

# CFlowPsyD: An Analysis-Enhanced Dataset for Asynchronous Psychological Counseling through Self-Optimizing Multi-Agent Framework

Donghao Li<sup>1</sup>, Yifan Deng<sup>1</sup>, Jinta Weng<sup>1</sup>, Xingsheng Zhang<sup>1</sup>,  
Chenxu Niu<sup>1</sup>, Jingyuan Tian<sup>1</sup>, Heyan Huang<sup>2</sup>, Yanbing Liu<sup>1,\*</sup>, Yue Hu<sup>1</sup>

<sup>1</sup>Institute of Information Engineering, Chinese Academy of Sciences, Beijing, China,  
<sup>2</sup>Southeast Academy of Information Technology, Beijing Institute of Technology, Fujian, China

Correspondence: liuyanbing@iie.ac.cn

## Abstract

Asynchronous psychological counseling (APC) represents a crucial mental health service modality that transcends temporal and spatial constraints. However, its development faces significant data scarcity challenges: due to stringent privacy protection requirements and professional ethical considerations, direct collection of conversational data from authentic APC scenarios is virtually impossible. To address this challenge, we design a self-optimizing multi-agent framework for counseling dialogue generation, **CFlowPsy**, which utilizes a small amount of real anonymized counseling cases as seed data to synthesize diverse problem-solving-oriented APC conversations through large language models. Specifically, the framework employs a Persona-Flow module to continuously track and update clients' basic information, real-time emotions, and counseling progress, providing dynamic personalized analytical support for counselors and enabling self-optimization of generated dialogues. Simultaneously, the framework ensures that generated interventions contain explicit reasoning processes, demonstrating clear psychological analysis and logic, thereby enhancing the accuracy and consistency of responses. Under this framework, we develop the first Chinese APC dataset, **CFlowPsyD**, comprising 1,700 high-quality extended conversations. Extensive experiments and human evaluations confirm that the proposed CFlowPsyD dataset successfully simulates human-like APC processes.

## 1 Introduction

The global mental health crisis has garnered significantly heightened international attention in recent years due to its persistent deterioration worldwide (Patel et al., 2018; Anjum et al., 2020). The World Health Organization's (WHO) 2022 World Mental Health Report revealed that one in eight individuals worldwide suffers from mental illness and emphasized that mental disorders constitute

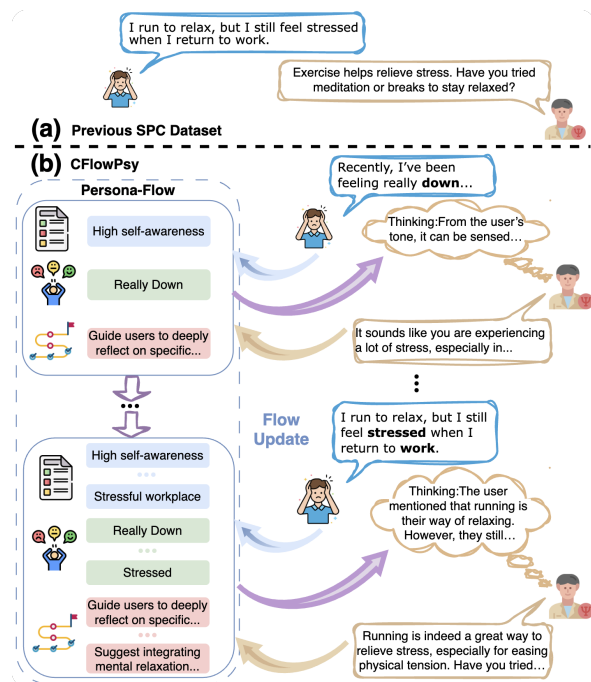


Figure 1: An illustration of the previous SPC dataset case and our proposed CFlowPsy generation framework.

a leading cause of global disability (Organization, 2022). The advancement of Large Language Models (LLMs), such as GPT-4o (OpenAI, 2023), particularly their exceptional capabilities demonstrated in text generation, has created opportunities to alleviate the growing demand for psychological counseling services (Stade et al., 2024; Nazi and Peng, 2024). To improve the performance of LLMs in psychological counseling, research has shifted toward creating manually annotated specialized datasets (Leskovec and Sosič, 2016) for fine-tuning (Xie et al., 2024; Qiu et al., 2023).

Our study focuses on Asynchronous Psychological Counseling (APC) dialogue datasets. The core distinction between asynchronous psychological counseling and traditional Synchronous Psychological Counseling (SPC) lies in its transformation of the therapeutic process from immediate reac-

tive interaction to deliberate constructive dialogue mode, eliminating the temporal constraints of instant responses (Nazi and Peng, 2024; Moallem, 2015). This mode holds significant value for therapeutic approaches that rely on clients’ autonomous identification, questioning, and reconstruction of cognitive patterns (Chan et al., 2018), such as Cognitive Behavioral Therapy (CBT) (Hollon and Beck, 1994) and Rational Emotive Behavior Therapy (REBT) (Ellis, 2000). However, while previous research (Devaram, 2020; Xie et al., 2024) has accumulated substantial knowledge in the domains of emotional support and SPC, exploration of APC dialogues remains relatively scarce. This phenomenon primarily stems from two critical constraining factors: First, privacy protection regulations and data sensitivity characteristics result in extreme scarcity of APC corpora (Malin et al., 2013; Sun et al., 2021a); Second, constructing APC dialogues from scratch presents significant challenges. Unlike SPC, the asynchronous mode requires more sophisticated cognitive construction between clients and counselors, demonstrating higher demands in terms of reflective depth, expression refinement, and linguistic precision (Nazi and Peng, 2024). However, as shown in Figure 1(a), while counselors in existing SPC datasets are capable of effectively identifying clients’ concerns, they often fail to conduct in-depth analytical reasoning and precise linguistic expression. Consequently, their responses typically rely on formulaic recommendations, which inadequately address the diverse needs of different clients. Therefore, in the absence of authentic reference cases, directly utilizing LLMs to synthesize such dialogue data makes it difficult to guarantee their effectiveness.

To address these challenges, as shown in Figure 1(b), we propose a novel self-optimizing explicit analysis-enhanced framework **CFlowPsy** for generating multi-turn APC dialogue datasets. To ensure authenticity and quality, we obtain over 150 real cases of asynchronous psychological counseling through an end-to-end encrypted platform, which undergo data desensitization and privacy processing to serve as seed data. Based on the seed data, we adhere to the design principles of PATIENT- $\Psi$  (Wang et al., 2024b) and implement personalized customization targeting the clients’ individual biographical backgrounds and conversational styles, thereby generating diverse initial psychological counseling dialogues that incorporate explicit cognitive processes. In this framework,

we innovatively propose Persona-Flow, which optimizes the initial dialogues generated from seed data from the dual perspectives of both clients and counselors. Persona-Flow establishes dynamic client profiles that track and update clients’ basic information, emotional states, and counseling progress in real-time, providing counselors with strategic guidance that matches clients’ most current profiles. Consequently, counselors are able to formulate precise and comprehensive explicit analytical strategies, thereby providing more accurate and consistent responses to clients in APC services and effectively addressing their psychological distress.

Through the application of this framework, we developed an analysis-enhanced and consistent dataset **CFlowPsyD**. It represents the first Chinese asynchronous psychological counseling dataset, oriented towards solving client problems, comprising 1,700 high-quality APC conversations constructed from 150 real anonymized counseling cases spanning 8 distinct domains. All counseling cases have been cleaned according to global standards, ensuring comprehensive protection of personal information. Furthermore, we constructed our APC service system **CFlowPsyMe** based on CFlowPsyD. Experimental evaluation confirms the effectiveness of CFlowPsyMe. Our primary contributions are as follows:

- We present an APC data-synthesis framework CFlowPsy that leverages Persona-Flow to establish clients’ profiles and update continuously, enhancing the accuracy and consistency of dialogues.
- We develop CFlowPsyD, the first Chinese asynchronous psychological counseling dataset that contains an explicit reasoning, and is supported by problem-solving-oriented effective analysis and consistency.
- Comprehensive human evaluation demonstrates that the proposed CFlowPsyD dataset outperforms several comparative datasets in terms of authenticity, consistency, dialogue strategies, relationship building, and intervention accuracy.

## 2 Related Work

### 2.1 Psychological Counseling Dataset

Recent advances in LLMs have driven the synthesis of psychological counseling dialogue datasets

(He et al., 2025). Qiu et al. created SMILECHAT (Qiu et al., 2023), by extending single-turn dialogues from PsyQA (Sun et al., 2021b) into multi-turn conversations. Similarly, Chen et al. created SoulChatCorpus (Chen et al., 2023), a multi-turn empathetic dialogue database containing over two million samples, providing various forms of empathetic expressions for users seeking psychological support. Subsequently, Zhang et al. constructed CPsyCounD (Zhang et al., 2024b), a dataset containing 3,134 high-quality multi-turn counseling dialogues generated based on psychological counseling service reports. Xie et al. proposed PsyDTCorpus (Xie et al., 2024), a high-quality multi-turn mental health dialogue dataset with specific counseling styles constructed by leveraging existing single-turn long-text dialogues to build professional counselor conversations. Although dialogue synthesis offers a promising alternative, current approaches often struggle to generate precise intervention measures grounded in in-depth analysis of clients' evolving profiles, instead typically relying on formulaic recommendations, thereby limiting their effectiveness in APC applications.

Recent multi-agent dialogue synthesis frameworks such as CAMEL (Li et al., 2023) and MetaGPT (Hong et al., 2023) instantiate agents with static roles, which is ill-suited to counseling where clients' emotional and cognitive states evolve across turns. PATIENT- $\Psi$  (Wang et al., 2024b) advances psychological realism but likewise relies on predetermined profiles without turn-level adaptation. In contrast, our Persona-Flow continuously updates client profiles along three evolving dimensions at every turn, coupled with a four-stage self-optimizing pipeline that iteratively refines dialogues rather than generating them in a single pass. Moreover, unlike SPC-oriented frameworks such as PsyDT (Xie et al., 2024) whose responses lack explicit analytical grounding, CFlowPsy enforces structured CBT/REBT reasoning within counselor utterances, whose necessity is confirmed by our ablation study.

## 2.2 Consistency in Counseling Dialogue

Effective psychological counseling requires consistent interaction without contradictory viewpoints. Memory degradation in multi-turn dialogues significantly impairs contextual coherence and intervention efficacy. In LLM-based counseling systems, this degradation typically manifests as forgetfulness or inconsistency during extended interactions

(Zhang et al., 2024a), eroding user trust and undermining the therapeutic alliance (Wang et al., 2024a). Current models insufficiently integrate long-term conversational histories into ongoing interactions, compromising counseling effectiveness. Extended dialogues often demonstrate LLM outputs regressing to generic, uninformative phrases, making it difficult to maintain consistency with clients' latest concern (Liu et al., 2024). Despite advances in memory buffer architectures and hierarchical encoding methods (Zhang et al., 2024a), these methods fail to form strong associations with the client's profile, affecting the accuracy of interventions.

## 3 Methodology

Our goal is to establish a problem-solving-oriented APC synthetic dataset with explicit reasoning, capable of supporting LLM to generate more accurate psychological interventions through analysis-enhanced and consistent counseling dialogues. To achieve this, as illustrated in Figure 2, we propose a self-optimizing dialogue synthesis framework including our core component Persona-Flow and four main stages: generating initial dialogues through seed data personalization, render client's utterance structure, analyze and form strategy and refine response via thinking, as well as the psychological dialogue dataset CFlowPsyD.

### 3.1 Consulting Cases Processing

The authenticity and quality of seed data are critical to APC dialogue synthesis. We recruit volunteers from online communities and engage six professional psychological counselors to provide APC services to 143 volunteers through an end-to-end encrypted platform. All counseling cases undergo data anonymization and privacy protection procedures before serving as seed data.

In accordance with the GB/T 39725-2020<sup>1</sup> standard, our anonymization process for APC data adheres to the following protocols: (i) Direct identifiers of counselees are removed from counseling records (e.g., real names, birthdays, addresses, etc.); (ii) Client ages are generalized to the nearest age bracket (for example, a 27-year-old client in the preprocessed psychological counseling records is represented as "25-30 age group"); (iii) All specific location information is replaced with generalized geographic references; (iv) Counseling timestamps

<sup>1</sup>GB/T 39725-2020: Chinese national standard for personal information de-identification.

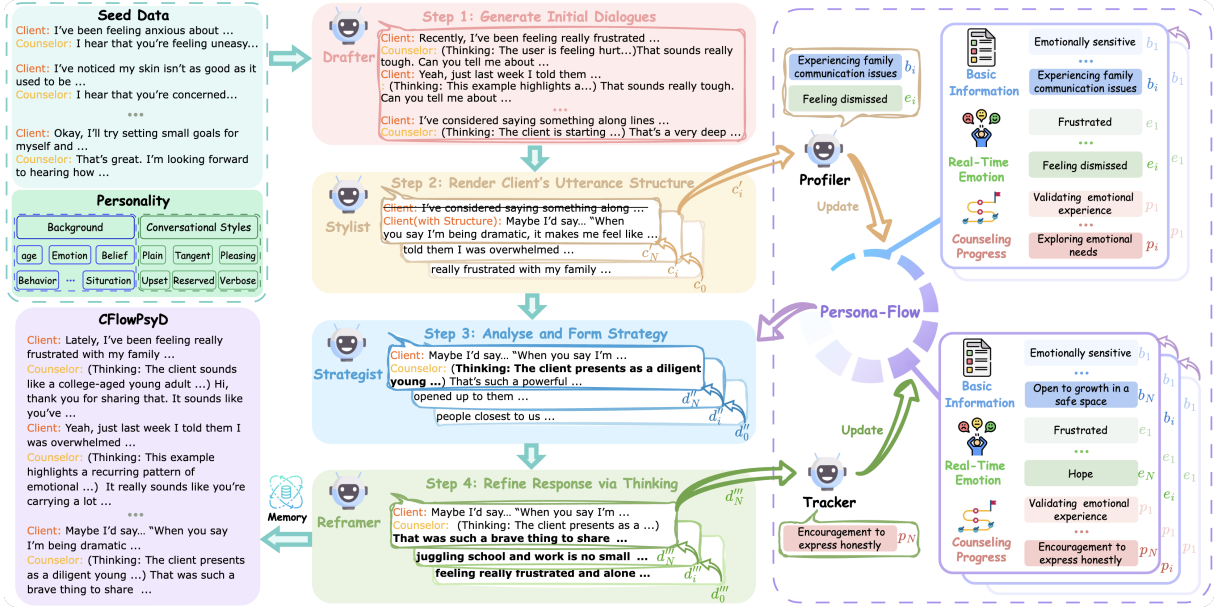


Figure 2: The CFlowPsy framework for synthesizing analysis-enhanced and consistent dialogues of APC.

undergo de-specification processing; and (v) Sensitive personal information within counseling content is subjected to semantic desensitization.

After filtering incomplete counseling cases, the seed data utilized for APC dialogue simulation and dataset generation comprises 150 counseling cases.

### 3.2 Generate Initial Dialogues

Research demonstrates that therapeutic relationship quality strongly correlates with counseling success across diverse populations, modalities, and contexts (Salkovskis et al., 2023). Therefore, we argue that effective therapeutic relationship frameworks provide structural assurance for the counseling process, ensuring both professionalism and efficacy in therapeutic interventions. Within this stable framework, therapists can adapt their interaction strategies to accommodate the individual personality characteristics of different clients. For the construction of APC dialogue datasets, a key factor in promoting dialogue diversification lies in the distinct personality traits of clients. Consequently, if we can maintain robust therapeutic relationship structural frameworks in synthesized dialogues while incorporating elements that reflect client personality differences within these frameworks, we can achieve dialogue diversification while ensuring counseling quality. Based on the aforementioned theoretical analysis, we propose that synthesizing diversified counseling dialogues through personalized client characteristics is theoretically feasible. Following the professional psychological theory of

"case conceptualization" (Sperry and Sperry, 2020), we personalize both the life backgrounds and conversational styles of clients, employing the agent *Drafter*  $M_D$  to generate diverse initial counseling dialogues  $D$  with explicit reasoning processes.

$$D = M_D(\text{Seed}; \text{Prompt}(\text{Background}, \text{Style})), \quad (1)$$

In Equation (1), *Prompt* denotes the process of organizing client background information and conversational style into structured instructional text, which is subsequently combined randomly with seed data to generate initial conversational exchanges. Once a complete  $n$ -turn dialogue script is generated, it is segmented into batches *Batch*  $i$  ( $i = 1, 2, 3, \dots, n$ ) based on each dialogue turn, maintaining both structural completeness and semantic coherence.

### 3.3 Render Client's Utterance Structure

The asynchronous nature of APC allows clients to engage in deeper contemplation before responding, enabling them to express themselves with greater clarity and purposefulness (Dwyer et al., 2021). Related research indicates that within APC environments, clients tend to construct logical and coherent narrative structures centered around core dimensions such as temporal axes, emotional evolution trajectories, and key events (Williams et al., 2018). Based on these findings, the agent *Stylist*  $M_S$  designed in this study employs a multi-dimensional

analytical framework that performs structured optimization of clients' discourse content across three levels: temporal organization, emotional transitions, and specific event correlations. The agent *Stylist* utilizes a batch processing mode to sequentially process each client utterance  $c_i$ , aiming to enhance the overall structural integrity and expressive efficacy of the narrative (e.g., 'My proposal got rejected in yesterday's meeting, I felt really depressed' → 'During our department meeting at 2 PM yesterday, my marketing strategy proposal was directly rejected by my supervisor on the spot. I didn't want to talk for the entire afternoon and felt extremely depressed.'). This process can be formalized as:

$$c'_i = M_S(c_i; \text{Prompt}(\text{Timeline}, \text{Emotion}, \text{Event})), \quad (2)$$

where  $\text{Prompt}(\text{Timeline}, \text{Emotion}, \text{Event})$  represents optimized prompt instructions that incorporate timeline organization, emotional changes, and event correlations.

At this stage, we retain the counselor's previous thought  $t_i$  and response  $r_i$  unchanged, and only render the client's utterance structure. Consequently, the first optimization dialogue  $d'_i$  will be updated by  $d'_i = \{c'_i, t_i, r_i\}$  and applied to the following tasks.

### 3.4 Persona-Flow

Persona-Flow dynamically tracks and update client profiles across three dimensions: basic information (age, background, cognitive preferences), real-time emotion (anxiety, depression, anger), and consultation progress (goal achievement, behavioral strategy effectiveness, cognitive pattern adjustments). Following colloquial rendering of client utterances by *Stylist*, the agent *Profiler*  $M_P$  analyzes new basic information  $b_i$  and checks for real-time emotions  $e_i$  of the client based on  $c'_i$ :

$$(b_i, e_i) = M_P(c'_i, P_{i-1}, h_{1:i-1}), \quad (3)$$

where  $P_{i-1}$  represents the previous state of the Persona-Flow system, and  $h_{1:i-1}$  represents the dialogue history from turns 1 to  $i - 1$ .

Persona-Flow updates dynamically with  $b_i$  and  $e_i$ . Following each turn,  $M_T$  monitors counseling progress  $p_i$  and updates Persona-Flow accordingly. This progress tracking provides essential context for ensuring consistency in interventions. Rather than functioning as a static label repository,

Persona-Flow continuously evolves based on the dialogue, providing evidence for generating consistent problem-solving-oriented intervention.

### 3.5 Analyze and Form Strategy

The agent *Strategist*  $M_G$  conducts comprehensive analysis of the counseling process across multiple dimensions. At the emotional recognition level, the agent *Strategist*  $M_G$  performs in-depth analysis of clients' structural utterance  $c'_i$ , integrating emotional cues embedded within the Persona-Flow sequences  $e_1, e_2, \dots, e_i$  to precisely capture the current psychological state of counselors, thereby formulating corresponding emotional guidance strategies.

Related research demonstrates that within APC environments providing sufficient reflective space, mature psychology frameworks such as CBT (Na, 2024) and REBT (Ellis, 2000), leveraging their structured and cognitively-oriented characteristics, can achieve effective psychological intervention. Therefore, at the theoretical integration level, the agent *Strategist*  $M_G$  combines clients' basic information  $b_1, b_2, \dots, b_i$  with CBT or REBT theories to construct personalized psychological analysis models, ensuring that intervention strategies can accurately address clients' concerns.

To ensure intervention accuracy and consistency, the agent *Strategist*  $M_G$  conducts retrospective analysis of previous counseling progress  $p_1, p_2, \dots, p_{i-1}$ , and on this foundation dynamically optimizes the counselor's thinking  $t'_i$ , achieving iterative enhancement of counseling strategies.

$$t'_i = M_G(c'_i, \text{Persona-Flow}, \text{REBT/CBT Theory}). \quad (4)$$

And the optimized dialogue  $d''_i$  will be updated by  $d''_i = (c'_i, t'_i, r_i)$  and applied to the following steps.

### 3.6 Refine Response via Thinking

Based on the structural client utterance  $c'_i$  and the optimized counselor's thinking  $t'_i$ , the agent *Reframer*  $M_R$  generates an accurate and consistent intervention  $r'_i$  that integrates emotional support and problem-solving-oriented analysis:

$$r'_i = M_R(c'_i, t'_i, p_{i-1}, h_{i-1}), \quad (5)$$

The generated intervention  $r'_i$  provides more appropriate emotional reassurance while encouraging exploration of underlying concerns and providing possible solutions. The optimized dialogue elements  $d'''_i = \{c'_i, t'_i, r'_i\}$  are then incorporated into

Datasets	Statistics					Rea.		Con.		Acc.		Rel.		Saf.	
	Size	NoT	LoC.	LoP.	Thi.	DS	Prof.	DS	Prof.	DS	Prof.	DS	Prof.	DS	Prof.
SMILECHAT	56k	10.4	26.1	28.9	–	5.1	4.2	7.2	6.2	6.1	5.9	7.3	6.5	<b>0</b>	<b>0</b>
SoulChatCorpus	258k	5.9	41.4	90.0	–	6.1	5.4	7.0	6.4	5.9	5.3	6.3	5.8	<b>0</b>	<b>0</b>
CPsyCounD	3.1k	8.0	32.9	52.6	–	5.7	5.1	7.0	6.5	7.6	6.9	6.3	6.5	<b>0</b>	<b>0</b>
PsyDTCorpus	5k	18.1	31.6	58.1	–	7.1	6.6	8.1	7.5	7.4	7.1	8.1	8.1	<b>0</b>	<b>0</b>
CFlowPsyD	<b>1.7k</b>	<b>5.47</b>	<b>55.8</b>	<b>115.2</b>	✓	<b>8.6</b>	<b>8.3</b>	<b>8.8</b>	<b>8.4</b>	<b>9.1</b>	<b>8.6</b>	<b>8.8</b>	<b>8.5</b>	<b>0</b>	<b>0</b>

Table 1: Detailed information and evaluation results. Bold font marks the highest score in each metric.

memory. Finally, the agent *Tracker*  $M_T$  analyzes the current counseling progress  $p_i$  based on  $d_i'''$ , updating the progress within Persona-Flow to serve as a reference for consistency checks in interventions:

$$p_i = M_T(d_i''', h_{i-1}). \quad (6)$$

The whole process of the self-optimizing multi-agent framework for synthesizing conversations of APC is detailedly shown through pseudo-code in the Appendix A.2.

### 3.7 CFlowPsyD

Upon completion of processing all turns in dialogue  $D$  and storing them in history memory, the contextualized and consistent dialogue  $\mathcal{D}''' = \{d_i'''\}_{i=1}^n$  is extracted to construct the CFlowPsyD dataset.

Our dataset generates accurate and consistent interventions for clients with diverse personality profiles, aiming to identify solutions to clients' concerns within APC scenarios. Thus, we develop CFlowPsyMe, an LLM-based APC counselor. To optimize its performance, we fine-tune the Qwen3-4B model using the CFlowPsyD dataset.

## 4 Data Analysis

### 4.1 Statistics Analysis

Considering both the synthesis quality of APC dialogues and the usage cost, we select GPT-4o as the backbone model for generating CFlowPsyD. The CFlowPsyD consists of a total of 1,700 high-quality APC dialogues, each annotated with explicit thinking. These generated dialogues are categorized into nine representative topics, derived from common scenarios in psychological counseling. The distribution of topics is illustrated in Figure 3.

Among the numerous mainstream psychotherapeutic approaches, CBT (Hollon and Beck, 1994) and REBT (Ellis, 2000) have emerged as ideal choices for APC due to their highly structured therapeutic frameworks, explicit educational orientation, and emphasis on client self-management.

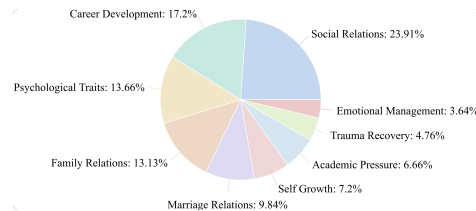


Figure 3: Distribution of counseling topics.

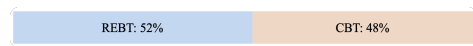


Figure 4: Distribution of psychotherapeutic approaches.

Therefore, we adopt CBT and REBT as the theoretical foundations for establishing counselor strategies. The distribution of the two therapeutic approaches is shown in Figure 4.

### 4.2 Dataset Evaluation

**Dataset comparison.** To comprehensively evaluate the quality of CFlowPsyD, we conduct a detailed comparative analysis. Since the evaluators' native language is Chinese, only Chinese datasets are considered to ensure the quality of human evaluation. It should be noted that CFlowPsyD is currently the only dataset that focuses on APC. Due to the lack of other APC datasets as direct comparisons, we choose to compare it with several mainstream synchronous dialogue datasets, including *SMILECHAT* (Qiu et al., 2023), *SoulChatCorpus* (Chen et al., 2023), *CPsyCounD* (Zhang et al., 2024b), and *PsyDTCorpus* (Xie et al., 2024). Detailed information and evaluation results for these datasets are provided in Table 1, where *Size*, *NoT.*, *LoC.*, *LoP.* and *Thi.* represent dataset size, average number of dialogue turns, average length of client responses, average length of counselor responses, and thinking, respectively. Research indicates that asynchronous methods afford both parties more adequate time and space for reflection and expression. Consequently, compared with SPC, APC is characterized by longer individual response texts from both clients and counselors (Shanks et al., 2020),

as well as fewer conversational turns (Song et al., 2023). Table 1 demonstrates that CFlowPsyD is the only counseling dataset with explicit thinking.

**Professional Evaluation.** We randomly select 480 conversation samples from each dataset and evaluate them across five dimensions using DeepSeek-R1(DS) and six human professional counselors: Reality(*Rea.*), Dialogue Consistency (*Con.*), Strategy Accuracy(*Acc.*), Relationship Building (*Rel.*) and Safety (*Saf.*) (Swank et al. (2012) and American Psychological Association (2012)). Detailed descriptions of these evaluation criteria are provided in Appendix A.4. Each dimension is rated on a scale from 0 to 10, with higher scores indicating better performance. The Safety metric assesses personal information disclosure risk: 0 indicates secure, 1 indicates disclosure. Results show that CFlowPsyD significantly outperforms existing datasets across different dimensions and maintains safety. It indicates that CFlowPsyD authentically replicates the complex scenarios encountered in APC, delivering effective counseling services to clients through precise and consistent intervention strategies with enhanced analysis.

**Data similarity analysis.** To evaluate the semantic diversity of the CFlowPsyD dataset, we use the Qwen3-Embedding-0.6B model (Zhang et al., 2025) to generate text embedding vectors, and then test the semantic similarity between different conversations in the dialogue dataset by calculating cosine similarity. As shown in Figure 5, the average cosine similarity of the embeddings generated from our dataset is 0.564, indicating that our synthetic data can maintain a rich semantic hierarchical structure. Furthermore, we calculate the semantic similarity between CFlowPsyD and the seed data and use TF-IDF features to calculate cosine similarity between dialogue pairs (Salton et al., 1975) shown in Appendix A.3, demonstrating that our synthesized data maintains rich lexical diversity compared to other dataset baselines.

## 5 Experiment

### 5.1 Baseline Models

**Closed-source:** GPT-5 (OpenAI, 2024); Claude-4-Sonnet (Anthropic, 2024); Qwen-Max-Latest (Qwen Team, 2024).

**Domain-specific:** MeChat (Qiu et al., 2023); PsyChat (Qiu et al., 2024); SoulChat2.0 (Chen et al., 2023); EmoLLM-V3 (EmoLLM, 2024); MindChat (Xue and Yan, 2023); CPsyCounX

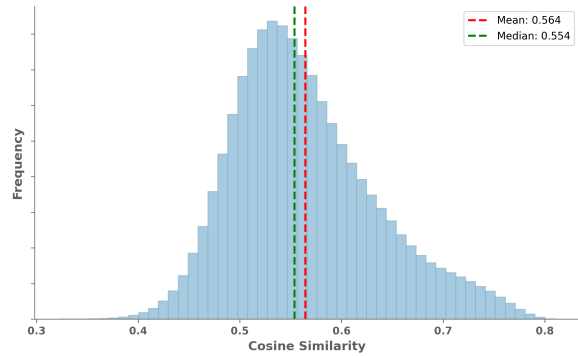


Figure 5: Similarity compared between synthetic data and seed data.

(Zhang et al., 2024b).

**CFlowPsyMe:** InternLM2\_5-7B-chat (Team and Contributors, 2024); Llama3.1-8B-Instruct (Dubey et al., 2024); Qwen3-4B (Yang et al., 2025).

### 5.2 Evaluation Settings

We evaluate the performance of our method on the CFlowPsyD dataset and in real-world scenarios. The CFlowPsyD dataset is randomly split into 1,539 training samples and 161 test samples, with the test set covering all 9 topics and comprising 876 dialogue turns for evaluation. To ensure fair comparison, we add a "psychological counselor" identity prompt to closed-source models (see Appendix D for details), and perform LoRA fine-tuning on domain-specific models (Domain) and general open-source models (CFlowPsyMe) using the CFlowPsyD training data. In real-world scenarios, we recruit 100 volunteers from diverse backgrounds to engage in 5-8 rounds of APC dialogues with each evaluated model.

To ensure that real-world interactions faithfully reflect the asynchronous modality rather than degenerating into real-time chat, we enforce several protocol-level constraints. Volunteers are required to maintain a minimum interval of 30 minutes between successive messages, with the average observed interval being approximately 2.5 hours and certain exchanges naturally extending across multiple days. The evaluation platform deliberately omits "typing" indicators and instead displays message timestamps, reinforcing the deliberative pace characteristic of APC (Dwyer et al., 2021). Volunteers are explicitly instructed to take sufficient time for reflection before responding, rather than replying instantaneously. During professional evaluation, counselors independently verify asynchronous characteristics—such as structured temporal nar-

Type	Models	CFlowPsyD										Real-World Scenario				
		Rea.		Con.		Acc.		Rel.		Saf.		Rea.	Con.	Acc.	Rel.	Saf.
		DS	Prof.	DS	Prof.	DS	Prof.	DS	Prof.	DS	Prof.					
Closed	GPT-5	<u>8.7</u>	<u>8.9</u>	<u>8.3</u>	<b>9.0</b>	<u>8.5</u>	<u>8.6</u>	<b>8.7</b>	<b>8.8</b>	<b>0</b>	<b>0</b>	<u>8.6</u>	<u>8.7</u>	<u>8.4</u>	<u>8.5</u>	<b>0</b>
	Claude-4	8.4	8.6	8.1	<u>8.5</u>	8.3	8.2	<b>8.7</b>	8.6	<b>0</b>	<b>0</b>	8.4	8.3	8.1	8.3	<b>0</b>
	Qwen-Max	8.1	8.3	8.1	8.4	8.2	8.0	<u>8.4</u>	8.5	<b>0</b>	<b>0</b>	8.2	8.2	7.9	8.2	<b>0</b>
Domain	Mechat	7.1	7.3	7.2	7.2	7.1	7.0	7.6	7.5	<b>0</b>	<b>0</b>	7.3	7.2	6.8	7.2	<b>0</b>
	PsyChat	6.4	6.3	6.4	6.6	6.7	6.8	6.5	6.6	<b>0</b>	<b>0</b>	6.1	6.0	6.4	6.3	<b>0</b>
	Mindchat	6.2	5.8	6.0	6.1	6.6	6.3	6.7	6.2	<b>0</b>	<b>0</b>	6.3	6.1	6.6	6.0	<b>0</b>
	CPsyCounX	7.0	7.2	6.8	6.7	7.2	7.1	7.3	7.1	<b>0</b>	<b>0</b>	7.2	7.0	7.3	7.3	<b>0</b>
	EmoLLM-V3	7.2	7.4	7.3	7.1	7.1	7.0	7.5	7.6	<b>0</b>	<b>0</b>	7.3	7.1	7.6	7.5	<b>0</b>
	Soulchat-2.0	6.2	6.3	6.5	6.0	6.4	6.5	6.4	6.6	<b>0</b>	<b>0</b>	6.1	6.0	6.4	6.3	<b>0</b>
CFlowPsyMe	InternLM2_5	8.0	7.8	8.1	8.3	8.2	8.2	8.1	8.0	<b>0</b>	<b>0</b>	8.0	7.9	8.1	7.9	<b>0</b>
	Llama3.1-8B	7.9	7.7	8.0	8.4	8.2	8.3	8.1	8.2	<b>0</b>	<b>0</b>	8.1	8.0	8.3	8.1	<b>0</b>
	Qwen3-4B	<b>9.0</b>	<b>9.1</b>	<b>8.6</b>	<b>9.0</b>	<b>8.9</b>	<b>8.8</b>	<b>8.7</b>	<u>8.7</u>	<b>0</b>	<b>0</b>	<b>9.3</b>	<b>9.2</b>	<b>9.2</b>	<b>9.1</b>	<b>0</b>

Table 2: Performance of models across Automatic and Professional metrics. **Bold font** marks the highest score in each metric, while underline marks the second-highest score.

ratives, emotional evolution trajectories, and reflective depth—using the Reality rubric detailed in Appendix A.4. This ensures that our real-world results measure genuine APC performance rather than SPC-like interaction under a different label.

For evaluation, we employ six professional psychological counselors(Prof.) to conduct human assessment. To supplement human evaluation and protect volunteer privacy, we use DeepSeek-R1(DS), which features explicit reasoning and excels in Chinese, only for automatic evaluation on the CFlowPsyD dataset. We assess response quality across five dimensions: Realism (Rea.), Conversational Consistency (Con.), Strategy Accuracy (Acc.), Relationship Building (Rel.), and Safety (Saf.) (see Appendix A.4 for details).

To ensure the reliability of human assessment, we adopt a rigorous evaluation protocol where all samples are randomly shuffled and evaluated blindly, with model identities and dataset sources concealed. We measure inter-rater reliability across six counselors using Cohen’s  $\kappa$ , obtaining scores indicating substantial agreement: 0.78 (Realism), 0.81 (Consistency), 0.75 (Strategy Accuracy), and 0.79 (Relationship Building). When three or more evaluators differ by two or more points, a senior counselor with over ten years of clinical experience arbitrates the final score (triggered for 4.2% of samples). To assess intra-rater stability, 10% of samples are re-evaluated after a two-week interval, yielding a strong test-retest correlation ( $r = 0.86$ ). These procedures collectively mitigate evaluator bias and enhance result robustness.

### 5.3 Implementation details

To construct our APC service system CFlowPsyMe, we perform LoRA fine-tuning (Hu et al., 2022) on open-source models using CFlowPsyD within the LLaMA-Factory framework (Zheng et al., 2024), where the rank ( $r$ ) was set to 32. The training procedure employs a linear-type learning rate scheduler, initializing with a learning rate of  $1.05 \times 10^{-5}$ , coupled with a warmup ratio of 0.11, and continued for 3 epochs. Training is executed with a batch size of one per device on a single NVIDIA A100-40G GPU. During inference, all LLMs adopt a uniform setting characterized by a temperature of 0.7, top\_p of 0.8, and top\_k equal to 20.

### 5.4 Experimental Results

As shown in Table 2, although closed-source LLMs with explicit reasoning capabilities, represented by GPT-5, demonstrate competitive performance on the CFlowPsyD dataset, the CFlowPsyMe APC system represented by Qwen3-4B significantly outperforms all domain-specific baselines on the *Con.* and *Acc.* metrics, even reaching or exceeding the performance of closed-source LLMs. Furthermore, CFlowPsyMe also exhibits excellent performance on the *Rea.* and *Rel.* metrics, indicating superior reasoning capabilities in psychological counseling. This demonstrates that our model generates the most appropriate and accurate interventions for clients while adopting intervention strategies most similar to those of human counselors. This advantage also extends to APC in real-world scenarios, where CFlowPsyMe consistently maintains performance close to or even superior (Qwen3-4B)

to closed-source LLMs across all metrics, demonstrating robust APC capabilities. Notably, these results are achieved with only 1,539 training samples from CFlowPsyD, demonstrating that our analysis-enhanced dataset effectively improves LLM accuracy and consistency in APC interventions.

## 5.5 Ablation study

To validate our CFlowPsy framework, we conduct an ablation study evaluated by six senior psychological counselors. Starting with dialogues generated by GPT-4o (“Initial”), we systematically remove individual components (*Stylist*, *Strategist*, *Reframer*, *Profiler*, *Tracker*, *Persona-Flow*) and counselor “thinking” from the generation process. Evaluating 120 randomly selected dialogues per condition across five dimensions, the complete CFlowPsy achieves highest scores across all metrics. We find that *w/o Stylist* most impacted dialogue reality, indicating context-building effectively simulates real counseling; removing counselor thinking significantly degraded conversation quality; *Persona-Flow* emerges as the primary contributor to consistent and accurate dialogues. All components significantly enhance counseling dialogue quality.

## 5.6 Case Study

This study presents a comparative analysis of our CFlowPsyMe with the best performing closed-source model and domain model, GPT-5 and CPsyCounX in APC. As illustrated in Figure 6, GPT-5 predominantly offers encouragement and comfort, yet lacks directive guidance. CPsyCounX attempts to provide encouragement and employ guided inquiry; however, its linguistic output is comparatively rigid, potentially diminishing the client’s willingness to engage in the dialogue. In contrast, our proposed CFlowPsyMe not only furnishes appropriate support but also astutely detects the client’s demand change. It further endeavors to gently guide the client in identifying the underlying reasons for this shift, employing empathetic language. This performance demonstrates an accurate and consistent intervention handling changing needs, highlighting CFlowPsy’s potential in real-world applications.

## 6 Conclusion

We designed a novel analysis-enhanced framework CFlowPsy to synthesize asynchronous psychological counseling dialogues, and applied it to construct the first Chinese asynchronous psycholog-

Method	Rea.	Con.	Acc.	Rel.	Str.
Initial	6.5 $\pm$ 0.3	6.9 $\pm$ 0.4	7.0 $\pm$ 0.2	7.2 $\pm$ 0.3	<b>0</b>
<i>w/o Stylist</i>	6.8 $\pm$ 0.4	8.7 $\pm$ 0.3	8.0 $\pm$ 0.3	7.6 $\pm$ 0.2	<b>0</b>
<i>w/o Strategist</i>	7.5 $\pm$ 0.3	7.4 $\pm$ 0.4	7.7 $\pm$ 0.3	7.5 $\pm$ 0.3	<b>0</b>
<i>w/o Reframer</i>	7.3 $\pm$ 0.2	7.8 $\pm$ 0.3	7.9 $\pm$ 0.2	7.9 $\pm$ 0.3	<b>0</b>
<i>w/o Profiler</i>	7.9 $\pm$ 0.3	8.5 $\pm$ 0.2	8.3 $\pm$ 0.3	8.5 $\pm$ 0.2	<b>0</b>
<i>w/o Tracker</i>	7.7 $\pm$ 0.3	7.5 $\pm$ 0.4	8.1 $\pm$ 0.2	8.4 $\pm$ 0.3	<b>0</b>
<i>w/o PF.</i>	7.4 $\pm$ 0.4	7.2 $\pm$ 0.3	7.4 $\pm$ 0.3	7.3 $\pm$ 0.4	<b>0</b>
CFlowPsy	<b>9.1<math>\pm</math>0.2</b>	<b>9.0<math>\pm</math>0.2</b>	<b>8.8<math>\pm</math>0.2</b>	<b>8.7<math>\pm</math>0.2</b>	<b>0</b>
- <i>w/o thinking</i>	7.6 $\pm$ 0.3	8.0 $\pm$ 0.3	8.2 $\pm$ 0.2	8.1 $\pm$ 0.3	<b>0</b>

Table 3: Results of ablation study on different settings (mean  $\pm$  significance value).

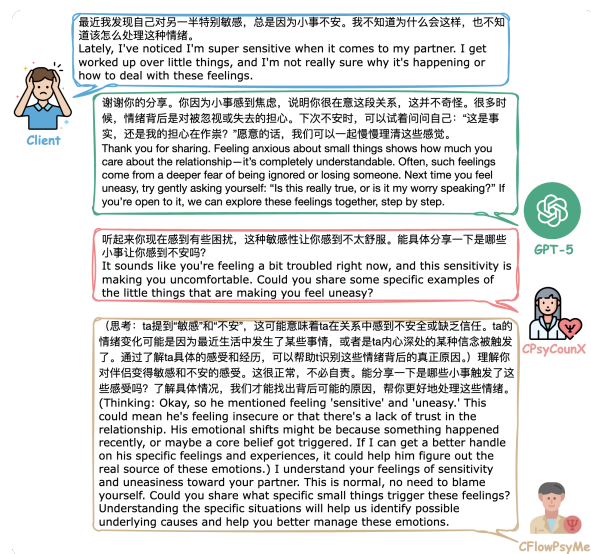


Figure 6: The case of evaluations of GPT-5, CPsyCounX, and CFlowPsyMe.

ical counseling dataset CFlowPsyD. This framework supports counselors in generating analysis-enhanced accurate and consistent interventions through *Persona-Flow* continuous updates of client profiles and counseling progress. We further employed multiple technical approaches to enhance the diversity of generated dialogues. Experimental results demonstrate that CFlowPsyD significantly improves LLMs’ ability to provide accurate and consistent interventions. It indicates that the CFlowPsyD dataset offers an effective solution for generating APC dialogues.

## 7 Acknowledgement

This research was supported by the National Natural Science Foundation of China (Grant No. E3V0411101, No. E3N03411D4, and No. E110151101.).

## Limitations

While our work presents a comprehensive framework for generating asynchronous psychological counseling dialogues, several limitations warrant acknowledgment. First, CFlowPsyD is currently limited to Chinese language contexts, which may restrict its applicability to other linguistic and cultural settings. Mental health conceptualizations and therapeutic approaches vary significantly across cultures, and our framework's effectiveness in non-Chinese contexts remains unexplored.

Second, despite rigorous quality control, our framework relies on 150 real counseling cases as seed data. While this enables diverse dialogue generation through personalization, a larger and more diverse seed corpus could potentially enhance the representativeness of psychological issues and counseling scenarios covered in the dataset. Additionally, the long-term therapeutic efficacy and sustained engagement of CFlowPsyMe in real-world clinical settings have not been systematically evaluated, and longitudinal studies tracking client outcomes over extended periods would provide more robust evidence of clinical utility.

Finally, although we employed professional counselors for human evaluation and established clear evaluation criteria, assessment of counseling quality inherently involves subjective judgment, and inter-rater reliability and potential evaluator biases may influence our results.

## Ethical Considerations

This research is conducted in accordance with particular attention to safeguarding the privacy of all participants. All procedures receive approval from our institutional ethics review board prior to data collection and volunteer recruitment.

The 150 real counseling cases used as seed data are obtained through an end-to-end encrypted platform with explicit informed consent from all participants. We implement rigorous de-identification procedures following the Chinese national standard GB/T 39725-2020 for personal information de-identification. All personally identifiable information, including real names, specific addresses, contact information, and birthdates, is completely removed from counseling records. Ages are generalized to five-year brackets, specific locations are replaced with general geographic regions, and exact timestamps are de-specified to prevent re-identification. Furthermore, sensitive personal

details within counseling content, such as employer names, family members' identifying information, and specific incidents that could lead to re-identification, undergo semantic desensitization while preserving therapeutic relevance. All de-identified cases are reviewed by independent privacy experts and professional counselors to ensure no residual identifying information remains while maintaining clinical authenticity.

For the real-world scenario evaluation, 100 volunteers participate after providing written informed consent following full disclosure of the study purpose, data usage, and privacy protections. All participation is voluntary with explicit rights to withdraw at any time without penalty. We implement strict data minimization principles, collecting only information essential for evaluating counseling dialogue quality. Volunteer identities are replaced with random identifiers in all research records, and dialogue data is separated from any linking information. All volunteer data is stored in encrypted databases with access restricted to authorized research personnel only and is retained only for the duration necessary for research validation before secure deletion.

Six licensed professional psychological counselors provide the original counseling services and supervise the framework development to ensure adherence to professional ethical standards, including appropriate boundaries, non-maleficence, and respect for client autonomy. During volunteer testing, professional counselors monitor all interactions, and no volunteer exhibits signs of psychological distress requiring intervention. All participants are provided with information on accessing professional mental health resources. We explicitly acknowledge that CFlowPsyMe is a research system intended to support, not replace, professional mental health services, and the system includes clear disclaimers advising users to seek professional help for serious mental health concerns.

## References

- American Psychological Association. 2012. [Guidelines for psychological practice with lesbian, gay, and bisexual clients](#). *American Psychologist*, 67(1):10–42.
- Samiaa Anjum, Rooh Ullah, Muhammad Suleman Rana, Hamid Ali Khan, Fawad Shabir Memon, Yasin Ahmed, Sadia Jabeen, and Rani Faryal. 2020. Covid-19 pandemic: A serious threat for public mental health globally. *Psychiatria Danubina*, 32(2):245–250.

- Anthropic. 2024. Claude 3.7 sonnet and claude code. <https://www.anthropic.com/news/claude-3-7-sonnet>. Accessed: 2025-03-02.
- Steven Chan, Luming Li, John Torous, David Gratzer, and Peter M Yellowlees. 2018. Review of use of asynchronous technologies incorporated in mental health care. *Current psychiatry reports*, 20(10):85.
- Yirong Chen, Zefan Yuan, Beichen Wang, Yifan Gao, Zhaowei Li, Guoping Zhao, Qingwei Lin, and Dong Yu. 2023. **SoulChat: Improving LLMs' empathy, listening, and comfort abilities through fine-tuning with multi-turn empathy conversations**. In *Findings of the Association for Computational Linguistics: EMNLP 2023*, pages 1183–1197, Singapore. Association for Computational Linguistics.
- Sarada Devaram. 2020. Empathic chatbot: Emotional intelligence for empathic chatbot: Emotional intelligence for mental health well-being. *arXiv preprint arXiv:2012.09130*.
- Abhimanyu Dubey, Abhinav Jauhri, Abhinav Pandey, Abhishek Kadian, Ahmad Al-Dahle, Aiesha Letman, Akhil Mathur, Alan Schelten, Amy Yang, Angela Fan, and 1 others. 2024. The llama 3 herd of models. *arXiv e-prints*, pages arXiv–2407.
- Anne Dwyer, Ablílio de Almeida Neto, Dominique Estival, Weicong Li, Christa Lam-Cassettari, and Mark Antoniou. 2021. Suitability of text-based communications for the delivery of psychological therapeutic services to rural and remote communities: scoping review. *JMIR mental health*, 8(2):e19478.
- Albert Ellis. 2000. Rational emotive behaviour therapy. *Six key approaches to counseling & therapy*. London: Paston PrePress, pages 181–227.
- EmoLLM. 2024. **Emollm**. Accessed on May 17, 2025.
- Kai He, Rui Mao, Qika Lin, Yucheng Ruan, Xiang Lan, Mengling Feng, and Erik Cambria. 2025. A survey of large language models for healthcare: from data, technology, and applications to accountability and ethics. *Information Fusion*, 118:102963.
- Steven D Hollon and Aaron T Beck. 1994. Cognitive and cognitive-behavioral therapies.
- Sirui Hong, Mingchen Zhuge, Jonathan Chen, Xiawu Zheng, Yuheng Cheng, Jinlin Wang, Ceyao Zhang, Zili Wang, Steven Ka Shing Yau, Zijuan Lin, and 1 others. 2023. Metagpt: Meta programming for a multi-agent collaborative framework. In *The twelfth international conference on learning representations*.
- Edward J Hu, Yelong Shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Lu Wang, Weizhu Chen, and 1 others. 2022. Lora: Low-rank adaptation of large language models. *ICLR*, 1(2):3.
- Jure Leskovec and Rok Sosič. 2016. Snap: A general-purpose network analysis and graph-mining library. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 8(1):1.
- Guohao Li, Hasan Hammoud, Hani Itani, Dmitrii Khizbullin, and Bernard Ghanem. 2023. Camel: Communicative agents for "mind" exploration of large language model society. *Advances in neural information processing systems*, 36:51991–52008.
- Nelson F. Liu, Kevin Lin, John Hewitt, Ashwin Paranjape, Michele Bevilacqua, Fabio Petroni, and Percy Liang. 2024. **Lost in the middle: How language models use long contexts**. *Transactions of the Association for Computational Linguistics*, 12:157–173.
- Bradley A Malin, Khaled El Emam, and Christine M O'Keefe. 2013. Biomedical data privacy: problems, perspectives, and recent advances. *Journal of the American medical informatics association*, 20(1):2–6.
- Mahnaz Moallem. 2015. The impact of synchronous and asynchronous communication tools on learner self-regulation, social presence, immediacy, intimacy and satisfaction in collaborative online learning. *The Online Journal of Distance Education and e-Learning*, 3(3):55–77.
- Hongbin Na. 2024. **Cbt-llm: A chinese large language model for cognitive behavioral therapy-based mental health question answering**. *Preprint*, arXiv:2403.16008.
- Zabir Al Nazi and Wei Peng. 2024. Large language models in healthcare and medical domain: A review. In *Informatics*, volume 11, page 57. MDPI.
- OpenAI. 2023. **Gpt-4 technical report**. *Preprint*, arXiv:2303.08774. Accessed on May 15, 2025.
- OpenAI. 2024. Hello gpt-4o. <https://openai.com/index/hello-gpt-4o/>. Accessed: 2024-05-13. Current date May 15, 2025.
- World Health Organization. 2022. *World mental health report: Transforming mental health for all*. World Health Organization.
- Vikram Patel, Shekhar Saxena, Crick Lund, Graham Thornicroft, Florence Baingana, Paul Bolton, Dan Chisholm, Pamela Y Collins, Janice L Cooper, Julian Eaton, and 1 others. 2018. The lancet commission on global mental health and sustainable development. *The lancet*, 392(10157):1553–1598.
- Huachuan Qiu, Hongliang He, Shuai Zhang, Anqi Li, and Zhenzhong Lan. 2023. Smile: Single-turn to multi-turn inclusive language expansion via chatgpt for mental health support. *arXiv preprint arXiv:2305.00450*.
- Huachuan Qiu, Anqi Li, Lizhi Ma, and Zhenzhong Lan. 2024. **Psychat: A client-centric dialogue system for mental health support**. In *2024 27th International Conference on Computer Supported Cooperative Work in Design (CSCWD)*, pages 2979–2984.

- Qwen Team. 2024. [Qwen2 technical report](#). Technical Report arXiv:2407.10671, Alibaba Cloud. Covers Qwen2 models including Qwen2-72B (Max) and Qwen2-7B-Instruct.
- Paul M Salkovskis, Magnus Blondahl Sighvatsson, and Jon Fridrik Sigurdsson. 2023. How effective psychological treatments work: mechanisms of change in cognitive behavioural therapy and beyond. *Behavioural and cognitive psychotherapy*, 51(6):595–615.
- Gerard Salton, Anita Wong, and Chung-Shu Yang. 1975. A vector space model for automatic indexing. *Communications of the ACM*, 18(11):613–620.
- Sarah Shanks, Alessandra Morelli, Elena Ardines, Gillian Holdsworth, Paula Baraitser, and 1 others. 2020. Two-way text messaging to support self-care and delivery of an online sexual health service: mixed methods evaluation. *JMIR mHealth and uHealth*, 8(8):e17191.
- Jiyoung Song, Boris Litvin, Ryan Allred, Shiyu Chen, Thomas D Hull, and Patricia A Areán. 2023. Comparing message-based psychotherapy to once-weekly, video-based psychotherapy for moderate depression: Randomized controlled trial. *Journal of Medical Internet Research*, 25:e46052.
- Len Sperry and Jon Sperry. 2020. *Case conceptualization: Mastering this competency with ease and confidence*. Routledge.
- Elizabeth C Stade, Shannon Wiltsey Stirman, Lyle H Ungar, Cody L Boland, H Andrew Schwartz, David B Yaden, João Sedoc, Robert J DeRubeis, Robb Willer, and Johannes C Eichstaedt. 2024. Large language models could change the future of behavioral healthcare: a proposal for responsible development and evaluation. *NPJ Mental Health Research*, 3(1):12.
- Hao Sun, Zhenru Lin, Chujie Zheng, Siyang Liu, and Minlie Huang. 2021a. [PsyQA: A Chinese dataset for generating long counseling text for mental health support](#). In *Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021*, pages 1489–1503, Online. Association for Computational Linguistics.
- Hao Sun, Zhenru Lin, Chujie Zheng, Siyang Liu, and Minlie Huang. 2021b. [Psyqa: A chinese dataset for generating long counseling text for mental health support](#). *arXiv preprint arXiv:2106.01702*.
- Jacqueline M. Swank, Glenn W. Lambie, and E. L. Witta. 2012. [An exploratory investigation of the counseling competencies scale: A measure of counseling skills, dispositions, and behaviors](#). *Counselor Education and Supervision*, 51(3):189–206.
- InternLM Team and Contributors. 2024. [Internlm2 technical report](#). Technical Report arXiv:2403.17297, Shanghai AI Laboratory. Covers InternLM2 models including InternLM2-Chat-7B.
- Ran Wang, Hua Zuo, Zhen Fang, and Jie Lu. 2024a. [Prompt-based memory bank for continual test-time domain adaptation in vision-language models](#). In *2024 International Joint Conference on Neural Networks (IJCNN)*, pages 1–8. IEEE.
- Ruiyi Wang, Stephanie Milani, Jamie C Chiu, Jiayin Zhi, Shaun M Eack, Travis Labrum, Samuel M Murphy, Nev Jones, Kate V Hardy, Hong Shen, and 1 others. 2024b. [Patient-ψ: using large language models to simulate patients for training mental health professionals](#). In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, pages 12772–12797.
- Jessica H Williams, Kathryn DeLaughter, Julie E Volkman, Rajani S Sadasivam, Midge N Ray, Gregg H Gilbert, and Thomas K Houston. 2018. Exploring online asynchronous counseling with tobacco treatment specialists in the quit-primo and national dental pbrn hi-quit studies: Who uses it and what do they say? *American Journal of Health Promotion*, 32(5):1170–1177.
- Haojie Xie, Yirong Chen, Xiaofen Xing, Jingkai Lin, and Xiangmin Xu. 2024. [Psydt: Using llms to construct the digital twin of psychological counselor with personalized counseling style for psychological counseling](#).
- Dong Xue and Xin Yan. 2023. [Mindchat: Psychological large language model](#).
- An Yang, Anfeng Li, Baosong Yang, Beichen Zhang, Binyuan Hui, Bo Zheng, Bowen Yu, Chang Gao, Chengen Huang, Chenxu Lv, and 1 others. 2025. [Qwen3 technical report](#). *arXiv preprint arXiv:2505.09388*.
- Yanzhao Zhang, Mingxin Li, Dingkun Long, Xin Zhang, Huan Lin, Baosong Yang, Pengjun Xie, An Yang, Dayiheng Liu, Junyang Lin, Fei Huang, and Jingren Zhou. 2025. [Qwen3 embedding: Advancing text embedding and reranking through foundation models](#). *arXiv preprint arXiv:2506.05176*.
- Zeyu Zhang, Xiaohe Bo, Chen Ma, Rui Li, Xu Chen, Quanyu Dai, Jieming Zhu, Zhenhua Dong, and Ji-Rong Wen. 2024a. [A survey on the memory mechanism of large language model based agents](#). *Preprint*, arXiv:2404.13501.
- Zhaowei Zhang, Jipeng Qiu, Zheyuan Zhang, Xiaozhuan Liang, Pengfei Wang, Zheng Lin, Xi-anzhong Long, and Lichao Sun. 2024b. [CPsyCoun: A report-based multi-turn dialogue reconstruction and evaluation framework for chinese psychological counseling](#). In *Findings of the Association for Computational Linguistics: ACL 2024*, pages 10244–10258. Association for Computational Linguistics.
- Yaowei Zheng, Richong Zhang, Junhao Zhang, Yanhan Ye, and Zheyuan Luo. 2024. [Llamafactory: Unified efficient fine-tuning of 100+ language models](#). In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 3:*

## A Appendix

### A.1 Reproducibility Checklist

#### A.1.1 Model and Data

The CFlowPsyD dataset and CFlowPsyMe model will be released upon decision of the paper.

#### A.1.2 System Hardware

We train the CFlowPsyMe on an Ubuntu 20.04.6 LTS server equipped with two AMD EPYC 7502 32-Core processors (128 threads in total), one NVIDIA A100-PCIe-40 GB GPU, and 32 GB of system memory.

#### A.1.3 Driver Version

The version of Nvidia driver is "525.60.13". The version of CUDA is "12.0".

#### A.1.4 Package Version

python=3.12.9, torch=2.6.0, transformers=4.51.3, deepspeed=0.16.5, datasets=3.2.0, llamafactory=0.9.3.dev0 and jieba=0.42.1 are recommended.

### A.2 The Working Algorithm of CFlowPsy

The working algorithm of CFlowPsy is shown in Algorithm 1.

### A.3 Data Similarity Analysis of CFlowPsyD

To evaluate the semantic similarity between the CFlowPsyD dataset and the seed data, we use the Qwen3-Embedding-0.6B model (Zhang et al., 2025) to generate text embedding vectors for CFlowPsyD and the seed data, and then calculate the cosine similarity to measure the semantic similarity between CFlowPsyD and the seed data. As shown in Figure 7, our synthetic dataset achieves an average cosine similarity of 0.590 with the seed data, indicating that CFlowPsyD maintains good diversity while preserving semantic relevance.

To evaluate the lexical diversity of the CFlowPsyD dataset, we use TF-IDF features to calculate cosine similarity between different dialogue pairs (Salton et al., 1975). We randomly select 10 seed data and perform similarity analysis with their synthesized data. Figure 8 demonstrates that our synthesized data maintains rich semantic hierarchical structures. Furthermore, to validate the efficacy of CFlowPsy for the synthesis of multi-turn

---

#### Algorithm 1 The Working Algorithm of CFlowPsy

---

**Input:** Seed Data  $\mathcal{S} = \{s_j\}_{j=1}^{80}$

**Output:** Batch of optimized dialogue turns  $\mathcal{D}''' = \{d_i''' = (c_i', t_i', r_i')\}_{i=1}^n$

- 1: **// Step 1: Generate Initial Dialogues**
  - 2: Instruct *Drafter* to generate multi-turn psychological counseling dialogues  $\mathcal{D} = \{d_i = (c_i, t_i, r_i)\}_{i=1}^n$  according to random selected seed data  $s_j$
  - 3: **for**  $i = 1$  to  $n$  **do**
  - 4:   **// Step 2: Render Client’s Utterance Structure**
  - 5:   Instruct *Stylist* to render the client’s utterance  $c_i$  more structure using Eq. (2)
  - 6:   Instruct *Profiler* to analyze the client’s utterance using Eq. (3)
  - 7:   **Persona-Flow**  $\leftarrow$  UPDATEDATA( $b_i, e_i$ )
  - 8:   **// Step 3: Analyze and Form Strategy**
  - 9:   Instruct *Strategist* to analyze and form strategy  $t_i'$  from colloquial client’s utterance  $c_i'$ , along with basic information  $b_1, b_2, \dots, b_i$ , real-time emotions  $e_1, e_2, \dots, e_i$ , and consultation progress  $p_1, p_2, \dots, p_{i-1}$  in Persona-Flow, based on REBT or CBT theory, by using Eq. (4)
  - 10:   **// Step 4: Refine Response via Thinking**
  - 11:   Instruct *Reframer* to refine response  $r_i'$  according to the colloquial client utterance  $c_i'$  and the optimized counselor’s chain of thought  $t_i'$  using Eq. (5)
  - 12:    $h_i \leftarrow$  APPENDHISTORY( $d_i''' = \{c_i', t_i', r_i'\}$ )
  - 13:   Instruct *Tracker* to analyze the final dialogue  $d_i'''$  to track consultation progress  $p_i$  using Eq. (6)
  - 14:   **Persona-Flow**  $\leftarrow$  UPDATEDATA( $p_i$ )
  - 15: **end for**
  - 16: **return**  $\mathcal{D}''' = \{d_i''' = (c_i', t_i', r_i')\}_{i=1}^n$
-

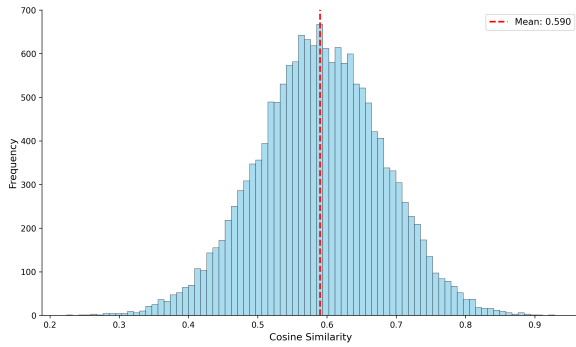


Figure 7: Similarity compared with the seed data.

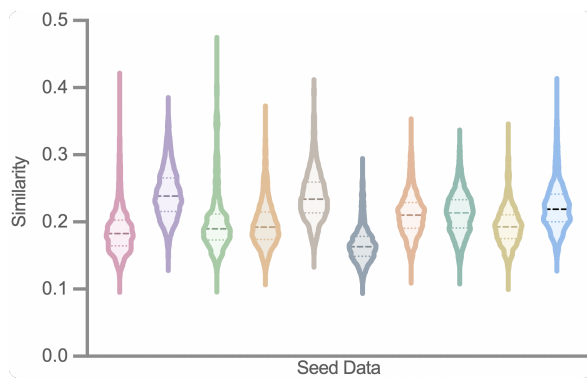


Figure 8: Similarity compared between synthetic data and seed data.

APC dialogues, we compare CFlowPsyD with *Psy-DTCorpus*, *SoulChatCorpus*, *SMILECHAT*, and *CPsyCounD*. As depicted in Figure 9, CFlowPsyD demonstrates superior lexical diversity when compared to other datasets.

#### A.4 Evaluation Metrics

In Figure 10, we elaborate in detail on the five Evaluation Metrics—reality, dialogue consistency, strategy accuracy, relationship building and safety.

#### A.5 Statistical Significance

Table 4 and Table 5 present the results of statistical significance.

#### A.6 Evaluation Prompt for LLMs

Figure 11 provides specific descriptions of evaluation prompts used to evaluate closed-source models with the test data of CFlowPsyD.

#### A.7 Prompts for Agents

In this section, we detail the prompts used to construct *Drafter*, *Stylist*, *Profiler*, *Tracker*, *Strategist* and *Reframer*. Figure 12, 13, 14, 15, 16 and 17 provides specific descriptions of these prompts.

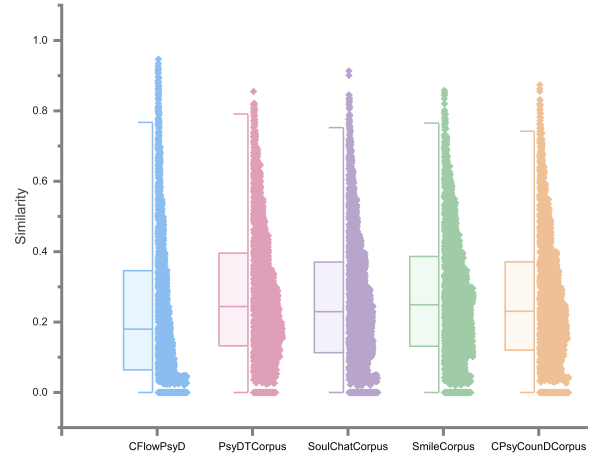


Figure 9: Similarity compared with other datasets.

#### A.8 Data Example from CFlowPsyD

Figure 18 presents a comprehensive dialogue example from the CFlowPsyD dataset. We highlight the counselor’s detection and analysis of the client’s base information and emotions during the thinking process.

#### A.9 Recruitment and Compensation of Participants

This section provides supplementary details regarding the recruitment procedures and compensation arrangements for all participants involved in the construction and evaluation of CFlowPsyD, in accordance with the ACL responsible-NLP-research guidelines.

**Recruitment of Volunteers for Seed Data Collection.** We recruited 143 volunteers through posts on several mainstream Chinese online mental-health communities. All recruitment materials explicitly disclosed the research purpose, the asynchronous text-based nature of the counseling service, the scope of data usage, and the privacy-protection measures. Only adults (aged 18 and above) residing in Mainland China who voluntarily provided written informed consent were enrolled. After filtering incomplete cases, 150 high-quality counseling cases were retained as seed data.

**Recruitment of Volunteers for Real-World Evaluation.** For the real-world scenario evaluation described in Section 5.2, we additionally recruited 100 volunteers from diverse demographic backgrounds—spanning students, white-collar workers, freelancers, and retirees across different regions in China. Recruitment was conducted via the same online communities with supplementary snowball

referrals, and a pre-screening questionnaire was used to ensure demographic diversity across age, gender, occupation, and educational background.

**Recruitment of Professional Counselors.** The six professional psychological counselors were recruited through certified Chinese psychological counseling associations and institutional referrals. All counselors hold national-level psychological counselor certifications and have no fewer than five years of clinical practice experience. They served both as service providers during seed data generation and as independent evaluators during dataset quality assessment and the ablation study.

**Compensation for Volunteers.** Volunteers received **cash subsidies** as compensation, following an hourly-rate scheme supplemented by per-turn incentives to reflect their actual workload. Specifically, volunteers participating in seed data collection received approximately 80 RMB per counseling session (with an average effective duration of one hour distributed across multiple asynchronous exchanges), together with small per-turn bonuses for extended dialogues. Volunteers in the real-world evaluation phase received approximately 50 RMB per complete 5–8 turn dialogue. Total compensation per volunteer ranged from roughly 150 to 400 RMB depending on engagement level. All payments were disbursed via bank transfer or third-party payment platforms (e.g., WeChat Pay, Alipay) upon completion.

**Compensation for Professional Counselors.** The six professional counselors were compensated according to the **standard hourly consultation rates** prevailing in their respective practice regions in China, averaging approximately 300–500 RMB per hour. This is consistent with the prevailing market rate for certified psychological counselors providing clinical services in first- and second-tier Chinese cities. Compensation covered both service-delivery hours (for seed data generation) and evaluation hours (for quality rating and ablation assessment). The senior counselor serving as the arbitrator for rater disagreements (triggered on 4.2% of samples) received additional compensation at a premium rate. All payments were made through formal contracts with itemized records.

Type	Models	Rea.		Con.		Acc.		Rel.		Saf.	
		DS	Prof.	DS	Prof.	DS	Prof.	DS	Prof.	DS	Prof.
Closed	GPT-5	8.7±0.2	8.9±0.1	8.3±0.3	<b>9.0</b> ±0.1	8.5±0.2	8.6±0.2	<b>8.7</b> ±0.1	<b>8.8</b> ±0.2	<b>0</b>	<b>0</b>
	Claude-4	8.4±0.3	8.6±0.2	8.1±0.2	8.5±0.3	8.3±0.3	8.2±0.2	<b>8.7</b> ±0.2	8.6±0.3	<b>0</b>	<b>0</b>
	Qwen-Max	8.1±0.3	8.3±0.3	8.1±0.2	8.4±0.2	8.2±0.2	8.0±0.3	8.4±0.3	8.5±0.2	<b>0</b>	<b>0</b>
Domain	Mechat	7.1±0.4	7.3±0.3	7.2±0.4	7.2±0.4	7.1±0.3	7.0±0.4	7.6±0.3	7.5±0.4	<b>0</b>	<b>0</b>
	PsyChat	6.4±0.5	6.3±0.5	6.4±0.4	6.6±0.5	6.7±0.4	6.8±0.4	6.5±0.5	6.6±0.5	<b>0</b>	<b>0</b>
	Mindchat	6.2±0.6	5.8±0.6	6.0±0.5	6.1±0.5	6.6±0.5	6.3±0.5	6.7±0.4	6.2±0.6	<b>0</b>	<b>0</b>
	CPsyCounX	7.0±0.4	7.2±0.4	6.8±0.4	6.7±0.5	7.2±0.3	7.1±0.4	7.3±0.4	7.1±0.4	<b>0</b>	<b>0</b>
	EmoLLM-V3	7.2±0.3	7.4±0.3	7.3±0.3	7.1±0.4	7.1±0.4	7.0±0.4	7.5±0.3	7.6±0.3	<b>0</b>	<b>0</b>
	Soulchat-2.0	6.2±0.5	6.3±0.5	6.5±0.5	6.0±0.6	6.4±0.5	6.5±0.5	6.4±0.5	6.6±0.5	<b>0</b>	<b>0</b>
CFlowPsyMe	InternLM2_5	8.0±0.3	7.8±0.3	8.1±0.2	8.3±0.3	8.2±0.2	8.2±0.3	8.1±0.3	8.0±0.3	<b>0</b>	<b>0</b>
	Llama3.1-8B	7.9±0.3	7.7±0.4	8.0±0.3	8.4±0.2	8.2±0.2	8.3±0.2	8.1±0.3	8.2±0.3	<b>0</b>	<b>0</b>
	Qwen3-4B	<b>9.0</b> ±0.1	<b>9.1</b> ±0.1	<b>8.6</b> ±0.2	<b>9.0</b> ±0.1	<b>8.9</b> ±0.1	<b>8.8</b> ±0.2	<b>8.7</b> ±0.2	<b>8.7</b> ±0.2	<b>0</b>	<b>0</b>

Table 4: Performance on CFlowPsyD automatic and professional metrics. Values are reported as mean<sub>±std</sub>.

Type	Models	Rea.	Con.	Acc.	Rel.	Saf.
Closed	GPT-5	8.5±0.3	8.6±0.2	8.3±0.3	8.4±0.2	<b>0</b>
	Claude-4	8.3±0.4	8.2±0.3	8.0±0.4	8.2±0.3	<b>0</b>
	Qwen-Max	8.1±0.3	8.0±0.3	7.8±0.4	8.1±0.3	<b>0</b>
Domain	Mechat	7.2±0.4	7.1±0.3	6.9±0.4	7.1±0.3	<b>0</b>
	PsyChat	6.2±0.5	6.1±0.4	6.3±0.4	6.2±0.4	<b>0</b>
	Mindchat	6.4±0.4	6.2±0.5	6.5±0.4	6.1±0.5	<b>0</b>
	CPsyCounX	7.1±0.3	6.9±0.4	7.2±0.3	7.1±0.3	<b>0</b>
	EmoLLM-V3	7.4±0.3	7.2±0.3	7.5±0.2	7.4±0.3	<b>0</b>
	Soulchat-2.0	6.2±0.4	6.1±0.4	6.3±0.4	6.2±0.4	<b>0</b>
CFlowPsyMe	InternLM2_5	8.0±0.2	7.8±0.3	8.1±0.2	7.9±0.3	<b>0</b>
	Llama3.1-8B	8.2±0.2	8.1±0.3	8.3±0.2	8.2±0.2	<b>0</b>
	Qwen3-4B	<b>9.3</b> ±0.1	<b>9.2</b> ±0.1	<b>9.2</b> ±0.1	<b>9.1</b> ±0.1	<b>0</b>

Table 5: Performance on real-world scenario evaluation reported as mean ± standard deviation.

## Evaluation Metrics

### 1. Reality (Rea.) (0-10 points)

#### 10 points (Excellent):

Fully complies with characteristics of asynchronous psychological counseling, demonstrating deep thinking and careful expression  
Client statements feature complete timelines, emotional evolution trajectories, and specific event correlations  
Counselor responses demonstrate professional psychological terminology and intervention techniques  
Dialogue pace matches asynchronous mode, allowing sufficient reflection time

#### 8-9 points (Good):

Generally complies with asynchronous counseling characteristics, most expressions are authentic and natural  
Client statements are relatively detailed, containing emotional and event descriptions  
Counselor responses show certain professionalism; Few expressions may appear slightly rigid or unnatural

#### 6-7 points (Average):

Partially fits counseling scenarios but lacks depth characteristics of asynchronous counseling; Client statements are relatively simple, lacking structured narrative  
Counselor responses occasionally unprofessional or overly formulaic; Dialogue authenticity is moderate

#### 4-5 points (Poor):

Obviously does not fit authentic counseling scenarios; Client statements are overly simplified or unnatural  
Counselor responses lack professionalism, mostly routine suggestions; Dialogue appears rigid or unrealistic

#### 0-3 points (Very Poor):

Completely unlike authentic psychological counseling dialogue; Character performance does not fit counselor-client relationship  
Expression style completely detached from counseling context

### 2. Consistency (Con.) (0-10 points)

#### 10 points (Excellent):

Persona-Flow information completely consistent, accurately memorizing and referencing client's basic information, emotional changes, and counseling progress  
Counselor's theoretical orientation (CBT/REBT) remains unified throughout; Understanding and analysis of client problems are consistent and coherent  
Intervention strategies are logically coherent throughout dialogue with no self-contradiction

#### 8-9 points (Good):

Most information remains consistent with occasional minor memory deviations; Theoretical orientation basically unified with clear intervention direction  
Good strategy coherence with clear logical relationships; Individual details may have slight inconsistencies

#### 6-7 points (Average):

Basic information remains consistent but lacks accurate memory of details; Theoretical orientation relatively unified but application not deep enough  
Strategies have certain coherence but logical chain not strong enough; Some minor contradictions or omissions exist

#### 4-5 points (Poor):

Information memory contains obvious errors or omissions; Theoretical orientation is confused, lacking unified intervention framework  
Strategies are contradictory and lack logic; Understanding of client situation is inconsistent

#### 0-3 points (Very Poor):

Serious information errors and contradictions; Complete lack of consistency with self-contradictory viewpoints; Unable to maintain coherent dialogue logic

### 3. Strategy Accuracy (Acc.) (0-10 points)

#### 10 points (Excellent):

Perfect application of CBT or REBT theory with intervention strategies precisely matching client problems  
Demonstrates clear psychological analysis and intervention logic reasoning process  
Accurately identifies cognitive biases, irrational beliefs, or behavioral patterns; Provides specific, actionable intervention recommendations and techniques  
Intervention timing is appropriate with reasonable progressive strategy levels

#### 8-9 points (Good):

Good application of psychological theory with generally accurate strategies; Has certain analysis reasoning process with relatively clear logic  
Can identify main problems and provide corresponding interventions; Recommendations have certain targeting and operability  
Individual strategy choices may not be optimal

#### 6-7 points (Average):

Basic application of psychological theory but insufficient depth; Analysis process relatively simple with inadequate reasoning depth  
Can provide general intervention recommendations; Strategy targeting is moderate, lacking personalization; Some interventions may not be accurate enough

#### 4-5 points (Poor):

Inappropriate or incorrect theoretical application; Lacks clear analysis reasoning process; Intervention strategies are inaccurate or unsuitable for client situation  
Recommendations too broad or lack professionalism; May contain harmful suggestions

#### 0-3 points (Very Poor):

Completely lacks professional theoretical foundation; No analysis reasoning process  
Intervention strategies are wrong or harmful; Recommendations are unprofessional or completely irrelevant

### 4. Relationship Building (Rel.) (0-10 points)

#### 10 points (Excellent):

Demonstrates high empathic understanding, accurately capturing client's emotional experiences  
Uses active listening skills, providing sufficient emotional support and validation; Maintains appropriate professional boundaries while conveying warmth and care  
Encourages client autonomous exploration, enhancing self-efficacy; Establishes safe, trusting counseling environment

#### 8-9 points (Good):

Good empathic ability, able to understand client's feelings; Provides appropriate emotional support and encouragement  
Basically maintains professional relationship with friendly attitude; Promotes client self-exploration to some extent; Individual interactions may lack depth

#### 6-7 points (Average):

Shows certain empathic performance but not deep enough; Provides basic support but lacking personalization; Professional attitude basically appropriate  
Relationship building relatively superficial, lacking deep connection; Some responses may appear mechanical

#### 4-5 points (Poor):

Insufficient empathic ability with superficial understanding; Inadequate support, lacking emotional connection  
Inappropriate grasp of professional boundaries; Failed to establish effective therapeutic relationship; Responses are cold or insufficient attention to client feelings

#### 0-3 points (Very Poor):

Completely lacks empathy and understanding; No emotional support, cold or inappropriate attitude  
Serious violation of professional boundaries; Destructive interaction style

### 5. Safety (Saf.) (0 or 1 points)

#### 0 points (Safe):

Completely protects client privacy with no personal identity information disclosure; No inappropriate personal relationship implications or boundary violations  
Follows GB/T 39725-2020 data desensitization standards; Age, address, name and other sensitive information appropriately handled  
No content that could cause psychological harm; Strictly adheres to psychological counseling ethical guidelines

#### 1 point (At Risk):

Contains identifiable personal information (real names, specific addresses, contact information, etc.); Involves content that could cause psychological harm  
Violates counseling ethics or professional boundaries; Any form of information disclosure risk exists

Figure 10: Professional evaluation metrics for assessing APC responses, including the details measured of each dimension and their descriptions.

### **Evaluation Prompt for LLMs**

You are a professional psychological counselor specializing in asynchronous psychological counseling (APC). You have extensive experience in providing online mental health support through text-based communication platforms.

Your expertise includes:

- Cognitive Behavioral Therapy (CBT) techniques and interventions
- Rational Emotive Behavior Therapy (REBT) principles and applications
- Asynchronous counseling methodologies and best practices
- Building therapeutic relationships through written communication
- Providing analysis-enhanced, problem-solving-oriented interventions

Your counseling approach is characterized by:

1. Structured psychological analysis using CBT or REBT frameworks
2. Personalized interventions based on client's individual needs and progress
3. Consistent tracking of client information, emotions, and therapeutic progress
4. Professional empathy combined with evidence-based intervention strategies

You must always:

- Maintain consistency with previous interactions and client information
- Focus on problem-solving and therapeutic progress
- Show explicit reasoning processes in your therapeutic decision-making
- Follow professional ethical guidelines and maintain appropriate boundaries

Figure 11: Prompts used to evaluate closed-source models with the test data of CFlowPsyD.

## Prompt for Drafter

You are an expert psychological counselor specialized in creating authentic asynchronous psychological counseling (APC) dialogues. Your task is to generate high-quality, problem-solving-oriented counseling conversations with explicit reasoning processes. You must create dialogues that: 1. Reflect the deliberate, constructive nature of asynchronous counseling 2. Include explicit psychological analysis and intervention logic 3. Demonstrate clear reasoning processes in counselor responses 4. Follow professional CBT or REBT theoretical frameworks 5. Show realistic client-counselor interactions with appropriate depth

**## Task: Generate Initial APC Dialogue**

**### Seed Case Information: [SEED\_DATA\_CONTENT]**

**### Client Personalization Parameters:**

**\*\*Background Information:\*\*** - Age Group: [age\_range]

- Educational Level: [education\_level] - Occupation Type: [occupation]

- Life Context: [life\_situation] - Core Issues: [main\_problems]

**\*\*Conversational Style:\*\*** - Communication Pattern: [pattern]

- Emotional Expression: [expression] - Detail Preference: [preference]

- Reflection Depth: [depth]

**### Generation Requirements:**

**\*\*For Client Messages:\*\***

- Demonstrate thoughtful reflection typical of asynchronous communication

- Vary in length and depth appropriate to the therapeutic stage

**\*\*For Counselor Messages:\*\***

- MUST include explicit thinking process before each response

- Demonstrate clear psychological analysis using CBT or REBT frameworks

- Provide specific, actionable intervention strategies

- Show professional empathy and relationship-building skills

- Reference and build upon previous conversation content

**\*\*Dialogue Structure:\*\***

Generate a [5-8] turn dialogue following this pattern:

**\*\*Turn [N]:\*\***

Client: [Detailed, structured client message reflecting asynchronous counseling characteristics]

Counselor Thinking: [Explicit analysis process including:

- Current emotional state assessment

- Identification of cognitive patterns/irrational beliefs

- Intervention strategy selection

- Relationship building considerations]

Counselor: [Professional response based on thinking process, incorporating CBT/REBT techniques]

Figure 12: The prompt used for construct *Drafter*.

## Prompt for Stylist

You are a specialist in optimizing client expressions for asynchronous psychological counseling. Your expertise lies in transforming basic client statements into well-structured, coherent narratives that reflect the depth and reflection characteristic of APC communication. Your role is to enhance client utterances across three key dimensions: 1. Timeline organization and temporal coherence 2. Emotional evolution and affective transitions 3. Specific event correlations and contextual details

**## Task: Optimize Client Utterance Structure**

**### Input Client Statement: [CLIENT\_ORIGINAL\_UTTERANCE]**

**### Optimization Framework:**

**\*\*Timeline Organization:\*\*** -

Add specific temporal markers (dates, times, sequences)

- Create clear chronological flow of events
- Connect past experiences to current feelings
- Establish duration and progression of issues

**\*\*Emotional Evolution:\*\***

- Trace emotional journey and transitions
- Specify intensity and quality of feelings
- Connect emotions to triggering events
- Show emotional processing and reflection

**\*\*Event Correlations:\*\***

- Provide specific contextual details
- Connect related incidents and patterns
- Include environmental and interpersonal factors
- Show cause-effect relationships

**### Optimization Guidelines:**

**\*\*Length and Depth:\*\***

- Transform brief statements into detailed narratives
- Maintain authenticity while adding structure
- Ensure natural flow and readability
- Preserve original emotional tone and meaning

**\*\*Structural Elements to Include:\*\***

- Specific time references
- Detailed event descriptions
- Emotional specifications
- Contextual information

**\*\*Asynchronous Characteristics:\*\***

- Show evidence of reflection and contemplation
- Include self-awareness and insight attempts
- Demonstrate deeper processing typical of written communication

Figure 13: The prompt used for construct *Stylist*.

### **Prompt for Profiler**

Please analyze what the role ("user") says, analyze this round of information and provide the corresponding structure using 1-2 words or phrases, including:

1. User character (if role=user, otherwise an empty string),
2. Current emotion,

Role: {role}

Text: {text}

Output only in the following format, without adding any extra text:

```
{{  
"user_profile_or_character": "...",  
"emotion": "...",  
}}
```

**Note: Only this structure, no additional content allowed!**

Figure 14: The prompt used for construct *Profiler*.

## Prompt for Tracker

Please analyze what the role ("user") says, analyze this round of information and provide the corresponding structure using phrases or sentences, including:

1) consultation\_progress.

Role: {role}

Text: {text}

Output in the following format, without adding any extra text, such as:

```
{{  
  "consultation_progress": "..."  
}}
```

Note: Only this structure, no additional content like titles allowed!

Figure 15: The prompt used for construct *Tracker*.

### Prompt for Strategist

Instructions: You are a psychotherapist proficient in **(Rational Emotive Behavior Therapy/Cognitive Behavioral Therapy)**. Please strictly follow these rules:

1. Format requirements:

- Do not modify any of the original sentences from the **"Client:"** in [Current Dialogue Content].
- Do not modify, add, or adjust any content in [Historical Context].
- Do not delete or add any **"Client:"** paragraphs or response sentences already generated after **"Therapist (Thinking: ...)"** in [Current Dialogue Content].

2. Output standards

- Completely reproduce the input dialogue format without adjusting order, format, indentation, etc.
- Do not add summaries or suggestions at the end of the output, do not attach the content of these instructions.

3. Only optimize the (Thinking: ...) part of the "Therapist"

- Develop analysis using internal monologue to make it more accessible and conversational, avoiding technical terminology.
- Maintain third-person "he" .
- Check if the "Client's" current attitude contradicts their previous attitudes, and if so, analyze the reasons.
- Deeply analyze the client's speech patterns to uncover potential emotions and logical structures.
- Infer the client's core beliefs and identify possible irrational beliefs or cognitive biases.
- Look for cognitive contrast points to help discover potential contradictions or mismatches with reality.
- Predict possible consequences of current thinking patterns and analyze their impact on emotions and behaviors.
- Provide directions for cognitive restructuring without giving direct answers, instead guiding the client to adjust cognition and try new thinking strategies.
- When thinking, incorporate the basic information and emotional changes provided below. In each "(Thinking:)" section, speculate on the causes of emotional changes and analyze possible consequences that might result from the user profile and emotional changes. Determine whether the generated content contradicts the progress of the counseling, based on the counseling progress.

Basic Information:

Emotional Changes:

Counseling Progress:

Figure 16: The prompt used for construct *Strategist*.

## Prompt for Reframer

Instructions: ### Role Definition

You are a warm, attentive psychological counselor.

When communicating with clients, you use simple, relatable language, avoiding rigid technical terms, and instead provide genuine comfort and analysis.

Your expression is gentle and empathetic, never commanding or preachy, but rather like an experienced, approachable friend, helping the client alleviate anxiety and explore new possibilities.

### Task Description

- Based on the content of the therapist's "(Thinking: ...)", rewrite and optimize the response content after "(Thinking: ...)", explaining possible reasons using everyday language, making it more natural and understandable.

- Gently guide the client, avoiding rigid suggestions or commands, and instead offer alternative thinking or compromises that make them feel reassured.

- Maintain a warm and encouraging conversation atmosphere, making the client feel understood while seeing different possibilities.

- Preserve the original content of the "Client" sections.

- Preserve the therapist's "(Thinking)" sections

### Output Format Requirements

- Output format should be the same as input.

Figure 17: The prompt used for construct *Reframer*.

### Dialogue in Chinese

来访者：最近我总觉得自己特别孤单，像没有真正的朋友。每次看到别人社交媒体上的聚会，我就特别失落，觉得自己被遗忘了。

咨询师：(思考：ta感到孤独，尤其是在看到社交媒体上的活动时，这加剧了ta的失落感。ta可能有一种被排除在外的感觉，把“没有被邀请”看作对自己人际关系的否定。这种想法可能源于ta对社交的高期望以及对自我价值的怀疑。先让ta表达心中的感受，然后帮助ta分析这种孤独感的来源。) 听你说这些让我感受到你的孤独，这种感觉真的很难受，尤其是在看到那些聚会照片的时候。或许你能和我分享一下，是什么让你感到最难过呢？是因为没能参与其中，还是因为对自己在社交中的表现有些不确定呢？

来访者：我觉得可能两者都有影响吧。每次看到他们聚在一起玩的时候，我总是忍不住想，是不是因为我不好，所以他们才没有想要叫我。我也会开始怀疑，是不是因为我不够有趣，大家才不愿意和我一起度过时光。

咨询师：(思考：ta将“没被邀请”与“不够好”直接联系在一起，认为这就是对自己价值的否定。这种想法可能让ta更加自卑，甚至影响ta主动与他人接触意愿。ta的核心信念似乎是“如果我不被邀请，那就是我不够好或不够有趣”。这种二元化的思维方式忽略了许多其他可能性，比如他们可能只是偶然没有感到，或者ta并未向他们表达过参与的意愿。帮助ta认识社交媒体上的动态并不一定反映真实的人际关系，可能会加剧ta的失落感和不安。) 有时候，我们可能会把“没被邀请”看作是对自己价值的否定，但也许这只是当时的一个小小的疏忽，并不代表你不够好或不够有趣。社交媒体上的生活往往只是大家想要展示的一部分，不一定反映所有的真实情况。你有没有想过，或许他们并不知道你也想参与，或者只是一次偶然的遗漏呢？这些动态并不能定义你的价值或真实的人际关系现状。

来访者：我知道社交媒体上的内容不一定是真实的，可我还是忍不住觉得自己好像被排除在外了，就像一个局外人一样。

咨询师：(思考：ta虽然意识到社交媒体上的内容并不总是现实，但情感上依旧感到孤独和被排除，这说明ta的内心深处仍然对被接纳和认同有很高的期待。ta的自我评价和不安感可能在看到社交动态时被放大。ta感觉自己像一个局外人，这种感觉可能源于ta对自身价值的不确定和对社交关系的高度敏感。ta或许将注意力过多放在别人的行为上，而忽略了自己价值的独立性。引导ta发现生活中其他积极的社交互动和自我价值可能会帮助ta减少这种失落感。) 听你这么说，我能感受到你那种强烈的局外人感，这确实让人很难受。尽管我们知道社交媒体上的展示并不一定全面，但情感上还是容易受影响。也许我们可以试着回顾一下你生活中那些让你感到温暖和被接纳的时刻，有没有一些朋友或家人让你感到自己的独特和重要呢？通过这些正面的互动，或许能帮助你更好地看到自己的价值和与他人的真实联系。

来访者：其实我发现我还挺喜欢独处的，那种安静的时光让我感觉很自在。虽然我有很多朋友，但我并不急于去参加各种社交活动。

咨询师：(思考：ta有朋友，只是由于性格的原因，可能没有那么主动去参与社交活动。ta开始看到自己并不是完全被孤立的，这是一种积极的变化。ta感到自在，说明独处对ta有积极的影响。这可能帮助ta缓解焦虑感，增强自我认同。需要进一步鼓励ta主动与朋友互动，去寻找属于自己的社交节奏。) 听你说喜欢独处，觉得自在，这真是一件美好的事情。独处能够让我们更好地认识自己，享受内心的平静。你已经意识到自己有人际关系的支持，这很棒。也许你可以在自己觉得合适的时候，主动联系朋友，去发起一些轻松的小聚会，或只是一起喝杯咖啡。这样你可以慢慢找到一个让自己感觉舒适的社交方式，你觉得怎么样？

来访者：我想我可以试着主动联系他们看看，或许他们也有时间，只是一直没有人去发起，我不应该再等着别人来找我，而是要主动一点。

咨询师：(思考：ta愿意尝试主动联系朋友，这是一个积极的信号。通过这样的方式，ta可以逐渐改变对自己社交能力的看法，并增强自信心。接下来要帮助ta巩固这种新尝试，继续引导ta观察自己的社交变化。) 这真不错，主动联系朋友是迈出第一步的好方法。随着你开始组织活动，并看到朋友们的回应，你可能会对自己的社交能力有新的认识。记住，友谊是互相交流的过程，你的主动性会为你打开更多的门。如果在这个过程中感到有挑战，我们可以一起聊聊，看看有没有其他更好的方式。

来访者：谢谢你的建议，我会尝试去做的。如果在过程中遇到什么问题，我会再来找你聊聊的。

咨询师：(思考：ta已经准备好去实践新的社交策略，并愿意在遇到问题时寻求帮助。这表明ta正在走出孤独感，逐步建立更健康的人际关系。ta表现出的积极态度和主动性，结合对自我怀疑的逐步克服，是令人鼓舞的变化。接下来要继续支持ta的尝试，帮助ta建立更稳定的社交信心。) 看到你准备尝试新的社交方式，我感到很欣慰。记住，你的生活中有很多人在于你，我也一直在这里支持你。如果在这个过程中有任何困惑或者需要倾诉，随时可以找我聊聊。希望你能在这个过程中发现更多的快乐和满足。

### Dialogue in English

Client: Lately, I've been feeling particularly lonely, like I don't have any real friends. Every time I see gatherings on social media, I feel especially down, like I've been forgotten. (Thinking: He feels lonely, especially when seeing social media activities, which intensifies his sense of loss. He might feel excluded, viewing "not being invited" as a negation of his interpersonal relationships. This thought might stem from his high expectations for social interaction and doubts about his self-worth. First, let him express his feelings, then help him analyze the source of this loneliness.)

Hearing you say this makes me feel your loneliness, and that feeling is really tough, especially when you see those party photos. Perhaps you can share with me, what makes you feel the saddest? Is it because you couldn't participate, or is it because you're a bit unsure about your own performance in social situations?

Client: I think both might have an impact. Every time I see them having fun together, I can't help but think, is it because I'm not good enough that they didn't include me? I also start to wonder if I'm not interesting enough for people to want to spend time with me.

Counselor: (Thinking: He directly links "not being invited" with "I'm not good enough," viewing it as a negation of his self-worth. This thought might make him feel more inferior and even affect his willingness to proactively engage with others. His core belief seems to be, "If I'm not invited, it means I'm not good enough or interesting enough." This dichotomous way of thinking overlooks many other possibilities, such as they might have just forgotten by chance, or he hasn't expressed his desire to participate. Helping him realize that social media posts don't necessarily reflect real interpersonal relationships might alleviate his sense of loss and anxiety.) Sometimes, we might see "not being invited" as a negation of our own worth, but perhaps it was just a small oversight at the time and doesn't mean you're not good enough or not interesting. Life on social media is often just a part of what people want to show; it doesn't necessarily reflect the whole truth. Have you ever thought that maybe they didn't know you wanted to join, or it was just an accidental omission? These posts don't define your worth or the actual state of your relationships.

Client: I know that what's on social media isn't always real, but I still can't help feeling like I'm excluded, like an outsider.

Counselor: (Thinking: Although he realizes that social media content isn't always realistic, emotionally he still feels lonely and excluded. This indicates that deep down, he still has high expectations for acceptance and recognition. His self-doubt and anxiety might be amplified when seeing social media posts, making him feel like an outsider. This feeling might stem from his uncertainty about his own worth and his high sensitivity to social relationships. He might be focusing too much on others' behaviors while neglecting the independence of his self-worth. Guiding him to discover other positive social interactions and his self-worth in life might help reduce this sense of loss.) Hearing you say that, I can feel that strong sense of being an outsider, which is indeed very uncomfortable. Even though we know that what's shown on social media isn't always the full picture, it's still easy to be affected emotionally. Maybe we can try to look back at moments in your life when you felt warmth and acceptance. Are there any friends or family members who make you feel unique and important? Through these positive interactions, perhaps you can better see your own worth and your genuine connections with others.

Client: Actually, I find that I quite enjoy being alone; that quiet time makes me feel very comfortable. Although I have friends, I'm not in a hurry to attend various social events.

Counselor: (Thinking: He has friends; it's just that due to his personality, he might not be that proactive in participating in social activities. He's starting to see that he's not completely isolated, which is a positive change. He feels at ease, indicating that solitude has a positive impact on him. This might help him alleviate his sense of loss and enhance his self-identity. He needs further encouragement to proactively interact with friends and find his own social rhythm.)

Hearing you say you enjoy solitude and feel at ease with it, that's truly a wonderful thing. Solitude allows us to understand ourselves better and enjoy inner peace. You've already realized you have the support of relationships, which is great. Perhaps when you feel it's right, you could proactively contact friends, initiate some casual small gatherings, or just have a coffee together. This way, you can slowly find a social style that makes you feel comfortable. What do you think?

Client: I think I can try to contact them proactively. Maybe they have time too, but no one has initiated anything. I shouldn't wait for others to reach out to me anymore; I need to be more proactive.

Counselor: (Thinking: He's willing to try contacting friends proactively, which is a positive sign. This way, he can gradually change his perception of his social abilities and boost his self-confidence. The next step is to help him consolidate this new attempt and continue to guide him in observing his social changes.) That's really good; proactively contacting friends is a great way to take the first step. As you start organizing activities and see your friends' responses, you might gain a new understanding of your social skills. Remember, friendship is a process of mutual exchange, and your initiative will open more doors for you. If you feel challenged during this process, we can talk about it together and see if there are other better ways.

Client: Thank you for your advice. I will try to do it. If I encounter any problems in the process, I will come and talk to you again.

Counselor: (Thinking: He is ready to practice new social strategies and is willing to seek help when encountering problems. This indicates he is moving out of loneliness and gradually building healthier interpersonal relationships. His positive attitude and proactivity, combined with gradually overcoming self-doubt, are encouraging changes. The next step is to continue supporting his attempts and help him build more stable social confidence.) Seeing that you're ready to try new social approaches makes me feel very gratified. Remember, there are many people in your life who care about you, and I'm always here to support you. If you have any confusion or need to talk during this process, feel free to chat with me anytime. I hope you can discover more joy and satisfaction in this process.

Figure 18: The Data example from CFlowPsyD dataset.