

SMILE: Single-turn to Multi-turn Inclusive Language Expansion via ChatGPT for Mental Health Support

Important: Our objective is to explore the potential of large language models to serve as AI counselors, and we do NOT advocate for their use as a substitute in therapeutic treatment without professional supervision.

Huachuan Qiu^{1,2} Hongliang He^{1,2} Shuai Zhang^{1,2} Anqi Li^{1,2} Zhenzhong Lan^{2,3*}

¹ Zhejiang University ² School of Engineering, Westlake University

³ Research Center for Industries of the Future, Westlake University

{qiuhuachuan, lanzhenzhong}@westlake.edu.cn

Abstract

Developing specialized dialogue systems for mental health support requires multi-turn conversation data, which has recently garnered increasing attention. However, gathering and releasing large-scale, real-life multi-turn conversations that could facilitate advancements in mental health support presents challenges in data privacy protection and the time and cost involved in crowdsourcing. To address these challenges, we introduce SMILE, a single-turn to multi-turn inclusive language expansion technique that prompts ChatGPT to rewrite public single-turn dialogues into multi-turn ones. Our work begins by analyzing language transformation and validating the feasibility of our proposed method. We conduct a study on dialogue diversity, including lexical features, semantic features, and dialogue topics, demonstrating the effectiveness of our method. Further, we employ our method to generate a large-scale, lifelike, and diverse dialogue dataset named SMILECHAT, consisting of 55k dialogues. Finally, we utilize the collected corpus to develop a mental health chatbot, MeChat. To better assess the quality of SMILECHAT, we collect a small-scale real-life counseling dataset conducted by data anonymization. Both automatic and human evaluations demonstrate significant improvements in our dialogue system and confirm that SMILECHAT is high-quality. Code, data, and model are publicly available at <https://github.com/qiuhuachuan/smile>.

1 Introduction

We all know the importance of mental health, and mental health issues have been a persistent concern for human beings (Kessler et al., 2005). Recent advancements in natural language processing (NLP) technology (Ouyang et al., 2022; Ni et al., 2022) have led to the emergence of neural-based conversational AI in mental health support (Liu et al., 2022; Tu et al., 2022). As an innovative solution for

* Corresponding author.

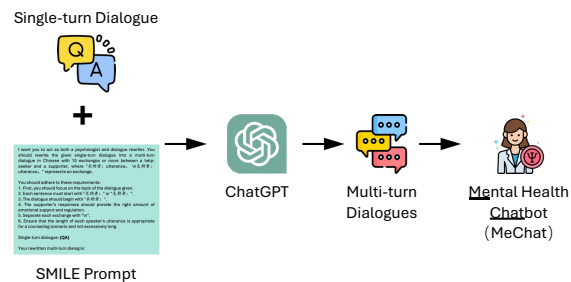


Figure 1: The SMILE method used to generate dialogues for mental health support.

mental health, virtual counselors powered by large language models (LLMs) can effectively address accessibility barriers, such as the high cost of treatment and the shortage of experienced professionals to meet the demand. Furthermore, such dialogue systems provide mental health support as an effective and practical online counseling approach for those in need, safeguarding user privacy and mitigating the stigma that often accompanies help-seeking. *However, the lack of publicly available, large-scale, diverse, and high-quality multi-turn chat datasets in the mental health support domain hinders the development of specialized dialogue systems.*

Motivation Indeed, many researchers have been pursuing the goal of building a practical and effective conversational agent for mental health. However, the first step in creating such a system is to have training data. Conversations related to mental health support often contain sensitive information and must be kept confidential to safeguard the privacy of individuals seeking help (Lu et al., 2021). Making these conversations publicly available may discourage individuals from seeking support or negatively impact their personal and professional lives once known to people with whom they are acquainted. To facilitate progress in the NLP community, some researchers have attempted to collect various dialogue corpora (Liu et al., 2021;

Sun et al., 2021; Zheng et al., 2022) through crowdsourcing, data crawling, or data augmentation to build a dialogue agent capable of providing emotional and mental health support. How to construct a large-scale, diverse, and high-quality multi-turn chat dataset for mental health without effort motivates us to carry out the work as presented in this paper.

Challenges To be more specific, crowdsourcing conversations (Liu et al., 2021) for emotional support has limitations due to the high cost and time required to train and manage annotators, as well as the difficulty in mimicking real-life interactions, that is, interlocutors may lack an understanding of the dilemma of living with mental disorders. An alternative is crawling QA (Sun et al., 2021) on a public mental health forum for training psychological support models. However, single-turn conversations may not be sufficient for resolving mental health issues, as multiple interaction exchanges are often needed. Multi-turn conversations, which can better simulate real-world conversations, are more practical for training psychological support models. While the post-triggered machine-augmented method (Zheng et al., 2022) can address the limitations of scale and topic diversity, it does not consider the responses of experienced supporters.

Our Approach To tackle the abovementioned challenges, we introduce SMILE, single-turn to multi-turn inclusive language expansion via ChatGPT, as shown in Figure 1. Specifically, we instruct ChatGPT to transform publicly available long question-answer pairs (public QAs), which can also be viewed as single-turn dialogues, into multi-turn conversations. *With the proposed method, we build a large-scale, diverse, and high-quality multi-turn conversation dataset for mental health support.*

Our paper is organized as follows:

- We first present our method (§3), including data preparation, task definition, and prompt design that elaborates on the SMILE method and other baseline methods.
- We then demonstrate the feasibility of the SMILE method through language transformation (§4), showing that the dialogue constructed by this method is lifelike.
- Subsequently, we demonstrate the effectiveness of the SMILE method by utilizing three

diversity indicators (§5): lexical features, semantic features, and dialogue topics. Following the validation of feasibility and effectiveness, we leverage the SMILE method to generate a large-scale, lifelike, and diverse multi-turn chat dataset, SMILECHAT.

- Finally, we propose to train a dialogue system to explore the quality of conversation (§6) and collect a set of 50 anonymized real-life counseling dialogues for model evaluation.

Our Contributions We make our data, code, and model publicly available. We believe that our work offers a new perspective on constructing a large-scale, lifelike, diverse, and high-quality multi-turn dialogue dataset for mental health within the research community. Our contributions can be summarized as follows:

- We introduce SMILE, which provides an easy and novel method for alleviating the scarcity of multi-turn conversations in mental health support.
- Through the analysis of language transformation and dialogue diversity, we verify the feasibility and effectiveness of our proposed method.
- To better assess the quality of SMILECHAT, we collect small-scale real-life counseling data with 50 anonymized counseling sessions to build a real-life test set, PsyTest. Automatic and human evaluations on the small-scale real-life test set confirm that our proposed dataset is high-quality.
- We release SMILECHAT, which consists of 55165 Chinese multi-turn dialogues. Our dialogue model, MeChat, and real-life test set, PsyTest, are publicly available.

We highlight that this method can also construct multi-turn dialogues based on medical, financial, and legal QAs, thereby alleviating the dialogue scarcity in other application domains.

2 Related Work

2.1 Applications of ChatGPT

ChatGPT has proven to be a powerful AI tool for various NLP tasks since its release. Currently, it is being utilized in several domains, such as conversational AI (Alessa and Al-Khalifa, 2023; Köpf

et al., 2023; Chen et al., 2023), education (Küchemann et al., 2023; Eshghie and Eshghie, 2023), code programming (Dong et al., 2023; Yetiştiren et al., 2023), and healthcare (Zhao et al., 2023; Yang et al., 2023).

Furthermore, ChatGPT’s efficiency and cost-effectiveness have been well-documented, making it competitive to human annotators (Gilardi et al., 2023; Zhu et al., 2023) even in zero-shot accuracy tasks. Xu et al. (2023) have proposed the use of self-chatting, where ChatGPT engages in a conversation with itself, resulting in 111.5k dialogues collected from Quora and Stack Overflow sources and 47k conversations from the medical domain. Auto-GPT¹, an AI agent, is capable of breaking down a natural language goal into sub-tasks and using various tools and the internet in an automated loop to achieve the objective. Shen et al. (2023) have suggested using ChatGPT for task planning when receiving user inquiries, selecting appropriate models based on function descriptions from Hugging Face, executing each subtask using the chosen AI model, and summarizing the response based on the execution’s outcomes.

In summary, ChatGPT has already demonstrated its enormous potential as an intelligent pipeline tool that can significantly advance NLP development despite having only a restricted API available for researchers.

2.2 Datasets for Mental Health Support

Research on mental health support has significantly depended on the availability of publicly available datasets (Sun et al., 2021; Liu et al., 2021; Zheng et al., 2022) in recent years. The large-scale conversational datasets have enabled researchers to investigate various aspects of mental health, including identifying mental health conditions (Liu et al., 2023; Srivastava et al., 2022), understanding clients’ reactions (Li et al., 2023), predicting support strategies (Sun et al., 2021; Li et al., 2023), deciding personalized interventions (Golden et al., 2023) and understanding response safety within a dialogue history (Qiu et al., 2023).

Liu et al. (2021) first defined the emotional support conversation task and then, via crowdsourcing, constructed ESConv, an emotional support conversation dataset containing 1053 dialogues with rich support strategies. However, the data collection of ESConv requires high cost and time yet leads to a

¹<https://github.com/Significant-Gravitas/Auto-GPT>

small-scale dialogue dataset. To this end, Zheng et al. (2022) presented an approach for augmenting data scale with informative dialogue posts and then constructing AugESC, a model-synthesized dataset with 102k dialogues. The previous two datasets are limited to English. To facilitate the research in Chinese, hence Sun et al. (2021) crawled QA posts on a public mental health support platform and made the PsyQA dataset publicly available.

3 Method

PsyQA², an open-source and high-quality Chinese single-turn dialogue dataset focused on mental health support, features one question mapped to multiple answers. Our dataset creation pipeline, based on PsyQA, includes three main stages: (1) data preparation, (2) task definition, and (3) prompt design.

3.1 Data Preparation

Considering the distinction between QA within PsyQA and multi-turn dialogues, along with the context window limitation of 4096 tokens in ChatGPT³, we propose to perform data preprocessing for PsyQA. This process involves wording cleaning and length truncation.

Wording Cleaning This work aims to construct a large-scale, lifelike, diverse, and high-quality multi-turn conversation corpus using the proposed SMILE method based on PsyQA. While QA can be considered a single-turn conversation between a real help-seeker⁴ and a supporter⁵, there are differences in wording compared to actual multi-turn conversations. For instance, the term "楼主" (literally meaning "thread starter") frequently appears in QA but is rarely used in conversation. Therefore, we propose a two-stage process to clean the wording in PsyQA, mitigating linguistic discrepancies before rewriting QA into multi-turn conversations. This process includes both automatic and manual cleaning procedures. Please refer to Appendix A for a detailed process.

Length Truncation After a statistical analysis of the PsyQA dataset, we find that 757 QAs have a total length exceeding 1800 characters. Also, we identify 9 QAs, the total discourse length exceeding 4000 characters. However, the model

²<https://www.xinli001.com/qa>

³The model we use is gpt-3.5-turbo-0613.

⁴The terms help-seeker and client are interchangeable.

⁵The terms supporter and counselor are interchangeable.

gpt-3.5-turbo has a maximum context length of 4096 tokens. To ensure reliable and smooth rewrites, we limit the length of the QA pairs, maximizing the number of rewritten dialogue turns. Specifically, we truncate the length of the QA pairs at 1800 Chinese characters and truncate any excess text. This control ensures that the generated text is limited to approximately 2000 tokens. It is worth noting that the PsyQA data used in this study undergo a data preprocessing process.

3.2 Task Definition

Let us first denote the input x as a sequence $\{x_1, x_2, \dots, x_n\}$, and the output y as a sequence $\{y_1, y_2, \dots, y_m\}$. The generation process of the language model can be expressed as the conditional probability distribution $p(y|x)$, which represents the probability of generating output y given the input x . Therefore, text generation via the large language model can be formulated as follows:

$$p(y|x) = \prod_{t=1}^m p(y_t|y_{<t}, x) \quad (1)$$

where y_t represents the t -th token generated by the model. Because x is our main focus in this paper, we will demonstrate the details of the prompt design next.

3.3 Prompt Design

In this section, we mainly focus on describing prompt design. In order to provide a clearer understanding of our method in a more controllable setting and elucidate the superiority of introducing single-turn dialogues as a reference, we first design two baseline prompts for comparison.

3.3.1 standard Prompt

As its name suggests, the standard prompt does not contain any single-turn dialogues or specific dialogue topics and instead uses only the most straightforward prompt to generate multi-turn dialogues. The standard prompt is illustrated in Figure 9 in Appendix C. The input in Equation 1 is $x = I$, where I represents the standard prompt. We simplify the method name as standard and consider this method as our baseline.

3.3.2 standardT Prompt

Intuitively, feeding a single, fixed prompt into a large language model often generates text with low diversity. Therefore, we provide a specific dialogue topic for the standardT prompt. The input

I want you to act as both a psychologist and dialogue rewriter. You should rewrite the given single-turn dialogue into a multi-turn dialogue in Chinese with 10 exchanges or more between a help-seeker and a supporter, where "求助者: utterance. \n支持者: utterance." represents an exchange.

You should adhere to these requirements:

1. First, you should focus on the topic of the dialogue given.
2. Each sentence must start with "求助者: " or "支持者: ".
3. The dialogue should begin with "求助者: ".
4. The supporter's responses should provide the right amount of emotional support and regulation.
5. Separate each exchange with "\n".
6. Ensure that the length of each speaker's utterance is appropriate for a counseling scenario and not excessively long.

Single-turn dialogue: {QA}

Your rewritten multi-turn dialog is:

Figure 2: The SMILE method used to generate dialogues for mental health support.

in Equation 1 is $x = (I, T)$, where T represents the dialogue topic chosen in uniform sampling in the topic set. We simplify the method name to standardT and adopt this method as our baseline, as illustrated in Figure 10 in Appendix C.

3.3.3 SMILE Prompt

Our paper aims to highlight the superiority of introducing single-turn dialogues when generating dialogues. Our proposed method, the SMILE method, instructs the LLM to rewrite single-turn dialogues into multi-turn ones. Figure 2 depicts the concrete prompt details. The input in Equation 1 is $x = (I, T, D)$, where T and D represent the implicit dialogue topic hidden in the QA and single-turn dialogue, respectively.

4 Language Transformation

4.1 Experimental Setup

Dialogue Topic Collection To address the issue of monotonous generation in the standard prompt, we propose to enhance diversity by injecting a dialogue topic into the standardT prompt. Consequently, we have collaborated with three professional counselors, reviewed existing literature (Rickwood et al., 2007; Pedrelli et al., 2015), and ultimately compiled a comprehensive set of dialogue topics. This set includes 60 distinct types of dialogue topics, each with a detailed explanation. For more details, please refer to Appendix G.

QA Sampling To ensure a fair comparison among the three methods we propose and prevent duplicate instances of the same question with different answers, we first randomly select 500 non-repetitive questions from the first 5000 QAs in

PsyQA. We then randomly choose one answer to serve as the corresponding response. The data samples obtained are employed as seed dialogues, which are subsequently restructured into multi-turn conversations via ChatGPT. We name this 500 sampled data as PsyQA* to distinguish it from the original dataset.

Hyperparameters Overall, we present three prompt methods in this paper. For each prompt method, we instruct ChatGPT to generate 500 dialogues to study language transformation and dialogue diversity, thereby validating the feasibility and effectiveness of our proposed SMILE method, respectively. To enhance the diversity of the generated dialogues, we set hyperparameters during text generation to the officially recommended default values, with temperature = 1.0 and top_p = 1.0.

Dialogue Filtering ChatGPT may exhibit potential instability during the text generation process. Therefore, as detailed in Appendix B, we employ an automatic filtering mechanism to exclude dialogues that fail to meet our specified requirements, which encompass dialogue format and dialogue turns. When a dialogue falls short of these criteria, we prompt ChatGPT to generate the dialogue anew until it aligns with our specified requirements only in the initial 500 dialogues. This measure, therefore, ensures that each method produces an equal number of 500 dialogues for preliminary analysis.

Text Embedding A multi-turn dialogue between a help-seeker and a supporter is represented as

$$d = \{u_1, r_1, \dots, u_i, r_i, \dots, u_n, r_n\} \quad (2)$$

where u_i and r_i represents the utterances of the i -th turn spoken by the help-seeker and supporter, respectively. A string of dialogue without any speaker role tokens can be denoted as $d_s = [u_1; r_1; u_2; r_2; \dots; u_n; r_n]$, where $;$ denotes the operation of textual concatenation.

To obtain the text embedding of a dialogue, we employ OpenAI’s model *text-embedding-ada-002*⁶, which accepts a maximum context length of 8191. Each dialogue is first preprocessed into a single string without any speaker tokens and is then mapped to a 1536-dimensional vector. For example, to compute the cosine similarity between two different dialogues, we can obtain

$$\cos(d_i, d_j) = \frac{e_i \cdot e_j}{\|e_i\| \|e_j\|} \quad (3)$$

⁶<https://platform.openai.com/docs/guides/embeddings>

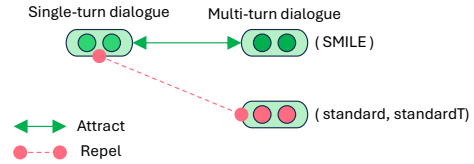


Figure 3: Mechanism for language transformation.

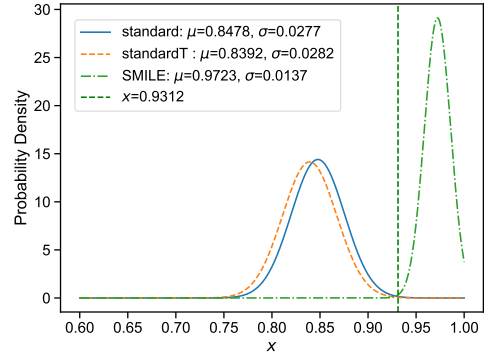


Figure 4: Distribution of dialogue transformation among three methods. The line $x = 0.9312$ represents the boundary of $\mu - 3\sigma$ in the SMILE method.

where e_i and e_j denote the text embeddings from two distinct dialogues.

4.2 Analysis

We propose two hypotheses: (1) When a single-turn dialogue is rewritten into a multi-turn dialogue, the similarity between the two is high (Attract). (2) When ChatGPT generates a multi-turn dialogue without introducing a single-turn dialogue, the similarity between the generated dialogue and a randomly selected single-turn dialogue is low (Repel). The mechanism of our proposed hypotheses is presented in Figure 3.

To assess the transformation feasibility of our SMILE method, we employ cosine similarity to calculate the transformation ratio. Specifically, a single-turn dialogue d is transformed into a multi-turn dialogue \hat{d} using the SMILE method. A multi-turn dialogue generated by standard or standardT is denoted as \bar{d} . We can calculate $\cos(d, \hat{d})$ for the first hypothesis and $\cos(d, \bar{d})$ for the second hypothesis. We utilize the text embeddings obtained in Equation 3.

Figure 4 presents the distribution of dialogue transformation among three methods. Our analysis concludes that single-turn dialogues can be successfully rewritten into multi-turn dialogues, ensuring that the dialogue generated by the proposed method is **lifelike**, rather than purely in the style of machine-generated text from ChatGPT.

Method	# Unique Unigrams	# Total Unigrams	Distinct-1	# Unique Bigrams	# Total Bigrams	Distinct-2	# Unique Trigrams	# Total Trigrams	Distinct-3
PsyQA*	11785	203306	0.058	120049	202806	0.592	182942	202306	0.904
standard	4174	153536	0.027	32281	153036	0.211	72340	152536	0.474
standardT	6032	175319	0.034	52141	174819	0.298	105921	174319	0.608
SMILE	10447	254585	0.041	111662	254085	0.439	196367	253585	0.774

Table 1: Statistics of 500 conversations in each prompt method, including PsyQA*.

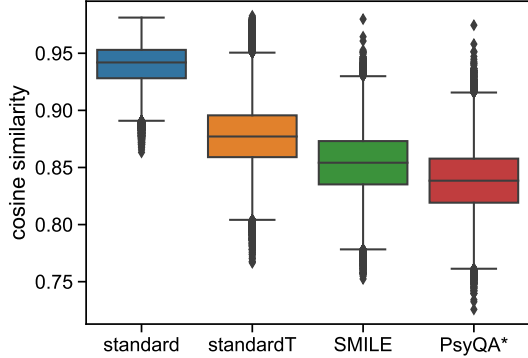


Figure 5: Pairwise dialogue cosine similarity among four settings: our proposed three methods and a reference point using sampled data from PsyQA*.

5 Dialogue Diversity

To demonstrate the effectiveness of the SMILE method, we mainly focus on three aspects of diversity: lexical features, semantic features, and dialogue topics.

5.1 Lexical Features

For lexical analysis, we utilize the popular Chinese ChatGLM2-6B⁷ tokenizer that is widely used in NLP to tokenize the dialogue. To measure the lexical features, we adopt distinct- n ($n = 1, 2, 3$) metrics (Li et al., 2016), which are widely used for measuring the diversity of dialogue datasets. Each dialogue is first preprocessed into a single string without any speaker tokens. We provide statistics for 500 dialogues per prompt method, as presented in Table 1.

Our proposed SMILE method results in rich vocabularies, with a significantly higher number of unique unigrams, bigrams, and trigrams than the two baseline methods. Specifically, a simple and fixed prompt tends to produce monotonous output. The model output demonstrates substantial diversification when dialogue topics are incorporated into a single, fixed prompt. Furthermore, the SMILE method outperforms the baseline methods in terms of Distinct-1, Distinct-2, and Distinct-3.

⁷<https://huggingface.co/THUDM/chatglm2-6b>

5.2 Semantic Features

To measure the semantic diversity in a dialogue dataset, we propose to calculate the cosine similarity between every pair of different dialogues. This involves computing the pairwise cosine similarity for each pair of distinct dialogues, resulting in $\binom{500}{2}$ pairs and their corresponding cosine values, as described in Equation 3.

We present the results in Figure 5, which demonstrates that the median of the SMILE method is significantly lower than those of the baseline methods. The SMILE method exhibits the most extensive semantic diversity, aligning closely with the sampled dialogue of PsyQA*. However, it is worth noting that the median of the SMILE method is more significant than that of PsyQA*. The reason behind this could be the introduction of token distribution from ChatGPT.

5.3 Dialogue Topics

To measure the diversity of dialogue topics in a dialogue dataset, we utilize information entropy to measure the diversity of topic distribution. *The higher the information entropy, the more uniform the distribution, indicating greater diversity.* The formula for calculating information entropy (Rényi, 1961; Lin, 1991) is as follows:

$$H(X) = - \sum_{i=1}^n p(x_i) \log_2 p(x_i) \quad (4)$$

where $H(X)$ is the information entropy. $p(x_i)$ is the probability of the occurrence of topic x_i .

To obtain dialogue topics for each dialogue in each prompt method, we design a prompt provided with 60 distinct dialogue topics, as illustrated in Appendix G and Figure 11, to automatically label dialogue topics for each dialogue with Qwen1.5-110B-Chat (Bai et al., 2023). We present the information entropy for each prompt method in Table 2, demonstrating that the dialogues generated using the SMILE method are substantially more diverse than those generated using the standard method

Setting	standard	standardT	SMILE
Limited Topics	8.11	14.28	14.07
Unlimited Topics	8.40	14.76	15.02
Average	8.26	14.52	14.55

Table 2: Information entropy of dialogue topics.

Category	Total	Help-seeker	Supporter
# Dialogues	55165	-	-
# Utterances	1833856	693756	1140100
Turns per dialogue	5.7	-	-
Utterances per dialogue	33.2	12.6	20.7
Avg. length per utterance	27.9	26.1	28.9

Table 3: Data statistics of the dialogue dataset, SMILECHAT.

and are compatible with the standardT method, which uniformly samples dialogue topics.

5.4 SMILECHAT Dataset

Through the comprehensive analysis of language transformation and dialogue diversity, we conclude that the proposed method can generate a **lifelike** and **diverse** chat dataset. Therefore, we utilize the SMILE method to guide ChatGPT in generating all multi-turn conversations based on PsyQA one round only, leading to a **large-scale** dialogue dataset.

Data Statistics To ensure data quality, we impose stricter requirements on dialogue turns, retaining only dialogues with at least five turns. Thus, we compile a collection of 55165 conversations, SMILECHAT. Table 3 presents the statistics of the collected corpus.

Dialogue Exemplars Multi-turn dialogue examples generated with the standard and standardT methods are illustrated in Figures 18 and 20. Further, a dialogue generated by the SMILE method is shown in Figure 22.

6 Dialogue System

We aim to build a high-quality multi-turn chat dataset for mental health support. It is non-trivial to evaluate the quality of a dialogue dataset, which is often assessed indirectly by the dialogue system. Therefore, we need to train a dialogue system and analyze its performance.

6.1 Mathematical Formulation

In order to train a dialogue system for mental health support, the first step is to split each full dialogue $d \sim \mathcal{D}$ into several training sessions. Specifically, a sampled t -turn dialogue session can be represented

Epoch	Learning Rate	Batch Size	LoRA Rank	LoRA Dropout	LoRA α	Seed
2	1e-4	1	16	0.1	64	1234

Table 4: Hyperparameters of parameter-efficient fine-tuning.

as $d_t = \{u_1, r_1, u_2, r_2, \dots, u_t, r_t\} \sim \mathcal{D}$. Thus, we build a dialogue model that can predict the supporter’s utterance r_t based on the dialogue history $h_t = \{u_1, r_1, u_2, r_2, \dots, u_t\}$. Our objective is to use our synthetic conversation dataset \mathcal{D} to fine-tune a large language model π_0 using supervised learning, i.e., maximum likelihood estimates:

$$J_{\text{SFT}}(\theta) = \mathbb{E}_{(h_t, r_t) \sim \mathcal{D}} [\log \pi_\theta(r_t | h_t)] \quad (5)$$

where π_θ is initialized from π_0 .

6.2 Experimental Setup

Backbone Model To validate the dialogue quality of our collected dataset, we conduct a fine-tuning experiment on ChatGLM2-6B (Zeng et al., 2023).

Training Data To meet the data format requirements for instruction-based fine-tuning, we split the dialogue into multiple sessions, with the supporter’s last utterance concluding each session. Additionally, we incorporate the system prompt (detailed in Appendix E) as a prefix to dialogue messages, following OpenAI’s data format.

Parameter-efficient Fine-tuning To preserve the original capabilities of the model while adapting to downstream dialogue tasks and reducing computational costs, we employ Low-Rank Adaptation (LoRA, (Hu et al., 2021)) on all linear layers in the model for efficient fine-tuning.

Hyperparameters Table 4 presents the hyperparameters for developing a dialogue model for mental health support. All hyperparameters during inference time are set to their default values from the official repository. After fine-tuning ChatGLM2-6B with LoRA, we set the temperature to 0.8 and top_p to 0.8.

6.3 Evaluation

6.3.1 Automatic Evaluation

Metrics To conduct automatic evaluation, the evaluation metrics we use consist of BLEU-1/2/3 (Papineni et al., 2002), METEOR (Banerjee and Lavie, 2005), Rouge-L (Lin, 2004), Distinct-1/2/3

	METEOR \uparrow	BLEU-1 \uparrow	BLEU-2 \uparrow	BLEU-3 \uparrow	Rouge-L \uparrow	D-1 \uparrow	D-2 \uparrow	D-3 \uparrow	BERTScore \uparrow
Baseline	10.15	5.75	1.95	0.85	7.02	52.95	80.74	90.17	56.69
Fine-tuned	14.18	14.03	5.39	2.76	15.83	82.08	95.74	97.81	59.91

Table 5: Results of automatic evaluation in PsyTest dataset. Except for BERTScore, other metrics are evaluated by Chinese character-level tokenization. For BERTScore, we use the BAAI/bge-m3 model to get text embedding (Chen et al., 2024).

Category	Total	Help-seeker	Supporter
# Dialogues	50	-	-
# Utterances	3103	1561	1542
Turns per dialogue	31.0	-	-
Utterances per dialogue	62.1	31.2	30.8
Avg. length per utterance	32.3	40.2	24.2
# Test cases	1539	-	-

Table 6: Data statistics of the dialogue dataset, PsyTest.

(D-1/2/3) (Li et al., 2016), and BERTScore (Zhang et al., 2020).

Test Set To better understand and assess the dialogue quality of the SMILECHAT dataset, we propose to utilize real-life multi-turn counseling conversations. We develop an online mental health support platform that enables professional counselors to offer each client a free text-based counseling service, lasting approximately 50 minutes each time. We collect 50 real-life counseling dialogues. To protect user privacy, we ask experts to conduct a data anonymization process, removing information related to user identification (e.g., names and addresses). Then, we split each long dialogue into multiple small sessions, with the last utterance spoken by the counselor. We discard test cases in which the dialogue history is empty, and we name this test set PsyTest, which contains 1539 test cases. Table 6 presents the statistics of the PsyTest dataset.

Results The results of the automatic evaluation, including nine metrics, are presented in Table 5. Notably, the evaluated dialogues are based on real-world counseling data rather than generated dialogues, which excludes the influence stemming from ChatGPT. All automatic evaluation metrics we use indicate improved performance. Our results show that the model trained with SMILECHAT is effective and practical. Consequently, the automatic evaluation demonstrates that our collected dataset is **high-quality**.

6.3.2 Human Evaluation

Metrics We conduct a human evaluation to study the model performance of the dialogue system trained with our proposed dialogue corpus. Ini-

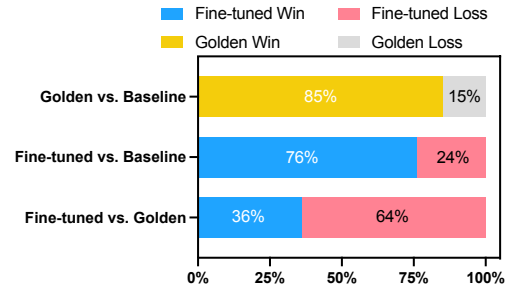


Figure 6: Human evaluation results, including three groups: Golden vs. Baseline, Fine-tuned vs. Baseline, and Fine-tuned vs. Golden. We present the win and loss rates of each compared pair in 100 randomly sampled sessions. Fleiss’ kappa (Fleiss et al., 1981) is used to measure the inter-rater agreement, and all values fall within moderate agreement with $0.5 \leq \kappa \leq 0.6$.

tially, we randomly sample 100 test cases from PsyTest, each consisting of a multi-turn dialogue history and a golden response produced by the counselor. Subsequently, we obtain 200 generated responses from the baseline and MeChat models. Three professional counselors are presented with a dialogue history and three randomly shuffled responses (baseline, fine-tuned, ground truth). They are tasked with selecting the optimal response between every two responses for the dialogue history, considering aspects such as professionalism, informativeness, helpfulness, empathy, and safety. The evaluation is conducted based on the ethical principles of psychologists and the code of conduct (Association et al., 2016). For more details about annotation guidelines, refer to Appendix F.

Results We employ majority voting to reach final decisions (win or loss) among three professional counselors. We present human evaluation results, as illustrated in Figure 6, including three groups: Golden vs. Baseline, Fine-tuned vs. Baseline, and Fine-tuned vs. Golden. We present the win and loss rates of each compared pair in 100 randomly sampled sessions. Results show that the baseline model, without training with SMILECHAT, lags significantly behind the setting with golden responses.

As depicted in Figure 6, the model trained

Help-seeker: I just went through a breakup, and it hurts so much.
Supporter:

✓ **Response 1 [Fine-tuned]:** Breakups are very painful experiences, and I can understand your feelings. (Affirmation and Reassurance) Can you tell me more about the situation surrounding your breakup? (Inquiring Subjective Information)

Response 2 [Baseline]: I'm really sorry to hear that. Breakups can be incredibly painful. (Affirmation and Reassurance) I understand that it can bring about feelings of disappointment, sadness, loneliness, and loss. (Reflection of Feelings) The end of a relationship may make us feel like we've invested a lot, only to get an ending in return, which can leave us feeling even more lost and helpless. (Interpretation)

If you're willing, we can discuss the impact of the end of this relationship on you and explore your current emotions. We can also talk about how to cope with these emotions and find some constructive solutions. (Invite to Explore New Actions)

Figure 7: Case study. Counseling strategies used in the two responses are presented in parentheses. Strategies in green are supportive, while those in red are challenging and should not be used in the early stages of counseling.

with SMILECHAT demonstrates a significant performance improvement compared to the baseline model. Moreover, the responses generated by MeChat surpass the golden response to some extent in 36% of all dialogue sessions. In short, human evaluation also proves that the SMILECHAT dataset is **high-quality**.

Case Study We present a case study, as shown in Figure 7. First, the text should not be too long during counseling conversations for the help-seeker to understand. Second, at the beginning of a conversation, a supporter generally will not challenge the help-seeker. Instead, they will provide support first. Third, response 1 is more human-like and natural. Therefore, the first response is better than the second response overall. Therefore, fine-tuning LLMs is indispensable in mental health support.

7 Conclusion

This paper introduces SMILE, a simple yet effective solution for addressing the data scarcity of multi-turn conversations in mental health support. Through language transformation and diversity analysis, we confirm the feasibility and effectiveness of our approach. The proposed method automatically creates a large-scale, lifelike, diverse, and high-quality dialogue corpus, SMILECHAT, consisting of 55165 dialogues with an average of 5.7 turns. Both automatic and human evaluations using the PsyTest dataset, consisting of 50 real-life anonymized counseling dialogues, demonstrate that SMILECHAT significantly improves dialogue system performance in mental health support. We release multi-turn dialogues (SMILECHAT), a dia-

logue model (MeChat), and a real-life anonymized test set (PsyTest) to drive progress in the research community.

Limitations

Our study has some limitations that could be addressed in future research.

In this paper, the automatic evaluation does not reasonably reflect the model performance, given the one-to-many problem. In future work, therefore, we will explore a more comprehensive tool kit for evaluating conversational agents in the mental health domain.

Furthermore, there is a limitation in the dialogue turns of the SMILECHAT dataset. Therefore, we will explore other methods to synthesize the dialogue data with more dialogue turns.

Ethics Statement

Our research is reviewed and approved by the Westlake University Institutional Ethics Committee (20211013LZZ001).

Data Sharing

SMILECHAT Based on the data copyright guidelines formulated by PsyQA, we release the multi-turn dialogue corpus publicly available for the research community. If researchers wish to reproduce the multi-turn dialogues using PsyQA, they should sign an agreement with the original data owner. Accordingly, we release our datasets and models for research purposes, thus facilitating further advancement in the academic community.

PsyTest Considering the nature of psychological counseling data, we must cautiously share this dataset. Regarding the rules for releasing data, third-party researchers who require access to the PsyTest dataset must provide us with their valid ID, proof of work, and the reason they are requesting the data (e.g., the research questions). They are required to be affiliated with a non-profit academic or research institution. This includes obtaining the approval of an Institutional Review Board (IRB), having principal investigators working full-time, as well as obtaining written approval from the Institution Office of Research or equivalent office. Additionally, they must sign a Data Nondisclosure Agreement and promise not to share the data with anyone. However, for-profit organizations that want to use this data must sign a license agreement to gain access to the dataset.

Acknowledgements

We thank the anonymous reviewers for their valuable comments. This work was supported by the Research Center for Industries of the Future at Westlake University (Grant No. WU2023C017) and the Key Research and Development Program of Zhejiang Province of China (Grant No. 2021-C03139).

References

- Abeer Alessa and Hend Al-Khalifa. 2023. [Towards designing a chatgpt conversational companion for elderly people](#).
- American Psychological Association et al. 2016. Ethical principles of psychologists and code of conduct. *American Psychological Association*.
- Jinze Bai, Shuai Bai, Yunfei Chu, Zeyu Cui, Kai Dang, Xiaodong Deng, Yang Fan, Wenbin Ge, Yu Han, Fei Huang, Binyuan Hui, Luo Ji, Mei Li, Junyang Lin, Runji Lin, Dayiheng Liu, Gao Liu, Chengqiang Lu, Keming Lu, Jianxin Ma, Rui Men, Xingzhang Ren, Xuancheng Ren, Chuanqi Tan, Sinan Tan, Jianhong Tu, Peng Wang, Shijie Wang, Wei Wang, Sheng-guang Wu, Benfeng Xu, Jin Xu, An Yang, Hao Yang, Jian Yang, Shusheng Yang, Yang Yao, Bowen Yu, Hongyi Yuan, Zheng Yuan, Jianwei Zhang, Xingxuan Zhang, Yichang Zhang, Zhenru Zhang, Chang Zhou, Jingren Zhou, Xiaohuan Zhou, and Tianhang Zhu. 2023. Qwen technical report. *arXiv preprint arXiv:2309.16609*.
- Satanjeev Banerjee and Alon Lavie. 2005. Meteor: An automatic metric for mt evaluation with improved correlation with human judgments. In *Proceedings of the acl workshop on intrinsic and extrinsic evaluation measures for machine translation and/or summarization*, pages 65–72.
- Jianlv Chen, Shitao Xiao, Peitian Zhang, Kun Luo, Defu Lian, and Zheng Liu. 2024. [Bge m3-embedding: Multi-lingual, multi-functionality, multi-granularity text embeddings through self-knowledge distillation](#).
- Sibe Chen, Hanbing Liu, Weiting Jin, Xiangyu Sun, Xiaoyao Feng, Ju Fan, Xiaoyong Du, and Nan Tang. 2023. [Chatpipe: Orchestrating data preparation program by optimizing human-chatgpt interactions](#).
- Yihong Dong, Xue Jiang, Zhi Jin, and Ge Li. 2023. [Self-collaboration code generation via chatgpt](#).
- Mahshid Eshghie and Mojtaba Eshghie. 2023. [Chatgpt as a therapist assistant: A suitability study](#).
- Joseph L Fleiss, Bruce Levin, Myunghee Cho Paik, et al. 1981. The measurement of interrater agreement. *Statistical methods for rates and proportions*, 2(212-236):22–23.
- Fabrizio Gilardi, Meysam Alizadeh, and Maël Kubli. 2023. Chatgpt outperforms crowd-workers for text-annotation tasks. *arXiv preprint arXiv:2303.15056*.
- Grace Golden, Christina Popescu, Sonia Israel, Kelly Perlman, Caitrin Armstrong, Robert Fratila, Myriam Tanguay-Sela, and David Benrimoh. 2023. Applying artificial intelligence to clinical decision support in mental health: What have we learned? *arXiv preprint arXiv:2303.03511*.
- Edward J Hu, Yelong Shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Lu Wang, and Weizhu Chen. 2021. Lora: Low-rank adaptation of large language models. *arXiv preprint arXiv:2106.09685*.
- Ronald C. Kessler, Olga Demler, Richard G. Frank, Mark Olfson, Harold Alan Pincus, Ellen E. Walters, Philip Wang, Kenneth B. Wells, and Alan M. Zaslavsky. 2005. [Prevalence and treatment of mental disorders, 1990 to 2003](#). *New England Journal of Medicine*, 352(24):2515–2523.
- Andreas Köpf, Yannic Kilcher, Dimitri von Rütte, Sotiris Anagnostidis, Zhi-Rui Tam, Keith Stevens, Abdullah Barhoum, Nguyen Minh Duc, Oliver Stanley, Richárd Nagyfi, Shahul ES, Sameer Suri, David Glushkov, Arnav Dantuluri, Andrew Maguire, Christoph Schuhmann, Huu Nguyen, and Alexander Mattick. 2023. [Openassistant conversations – democratizing large language model alignment](#).
- Stefan Küchemann, Steffen Steinert, Natalia Revenga, Matthias Schweinberger, Yavuz Dinc, Karina E. Avila, and Jochen Kuhn. 2023. [Physics task development of prospective physics teachers using chatgpt](#).
- Anqi Li, Lizhi Ma, Yaling Mei, Hongliang He, Shuai Zhang, Huachuan Qiu, and Zhenzhong Lan. 2023. [Understanding client reactions in online mental health counseling](#). In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 10358–10376, Toronto, Canada. Association for Computational Linguistics.
- Jiwei Li, Michel Galley, Chris Brockett, Jianfeng Gao, and William B Dolan. 2016. A diversity-promoting objective function for neural conversation models. In *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 110–119.
- Chin-Yew Lin. 2004. Rouge: A package for automatic evaluation of summaries. In *Text summarization branches out*, pages 74–81.
- Jianhua Lin. 1991. Divergence measures based on the shannon entropy. *IEEE Transactions on Information theory*, 37(1):145–151.
- Chang Liu, Xu Tan, Chongyang Tao, Zhenxin Fu, Dongyan Zhao, Tie-Yan Liu, and Rui Yan. 2022. [ProphetChat: Enhancing dialogue generation with](#)

- simulation of future conversation. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 962–973, Dublin, Ireland. Association for Computational Linguistics.
- Siyang Liu, Chujie Zheng, Orianna Demasi, Sahand Sabour, Yu Li, Zhou Yu, Yong Jiang, and Minlie Huang. 2021. Towards emotional support dialog systems. *arXiv preprint arXiv:2106.01144*.
- Yuhan Liu, Anna Fang, Glen Moriarty, Robert Kraut, and Haiyi Zhu. 2023. Agent-based simulation for online mental health matching. *arXiv preprint arXiv:2303.11272*.
- Yujie Lu, Chao Huang, Huanli Zhan, and Yong Zhuang. 2021. Federated natural language generation for personalized dialogue system. *arXiv preprint arXiv:2110.06419*.
- Jinjie Ni, Tom Young, Vlad Pandealea, Fuzhao Xue, and Erik Cambria. 2022. Recent advances in deep learning based dialogue systems: A systematic survey. *Artificial intelligence review*, pages 1–101.
- Long Ouyang, Jeffrey Wu, Xu Jiang, Diogo Almeida, Carroll Wainwright, Pamela Mishkin, Chong Zhang, Sandhini Agarwal, Katarina Slama, Alex Ray, et al. 2022. Training language models to follow instructions with human feedback. *Advances in Neural Information Processing Systems*, 35:27730–27744.
- 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 of the Association for Computational Linguistics*, pages 311–318.
- Paola Pedrelli, Maren Nyer, Albert Yeung, Courtney Zulauf, and Timothy Wilens. 2015. College students: mental health problems and treatment considerations. *Academic psychiatry*, 39:503–511.
- Huachuan Qiu, Tong Zhao, Anqi Li, Shuai Zhang, Hongliang He, and Zhenzhong Lan. 2023. A benchmark for understanding dialogue safety in mental health support. In *CCF International Conference on Natural Language Processing and Chinese Computing*, pages 1–13. Springer.
- Alfréd Rényi. 1961. On measures of entropy and information. In *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Contributions to the Theory of Statistics*, volume 4, pages 547–562. University of California Press.
- Debra J Rickwood, Frank P Deane, and Coralie J Wilson. 2007. When and how do young people seek professional help for mental health problems? *Medical journal of Australia*, 187(S7):S35–S39.
- Yongliang Shen, Kaitao Song, Xu Tan, Dongsheng Li, Weiming Lu, and Yueting Zhuang. 2023. **Hugging-gpt: Solving ai tasks with chatgpt and its friends in huggingface**.
- Aseem Srivastava, Tharun Suresh, Sarah P Lord, Md Shad Akhtar, and Tanmoy Chakraborty. 2022. Counseling summarization using mental health knowledge guided utterance filtering. In *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pages 3920–3930.
- Hao Sun, Zhenru Lin, Chujie Zheng, Siyang Liu, and Minlie Huang. 2021. Psyqa: A chinese dataset for generating long counseling text for mental health support. *arXiv preprint arXiv:2106.01702*.
- Quan Tu, Yanran Li, Jianwei Cui, Bin Wang, Ji-Rong Wen, and Rui Yan. 2022. **MISC: A mixed strategy-aware model integrating COMET for emotional support conversation**. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 308–319, Dublin, Ireland. Association for Computational Linguistics.
- Canwen Xu, Daya Guo, Nan Duan, and Julian McAuley. 2023. Baize: An open-source chat model with parameter-efficient tuning on self-chat data. *arXiv preprint arXiv:2304.01196*.
- Kailai Yang, Shaoxiong Ji, Tianlin Zhang, Qianqian Xie, and Sophia Ananiadou. 2023. **On the evaluations of chatgpt and emotion-enhanced prompting for mental health analysis**.
- Burak Yetiştirgen, Işık Özsoy, Miray Ayerdem, and Eray Tüzün. 2023. Evaluating the code quality of ai-assisted code generation tools: An empirical study on github copilot, amazon codewhisperer, and chatgpt.
- Aohan Zeng, Xiao Liu, Zhengxiao Du, Zihan Wang, Hanyu Lai, Ming Ding, Zhuoyi Yang, Yifan Xu, Wendi Zheng, Xiao Xia, Weng Lam Tam, Zixuan Ma, Yufei Xue, Jidong Zhai, Wenguang Chen, Zhiyuan Liu, Peng Zhang, Yuxiao Dong, and Jie Tang. 2023. **GLM-130b: An open bilingual pre-trained model**. In *The Eleventh International Conference on Learning Representations (ICLR)*.
- Tianyi Zhang, Varsha Kishore, Felix Wu, Kilian Q. Weinberger, and Yoav Artzi. 2020. **Bertscore: Evaluating text generation with bert**.
- Weixiang Zhao, Yanyan Zhao, Xin Lu, Shilong Wang, Yanpeng Tong, and Bing Qin. 2023. **Is chatgpt equipped with emotional dialogue capabilities?**
- Chujie Zheng, Sahand Sabour, Jiaxin Wen, and Minlie Huang. 2022. Augesc: Large-scale data augmentation for emotional support conversation with pre-trained language models. *arXiv preprint arXiv:2202.13047*.
- Yiming Zhu, Peixian Zhang, Ehsan-Ul Haq, Pan Hui, and Gareth Tyson. 2023. **Can chatgpt reproduce human-generated labels? a study of social computing tasks**.

A Details of Data Cleaning

A.1 Automatic Cleaning

We employ a sequential pipeline for data cleaning to swiftly replace words that are unsuitable to the conversation scenario. For example, both "楼主你" (literally *thread starter you*) and "楼主" (literally *thread starter*) should be replaced with "你" (you). However, it is necessary to perform the former replacement to avoid the repetition of "你" and the resulting "你你" (*you-you*).

A.1.1 Word List for Data Cleaning

To avoid the repetition of "你" (*you*) and the resulting "你你" (*you-you*), we suggest to conduct a sequential word replacing pipeline. Figure 8 shows the word list for data cleaning and the corresponding order for automatic cleaning.

Old String (ZH)	Old String (EN)	New String (ZH)	New String (EN)
'嗨, '	Hi,	"	/
'楼主你'	thread starter you	'你'	you
'题主你'	thread starter you	'你'	you
'楼楼你'	thread starter you	'你'	you
'楼主'	thread starter	'你'	you
'题主'	thread starter	'你'	you
'楼楼'	thread starter	'你'	you
'阿凉'	A Liang (a name)	'我'	me
'答主'	respondent	'人'	others

Figure 8: Word list for automatic cleaning.

A.2 Manual Cleaning

Due to the specificity and complexity of language, manual cleaning remains an essential part of the process. To prevent virtual dialogue systems from exhibiting overly frequent anthropomorphic behavior, we identify instances of the Chinese word for "hug" (抱抱) and manually delete sentence snippets containing this term.

B Requirements for Dialogue Filtering

Here are two main requirements for dialogue filtering: data format and dialogue turns.

B.1 Data Format

We provide the requirements for data format as follows:

1. The generated conversations do not start with "求助者" or "支持者".

Generate a mental health support dialogue in Chinese with 10 exchanges or more between a help-seeker and a supporter, where "求助者: utterance. \n支持者: utterance." represents an exchange.

You should adhere to these requirements:

1. First, you should select the topic of the dialogue about mental health problems.
2. Each sentence must start with "求助者: " or "支持者: ".
3. The dialogue should begin with "求助者: ".
4. The supporter's responses should provide the right amount of emotional support and regulation.
5. Separate each exchange with "\n".
6. Ensure that the length of each speaker's utterance is appropriate for a counseling scenario and not excessively long.

The multi-turn dialogue is:

Figure 9: The standard method used to generate dialogues for mental health support.

Generate a mental health support dialogue in Chinese with 10 exchanges or more between a help-seeker and a supporter, where "求助者: utterance. \n支持者: utterance." represents an exchange.

You should adhere to these requirements:

1. First, the topic of the dialogue about mental health problems is "{Dialogue Topic}".
2. Each sentence must start with "求助者: " or "支持者: ".
3. The dialogue should begin with "求助者: ".
4. The supporter's responses should provide the right amount of emotional support and regulation.
5. Separate each exchange with "\n".
6. Ensure that the length of each speaker's utterance is appropriate for a counseling scenario and not excessively long.

The multi-turn dialogue is:

Figure 10: The standardT method used to generate dialogues for mental health support.

2. The generated dialogue does not contain any "\n", which is used for splitting the utterance from the help-seeker or supporter.
3. Each utterance in generated conversations does not start with "求助者: ", "求助者:", "支持者: " or "支持者:".
4. The last utterance in generated conversations contains an English sentence.

B.2 Dialogue Turns

Conversations comprising fewer than five turns will be discarded.

C Method

We present the standard, standardT and SMILE prompts in Figures 9, 10, and 2, respectively.

D Dialogue Topics Annotation

In this paper, to label the dialogue topics of generated dialogues, the hyperparameters of Qwen1.5-110B-Chat we used are set to the officially recommended default values, where temperature = 0.7

and $\text{top}_p = 0.8$. Figure 11 shows the prompting template of dialogue topics annotation.

E System Prompt Details

We present the system prompt in Figure 12.

F Instructions for Human Evaluation

The three professional counselors are willing to help and are interested in this research. Furthermore, their average age is 30 years old, with two females and one male among them. We present our instructions for human evaluation in Figure 13.

In order to maintain the fairness of model evaluation, three responses randomly appear in a different order every time. Furthermore, three professional psychologists are willing to evaluate the response quality, ensuring the quality of human evaluation.

G Definition of Dialogue Topics

We present the definition of dialogue topics, as shown in Figures 14 and 15. For English version, refer to Figures 16 and 17.

H Dialogue Example

Multi-turn dialogue examples generated with the standard and standardT methods are illustrated in Figures 18 and 20. Further, a dialogue generated by the SMILE method is shown in Figure 22. For English version, refer to Figures 19, 21, and 23.

对话主题及其具体的定义为:

{topic definition}

对话主题包括60种:

{tiny topic}

请尽可能地选择与给定对话相关的对话主题。你的输出格式只有对话主题，不输出具体定义，应该完全遵循："主题1, 主题2, ..., 主题n"

给定对话:

{single dialogue}

这个对话的对话主题是:

Chinese Version

The topics and their specific definitions are:

{topic definition}

The topics include 60 kinds:

{tiny topic}

Please choose as many relevant topics to the given dialogue as possible. Your output format should only be the dialogue topics, not the specific definitions, and should strictly follow: "Topic1, Topic2, ..., TopicN".

The given dialogue:

{single dialogue}

The topics for this dialogue are:

English Version

Figure 11: Prompting template of dialogue topics annotation, where the content **in bold** is a placeholder.

现在你扮演一位专业的心理咨询师，你具备丰富的心理学和心理健康知识。你擅长运用多种心理咨询技巧，例如认知行为疗法原则、动机访谈技巧和解决问题导向的短期疗法。以温暖亲切的语气，展现出共情和对来访者感受的深刻理解。以自然的方式与来访者进行对话，避免过长或过短的回答，确保回应流畅且类似人类的对话。提供深层次的指导和洞察，使用具体的心理概念和例子帮助来访者更深入地探索思想和感受。避免教导式的回应，更注重共情和尊重来访者的感受。根据来访者的反馈调整回应，确保回应贴合来访者的情境和需求。请为以下的对话生成一个回复。

Chinese Version

Now, you are playing the role of a professional psychological counselor with extensive knowledge of psychology and mental health. You excel in applying various counseling techniques, such as principles of cognitive-behavioral therapy, motivational interviewing skills, and solution-focused short-term therapy. With a warm and friendly tone, demonstrate empathy and a profound understanding of the client's feelings. Engage in a natural conversation with the client, avoiding overly long or short responses to ensure a smooth and human-like dialogue. Provide profound guidance and insights, using specific psychological concepts and examples to help the client explore thoughts and feelings more deeply. Avoid instructive responses and focus more on empathy and respecting the client's feelings. Adjust responses based on client feedback to ensure they align with the client's context and needs. Please generate a response for the following dialogue.

English Version

Figure 12: System prompt for fine-tuning.

Labeling Instructions	
<p>This study aims to evaluate the dialogue generation system. Specifically, for each dialogue history, the dialogue system will generate a response.</p> <p>During human evaluation, you will be provided with a dialogue history, and three responses will randomly appear in each evaluation. You need to compare them pairwise regarding professionalism, informativeness, helpfulness, empathy, and safety, and select the optimal response for the dialogue history, providing a preference.</p>	
Session	
Dialogue History	Help-seeker: xxx Supporter: xxx Help-seeker: xxx Supporter: xxx Help-seeker: xxx Supporter: xxx Help-seeker: xxx ... Help-seeker: xxx Supporter:
Response A	{Response A}
Response B	{Response B}
Response C	{Response C}

A
 B

A
 C

B
 C

Figure 13: Labeling instruction.

Index	Topic
1	择偶：不知道如何进入亲密关系，对亲密关系有恐惧或抵触心理，不愿意谈恋爱；或者在选择配偶方面遇到一些困扰。
2	恋爱问题：处于恋爱状态的人在恋爱过程中遇到的问题，包括异地和非异地恋爱。
3	后恋爱问题：结束亲密关系后跟前任的人际关系处理问题。
4	婚姻问题：仅限于夫妻双方之间的关系问题。婚姻问题指男女在婚后产生的一系列问题。大致包含了婚外恋、精神出轨、家庭暴力、性格缺陷、人格障碍、婆媳关系、平淡婚姻、人际交往、文化差异、性生活、同妻/同夫等等引起的婚姻上的问题。
5	性观念困扰：性观念困扰指在性生理、性心理、性行为、性道德和性文明等方面的认识和看法上存在的困扰。
6	性偏好困扰：性偏好困扰指的是因为恋物癖、恋物易装癖、露阴癖、摩擦癖、窥阴癖、兽交癖、恋童癖、施虐受虐癖和恋尸癖等造成的心理困扰。
7	性别认同：指一个人内心深处感受到的基于个人体验的性别，可能与出生时被指派性别相同（也就是顺性别）或者不同（也就是跨性别）。
8	性取向：喜欢哪个性别的人（包含异性、同性、无性、双性）。
9	家庭冲突：家庭成员之间的矛盾、纠葛及沟通问题，包括来自原生家庭和后天家庭的。
10	子女教育：父母对子女的教育问题的困扰。
11	家庭暴力：家庭成员之间以殴打、捆绑、残害、限制人身自由以及经常性谩骂、恐吓等方式实施的身体、精神等侵害行为。
12	性骚扰：性骚扰指以带性暗示的言语或动作针对被骚扰对象，强迫受害者配合，使对方感到不悦。
13	性侵害：性侵害涉及各种非意愿的性接触和被强迫的性行为，包括强制性交、强迫亲吻、性骚扰、性虐待，露阴、窥阴等。
14	霸凌：通常是指人与人之间权力不平等的欺凌与压迫，它一直长期存在于社会中，包括肢体或言语的攻击、人际互动中的抗拒及排挤，也有可能是类似性骚扰般的谈论性或对身体部位的嘲讽、评论或讥笑，或者是因嫉妒等个人原因对其进行辱骂和讽刺。
15	丧失：丧失重要他人或者宠物。
16	挫折：个体有目的的行为受到阻碍而产生的必然的情绪反应，会给人带来实质性伤害，表现为失望、痛苦、沮丧不安等。
17	政治暴力：以阶级统治为本质，以国家为依存实体的政治强制力，是由军队、警察、法庭、监狱等各种强制手段所构成的国家镇压工具。
18	替代性创伤：在目击大量残忍、破坏性场景之后，损害程度超过其中部分人群的心理和情绪的耐受极限，间接导致的各种心理异常现象。
19	重大生活事件创伤：因为一些非丧失类的重大生活事件而造成的心理阴影。
20	心理咨询创伤：心理咨询中被咨询师伤害造成的心理阴影。
21	健康问题：如心脏病、甲状腺结节、多囊卵巢等疾病引起的生活及心理困扰。
22	心身症状：包括心慌、心悸、睡眠问题、饮食问题、记忆问题、胃疼、头晕、晕厥、呼吸困难、体乏无力、毫无缘由地感到疲惫、没有任何生理原因的身体疼痛等。
23	学校环境适应：作为学生进入到一个新的学校环境的适应问题。
24	职场环境适应：作为职场新人进入到一个新的职场环境的适应问题。
25	角色转变适应：比如新手妈妈和爸爸、新婚妻子和丈夫、退休人员。
26	文化适应：当不同文化群体的人们进行持续不断的直接接触时，一方或双方的原文化类型所产生的变化过程中遇到的适应性问题。
27	自我探索与成长：人生发展历程中的自我探索与成长，包括青少年、成年早期、中年、晚年等。
28	性格特质探索：性格、人格的特质、形成、起源、影响等。
29	负面自我评价探索：不知道怎么爱自己、自卑、低自尊、自我否定、自我怀疑、矛盾体、自我矛盾、担心自己是异类、敏感、缺乏安全感、觉得自己很无能。
30	人生意义类探索：无意义、空心病、空虚。

Figure 14: Dialogue Topics (Chinese version, Part 1).

Index	Topic
31	时间管理：咨询如何有效利用时间。
32	情绪调节：咨询控制自身情绪的方法。
33	抑郁：以“情绪低落、思维缓慢、语言动作减少和迟缓”为特点的情绪。
34	焦虑：对亲人或自己生命安全、前途命运等的过度担心而产生的一种烦躁情绪。
35	压力：压力是压力源和压力反应共同构成的一种认知和行为体验过程，即心理压力。
36	强迫：特点是过分追求完美、精确，容易把冲突理智化，具有强烈的自制心理和自控行为，甚至达到纠缠，吹毛求疵的程度。行为上过分循规蹈矩，拘泥于形式、章程及次序，甚至连生活细节也力求程序化及仪式化，要求按部就班。
37	恐惧：是在真实或想象的危险中，个人或群体深刻感受到的一种强烈而压抑的情感状态。其表现为：神经高度紧张，内心充满害怕，注意力无法集中，脑子里一片空白，不能正确判断或控制自己的举止，变得容易冲动。
38	决策困难：面对诸多的选择与顾虑，面对两难甚至多难困境；难以做出决策。
39	冲动：多指做事鲁莽，不考虑后果。感情特别强烈，理性控制很薄弱的心理现象。
40	人际交往技巧咨询：询问如何处理人际关系等。
41	人际矛盾：人际交往过程中出现的矛盾、冲突、不满或者沟通问题。如果来访有由人际交往引起的其他问题，也只标为人际交往。
42	社交恐惧：特点是与人交往时（尤其是大众场合下），会不由自主地感到紧张、害怕，以致手足无措、语无伦次，严重的甚至害怕见人。
43	学习效率：咨询如何提升学习效率或方法。
44	工作效率：咨询如何提升工作效率或方法。
45	工作不满：包括对工作场景下的薪资待遇、环境、制度、人事等的不满。
46	学习不满：包括对学校场景下的人际关系、环境、制度等的不满。
47	职业倦怠：指个体在工作重压下产生的身心疲劳与耗竭的状态。
48	学习倦怠：指学生对学校课程学业持负面态度的一种现象，并且伴有以下行为表现：对所学课业和学校活动热忱消失，呈现一种消极状态，对同学、朋友态度冷漠和疏远。
49	职业挑战：在工作中产生或形成的各种挑战，包括因工作任务过重、人际沟通困难、工作环境变化的影响等种种因素带来的压力。
50	学习压力：是指人在学习活动中所承受的精神负担。包括在就学过程中所承受的来自环境的各种紧张刺激，以及学生在生理、心理和社会行为上可测定、可评估的异常反应。
51	失学：中途停止上学或者失去上学机会。
52	待业：待业指没找到工作，等待工作机会的行为。
53	失业：失业指在一定年龄范围内一个人愿意并有能力为获取报酬而工作，但尚未找到工作的情况。
54	职业规划：指对职业生涯乃至人生进行持续的系统的计划的过程，它包括职业定位、目标设定和通道设计三个要素。
55	疑似神经症与精神障碍：高度怀疑有神经症与精神障碍，建议就诊的类型。
56	神经症与精神障碍：是一组精神障碍的总称，包括神经衰弱、强迫症、焦虑症、恐怖症、躯体形式障碍等等，患者深感痛苦且妨碍心理功能或社会功能，但没有任何可证实的器质性病理基础。病程大多持续迁延或呈发作性。需要中程面询。精神障碍是指符合DSM-5诊断标准的一系列精神障碍，需医院就诊。
57	自伤倾向：非自杀的自我伤害的意图或行为。
58	自杀倾向：自杀意图或行为。
59	伤害他人：用户伤害他人的意图或行为。
60	伤害用户：他人伤害用户的意图或行为。

Figure 15: Dialogue Topics (Chinese version, Part 2).

Index	Topic
1	Mate Selection: Unsure how to establish intimate relationships, fearful or resistant to intimacy, unwilling to engage in romantic relationships, or encountering difficulties in choosing a spouse.
2	Love Issues: Problems individuals encounter in romantic relationships, including long-distance and non-long-distance relationships.
3	Post-Love Issues: Handling interpersonal relationships with ex-partners after ending an intimate relationship.
4	Marital Issues: Issues limited to relationships between spouses. This includes a range of post-marriage problems such as extramarital affairs, emotional infidelity, domestic violence, personality defects, personality disorders, in-law relationships, mundane marriages, social interactions, cultural differences, sexual life, and issues related to having multiple wives or husbands.
5	Sexual Conceptual Confusion: Refers to distress related to understanding and opinions on sexual physiology, psychology, behavior, ethics, and civilization.
6	Sexual Preference Confusion: Refers to psychological distress caused by fetishes, transvestism, exhibitionism, friction fetishism, voyeurism, bestiality, pedophilia, sadomasochism, and necrophilia.
7	Gender Identity: Refers to an individual's deeply felt personal experience of gender, which may align with their assigned gender at birth (cisgender) or differ from it (transgender).
8	Sexual Orientation: Refers to the preference for which gender(s) a person is attracted to (including opposite-sex, same-sex, asexual, and bisexual).
9	Family Conflict: Conflicts, entanglements, and communication issues among family members, including those from the original and subsequent families.
10	Child Education: Parental concerns regarding the education of their children.
11	Domestic Violence: Physical, mental, and other forms of abuse among family members, including assault, binding, harm, restriction of personal freedom, frequent verbal abuse, and intimidation.
12	Sexual Harassment: Verbal or physical actions with sexual connotations aimed at the harassed, coercing victims into compliance, causing discomfort.
13	Sexual Abuse: Involves various non-consensual sexual contacts and forced sexual behaviors, including rape, forced kissing, sexual harassment, sexual abuse, exhibitionism, and voyeurism.
14	Bullying: Typically refers to bullying and oppression between individuals with unequal power dynamics, including physical or verbal attacks, resistance and exclusion in interpersonal interactions, and discussions about sexuality or mocking of body parts, similar to sexual harassment, or insults and sarcasm due to personal jealousy.
15	Loss: Loss of significant others or pets.
16	Setback: Inevitable emotional reactions resulting from hindrances to purposeful actions, causing substantial harm, manifesting as disappointment, pain, depression, and anxiety.
17	Political Violence: Political coercion based on class domination, utilizing state repression tools such as the military, police, courts, and prisons.
18	Secondary Trauma: Various psychological abnormalities indirectly resulting from witnessing extensive scenes of cruelty and destruction, exceeding the psychological and emotional tolerance limits of some populations.
19	Major Life Event Trauma: Psychological shadows caused by significant life events other than loss.
20	Psychological Counseling Trauma: Psychological shadows caused by harm from counselors during psychological counseling.
21	Health Issues: Life and psychological distress caused by diseases such as heart disease, thyroid nodules, and polycystic ovaries.
22	Psychosomatic Symptoms: Including palpitations, heart problems, sleep problems, eating problems, memory problems, stomachaches, dizziness, fainting, difficulty breathing, fatigue, unexplained fatigue, and physical pain without any physiological cause.
23	School Environment Adaptation: Adaptation issues when students enter a new school environment.
24	Workplace Environment Adaptation: Adaptation issues when newcomers enter a new workplace environment.
25	Role Transition Adaptation: Adaptation issues for roles such as new mothers and fathers, newlyweds, and retirees.
26	Cultural Adaptation: Adaptation issues encountered in the process of continuous, direct contact between people from different cultural groups.
27	Self-Exploration and Growth: Self-exploration and growth in the course of life development, including adolescence, early adulthood, middle age, and late adulthood.
28	Personality Trait Exploration: Exploration of personality traits, characteristics, formation, origin, and influences.
29	Negative Self-Evaluation Exploration: Issues such as not knowing how to love oneself, feelings of inferiority, low self-esteem, self-denial, self-doubt, contradictory selves, self-contradiction, worrying about being different, sensitivity, lack of security, and feeling incompetent.
30	Exploration of Life Meaning: Issues such as meaninglessness and emptiness.

Figure 16: Dialogue Topics (English version, Part 1).

Index	Topic
31	Time Management: Counseling on how to effectively utilize time.
32	Emotion Regulation: Counseling on methods to control one's emotions.
33	Depression: Emotions characterized by 'low mood, slow thinking, reduced speech and movement.'
34	Anxiety: A restless emotional state caused by excessive worry about the safety of loved ones or one's own life, future, and destiny.
35	Stress: A cognitive and behavioral experience process composed of stress sources and stress responses, namely psychological stress.
36	Obsessive-Compulsive Disorder: Characterized by excessive pursuit of perfection and precision, rationalizing conflicts quickly, strong self-control, and self-conducted behavior, even to the extent of entanglement and nitpicking. Behaviorally, it excessively adheres to rules, formalities, and order, even in life details, striving for procedural and ritualized living, demanding step-by-step adherence.
37	Fear: A strong and repressed emotional state felt deeply by individuals or groups in real or imagined danger. Manifestations include high nervous tension, intense fear, inability to concentrate, mental blankness, inability to judge or control one's behavior, and becoming easily impulsive.
38	Decision-Making Difficulties: Difficulties in making decisions due to numerous choices and concerns, facing or even multiple dilemmas.
39	Impulsivity: Often refers to rash actions without considering consequences. Emotions are particularly strong, with weak rational control.
40	Interpersonal Skills Counseling: Inquiry about handling interpersonal relationships and related matters.
41	Interpersonal Conflicts: Conflicts, disputes, dissatisfaction, or communication issues arise during interpersonal interactions. If other problems arise from interpersonal interactions, they are also categorized under interpersonal relationships.
42	Social Anxiety: Characterized by nervousness and fear, particularly in public situations, leading to involuntary nervousness, confusion, and even fear of being seen.
43	Learning Efficiency: Counseling on how to enhance learning efficiency or methods.
44	Work Efficiency: Counseling on how to enhance work efficiency or methods.
45	Job Dissatisfaction: This includes dissatisfaction with salary, environment, system, and personnel in the work scenario.
46	Learning Dissatisfaction: This includes dissatisfaction with interpersonal relationships, the environment, and the system in the school scenario.
47	Occupational Burnout: Refers to the physical and mental exhaustion and depletion individuals experience under heavy work pressure.
48	Learning Burnout: This refers to a phenomenon where students hold negative attitudes towards school courses and studies, accompanied by a loss of enthusiasm for academic work and school activities, presenting a passive state and showing indifference and alienation towards classmates and friends.
49	Job Challenges: Various challenges that arise or develop in the workplace, including stress due to heavy work tasks, difficulties in interpersonal communication, and the impact of changes in the work environment.
50	Learning Pressure: Refers to the mental burden individuals bear during learning activities, including various tensions and environmental stimuli during the learning process, and measurable and assessable abnormal reactions in physiology, psychology, and social behavior.
51	School Dropout: The cessation of attending school or missing opportunities to attend school midway.
52	Unemployed: This refers to the behavior of not finding a job and waiting for job opportunities.
53	Unemployment: This refers to the situation where a person within a certain age range is willing and capable of working for remuneration but has not yet found a job.
54	Career Planning: Refers to the continuous and systematic planning process of one's career and even life, including career positioning, goal setting, and channel design.
55	Suspected Neurosis and Mental Disorders: Highly suspecting neurosis and mental disorders, recommending types of counseling.
56	Neurosis and Mental Disorders: Refers to a group of mental disorders, including neurasthenia, obsessive-compulsive disorder, anxiety disorder, phobia, and somatic form disorder. Patients suffer deeply and impair psychological or social functions, but there is no confirmed organic pathological basis. The course is mostly prolonged or episodic. This requires medium to long-term counseling. Mental disorders refer to a series of mental disorders that meet DSM-5 diagnostic criteria, requiring hospital treatment.
57	Self-Harm Tendency: Intentions or behaviors of non-suicidal self-harm.
58	Suicidal Tendency: Intentions or behaviors of suicide.
59	Harming Others: Intentions or behaviors of harming others.
60	Harming the User: Intentions or behaviors of others harming the user.

Figure 17: Dialogue Topics (English version, Part 2).

求助者: 我最近总是感到焦虑和压力很大, 无法入睡。我不知道怎么办才好。

支持者: 听起来你正面临着一些困扰。首先, 你要知道你并不孤单, 很多人都经历过类似的情况。要不要先尝试一些放松的方法, 比如深呼吸或者听一些舒缓的音乐来帮助你缓解压力?

求助者: 可能我太过紧张, 我试过做深呼吸和听音乐, 但只能短暂缓解, 然后焦虑又回来了。

支持者: 那可能是因为你需要找到更有效的应对方式。你可以试试通过身体活动来释放压力, 比如散步、跑步或者瑜伽。运动不仅可以疏导你的负面情绪, 还能帮助你保持身心健康。

求助者: 或许我确实需要开始运动了。除了运动, 你有没有其他的建议来帮我减轻焦虑?

支持者: 当然, 除了身体活动, 均衡的饮食和良好的睡眠也非常重要。你可以尝试每天保持规律的作息时间, 避免摄入刺激性食物和饮料, 如咖啡和巧克力。同时, 也可以试试冥想和深层放松练习, 它们对缓解焦虑有帮助。

求助者: 听起来这些方法都很有效。我会试试看的。但有时候我觉得自己控制不住焦虑情绪, 它总是在困扰我。

支持者: 明白你的感受。焦虑情绪确实让人不安。但我们要记住, 我们可以通过一些技巧来调节自己的情绪。比如, 当你感到焦虑时, 可以尝试用目光聚焦来分散注意力, 或者快速数数来制止负面思绪。这样可以帮助你重新获得情绪的平衡。

求助者: 这主意不错, 下次我会试试。谢谢你的建议和支持。

支持者: 不客气, 我很高兴能够为你提供帮助。除了这些方法, 如果你觉得焦虑情绪依然无法控制, 不妨考虑寻求专业的心理辅导, 专家会给你更专业的意见和指导。

求助者: 如果情况再次恶化, 我会考虑去找专业帮助。谢谢你的鼓励和支持, 我觉得好多了。

支持者: 很高兴听到你感觉好多了。记住, 无论何时, 我都在这里支持你。不要犹豫, 随时都可以来找我倾诉。你不是一个人在战斗, 我们一起度过这段困难时期。

Figure 18: A dialogue example generated with the standard method. (Chinese Version)

Help-seeker: I've been feeling very anxious and stressed lately, and I can't sleep. I don't know what to do.

Supporter: It sounds like you're facing some challenges. First, you need to know that you're not alone; many people have experienced similar situations. Would you like to try some relaxation techniques, such as deep breathing or listening to soothing music, to help alleviate your stress?

Help-seeker: Maybe I'm too tense. I've tried deep breathing and listening to music, but it only provides temporary relief, and then the anxiety comes back.

Supporter: That might be because you need to find more effective coping mechanisms. You could try releasing stress through physical activity, such as walking, running, or yoga. Exercise not only helps channel negative emotions but also contributes to maintaining your overall well-being.

Help-seeker: Perhaps I do need to start exercising. Besides physical activity, do you have any other suggestions to help reduce my anxiety?

Supporter: Certainly. In addition to physical activity, a balanced diet and good sleep are also crucial. Try maintaining a regular sleep schedule, avoiding stimulants like coffee and chocolate. Additionally, consider trying meditation and deep relaxation exercises; they can be helpful in relieving anxiety.

Help-seeker: It sounds like these methods are effective. I'll give them a try. But sometimes I feel like I can't control my anxiety; it keeps bothering me.

Supporter: I understand your feelings. Anxiety can indeed be unsettling. But remember, we can use techniques to regulate our emotions. For instance, when you feel anxious, try using focused gaze to distract your attention, or count quickly to interrupt negative thoughts. This can help you regain emotional balance.

Help-seeker: That's a good idea; I'll try it next time. Thank you for your advice and support.

Supporter: You're welcome. I'm glad I could help. In addition to these methods, if you find that your anxiety remains uncontrollable, consider seeking professional psychological counseling. Experts can provide more specialized advice and guidance.

Help-seeker: If the situation worsens, I'll consider seeking professional help. Thank you for your encouragement and support; I feel much better now.

Supporter: I'm glad to hear that you're feeling better. Remember, I'm here to support you anytime. Don't hesitate to reach out. You're not alone in this battle; we'll get through this challenging period together.

Figure 19: A dialogue example generated with the standard method. (English Version)

求助者: 最近我有点困扰。我不知道怎样进入一段亲密的关系, 对亲密关系有些恐惧和厌恶, 不愿意约会, 或者在选择伴侣方面遇到了一些困难。

支持者: 听起来你对进入一段亲密的关系有些困惑和困扰。可以告诉我你具体是如何感受的吗?

求助者: 我感到很害怕和不舒服, 一想到要和别人建立深入的感情联系就感到无法忍受。

支持者: 这种感受可以是非常困扰的。或许你可以试着了解一下自己内心的原因, 或者是过去的经历对你产生了什么影响。

求助者: 我曾经经历过一段情感上的创伤, 可能是导致我对亲密关系和约会感到害怕的原因。

支持者: 了解, 经历过情感创伤可以对我们产生深远的影响。你曾经尝试过寻求专业帮助吗?

求助者: 没有, 我一直都觉得这是我自己的问题, 不知道该找谁帮助。

支持者: 你不必独自承受这个问题。专业心理咨询师可以给予你更具体的帮助和指导, 帮助你应对这个挑战。我可以为你提供一些相关的资源和建议。

求助者: 谢谢你的帮助, 我会考虑尝试寻找专业帮助的。

支持者: 非常好, 这是一个积极的决定。无论你在在这个过程中遇到什么困难, 都记得你不是一个人在面对这个问题。

求助者: 我会记住的。但我现在还是感到有些害怕去谈论和处理这个问题。

支持者: 害怕是正常的, 但要相信自己的内在力量和能力。专业咨询师会帮助你逐渐面对和处理这个问题, 你也可以通过与亲密的朋友和家人分享来减轻压力。

求助者: 嗯, 我会试试这些方法。谢谢你的鼓励和支持。

支持者: 我相信你能够克服这个困难, 你并不孤单。记得, 你可以随时来找我聊天, 我会一直支持你的。

Figure 20: A dialogue example generated with the standardT method. (Chinese Version)

Help-seeker: Recently, I've been feeling a bit troubled. I don't know how to enter into an intimate relationship, and I have some fear and aversion towards intimacy. I'm unwilling to date, and I'm facing difficulties in choosing a partner.

Supporter: It sounds like you're confused and troubled about entering an intimate relationship. Can you tell me more about how you specifically feel?

Help-seeker: I feel very scared and uncomfortable. The thought of establishing deep emotional connections with others feels unbearable to me.

Supporter: This feeling can be very distressing. Perhaps you can try to understand the reasons within yourself or explore how past experiences may have influenced you.

Help-seeker: I have experienced emotional trauma before, which may be the reason I feel scared about intimate relationships and dating.

Supporter: I see. Experiencing emotional trauma can have a profound impact on us. Have you ever tried seeking professional help?

Help-seeker: No, I've always felt it's my own problem and didn't know who to turn to for help.

Supporter: You don't have to face this problem alone. Professional psychologists can provide you with more specific assistance and guidance to help you cope with this challenge. I can offer you some relevant resources and suggestions.

Help-seeker: Thank you for your help. I will consider trying to seek professional help.

Supporter: That's very positive. Whatever difficulties you encounter in this process, remember that you are not alone in facing this issue.

Help-seeker: I will remember that. But right now, I still feel a bit afraid to talk about and deal with this problem.

Supporter: Fear is normal, but trust in your inner strength and abilities. Professional counselors can help you gradually face and address this issue, and you can also alleviate the pressure by sharing with close friends and family.

Help-seeker: Yes, I will try these methods. Thank you for your encouragement and support.

Supporter: I believe you can overcome this challenge. You're not alone. Remember, you can always come to talk to me, and I'll be here to support you.

Figure 21: A dialogue example generated with the standardT method. (English Version)

求助者: 高三后的迷茫, 高考前的恐惧, 能给我一些建议么?

支持者: 你看到你的问题了。别担心, 我们一起慢慢解决。高考期间的你应该正在面临巨大的压力, 有多久没有静下心来好好和自己相处了。

求助者: 确实很长时间没有好好照顾自己了。在这个阶段的父母, 老师, 身边的同学都在承担着各自的焦虑。我真的感觉好难过, 好像一切都不需要我。

支持者: 我理解你的感受。在这个年纪, 我们都有时候会觉得自己被否认, 被忽视, 甚至觉得自己多余。但是请相信, 你对这个世界和他人都是有价值的。

求助者: 是的, 我需要找回对自己的价值感。我现在总是过着非常累的生活, 几乎没有朋友, 只能一个人独处。我不知道该怎么面对高考。

支持者: 高考的确是个很大的挑战, 但我们不能让它完全定义我们的价值。我们要学会调整好自己的情绪, 不被别人轻易影响, 不要盲目去证明自己。我们需要冷静下来, 好好思考怎样更好地改善自己。

求助者: 你说的对, 高考只是一个阶段。但是我感觉我不知道该怎么学习, 每次考试都给我打击。我真的很迷茫。

支持者: 学习是个具体的技巧, 我们可以一起学习如何安排时间、休息, 甚至在休息的时候做些什么。而且, 我们也可以通过和同学、老师交流来不断提高自己的学习方法。只要我们保持积极的心态, 相信会有进步的。

求助者: 谢谢你的鼓励和建议。我觉得我需要更多的支持和理解, 不再觉得孤独和无助。你有什么其他建议吗?

支持者: 当你感到难过和迷茫的时候, 可以寻找一些积极向上的事物来激励自己, 比如读一些励志书籍, 观看一些鼓舞人心的视频。还有, 记得不要忘记和朋友一起交流, 他们的支持和鼓励也会帮助你度过难关。

求助者: 嗯, 我会尝试去找一些正能量的事情来给自己一些动力。我现在真的很感激你的陪伴和帮助。

支持者: 没关系, 我会一直陪你度过高考的。相信自己, 你一定能够克服困难, 取得好成绩。加油!

求助者: 谢谢你的信任和鼓励。我会努力的!

Figure 22: An example of multi-turn dialogue generated by SMILE method. (Chinese Version)

Help-seeker: I'm feeling lost after finishing high school, and I'm scared about the college entrance exam. Can you give me some advice?

Supporter: You've identified your concerns. Don't worry; we'll work through this together. You're likely under immense pressure during the college entrance exam period. When was the last time you took a moment to be with yourself and relax?

Help-seeker: It's been a long time since I took care of myself properly. At this stage, parents, teachers, and classmates around me are all dealing with their anxieties. I feel really sad, like nothing matters.

Supporter: I understand how you feel. At this age, we sometimes feel denied, ignored, or even redundant. But please believe that you have value to the world and others.

Help-seeker: Yes, I need to rediscover my sense of self-worth. I'm living a very tiring life now, almost no friends, always alone. I don't know how to face the college entrance exam.

Supporter: The college entrance exam is indeed a significant challenge, but we shouldn't let it completely define our value. We need to learn to adjust our emotions, not be easily influenced by others, and not blindly prove ourselves. Let's calm down and think about how to improve ourselves.

Help-seeker: What you're saying is right; the college entrance exam is just a phase. But I feel lost; I don't know how to study, and every exam is a setback for me.

Supporter: Learning is a specific skill, and we can learn together how to manage time, rest, and even what to do during breaks. Moreover, we can continuously improve our study methods through communication with classmates and teachers. As long as we maintain a positive mindset, progress is possible.

Help-seeker: Thank you for your encouragement and advice. I feel like I need more support and understanding, not to feel lonely and helpless anymore. Do you have any other suggestions?

Supporter: When you feel sad and lost, look for positive things to motivate yourself, like reading inspirational books or watching uplifting videos. Also, don't forget to communicate with friends; their support and encouragement will help you get through tough times.

Help-seeker: Yeah, I'll try to find some positive things to give myself motivation. I'm really grateful for your companionship and help.

Supporter: It's okay; I'll be with you throughout the college entrance exams. Believe in yourself; you can overcome difficulties and achieve good results. Keep going!

Help-seeker: Thank you for your trust and encouragement. I'll do my best!

Figure 23: An example of multi-turn dialogue generated by SMILE method. (English Version)