and similarity scores. By denoting by $S_{c_n,SP} = [S_{c_n,sp_1}, \ldots, S_{c_n,sp_K}]$ the similarity scores between c_n and each sp_i in SP, the prompt selection loss is formulated as

$$\begin{aligned} \mathcal{L}_{normed}^{LLM} &= \text{Softmax}(-\mathcal{L}^{LLM}/\tau_g), \\ \mathcal{L}_{selection} &= \text{KL}(S_{c_n,SP}, \mathcal{L}_{normed}^{LLM}), \end{aligned} \tag{3}$$

where Softmax$(\cdot)$ denotes the softmax function, τ_g is a temperature hyper-parameter, and KL$(\cdot,\cdot)$ denotes the KL divergence.This loss is pivotal in ensuring the selections of the dense retriever are informed and coherent with the LLM, effectively mirroring the performance of soft prompts in generating contextually relevant and engaging responses.

3.5 Context-Prompt Contrastive Learning

While the aforementioned losses aid in training, there is a risk that the retriever often retrieves a single prompt and stagnates in such sub-optimal states. To alleviate this and foster prompt diversity to retrieve more prompts, we propose a context-prompt contrastive loss. This loss refines prompt selection by adjusting similarity scores based on the textual similarity of distinct contexts, thereby preventing to always select a single soft prompt and promoting varied selections. Specifically, the context-prompt contrastive loss dynamically recalibrates the similarity scores between pairs of context contents, considering their textual resemblance. Mathematically, the context-prompt contrastive loss is formulated as

$$\mathcal{L}_{con}(s_{c_i}, s_{c_j}) = \begin{cases} 1 - \cos(s_{c_i}, s_{c_j}) & \text{if } M(c_i, c_j) > \Gamma \\ \max(0, \cos(s_{c_i}, s_{c_j})) & \text{otherwise} \end{cases} \tag{4}$$

where $M(\cdot,\cdot)$ denotes a distance function such as BLEU (Papineni et al., 2002), Γ denotes a threshold, s_{c_i} denotes a vector of cosine similarity scores between a context c_i and soft prompts in the soft prompt group, and $\cos(\cdot,\cdot)$ denotes the cosine similarity.

The function $\mathcal{L}_{con}$ amplifies the cosine similarity for similar context pairs (i.e., $M(c_i, c_j) > \Gamma$) and dampens it for dissimilar pairs (i.e., $M(c_i, c_j) \leq \Gamma$). This contrastive strategy not only ensures the retriever's alignment with the LLM's evaluations but also fosters a rich diversity and distinctiveness among different dialogue contexts, significantly bolstering the framework's overall adaptability.

3.6 Prompt Fusion Learning

To optimize the effectiveness of the soft prompts, we introduce a prompt fusion learning loss. This loss averages the predictive probabilities from all the soft prompts in the soft prompt group, aiming to aggregate a unified outcome that closely aligns with the desired output. The averaging operation in this loss smooths out variances and biases from individual prompts, thus improving the overall prediction accuracy and reliability. Formally, this loss is formulated as

$$\begin{aligned} p_{fused} &= \frac{1}{K}\sum_{i=1}^{K} \text{LLM}(\text{concat}(sp_i, c_n)) \\ \mathcal{L}_{fusion} &= \text{NLL}(p_{fused}, target_n). \end{aligned} \tag{5}$$

By utilizing the collective strengths of diverse prompts, this loss enhances the model's ability to generate context-appropriate responses.

3.7 Overall Objective Function

The SPT framework hinges on the harmonious integration of the aforementioned loss functions, where each addresses a distinct aspect. The soft prompt loss (i.e., $\mathcal{L}^{LLM}$) ensures the LLM fidelity, the prompt selection loss (i.e., $\mathcal{L}_{selection}$) aligns the retriever's similarity assessment with the LLM's output, the context-prompt contrastive loss (i.e., $\mathcal{L}_{con}$) promotes diversity in prompt selection, and the prompt fusion learning loss (i.e., $\mathcal{L}_{fusion}$) enhance the overall performance for all the soft prompts. The overall objective of the SPT method is to minimize a composite loss function that encapsulates these individual components. Formally, the overall objective function $\mathcal{L}_{Total}$ for the SPT framework is formulated as

$$\mathcal{L}_{Total} = \sum_{i=1}^{K} \mathcal{L}_i^{LLM} + \lambda_1 \sum_{\substack{i,j=1 \\ i \neq j}}^{K} \mathcal{L}_{con}(s_{c_i}, s_{c_j}) + \lambda_2 \mathcal{L}_{selection} + \lambda_3 \mathcal{L}_{fusion}, \quad (6)$$

where λ_1, λ_2, and λ_3 are hyperparameters that control the relative contribution of each loss component. In our experiments, we simply set λ_1, λ_2, and λ_3 to be 1, which could achieve good performance.

By minimizing $\mathcal{L}_{Total}$ during training, the SPT framework effectively balances the fidelity to the LLM, the accuracy of the retriever, and the diversity in prompt selection, leading to an adaptive dialogue generation system.

3.8 Inference

During inference, the dense retriever selects the most appropriate soft prompt from the soft prompt group based on the given context. This selected prompt, along with the context, is then fed into the LLM to decode the final result. Formally, for a given context C, soft prompt group SP, and dense retriever Ret, the inference process proceeds as

$$\begin{aligned} i^* &= \arg\max_{1 \leq i \leq K} Ret(C, SP), \\ pred &= \text{LLM}(\text{concat}(sp_{i^*}, C)), \end{aligned} \quad (7)$$

where sp_{i^*} denotes the selected soft prompt with index i^* and $pred$ denotes the response generated by the LLM.

4 Experiments

In this section, we empirically evaluate the proposed SPT model.

4.1 Dataset

We conduct experiments on the ConvAI2 dataset (Dinan et al., 2019), a benchmark for personalized dialogue generation. It comprises 8,939 training and 1,000 validation multi-turn conversations sourced from crowdworkers. Each dialogue includes persona profiles, each of which has four to five sentences to describe the background of each speaker, and the conversational history between the two interlocutors. By following (Liu et al., 2020; Huang et al., 2023a), our experiments employ a self-persona setting where only the speaking interlocutor's persona is revealed, maintaining the other's persona as obscured.

4.2 Experimental Setup

All experiments are based on two LLMs, including OPT (Zhang et al., 2022) and Llama2 (Touvron et al., 2023) of different sizes, which serve as the foundation model for the proposed SPT method. We randomly initialize soft prompts using a standard Gaussian distribution. For OPT models, we set the soft prompt token length to 8, and for the Llama2 model, we use a token length of 1. The soft prompt group consists of $K = 4$ candidates. Learning rates of different LLMs are recorded in Table 6 in the Appendix. The threshold Γ in Eq. (4) is set to 20.

4.3 Evaluation Metrics

We evaluate our model using a suite of established metrics for persona-based dialogue generation, including Unigram F1, BLEU, ROUGE, BERT Score, and textual unigram/bigram distinctness (denoted by DIST-1 and DIST-2). Unigram F1 measures the harmonic mean of precision and recall at the token level. BLEU (Papineni et al., 2002) and ROUGE (Lin, 2004) evaluate the overlap of n-grams between the generated text and target reference. BERT score (Zhang et al., 2019), using the *deberta-xlarge-mnli* model[2] as recommended for its improved performance over *roberta-large*, captures the semantic similarity of text pairs. Unigram and bigram distinctness (denoted by DIST-1 and DIST-2) gauge the diversity of the generated text, where DIST$_{AVG}$ denotes the average of DIST-1 and DIST-2.

4.4 Results

Table 1 illustrates that the proposed SPT consistently outperforms the baseline models across various metrics. Notably, the OPT-2.7B-SPT and Llama2-7B-SPT models exhibit significant performance improvements (i.e., 33.04% and 26.26%, respectively). Those improvements affirm the effec-

[2] https://github.com/Tiiiger/bert_score

Model	F1	BLEU	ROUGE-1	ROUGE-2	ROUGE-L	BERT$_{F1}$	BERT$_{P}$	BERT$_{R}$	DIST-1	DIST-2	AVG$_{\uparrow}$
OPT-125M-PT	10.79	1.61	14.36	2.67	13.25	53.15	53.90	52.91	3.94	13.67	-
OPT-125M-SPT	11.06	2.22	16.45	3.60	15.42	54.86	56.23	53.91	4.87	17.38	16.60%
OPT-1.3B-PT	8.16	1.82	11.48	2.22	10.29	55.31	57.12	53.93	4.87	17.19	-
OPT-1.3B-SPT	9.94	2.66	13.74	3.24	12.38	56.34	58.08	54.99	4.93	17.76	16.43%
OPT-2.7B-PT	8.67	1.77	11.84	2.30	10.61	56.25	58.48	54.49	5.18	18.61	-
OPT-2.7B-SPT	12.23	**3.11**	16.97	4.37	**15.61**	**57.96**	**59.92**	56.45	**5.84**	20.76	33.04%
Llama2-7B-PT	17.12	1.99	15.74	4.07	13.72	52.30	48.57	57.11	2.80	12.91	-
Llama2-7B-SPT	**17.49**	2.80	**17.02**	**4.48**	15.24	54.66	53.02	**57.14**	5.69	**22.86**	26.62%

Table 1: Performance comparison of different LLMs across different model sizes. BERT$_{F1}$, BERT$_{P}$, and BERT$_{R}$ denote the BERT Score F1, Precision, and Recall. AVG$_{\uparrow}$ indicates the average improvement over the corresponding baseline method. Models appended with '-SPT' indicate the combination of the proposed SPT method with the corresponding LLM, while '-PT' indicates the conventional prompt tuning method. The best performance in each metric is in bold.

Model	F1	BLEU	ROUGE-L	BERT$_{F1}$	DIST$_{AVG}$
Llama-7B-SPT	**17.49**	**2.80**	**15.24**	**54.66**	14.27
w/o CL	15.95	2.00	13.17	52.80	14.23
w/o FUSION	16.02	1.90	13.24	52.89	**14.69**
w/o SL	16.39	1.93	13.71	53.75	13.06

Table 2: The ablation study on the training losses. 'w/o CL', 'w/o FUSION', and 'w/o SL' denote no context-prompt contrastive loss, no prompt fusion learning loss, and no prompt selection loss, respectively.

tiveness of the proposed SPT method in fostering more diverse and personalized responses.

For baseline models, we can see that there exists a common trade-off between linguistic quality and diversity. Specifically, the Llama2-7B model scores 17.12 in F1 and 1.99 in BLEU, but its diversity seems not so good (i.e., 2.80 in DIST-1 and 12.91 in DIST-2). This is in contrast to the OPT-125M model, which has lower linguistic scores (i.e., 10.79 in F1 and 1.61 in BLEU) but higher distinctness (i.e., 3.94 in DIST-1 and 13.67 in DIST-2). Different from those models, the proposed SPT method significantly enhances both diversity and linguistic quality, thereby avoiding the common compromise between linguistic enhancement and diversity.

5 Ablation Studies

In this section, we conduct ablation studies for the proposed SPT method.

5.1 Training Losses

Table 2 reveals the impact of different training losses on performance. Omitting the prompt fusion loss slightly increases the prediction diversity in terms of DIST$_{AVG}$ but reduces the overall performance in terms of F1, BLEU, ROUGE, and BERT Score. One possible reason is that the prompt fusion loss contributes to the linguistic quality at the cost of the diversity. Excluding the context-prompt contrastive loss leads to a decline in all the evaluated metrics, which shows the effectiveness of the context-prompt contrastive loss. The absence of the prompt selection loss significantly affects the prediction diversity, causing the model to favor a single soft prompt, akin to utilizing a single prompt. The above results underscore the importance of each loss in enhancing the model performance and response diversity.

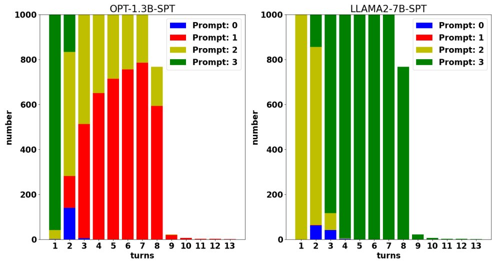


Figure 2: Analysis of the usage of each soft prompt cross the conversational process, where the horizontal axis represents the index of the conversational turn and the vertical axis denotes the times that each soft prompt is chosen.

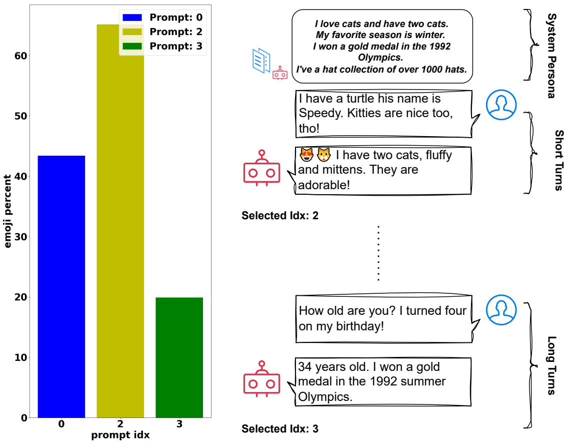


Figure 3: The varied response styles of the Llama2-7B-SPT model, highlighting its tendency to incorporate emojis into responses during initial conversational turns.

5.2 Prompt Usage in Varied Contexts

To see the prompt usage during the conversational process, we plot in Figure 2 the times each soft

K	F1	BLEU	ROUGE-L	BERT$_{F1}$	DIST$_{AVG}$
1	17.76	1.76	15.21	54.86	14.15
2	17.71	2.55	**15.63**	**55.52**	14.29
3	17.34	2.45	15.09	55.31	**15.23**
4	17.49	**2.80**	15.24	54.66	14.27
5	16.07	2.42	12.88	47.12	13.99
6	17.46	2.21	14.94	54.43	15.12
7	**17.76**	2.42	15.42	54.96	13.51
8	17.48	2.32	15.29	54.87	13.89

Table 3: The effect of the size of the soft prompt group (i.e., K) to the performance of Llama2-7B-SPT.

prompt is chosen during the entire conversation. According to Figure 2, we can see that in the OPT-1.3B-SPT model, prompt sp_3 is predominantly utilized for the initial stage in the conversation, sp_2 for the middle stage of the conversation, and sp_1 for the later stage of the conversation. For the Llama2-7B-SPT model, we have similar observations, indicating that soft prompts have functionalities in different stages of the conversation.

Moreover, Figure 3 explores the stylistic aspects of responses generated by different prompts, i.e., emojis in the generated responses. In the Llama2-7B-SPT model, sp_2, which is often used in the initial stage of the conversations, tends to generate emojis in the generated response. Differently, sp_3, often used in the late stage of the conversation, tends to generate few emoji in decoded responses. This phenomenon suggests a strategic use of emojis at different stages of the conversation.

5.3 Number of Soft Prompt Candidates

Table 3 shows the effect of the number of soft prompts (i.e., K) to the model performance in terms of different metrics. Though the best performance occurs at different K's for different performance metrics, the best performance for different metrics usually occurs when $K \leq 4$, which is likely due to the sizes of both the CONVAI2 dataset and the LLM used. Hence, in all the experiments, K is set to be 4 by default.

5.4 Comparison to Longer Prompt Tuning

As shown in Table 4, the SPT method with four single-token soft prompts outperforms the four-token prompt tuning method, highlighting effectiveness of the proposed SPT method. Moreover, SPT excels the eight-token prompt tuning method in terms of BLEU, ROUGE, and DIST$_{AVG}$, showing its effectiveness despite fewer trainable parameters.

Model	F1	BLEU	ROUGE-L	BERT$_{F1}$	DIST$_{AVG}$
Llama2-7B-SPT	17.49	**2.80**	**15.24**	54.66	14.27
Llama2-7B-4-PT-TOKEN	16.47	1.78	13.64	52.65	9.52
Llama2-7B-8-PT-TOKEN	**17.64**	2.13	14.69	**55.85**	13.33
Llama2-7B-LoRA	15.61	2.20	11.66	47.46	10.21
GPT-3.5-ICL	6.78	0.77	0.09	47.96	**23.24**

Table 4: Performance comparison across varying prompt token lengths as well as LoRA and In-Context Learning on GPT-3.5 Turbo. '-SPT' denotes the proposed SPT model with a single token length per prompt, while Llama2-7B-4-PT-TOKEN and Llama2-7B-8-PT-TOKEN have token lengths of 4 and 8, respectively.

Model	BLEU	BLEU-1	BLEU-2	BLEU-3	BLEU-4
Llama2-7B-PT	8.79	43.42	20.44	13.51	10.06
Llama2-7B-SPT	2.07	41.99	16.62	10.48	6.95

Table 5: Comparison of text overlapping between the prediction of different models and the persona.

5.5 Comparison to LoRA

As LoRA (Hu et al., 2022) is another type of parameter-efficient finetuning method and has shown to be effective to utilize LLMs for different applications, we compare the proposed SPT method with it based on the Llama2-7B model under the condition that they have comparable numbers of trainable parameters. As shown in Table 4, LoRA exhibits improvements in the BLEU score and DIST$_{AVG}$ but has lower ROUGE-L, BERT$_{F1}$, and F1 scores compared with the four-token prompt tuning method. Moreover, the proposed SPT method surpasses LoRA across all the evaluation metrics, highlighting its superior performance and affirming its effectiveness under the condition of comparable numbers of trainable parameters.

5.6 Comparison to In-Context Learning

To compare the performance with In-Context Learning (ICL) on LLMs, we compare the SPT method with the zero-shot GPT-3.5 turbo with instructions. According to results shown in Table 4, we can see that ICL gains a higher diversity score (i.e., DIST$_{AVG}$) but lower scores in terms of other metrics. This implies that simply prompting a more powerful LLM without proper tuning is hard to gain comparable performance to tuning methods.

5.7 Text Overlap Between Prediction and Persona

Table 5 presents BLEU scores between the model's predictions and the system's persona descriptions for different models. We can see that the prompt tuning method exhibit larger text overlap with the system's persona, often leading to repetitive responses aligned with the persona. In contrast, the proposed SPT method has lower linguistic similarities to the persona, which results in more diverse and effective responses. This suggests that the proposed SPT method effectively balances the persona consistency and response diversity, avoiding the pitfalls of over-repetition.

6 Conclusion

In this paper, we introduce SPT, a strategic approach for personalized dialogue generation through selective prompt tuning. By jointly training a soft prompt group and a dense retriever, SPT adeptly navigates various conversational scenarios automatically, enriching response diversity while improving both linguistic and neural-based metrics. Experiments on the CONVAI2 dataset highlights the capacity of SPT to identify intrinsic conversational settings, showing its effectiveness in generating contextually appropriate dialogues.

Acknowledgements

This work is supported by NSFC general grant 62076118 and Shenzhen fundamental research program JCYJ20210324105000003.

Limitations

This paper has introduced the selective prompt tuning in personalized dialogue generation. Through diverse prompting, the LLMs can generate more diverse and engaged responses when compared with single prompt tuning. However, despite the context-prompt contrastive mechanism and prompt selection loss, there is still a risk for the retriever to fall into a narrow selection of soft prompts (e.g., given $K = 4$ in Llama2-7B, there is still one soft prompt that is selected only once during inference). This limitation may caused by a larger K used, making the determination of K important. Meanwhile, in the context-prompt contrastive loss, simply using BLEU to measure text similarity may not be sufficient to distinguish the difference between two dialogues, which could be enhanced by neural metrics powered by LLMs that could distinguish texts from both semantic and linguistic perspectives. Additionally, in the decoded text of Llama2-7B, the existence of emoji is not designed in the PersonaChat dataset, which is worth further investigation.

Ethic Statement

This research confines the use of personal data to fictional persona profiles in the CONVAI2 dataset, avoiding the handling or storage of real personal data. All the soft prompts within the SPT are vector-based parameters without directly encoding or representing any individual's personal information. When applying to real-world applications, it is vital to prioritize data privacy, ensuring that personal information for personalized dialogues is ethically sourced and used with informed consent.

References

Tom B. Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, J. Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, Sandhini Agarwal, Ariel Herbert-Voss, Gretchen Krueger, T. Henighan, R. Child, A. Ramesh, Daniel M. Ziegler, Jeff Wu, Clemens Winter, Christopher Hesse, Mark Chen, Eric Sigler, Mateusz Litwin, Scott Gray, Benjamin Chess, Jack Clark, Christopher Berner, Sam McCandlish, Alec Radford, Ilya Sutskever, and Dario Amodei. 2020. Language models are few-shot learners. *arXiv:2005.14165*.

Yu Cao, Wei Bi, Meng Fang, Shuming Shi, and Dacheng Tao. 2022. A model-agnostic data manipulation method for persona-based dialogue generation. *arXiv:2204.09867*.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of deep bidirectional transformers for language understanding. In *North American Chapter of the Association for Computational Linguistics*, pages 4171–4186.

Emily Dinan, Varvara Logacheva, Valentin Malykh, Alexander H. Miller, Kurt Shuster, Jack Urbanek, Douwe Kiela, Arthur D. Szlam, Iulian Serban, Ryan Lowe, Shrimai Prabhumoye, Alan W. Black, Alexander I. Rudnicky, Jason Williams, Joelle Pineau, Mikhail S. Burtsev, and Jason Weston. 2019. The second conversational intelligence challenge (convai2). *arXiv:1902.00098*.

Edward J Hu, Yelong Shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Lu Wang, and Weizhu Chen. 2022. LoRA: Low-rank adaptation of large language models. In *International Conference on Learning Representations*.

Qiushi Huang, Shuai Fu, Xubo Liu, Wenwu Wang, Tom Ko, Yu Zhang, and Lilian Tang. 2023a. Learning retrieval augmentation for personalized dialogue generation. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, pages 2523–2540, Singapore. Association for Computational Linguistics.

Qiushi Huang, Yu Zhang, Tom Ko, Xubo Liu, Bo Wu, Wenwu Wang, and H Tang. 2023b. Personalized dialogue generation with persona-adaptive attention. *Proceedings of the AAAI Conference on Artificial Intelligence*, 37(11):12916–12923.

Brian Lester, Rami Al-Rfou, and Noah Constant. 2021. The power of scale for parameter-efficient prompt tuning. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pages 3045–3059, Online and Punta Cana, Dominican Republic. Association for Computational Linguistics.

Chin-Yew Lin. 2004. ROUGE: A package for automatic evaluation of summaries. In *Text Summarization Branches Out*, pages 74–81, Barcelona, Spain. Association for Computational Linguistics.

Nelson F. Liu, Kevin Lin, John Hewitt, Ashwin Paranjape, Michele Bevilacqua, Fabio Petroni, and Percy Liang. 2023. Lost in the middle: How language models use long contexts. *ArXiv*, abs/2307.03172.

Qian Liu, Yihong Chen, Bei Chen, Jian-Guang Lou, Zixuan Chen, Bin Zhou, and Dongmei Zhang. 2020. You impress me: Dialogue generation via mutual persona perception. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 1417–1427, Online. Association for Computational Linguistics.

Long Ouyang, Jeff Wu, Xu Jiang, Diogo Almeida, Carroll L. Wainwright, Pamela Mishkin, Chong Zhang, Sandhini Agarwal, Katarina Slama, Alex Ray, John Schulman, Jacob Hilton, Fraser Kelton, Luke E. Miller, Maddie Simens, Amanda Askell, Peter Welinder, Paul Francis Christiano, Jan Leike, and Ryan J. Lowe. 2022. Training language models to follow instructions with human feedback. *ArXiv*, abs/2203.02155.

Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu. 2002. Bleu: A method for automatic evaluation of machine translation. In *Proceedings of the 40th Annual Meeting on Association for Computational Linguistics*, Association for Computational Linguistics, page 311–318, USA. Association for Computational Linguistics.

Alec Radford, Jeff Wu, Rewon Child, David Luan, Dario Amodei, and Ilya Sutskever. 2019. Language models are unsupervised multitask learners.

Haoyu Song, Yan Wang, Kaiyan Zhang, Weinan Zhang, and Ting Liu. 2021. Bob: Bert over bert for training persona-based dialogue models from limited personalized data. In *Association for Computational Linguistics*.

Hugo Touvron, Louis Martin, Kevin Stone, Peter Albert, Amjad Almahairi, Yasmine Babaei, Nikolay Bashlykov, Soumya Batra, Prajjwal Bhargava, Shruti Bhosale, et al. 2023. Llama 2: Open foundation and fine-tuned chat models. *arXiv preprint arXiv:2307.09288*.

Thomas Wolf, Victor Sanh, Julien Chaumond, and Clement Delangue. 2019. Transfertransfo: A transfer learning approach for neural network based conversational agents. *arXiv:1901.08149*.

Saizheng Zhang, Emily Dinan, Jack Urbanek, Arthur D. Szlam, Douwe Kiela, and Jason Weston. 2018. Personalizing dialogue agents: I have a dog, do you have pets too? In *Association for Computational Linguistics*.

Susan Zhang, Stephen Roller, Naman Goyal, Mikel Artetxe, Moya Chen, Shuohui Chen, Christopher Dewan, Mona Diab, Xian Li, Xi Victoria Lin, et al. 2022. Opt: Open pre-trained transformer language models. *arXiv preprint arXiv:2205.01068*.

Tianyi Zhang, Varsha Kishore, Felix Wu, Kilian Q Weinberger, and Yoav Artzi. 2019. Bertscore: Evaluating text generation with bert. *arXiv preprint arXiv:1904.09675*.

A Appendix

A.1 Complete Training Procedure

The full training procedure is described at Algorithm 1.

A.2 Detailed Settings for SPT Training

Shared Parameters	
HyperParameter	Value
K	4
Optimizer	Adam
τ_g	1
λ_1	1
λ_2	1
λ_3	1
λ_4	1
Llama2-7B-SPT	
Prompt Length	1
Learning Rate	0.01
OPT-2.7B	
Prompt Length	8
Learning Rate	0.001
OPT-1.3B	
Prompt Length	8
Learning Rate	0.01
OPT-125M	
Prompt Length	8
Learning Rate	0.01

Table 6: The hyper-parameters for the SPT training.

Table 6 lists the detailed hyper-parameters for training SPT. The share parameters are used for all model training. Meanwhile, the Llama2-7B-SPT, OPT-2.7B, OPT-1.3B, and OPT-125M indicate the specific hyper-parameters used in the specific model training. We trained the SPT models on eight Tesla-V100 32GB GPUs. For each SPT model except OPT-125M-SPT, we train one epoch and then do the evaluation. For OPT-125M-SPT, we train for 15 epochs until it converges.

A.3 Details for Ablation Study

Table 8 details our ablation study's findings. Selective Prompt Tuning (SPT) with four one-token soft prompts demonstrates superior performance over both the traditional four-token and eight-token soft prompt tuning approaches, highlighting our method's effectiveness. In a comparative analysis with LoRA under a similar parameter setup, SPT outperforms in all evaluated metrics, reinforcing its efficiency. Furthermore, compared to GPT-3.5 Turbo's In-Context Learning (ICL), SPT shows significant improvements in F1 and BLEU scores, indicating challenges with ICL's alignment to target responses despite its higher diversity in textual outputs.

	Persona Consistency	Dialogue Consistency	Engageness
Llama2-7B-SPT	**1.89**	**1.29**	**1.34**
Llama2-7B-PT	1.33	1.13	1.29

Table 7: Human evaluation over Llama2-7B-SPT and Llama2-7B-PT.

A.4 Human Evaluation

We conducted human evaluation on three metrics, persona consistency, context consistency, and engagingness. Each metric is ranked for three scores: 0, 1, 2. For persona consistency, 0 means contradicts the persona, 1 means not relevant to the persona, and 2 means consistent to the persona. For context consistency, 0 means contradicts previous dialogue history, 1 means not relevant to the previous dialogue, and 2 means consistent to the previous dialogue. For engagingness, 0 means a boring response, 1 means a safe but bland response, and 2 means an interesting response. We randomly sampled 100 responses from Llama2-7B-SPT and Llama2-7B-PT. The results are displayed in Table 7. Our proposed SPT outperforms PT over all three metrics, indicating the effectiveness of our approach in both three perspectives.

A.5 Experimental Results on Larger Dataset

To further evaluate the efficiency and scalability of the SPT framework. We conducted additional experiments on the DailyDialog dataset, a more extensive and complex dialogue dataset than PersonaChat. Notably, the DailyDialog dataset lacks explicit persona descriptions in its entries, presenting a unique challenge for personalization techniques. The results of the DailyDalog are shown as Table 9.

Result Analysis : The experimental setup involved executing four separate runs using both soft prompt tuning (PT) and SPT strategies on the DailyDialog dataset. The empirical evidence clearly demonstrates the superiority of the SPT framework over the conventional PT approach across all evaluated metrics. Specifically, the SPT method exhibits significant performance improvements, showcasing its adaptability and effectiveness in handling

Algorithm 1 SPT Training

```
Input: Input context $C = \{c_1, .., c_N\}$, Input context batch $C_{batch} = \{c_i, .., c_{i+batchsize}\} \subset C$, ground truth response $Y = \{y_1, ..., y_N\}$, a soft prompt group $SP = \{sp_1, .., sp_K\}$, a dense retriever $Ret$, textual similarity threshold $\Gamma$, a text similarity metric $M$, and a large language model $LLM$
Output: A tuned soft prompt group $SP$ and a tuned dense retriever $Ret$
for $C_{batch}$ in $C$ do
    Initialize batch soft prompt loss $\mathcal{L}_{batch}^{LLM} = 0$, batch prompt selection loss $\mathcal{L}_{selection}^{batch} = 0$,
    Initialize batch prompt fusion loss $\mathcal{L}_{fusion}^{batch} = 0$, batch context-prompt contrastive loss $\mathcal{L}_{con}^{batch} = 0$
    for Input Context $c_n$ in $C_{batch}$ do
        Compute one soft prompt $\mathcal{L}_i^{LLM} = NLLLoss(concat(sp_i, c_n), y_n)$
        Obtain $K$ soft prompt loss $\mathcal{L}^{LLM} = \{\mathcal{L}_1^{LLM}, ..., \mathcal{L}_K^{LLM}\}$ with above computation
        Normalized negative soft prompt loss $\mathcal{L}_{normed}^{LLM} = \text{Softmax}(-\mathcal{L}^{LLM}/\tau_g)$ for retriever update
        Compute retriever score between context $c_n$ and soft prompt $sp_i$ as $s_{c_n,sp_i} = Ret(sp_i, c_n)$
        Obtain $K$ retriever scores by $s_{c_n,SP} = \{s_{c_n,sp_1}, ..., s_{c_n,sp_K}\}$
        Compute prompt selection loss using KL Divergence by $\mathcal{L}_{selection} = \text{KL}(s_{c_n,SP}, \mathcal{L}_{normed}^{LLM})$
        Aggregate $K$ predictions from LLM given $c_n$ and $SP$ as $p_{fused}$
        Compute prompt fusion loss as $\mathcal{L}_{fusion} = \text{NLL}(p_{fused}, y_n)$
        Sum soft prompt loss, prompt selection loss, and prompt fusion loss to their batch opponents
        $\mathcal{L}_{batch}^{LLM} = \mathcal{L}_{batch}^{LLM} + \mathcal{L}^{LLM}$, $\mathcal{L}_{selection}^{batch} = \mathcal{L}_{selection}^{batch} + \mathcal{L}_{selection}$, $\mathcal{L}_{fusion}^{batch} = \mathcal{L}_{fusion}^{batch} + \mathcal{L}_{fusion}$
    end for
    for Input Context $c_i, c_j$ in $C_{batch}$ do
        Compute textual similarity $T = M(c_i, c_j)$
        Compute retriever score for $c_i, c_j$ as $s_{c_i,SP}, s_{c_j,SP}$
        Compute context-prompt contrastive loss:
        if then$T > \Gamma$
            $\mathcal{L}_{con} = 1 - cos(s_{c_i,SP}, s_{c_j,SP})$
        else
            $\mathcal{L}_{con} = \max(0, \cos(s_{c_i,SP}, s_{c_j,SP}))$
        end if
        Sum context-prompt contrastive loss to batch context-prompt contrastive loss
        $\mathcal{L}_{con}^{batch} = \mathcal{L}_{con}^{batch} + \mathcal{L}_{con}$
    end for
    Sum all objective together: $\mathcal{L}_{Total} = \mathcal{L}_{batch}^{LLM} + \mathcal{L}_{selection}^{batch} + \mathcal{L}_{fusion}^{batch} + \mathcal{L}_{con}^{batch}$
    Update soft prompts and retriever via back-propagation with $\mathcal{L}_{Total}$
end for
```

more complex and extensive datasets. The evaluation metrics are summarized in the table below, where we observe notable enhancements in key areas such as F1 score, BLEU, ROUGE, and BERT-based metrics, underlining SPT's potential applicability across diverse conversational tasks.

A.6 Comparison to RAG (Retrieval Augmented Generation)

Conceptual Differences: RAG and SPT fundamentally differ in their approaches. RAG enhances inputs by incorporating external information from a database, focusing on the value of external data. In contrast, SPT focuses on selecting the optimal soft prompt based on given context input. While they operate differently, they aren't inherently conflicting and could be seen as complementary since SPT can treat the retrieval-augmented input as context as a whole. SPT has the potential to integrate RAG's enriched inputs comprehensively. The exploration of combining RAG and SPT falls beyond the scope of this work and is reserved for future research.

RAG Experimentation: We experimented with the RAG framework under the Llama2-7B model to compare SPT with RAG. We observed that the choice of K (number of retrieval contents) is crucial due to the RAG's reliance on the training set for retrieval. A large K value can lead to the concate-

Model	F1	BLEU	ROUGE-1	ROUGE-2	ROUGE-L	BERT_{F1}	BERT_{P}	BERT_{R}	DIST-1	DIST-2
Llama-7B-SPT	17.49	2.80	**17.02**	**4.48**	**15.24**	54.66	53.02	**57.14**	**5.69**	**22.86**
Llama2-7B-4-PT-TOKEN	16.47	1.78	15.74	3.74	13.64	52.65	49.09	57.18	3.35	15.70
Llama2-7B-8-PT-TOKEN	**17.64**	2.13	16.49	4.01	14.69	**55.85**	**54.98**	57.34	4.75	21.91
LoRA	15.61	2.20	13.09	2.95	11.66	47.46	47.78	47.48	3.35	17.08
GPT-3.5-ICL	6.78	0.77	0.00	0.00	0.09	47.96	46.73	49.77	**8.03**	**38.45**

Table 8: Detailed results for the ablation study.

Model	F1	BLEU	ROUGE1	ROUGE2	ROUGEL	BERT F1	BERT Precision	BERT Recall	DIST-1	DIST-2
Llama2-7B-PT-LR=0.001	18.03	0.18	15.66	4.41	14.13	55.90	55.97	56.99	7.71	35.40
Llama2-7B-PT-LR=0.01	17.06	0.21	14.46	4.09	13.00	54.58	54.71	55.60	7.41	34.83
Llama2-7B-SPT-LR=0.001	**18.38**	**0.31**	**15.87**	**4.49**	**14.37**	**56.95**	**57.21**	**57.73**	**7.97**	**36.89**
Llama2-7B-SPT-LR=0.01	17.40	0.08	15.04	4.57	13.68	53.72	53.73	55.16	7.53	34.68

Table 9: Detailed results for DailyDialog.

nated content overwhelming the context window size, thus significantly increasing computational resource demands.

Efficient Training Setup for RAG: For efficiency, we set $K = 1$ for our RAG experiment, focusing on retrieving the most semantically similar dialogue to augment the current context. The retriever used is the Contriever from Facebook, which is known for its ability to retrieve highly relevant content based on textual semantics. This setup allowed us to directly compare the efficiency and scalability of RAG and SPT under similar computational constraints.

Comparative Results: The training time for an epoch under the RAG setup was approximately 14 hours, compared to 7 hours for SPT. This underscores SPT's efficiency and scalability, especially in resource-constrained environments. Detailed results are displayed in the table 10. In terms of performance, SPT outperformed RAG in nearly all the metrics. This shows that SPT is not only faster but can also produce better results. The only area where RAG did slightly better was in creating more diverse responses (DIST-1 and DIST-2 metrics). This comparison shows that SPT is more efficient and often more effective than RAG. However, these two approaches do not necessarily contradict each other. Instead, combining these two methods could lead to even better performance. We might create more accurate and engaging dialogues by using RAG to get the proper context and SPT to fine-tune the response. This approach has a lot of potential for improving conversational AI systems.

A.7 SPT Stability Experiment

To evaluate the stability of the SPT, we further conducted additional experiments designed to test the system's resilience to disruptions. Specifically, we introduced Gaussian noises with the mean as 0 and the standard deviation as 1 to the similarity scores during inference to simulate the effect of inaccuracies in the soft prompt selection process. Additionally, we add a parameter α to control the strength of the noise. Formally, the disrupted selection score would become $score = score + \alpha * noise$. The objective of this experiment is to observe the stability of our retriever under less-than-ideal conditions. Detailed results of these experiments will be included in our revision.

Result Analysis : The results presented in Table 11 demonstrate the impact of noise on retrieval performance. The introduction of mild noise (e.g., 0.01 to 0.1) results in negligible performance degradation, with some metrics showing slight improvements. However, as noise levels increase to 1.0, a deterioration in performance is observed despite a noticeable increase in DIST-2. This pattern suggests that while our SPT framework exhibits good stability to minor disturbances, its performance is adversely affected by severe interference.

A.8 Retriever Stability Experiment

To evaluate the robustness of our dense passage retrieval system, we introduced Gaussian noise with standard deviation. Specifically, we apply noise with a varying strength α, choosing from [0.001, 0.01, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0], to the L^{LLM}_{normed} loss during the retriever's training phase. Therefore, the disrupted L^{LLM}_{normed} will become $L^{LLM}_{normed} + \alpha * noise$. This approach aimed to simulate potential disruptions in the soft prompt selection process, thereby testing the stability and resilience of our retriever under adversarial conditions.

Model	F1	BLEU	ROUGE-1	ROUGE-2	ROUGE-L	BERT-F1	BERT-Precision	BERT-Recall	DIST-1	DIST-2
RAG-Contriever-LR=0.01	14.93	2.18	9.36	2.03	8.55	45.01	45.87	44.72	4.95	24.47
RAG-Contriever-LR=0.001	12.66	2.59	9.66	2.48	8.89	40.72	41.60	40.35	5.13	24.81
RAG-Contriever-LR=0.0001	15.16	2.53	11.53	2.86	10.56	50.52	50.70	51.12	**5.75**	**26.83**
Llama2-7B-SPT	**17.49**	**2.80**	**17.02**	**4.48**	**15.24**	**54.66**	**53.02**	**57.14**	5.69	22.86

Table 10: The comparison between RAG and SPT.

Model	F1	bleu	rouge1	rouge2	rougel	BERT f1	BERT precision	BERT recall	dist-1	dist-2
SPT (Noise=0)	**17.49**	2.80	**17.02**	4.48	**15.24**	54.66	53.02	**57.14**	**5.69**	22.86
Noise=0.01	17.42	2.97	16.93	**4.49**	15.18	54.70	53.45	56.71	5.42	22.55
Noise=0.05	17.41	2.98	16.91	4.48	15.16	54.68	53.41	56.71	5.41	22.56
Noise=0.1	17.42	2.93	16.91	4.44	15.17	54.75	53.46	56.78	5.46	22.86
Noise=0.5	17.41	**2.99**	16.92	**4.49**	15.16	**54.93**	**53.56**	57.03	5.61	24.04
Noise=1.0	17.43	2.76	16.72	4.32	15.03	54.82	53.38	57.00	5.56	**24.15**

Table 11: The experiment on the stability of the SPT.

Model	F1	BLEU	ROUGE-1	ROUGE-2	ROUGE-L	BERT f1	BERT precision	BERT recall	DIST-1	DIST-2
SPT (Noise=0)	17.49	2.80	17.02	4.48	15.24	54.66	53.02	57.14	5.69	22.86
Noise=0.001	17.69	2.75	17.11	4.46	15.38	54.94	53.41	57.19	5.54	23.83
Noise=0.01 1.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN

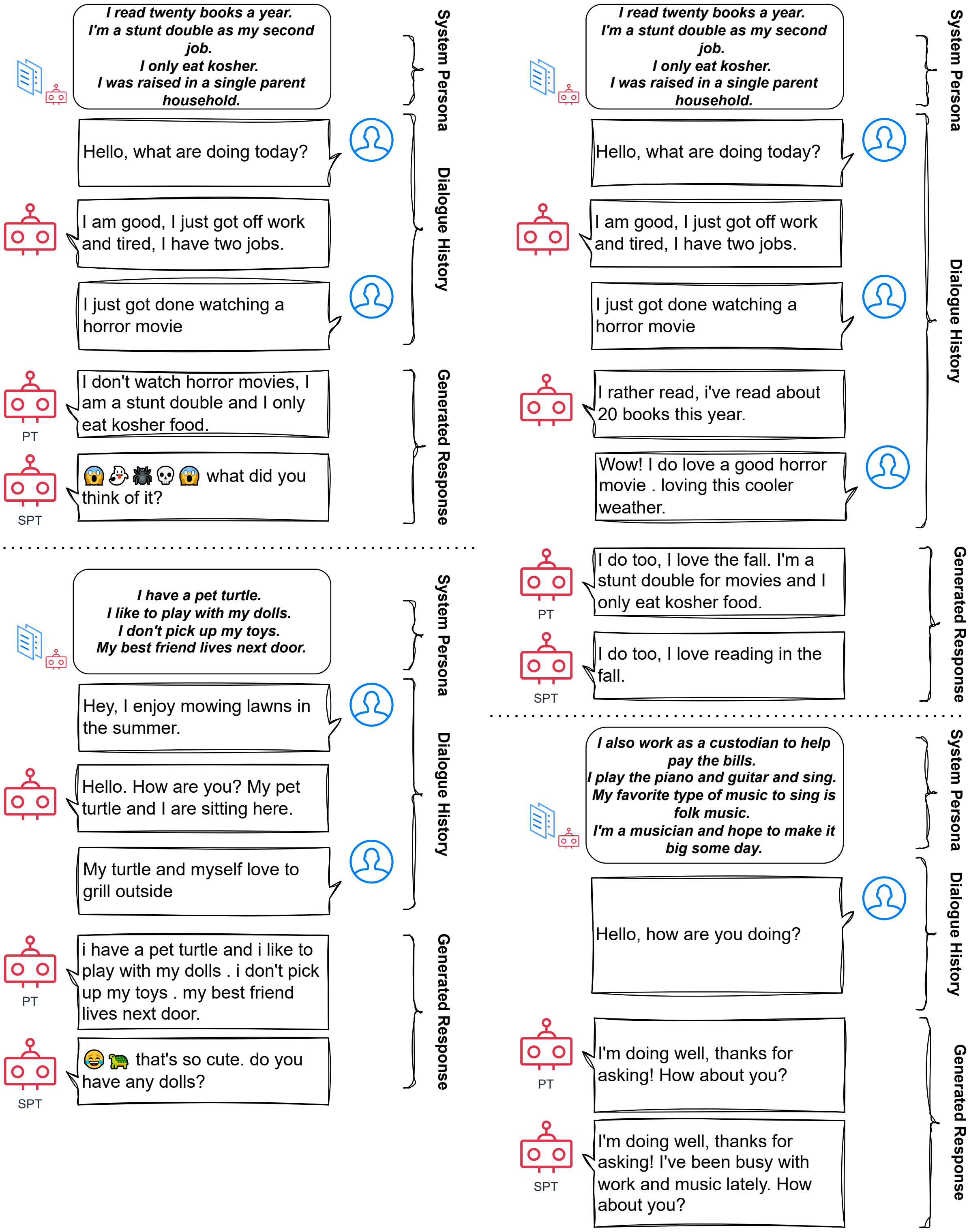
Table 12: The experiment on dense retriever stability.

Adversarial Noise Impact on Retriever Robustness: The introduction of Gaussian noise served as a means to disturb the updating process of the retriever, allowing us to observe its behaviour and adaptability in the interference. Specifically, we add the noise the $\mathcal{L}^{LLM}_{normed}$ to make the KL Divergence update become noisy. The varying levels of noise strength were chosen to represent a wide spectrum of potential adversarial impacts, from mild to severe disruptions, i.e., [0.001, 0.01, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0]

Results and Insights: According to the Table 12, introducing the mildest level of noise (0.001) yielded improved performance across several key metrics, including F1, ROUGE-1, ROUGE-L, BERT Score, and DIST-2. This improvement suggests that slight perturbations may act as a beneficial regularizer within the training process, thereby enhancing performance. In contrast, levels of noise beyond the mildest introduced numerical instability (manifesting as overflow or underflow, particularly as we utilize fp16 for SPT training). This instability disrupts the training process, leading to outcomes marked as NaN (Not a Number).

A.9 Case Study

Figure 4 shows a comparison between SPT and a prompt-tuned model. SPT uniquely incorporates horror-related emojis in a conversation about horror movies, while the prompt-tuned model tends to repeat persona profile content. This trend continues in subsequent dialogues. In the last case, SPT adeptly weaves persona details into its responses, offering a more engaging and personalized conversational experience compared to the more generic replies of the prompt-tuned model.



I read twenty books a year.
I'm a stunt double as my second job.
I only eat kosher.
I was raised in a single parent household.

Hello, what are doing today?

I am good, I just got off work and tired, I have two jobs.

I just got done watching a horror movie

I don't watch horror movies, I am a stunt double and I only eat kosher food.

😱👻🕷️💀😱 what did you think of it?

I have a pet turtle.
I like to play with my dolls.
I don't pick up my toys.
My best friend lives next door.

Hey, I enjoy mowing lawns in the summer.

Hello. How are you? My pet turtle and I are sitting here.

My turtle and myself love to grill outside

i have a pet turtle and i like to play with my dolls . i don't pick up my toys . my best friend lives next door.

😂🐢 that's so cute. do you have any dolls?

I read twenty books a year.
I'm a stunt double as my second job.
I only eat kosher.
I was raised in a single parent household.

Hello, what are doing today?

I am good, I just got off work and tired, I have two jobs.

I just got done watching a horror movie

I rather read, i've read about 20 books this year.

Wow! I do love a good horror movie . loving this cooler weather.

I do too, I love the fall. I'm a stunt double for movies and I only eat kosher food.

I do too, I love reading in the fall.

I also work as a custodian to help pay the bills.
I play the piano and guitar and sing.
My favorite type of music to sing is folk music.
I'm a musician and hope to make it big some day.

Hello, how are you doing?

I'm doing well, thanks for asking! How about you?

I'm doing well, thanks for asking! I've been busy with work and music lately. How about you?

Figure 4: Four case studies, where PT denotes the prompt tuning method (Lester et al., 2021).