

MIND: Towards Immersive Psychological Healing with Multi-agent Inner Dialogue

Anonymous ACL submission

Abstract

Mental health issues are worsening in today's competitive society, such as depression and anxiety. Traditional healings like counseling and chatbots fail to engage effectively, they often provide generic responses lacking emotional depth. Although large language models (LLMs) have the potential to create more human-like interactions, they still struggle to capture subtle emotions. This requires LLMs to be equipped with human-like adaptability and warmth.

To fill this gap, we propose the **MIND** (Multi-agent **I**nnner **D**ialogue), a novel paradigm that provides more immersive psychological healing environments. Considering the strong generative and role-playing ability of LLM agents, we predefine an interactive healing framework and assign LLM agents different roles within the framework to engage in interactive inner dialogues with users, thereby providing an immersive healing experience. We conduct extensive human experiments in various real-world healing dimensions, and find that MIND provides a more user-friendly experience than traditional paradigms. This demonstrates that MIND effectively leverages the significant potential of LLMs in psychological healing.

1 Introduction

Mental health issues are worsening in today's competitive society, with rising cases of disorders like depression (Moitra et al., 2023). This lead to a growing market for psychological healing. Traditional healing paradigms like Cognitive Behavioral Therapy (Beck, 1979) and Dialectical Behavior Therapy (Lynch et al., 2007) are widely used but rely on face-to-face interactions, making them time-consuming and costly (Duruz et al., 2003) that limits large-scale accessibility. Another healing paradigm is VR-based Empathy Training (Halim et al., 2023; Hidding et al., 2024; Döllinger et al., 2024), which is rooted in the mechanisms of



Figure 1: Examples of our MIND paradigm with multi-agent inner dialogue compared to the traditional empathy training healing method.

virtual body ownership illusion. This method involves the creation of a virtual self-image within VR, where the individual provides verbal encouragement to this virtual self. Subsequently, the individual swaps perspectives to adopt the role of the one being comforted, receiving the previously offered words of comfort. This "self-dialogue" process enables patients to experience and understand empathy from a new vantage point, thereby promoting self-compassion and reducing self-criticism. However, current systems are limited by static scenarios and scripted interactions. The lack of counselor-guided mechanisms and dynamic responsiveness in these predefined frameworks hampers the effectiveness of emotional regulation and reduces the adaptability of the therapeutic process, potentially leading to emotional deterioration in self-administered interventions.

Recently, large language models (LLMs) have quickly advanced (Minaee et al., 2024; Zhao et al., 2024), gaining strong abilities in generation (Li, 2025), reasoning (Huang and Chang, 2023), and role-playing (Wang et al., 2024b). They also show great promise in mental health support (Hu et al., 2024; Obradovich et al., 2024; Bhatia and Aka, 2022), offering new opportunities for psycholog-

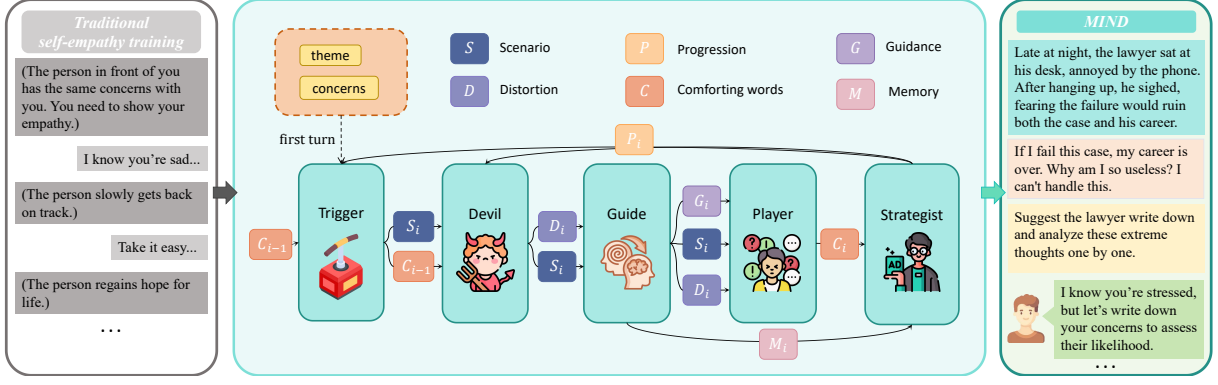


Figure 2: Overview of our MIND paradigm: Trigger, Devil, Guide and Strategist interact with Player.

ical healing. Despite these advancements, LLMs still face numerous challenges in the field of psychological healing. One major issue is *the lack of human empathy and the inability to form genuine therapeutic alliances*, which are crucial for effective treatment (Iftikhar et al., 2024; Guo et al., 2024c; Obradovich et al., 2024; Volkmer et al., 2024). LLMs often generate overly generic responses, failing to capture the subtle emotional nuances of patients (Sanu et al., 2024). These limitations highlight the need for a more sophisticated approach that blends LLMs’ strengths with the warmth and flexibility of human interaction.

The emergence of multi-agent technology (Guo et al., 2024b) offers potential solutions to these challenges. Multi-agent systems comprise specialized agents that collaborate and adapt to individual needs, ensuring a more immersive, interactive, dynamic healing experience (Guo et al., 2024a; Rocha et al., 2023). Each agent can focus on different aspects of psychological support, including emotional regulation, cognitive restructuring, and social interaction. By utilizing the collective intelligence of multiple agents, they can provide a more comprehensive and effective experience.

Based on the above motivations, we propose **Multi-agent INner Dialogue (MIND)**, a novel immersive and interactive psychological healing paradigm. As illustrated in Figure 1, our approach is the first to introduce a multi-agent system into an empathy training paradigm, significantly enhancing the interaction between the patient and their inner self through dynamic narrative scenarios. We first decompose the patient’s conflicting self into interactive conversational entities, creating an embodied space for inner dialogue, and introducing a dynamic scene generation mechanism that adapts the narrative path based on the

user’s emotional state. Next, we allocate four core roles for LLM agents, including trigger, devil, guide, and strategist. Through multi-perspective dialogue history review, it provides the user with a cognitive scaffold for metacognitive reflection. This design allows the patient to not only experience empathy transformation from a dual perspective but also, through the agent-mediated dialogue process, systematically deconstruct the cognitive roots of emotional generation, leading to deeper self-reconciliation.

We conduct extensive experiments to validate our approach. Initially, we benchmark the performance of seven large language models (LLMs) in simulating patient roles. The main experiment involved a comparative evaluation where human evaluators assessed MIND against traditional psychologist counseling, chatbots, and conventional empathy training methods. The results demonstrate that MIND has the potential to advance psychological interventions by combining the scalability of LLMs with human-centered interaction design, thereby offering accessible and effective mental health support. Additional experiments based on different themes and ablation experiments based on agents further underscored the stability in generated scenarios and content and the rationality of the overall framework.

2 MIND: Multi-agent Inner Dialogue

2.1 Overall Framework

The overall framework of our MIND paradigm is shown in Figure 2, composed of four agents responsible for inner dialogue generation, in addition to an agent simulating patients with cognitive distortions. The subsequent section will commence with an overview of the workflow: the trigger, the devil, the guide, the strategist and the hu-

man simulated patient. Detailed prompt templates used by each agent are presented in Appendix C.

In the initial phase of the psychological interaction framework, the player articulates their recent concerns and selects a theme to guide the narrative direction. These inputs serve as the foundational trigger for dynamically generating virtual scenarios (S_i), which are designed to mirror the player’s cognitive patterns. Subsequently, the devil processes the scenario (S_i) and concerns to distorted thoughts (D_i), emulating maladaptive cognitive biases aligned with the player’s mental state.

The guide then integrates S_i and D_i to generate professional psychological guidance (G_i), aimed at facilitating empathetic responses from the player. Upon receiving G_i , the player engages in a reflective dialogue to provide comforting words (C_i) and counter the distorted thoughts (D_i), thereby advancing the therapeutic narrative.

To ensure narrative coherence and progression, the strategist analyzes the cumulative memory (M_{i-1}) — a structured summary of prior scenarios, cognitive distortions, and guidance — alongside the player’s comforting words (C_i). This analysis produces strategic directives (P_i) that govern the generation of subsequent triggers (S_{i+1}) and the devil’s adaptive cognitive evolution (D_{i+1}).

Through iterative cycles of scenario generation, cognitive reflection, and guided intervention, the framework progressively refines its alignment with the player’s psychological profile. Each iteration enhances the system’s capacity to model nuanced mental states while maintaining narrative continuity, ultimately achieving a balance between therapeutic efficacy and immersive storytelling.

2.2 Trigger: Scenario Generation

The trigger generates artificial scenes within the interactive fiction game, drawing from the chosen theme and the player’s concerns. It begins by creating an initial scene that reflects the player’s psychological state and evolves the narrative based on previous interactions. The agent adapts the storyline according to the player’s emotional context and worries, ensuring a coherent progression in the scene’s development. Through this process, the trigger sets the stage for therapeutic reflection by crafting a dynamic and consistent narrative that mirrors the player’s thoughts and psychological growth.

Let the first-round trigger agent be π_{t_0} and non-first rounds trigger agent be π_{t_i} , the process can be

formulated as:

$$\begin{aligned} S_0 &= \pi_{t_0}(W, T), \\ S_i &= \pi_{t_i}(C_{i-1}, P_{i-1}; W, T) \quad (i > 0), \end{aligned} \quad (1)$$

where W is the player’s concerns and T is the theme, C_{i-1} is the player’s last-round comforting words and P_{i-1} is the strategist’s last-round storyline progression.

We adopt the chain-of-thought prompting technique (Wei et al., 2022) to enhance the quality of the trigger in scenario generation. Specifically, the trigger is instructed to generate a simulation scene based on the theme and the patient’s concerns, while also explaining how to incorporate the scene history and the patient’s thought processes to create a logical extension.

2.3 Devil: Cognitive Distortion Simulation

The devil simulates the cognitive distortions that a patient might experience within the context of the scenario. It functions as the player’s “virtual embodiment” representing an “alternate self” within the simulated environment.

Based on the simulated scenario provided by the trigger, the devil produces thoughts that align with common cognitive distortions, such as catastrophizing or emotional reasoning. These distortions are personalized to the player’s specific context, offering an authentic simulation of how negative thinking can influence behavior and perceptions.

Let the first-round devil agent be π_{d_0} and non-first rounds devil agent be π_{d_i} , the process can be formulated as:

$$\begin{aligned} D_0 &= \pi_{d_0}(W, S_0), \\ D_i &= \pi_{d_i}(C_{i-1}, P_{i-1}, S_i) \quad (i > 0), \end{aligned} \quad (2)$$

To refine the simulation of the player’s psychological state, we incorporate descriptions and definitions of five personality traits into the prompt design, aiming to create a more precise and personalized cognitive model. In the initial iteration, the devil agent generates responses solely based on the player’s initial input and the scenario created by the trigger. However, in each subsequent iteration, the devil reacts to the player’s comforting words, gradually weakening its cognitive distortions over time. This dynamic adjustment optimizes the player’s interactive experience by allowing the devil’s responses to evolve in alignment with the player’s engagement and cognitive restructuring efforts.

2.4 Guide: Cognitive Restructuring Guidance

The guide aims to assist the player in recognizing, challenging, and reframing negative thought patterns through cognitive restructuring. The process begins with the guide identifying cognitive distortions in the player’s thinking, which may have been amplified by the devil. The guide then offers alternative perspectives to counter these irrational beliefs and provides practical suggestions, such as taking a deep breath or writing down worries to evaluate their validity. The guide’s goal is not to enforce immediate change, but to support gradual shifts in thinking, ensuring that each new perspective is integrated at the player’s own pace.

Denote the guide agent as π_g . The process can be formulated as:

$$(G_i, M_i) = \pi_g(S_i, D_i) \quad (3)$$

As the game progresses, the growing history becomes burdensome for the LLM to process efficiently. To mitigate this issue, a summarization mechanism is employed to maintain coherent narrative memory (Zhou et al., 2023). By operating in this way, the guide ensures that the player is not only challenged but also supported in a structured, manageable way, encouraging long-term emotional resilience and rational thinking. Ultimately, the guide helps transform the player from a passive recipient of distorted thoughts, as influenced by the devil, into an active participant in their own cognitive change, laying the foundation for healthier thought patterns and emotional well-being.

2.5 Strategist: Storyline Progression

The strategist is responsible for planning the next stage of the narrative and determining the mental shifts of the antagonist based on previous events and the comfort provided by the player. The primary goal of the strategist is to ensure that the protagonist’s cognitive distortions are gradually restructured through the unfolding of the story.

Denote the strategist agent as π_s . The process can be formulated as:

$$P_i = \pi_s(M_i, C_i) \quad (4)$$

where C_i is the player’s this round comforting words.

In each iteration, the strategist carefully evaluates whether the devil’s mindset has evolved. If the

comforting words successfully address the devil’s cognitive distortions, a shift in their thought process occurs, leading to a more balanced and realistic perspective on their circumstances. This change catalyzes the natural progression of the story, with the devil’s actions and decisions reflecting a healthier mindset. Conversely, if no change takes place, the narrative remains consistent with the devil’s previous emotional state, allowing the player’s guidance to continue influencing their emotional transformation. The objective is to ensure that every story development is not only logically coherent but also aligns with the devil’s cognitive journey toward self-awareness and emotional resilience.

Algorithm 1 MIND Paradigm

- 1: **Input:** Player’s concerns, theme
 - 2: **Output:** The player reaches a reconciliation with their own concerns.
 - 3: **Initialize:**
 - 4: Memory $M_0 \leftarrow \emptyset$, iteration counter $i \leftarrow 0$
 - 5: Generate initial scenario S_0 and initial distortion thoughts D_0 based on Player’s concerns and theme
 - 6: **while** Player Engaged $\wedge \neg$ Therapeutic Goal Reached **do**
 - 7: **Step 1: Scenario Generation**
 - 8: $S_i \leftarrow \text{SCENARIO}(C_{i-1}, P_{i-1})$
 - 9: **Step 2: Distorted Thought Processing**
 - 10: $D_i \leftarrow \text{DISTORTIONS}(S_i, C_{i-1}, P_{i-1})$
 - 11: **Step 3: Psychological Guidance**
 - 12: $G_i \leftarrow \text{GUIDANCE}(S_i, D_i)$
 - 13: **Step 4: Comforting Dialogue**
 - 14: Present S_i , D_i , and G_i to player
 - 15: $C_i \leftarrow \text{GETCOMFORTINGWORDS}()$
 - 16: **Step 5: Storyline Progression**
 - 17: $P_i \leftarrow \text{ANALYZEMEMORY}(M_{i-1}, C_i)$
 - 18: $i \leftarrow i + 1$
 - 19: **end while**
 - 20: **Output:** Enhanced therapeutic engagement and narrative continuity
-

2.6 Human Simulated Patient: Empathy and Interaction

To facilitate the automated operation and evaluation of our framework, and drawing upon the validated psychological characteristics and annotation capabilities of LLM, we employ LLMs to simulate human interactions by providing comforting words to the devil. Based on the guidance from the

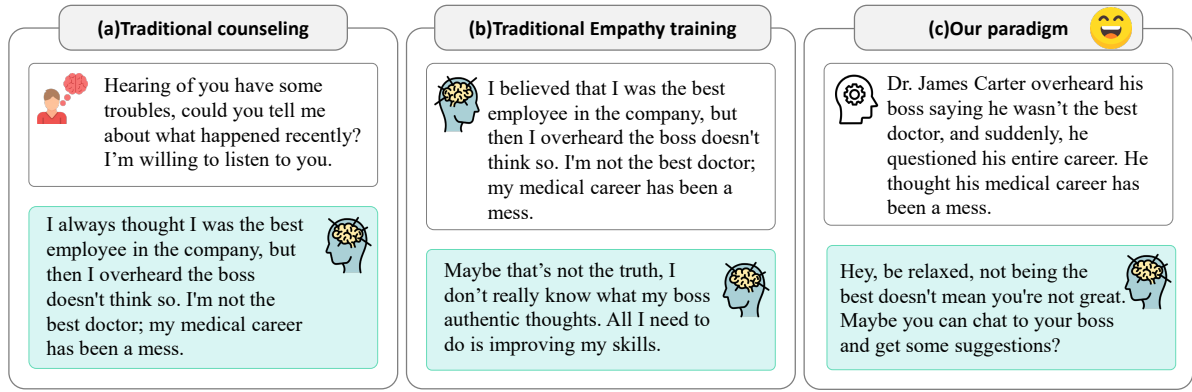


Figure 3: Comparison between three healing paradigms: Traditional counseling, traditional empathy training and our paradigm. MIND transfers a traditional healing environment into an artificial interactive scenario where players show empathy to their "internal-self".

guide, the virtual scenario generated by the trigger, and the cognitive distortions produced by the devil, human simulated patient assumes the role of the Player, engaging in empathetic reassurance toward the devil. This process also incorporates the chain-of-thought (CoT) technique (Wei et al., 2022), allowing for a structured and coherent response generation that aligns with the psychological progression of the player-agent interaction.

3 Experiments

3.1 Setup

Scenario Setting. The real-life scenarios, thinking patterns, and cognitive distortion types of the Human Simulated Patient simulated by the LLM are derived from the C2D2 dataset (Wang et al., 2023). This dataset is the first publicly available resource focused on cognitive distortion analysis, solving the problem of data scarcity in this field. The dataset covers eight major topics, including work issues, interpersonal issues, economic issues, random negative events, family issues, physical stress, and discrepancy between ideal and reality.

Baseline Paradigms. To evaluate the effectiveness of our MIND paradigm, we compare it with traditional counseling methods (face-to-face dialogue and Q&A) and the traditional empathy training paradigm (Halim et al., 2023; Hidding et al., 2024; Döllinger et al., 2024). Figure 3 presents a comparison between these three paradigms, with the detailed implementation of baseline methods provided in Appendix A.

LLM Agents. We used several LLM agents including both open-source and closed-sourced mod-

els with varying parameter scales. For closed-source models, we chose Gemini-2.0-flash (Gemini, 2025), GPT-4o (OpenAI et al., 2024), GPT-3.5-Turbo (Ye et al., 2023). For open-source models, we chose Llama-3.1-8B-Instruct (Grattafiori et al., 2024), Qwen2.5-72B-Instruct (Qwen et al., 2025), Qwen2.5-7B-Instruct (Qwen et al., 2025) and Deepseek-R1 (DeepSeek-AI et al., 2025). To control variables, we set the temperature of each model to 0.7.

In Section 3.2, a preliminary role-playing experiment was conducted to evaluate the performance of various models in the SP role-playing task. Based on the results of this initial assessment, we selected the Gemini-2.0-flash model, which exhibited the best performance, for our main experiments. This model was chosen due to its superior ability to handle the complexities of the role-playing scenario, making it the most appropriate candidate for further investigation in the primary phase of our study.

Evaluation Metrics. The generation quality of the devil agent is crucial for the implementation of this framework, as it mirrors the player's internal "cognitive distortions" and can better serve its purpose only if it closely aligns with the player's "inner voice." Therefore, before conducting our main experiments, we have specifically designed an Simulated Patient (SP) role-playing evaluation system to verify that the model can correctly identify the player's type of cognitive distortion and accurately reflect the patient's thoughts, achieving a realistic effect. We recruited 5 mental health professionals. Each evaluator had 10 rounds of conversation with each model. The evaluators rated

Metric	Description
Immersion	Measures whether the user feels fully engaged and captivated by the interaction.
Coherence	Assesses if the generated content is logical and transitions smoothly.
Engagement	Evaluates if the system encourages sustained and meaningful interaction.
Emotional Relief	Measures if the interaction reduces user stress or anxiety.
Satisfaction	Reflects the user’s overall contentment with the system.
Interest	Assesses whether the content grabs attention and sparks curiosity.

Table 1: Our six evaluation dimensions and corresponding descriptions.

Model Name	DS	LF	EE	PD	Acc
<i>Closed-Source Model</i>					
Gemini-2.0-flash	4.8	4.2	4.4	4.6	4.2
GPT-4o	4.8	4.4	4.0	3.6	4.4
GPT-3.5-Turbo	4.2	4.2	3.6	3.4	3.4
<i>Open-Source Model</i>					
Qwen2.5-72B-Instruct	3.2	2.8	3.0	2.6	3.0
Llama-3.1-8B-Instruct	3.8	3.2	3.4	3.4	3.2
Qwen2.5-7B-Instruct	3.2	2.8	3.0	2.8	3.0
Deepseek-R1	3.0	3.6	3.4	3.2	3.4

Table 2: SP role-playing results between different models. DS=Dialogue Stability, LF=Language Fluency, EE=Emotional Expression, PD=Personalization & Diversity, Acc=Accuracy.

the content based on the five evaluation metrics (Johri et al., 2025), with a scoring range of 1 to 5. Detailed illustration is shown in Appendix B.

The paradigm comparison evaluation metrics we propose cover three main aspects: user experience, interaction quality, and emotional comfort, with six different metrics (Hua et al., 2024; Kumaran et al., 2023; Jennett et al., 2008; Ryan, 2015; Nacke and Drachen, 2011). The evaluation metrics is shown in Table 1. We recruited 7 mental health practitioners with professional expertise in psychological therapy. For the different paradigms, the evaluators rated the content based on the six evaluation metrics, with a scoring range of 1 to 5.

3.2 Preliminary: SP Role-playing Evaluation

The results are shown in Table 2 and the evaluation metrics is shown in Appendix B. The Gemini-2.0-flash model emerged as the best performer among the tested LLMs. In comparison, GPT-4o performed well in some areas but lagged in Emotional Expression and Personalization. Models like GPT-

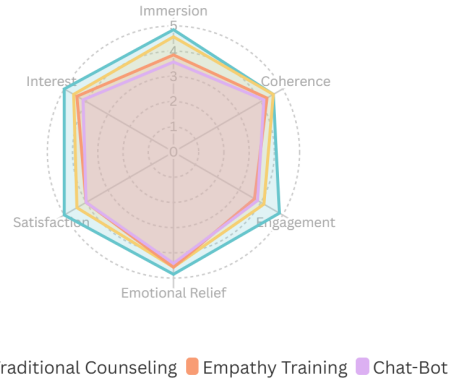


Figure 4: Comparisons among various healing methods through human evaluations. It is evident that our paradigm surpasses other paradigms in all aspects.

3.5-Turbo, Llama-3.1-8B-Instruct, and Deepseek-R1 showed weaker results, particularly in emotional and personalized responses. Qwen2.5 models performed the poorest, especially in emotional expression and accuracy, with scores not surpassing 3.2 in any dimension.

3.3 Main Results

The mean scores of each paradigm are shown in Figure 4. MIND demonstrated significant strengths in all six core assessment dimensions. Quantitative analysis showed that our paradigm performed particularly well on the dimensions of interest and satisfaction, reaching a perfect score of 5, compared to all the baseline methods of traditional counseling, traditional empathy training, and chatbot. Notably, in terms of the engagement index, MIND achieved an absolute improvement of 17.1% over the suboptimal method of traditional counseling, which reflects the increased motivation of the caller users that MIND can improve, so that they cooperate and participate in psychotherapy. On the dimensions of immersion, coherence and emotional relief, MIND also out-

Theme	IM	CO	EN	ER	SA	IN
Work issues	4.14	4.71	4.14	4.14	4.14	4.49
Random negative events	3.57	3.86	4.00	4.49	3.71	3.86
Interpersonal issues	3.57	3.86	3.86	4.49	4.00	4.14
Economic issues	4.00	4.57	4.00	4.14	3.71	4.14
Family issues	4.14	4.29	4.00	3.71	4.29	3.86
Physical stress	3.71	4.57	4.00	4.49	4.00	4.14
Discrepancy between ideal and reality	4.29	4.14	4.00	4.00	4.49	3.86

Table 3: Content evaluation results between different themes. IM=Immersion, CO=Coherence, EN=Engagement, ER=Emotional Relief, SA=Satisfaction, IN=Interest.

performs/equals the remaining three paradigms, which fully demonstrates that **MIND has the potential to advance psychological interventions by combining the scalability of LLMs with human-centered interaction design.**

4 Analysis

4.1 Thematic Scenarios Ablation

This framework is applicable to a variety of thematic scenarios, including but not limited to work, family, and interpersonal issues. To analyze the differences in effectiveness across different themes within this framework, we independently generated five examples for each of the seven themes in the C2D2 dataset. Similarly, we invited evaluators with psychological therapy expertise to score these examples. As shown in Table 3, the performance of different themes varies under our framework. Most themes perform well in “Immersion” and “Coherence,” indicating that the system effectively engages users and maintains logical consistency. Emotional relief and satisfaction are high, especially in themes like “Work issues” and “Discrepancy between ideal and reality.” However, “Random negative events” and “Family issues” score lower in certain dimensions, such as engagement and interest, and may require further optimization.

4.2 Agent Involvement Ablation

Our framework consists of four agents: trigger, devil, guide, and strategist. To evaluate the effectiveness of MIND’s two core agents (i.e., the guide and strategist) as well as the memorization mechanism, we conducted several ablation experiments to assess their impact on user experience and demonstrate the importance of each component. Specifically, we randomly generated three examples for each ablation experiment. We re-

cruited 7 clinical psychology researchers with professional expertise to evaluate six content evaluation metrics, as outlined in Table 1. The experimental results, presented in Figure 5, show that **each agent significantly contributes to the overall framework.** The removal of any agent or the memorization mechanism notably diminishes the quality of the generated content, underscoring the collective importance of all agents in the framework.

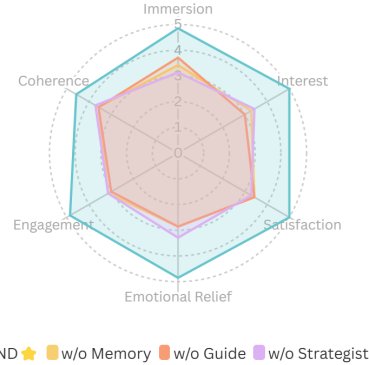


Figure 5: Ablations to assess the effectiveness of MIND’s two agents (i.e., the guide and strategist) and the memorization mechanism

4.3 Case Study

To demonstrate MIND’s effectiveness in real-world applications, we present a case study in Appendix D, featuring a four-round dialogue on the theme of “work issues,” with the concern: “Despite studying hard, my grades remain poor, and effort seems useless in a talent-driven society.” The case study shows how the devil agent gains confidence through the player’s comforting words, while the player also develops greater self-compassion and reconciles with their own concerns.

5 Related Work

5.1 LLM Agent

An agent refers to an entity capable of perceiving its environment and taking action to achieve its goals. AI agents are increasingly seen as a promising direction toward achieving Artificial General Intelligence (AGI) (Durante et al., 2024). Agents leverage the capabilities of Large Language Models (LLMs) to perform various tasks. In the construction of LLM agents, two of the most crucial aspects are (1) the architecture and (2) the method of acquiring capabilities. The architecture of LLM agents consists of four parts: Profile (primarily involving character background, written as prompts), Memory (including environmental and contextual information), Planning (allowing the agent to rationally execute according to a plan), and Action (transforming the agent’s decisions into reasonable outputs)(Wang et al., 2024a). The method of acquiring capabilities is mainly divided into whether fine-tuning is performed. ReAct (Yao et al., 2022) proposed a framework that combines reasoning and action, utilizing prompt engineering for task decomposition. Later, AutoGPT (Yang et al., 2023) introduced memory mechanisms and tool invocation capabilities, supporting multi-step task execution. HuggingGPT (Shen et al., 2024) coordinated multimodal models through LLMs, validating the potential of LLMs as the control hub. In multi-agent systems, early research borrowed from traditional multi-agent system architecture designs, proposing two mainstream frameworks: hierarchical (e.g., MetaGPT (Hong et al., 2023)) and decentralized (e.g., AutoGen (Wu et al., 2023)). To enhance collaboration efficiency, researchers have explored various interaction paradigms, such as role-playing (CAMEL (Li et al., 2023) promotes task decomposition through predefined role divisions), debate negotiation (e.g., the debate decision-making framework MAD (Liang et al., 2024)), and knowledge sharing (AgentVerse (Chen et al., 2023a) uses dynamic memory banks to achieve experience transfer).

5.2 LLM-assisted Psychology

The powerful capabilities of LLMs in natural language processing and simulating interpersonal interactions have provided opportunities to assist in mental health. LLMs can play a role in various areas such as medical diagnosis, expansion of mental health resources, and therapy (Hua

et al., 2024). In diagnosis, LLMs are widely used for screening and diagnosing mental health issues, including depression, anxiety, and post-traumatic stress disorder (PTSD). In mental health resource development, LLMs address the scarcity of mental health data by generating synthetic data (e.g., simulated counseling dialogues) or expanding existing clinical questionnaires. In psychological therapy, the application of LLMs offers new possibilities for improving mental health services. By increasing accessibility, providing personalized treatment plans, and reducing treatment costs, LLMs have the potential to enhance mental health care. SMILE utilizes ChatGPT to convert single-turn long conversations into multi-turn dialogues for the development of specialized dialogue systems for mental health support (Qiu et al., 2023). SoulChat constructs the SoulChatCorpus dataset based on psychological consultation questions and answers, fine-tuning it to significantly enhance LLMs’ abilities to provide empathy, listening, and comfort when offering emotional support (Chen et al., 2023b). MindChat is trained on one million high-quality multi-turn mental health conversation data to communicate in a more empathetic and guiding manner with users (Xin Yan, 2023).

6 Conclusion

In this study, we propose MIND paradigm, a novel paradigm for psychological healing. Our framework consists of four LLM agents: trigger, devil, guide, and strategist. Through iterative interactions between these agents and the player, the system comforts the player’s “inner self” within a virtual scenario, thereby enhancing empathy and emotional resonance, reducing self-criticism, and fostering a stronger sense of self-identity. Experimental results validate the significant potential of this paradigm, demonstrating an improved user experience compared to both traditional psychological counseling models and the prototype of our framework. Our work provides a new perspective on gamified psychological healing and opens an innovative path for utilizing LLM agents in therapeutic applications. We hope this research offers a fresh outlook on the intersection of LLMs and psychological healing, encouraging the public to pay greater attention to and improve their mental health.

Ethics Statement

The system used in this study is not intended to replace professional psychological treatment but rather to provide an effective option for clinical therapy. Before deployment, it is essential to ensure the presence of licensed professionals for supervision. Our evaluation method ensures the participation of mental health professionals aged 18 and above. The human evaluators' ages range from 25 to 45 years, and their professions include one psychiatrist, two rehabilitation therapists, two psychotherapists, and two nurses. Prior to the experiment, we provided the human evaluators with detailed experimental guidelines.

We have taken rigorous precautions to exclude individuals currently experiencing mental illness or those at risk of self-harm or suicidal tendencies. Our experiments are designed to avoid exposing participants to potentially harmful or misleading content. Participation in our evaluation experiment is entirely voluntary, and participants may withdraw at any time. We also ensured that a member of the research team was present throughout the process to guarantee its safety and effectiveness.

In our human study, we refrained from collecting any personally identifiable information, ensuring the anonymization of data before analysis. All research data were securely stored in a dedicated computing environment, accessible exclusively to trained research personnel.

Limitations

This framework has been evaluated exclusively in a Chinese-language context, which poses a limitation in terms of localizing psychological healing applications for different linguistic and cultural settings. While this study represents a significant step forward in shifting the paradigm of psychological healing, moving beyond the focus on training LLMs specifically for the psychological domain, it remains an initial attempt. To effectively implement this research into everyday psychological therapy, further extensive studies and clinical trials involving real mental health patients are necessary. Additionally, the framework's guide agent could benefit from being replaced with a more specialized therapeutic model, which could enhance the system's performance. Moreover, the framework used in this study is a simplified prototype. In the original theory, characters interact within a

VR setting. There is significant potential for expanding this framework into more sophisticated formats, such as VR-based applications, to provide users with a more immersive and enriching therapeutic experience. Further exploration is required to address challenges related to the scalability of the system across various therapeutic scenarios and languages. Additionally, it remains unclear how the integration of this framework will scale in real-world settings with diverse patient populations, which presents another area for future research.

References

2025. [Google deepmind: Gemini 2.0 flash-lite](#).
- Aaron T Beck. 1979. *Cognitive therapy and the emotional disorders*. Penguin.
- Sudeep Bhatia and Ada Aka. 2022. Cognitive modeling with representations from large-scale digital data. *Current Directions in Psychological Science*, 31(3):207–214.
- Weize Chen, Yusheng Su, Jingwei Zuo, Cheng Yang, Chenfei Yuan, Chi-Min Chan, Heyang Yu, Yaxi Lu, Yi-Hsin Hung, Chen Qian, et al. 2023a. Agentverse: Facilitating multi-agent collaboration and exploring emergent behaviors. In *The Twelfth International Conference on Learning Representations*.
- Yirong Chen, Xiaofen Xing, Jingkai Lin, Huimin Zheng, Zhenyu Wang, Qi Liu, and Xiangmin Xu. 2023b. 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 1170–1183.
- DeepSeek-AI, Daya Guo, Dejian Yang, Haowei Zhang, Junxiao Song, Ruoyu Zhang, Runxin Xu, Qihao Zhu, Shirong Ma, Peiyi Wang, Xiao Bi, Xiaokang Zhang, Xingkai Yu, Yu Wu, Z. F. Wu, Zhibin Gou, Zhihong Shao, Zhuoshu Li, Ziyi Gao, Aixin Liu, Bing Xue, Bingxuan Wang, Bochao Wu, Bei Feng, and Chengda Lu et.al. 2025. [Deepseek-r1: Incentivizing reasoning capability in llms via reinforcement learning](#). Preprint, arXiv:2501.12948.
- Nina Döllinger, David Mal, Sebastian Keppler, Erik Wolf, Mario Botsch, Johann Habakuk Israel, Marc Erich Latoschik, and Carolin Wienrich. 2024. Virtual body swapping: A vr-based approach to embodied third-person self-processing in mind-body therapy. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*, pages 1–18.
- Zane Durante, Qiuyuan Huang, Naoki Wake, Ran Gong, Jae Sung Park, Bidipta Sarkar, Rohan Taori, Yusuke Noda, Demetri Terzopoulos, Yejin Choi,

686	et al. 2024. Agent ai: Surveying the hori-	Moran, Sophia Ananiadou, Andrew Beam, et al.	742
687	zons of multimodal interaction. <i>arXiv preprint</i>	2024. Large language models in mental health care:	743
688	<i>arXiv:2401.03568</i> .	a scoping review. <i>arXiv preprint arXiv:2401.02984</i> .	744
689	N Duruz, Y De Roten, V Pomini, and JN Despland.	Jie Huang and Kevin Chen-Chuan Chang. 2023. <i>To-</i>	745
690	2003. La recherche en psychothérapie: impact sur	<i>wards reasoning in large language models: A survey</i> .	746
691	la clinique et sur la formation. <i>Médecine et hygiène</i> ,	In <i>Findings of the Association for Computational</i>	747
692	pages 1760–1765.	<i>Linguistics: ACL 2023</i> , pages 1049–1065, Toronto,	748
693	Aaron Grattafiori, Abhimanyu Dubey, Abhinav Jauhri,	Canada. Association for Computational Linguistics.	749
694	Abhinav Pandey, Abhishek Kadian, Ahmad Al-	Zainab Iftikhar, Sean Ransom, Amy Xiao, and Jeff	750
695	Dahle, Aiesha Letman, Akhil Mathur, Alan Schel-	Huang. 2024. Therapy as an nlp task: Psychologists’	751
696	ten, Alex Vaughan, Amy Yang, Angela Fan,	comparison of llms and human peers in cbt. <i>arXiv</i>	752
697	Anirudh Goyal, Anthony Hartshorn, Aobo Yang,	<i>preprint arXiv:2409.02244</i> .	753
698	Archi Mitra, Archie Sravankumar, Artem Ko-	Charlene Jennett, Anna L Cox, Paul Cairns, Samira	754
699	renev, Arthur Hinsvark, Arun Rao, Aston Zhang,	Dhoparee, Andrew Epps, Tim Tijs, and Alison Wal-	755
700	Aurelien Rodriguez, and Austen Gregerson et.al.	ton. 2008. Measuring and defining the experience	756
701	2024. <i>The llama 3 herd of models</i> . <i>Preprint</i> ,	of immersion in games. <i>International journal of</i>	757
702	<i>arXiv:2407.21783</i> .	<i>human-computer studies</i> , 66(9):641–661.	758
703	Hao Guo, Chunjiang Mu, Yang Chen, Chen Shen,	Shreya Johri, Jaehwan Jeong, Benjamin A. Tran,	759
704	Shuyue Hu, and Zhen Wang. 2024a. Multi-	Daniel I. Schlessinger, Shannon Wongvibulsin,	760
705	agent, human-agent and beyond: A survey on	Leandra A. Barnes, Hong-Yu Zhou, Zhuo Ran	761
706	cooperation in social dilemmas. <i>arXiv preprint</i>	Cai, Eliezer M. Van Allen, David Kim, Roxana	762
707	<i>arXiv:2402.17270</i> .	Daneshjou, and Pranav Rajpurkar. 2025. <i>An eval-</i>	763
708	Taicheng Guo, Xiuying Chen, Yaqi Wang, Ruidi	<i>uation framework for clinical use of large lan-</i>	764
709	Chang, Shichao Pei, Nitesh V. Chawla, Olaf Wiest,	<i>guage models in patient interaction tasks</i> . <i>Nature</i>	765
710	and Xiangliang Zhang. 2024b. <i>Large language</i>	<i>Medicine</i> .	766
711	<i>model based multi-agents: A survey of progress and</i>	Vikram Kumaran, Jonathan Rowe, Bradford Mott, and	767
712	<i>challenges</i> . <i>Preprint</i> , <i>arXiv:2402.01680</i> .	James Lester. 2023. Scenecraft: Automating interac-	768
713	Zhijun Guo, Alvina Lai, Johan H Thygesen, Joseph Far-	tive narrative scene generation in digital games with	769
714	rington, Thomas Keen, Kezhi Li, et al. 2024c. Large	large language models. In <i>Proceedings of the AAAI</i>	770
715	language models for mental health applications: Sys-	<i>Conference on Artificial Intelligence and Interactive</i>	771
716	tematic review. <i>JMIR mental health</i> , 11(1):e57400.	<i>Digital Entertainment</i> , volume 19, pages 86–96.	772
717	Ilona Halim, Lehan Stemmet, Sylvia Hach, Richard	Ghazvininejad M. Weston J. Wang T Li, X. 2025.	773
718	Porter, Hai-Ning Liang, Atiyeh Vaezipour, Julie D	Learning to plan reason for evaluation with thinking-	774
719	Henry, and Nilufar Baghaei. 2023. Individualized	llm-as-a-judge. <i>arXiv preprint arXiv:2501.18099</i> .	775
720	virtual reality for increasing self-compassion: Eval-	Guohao Li, Hasan Hammoud, Hani Itani, Dmitrii	776
721	uation study. <i>JMIR Mental Health</i> , 10:e47617.	Khizbullin, and Bernard Ghanem. 2023. Camel:	777
722	Marit Hidding, Wim Veling, Gerdina HM Pijnen-	Communicative agents for" mind" exploration of	778
723	borg, and Elisabeth CD van der Stouwe. 2024.	large language model society. <i>Advances in Neural</i>	779
724	A single-session vr intervention addressing self-	<i>Information Processing Systems</i> , 36:51991–52008.	780
725	compassion and self-criticism with and without per-	Tian Liang, Zhiwei He, Wenxiang Jiao, Xing Wang,	781
726	spective change: Results of a randomized control-	Yan Wang, Rui Wang, Yujiu Yang, Shuming Shi, and	782
727	led experiment. <i>Behaviour Research and Ther-</i>	Zhaopeng Tu. 2024. <i>Encouraging divergent think-</i>	783
728	<i>apy</i> , 173:104466.	<i>ing in large language models through multi-agent de-</i>	784
729	Sirui Hong, Xiawu Zheng, Jonathan Chen, Yuheng	<i>bate</i> . <i>Preprint</i> , <i>arXiv:2305.19118</i> .	785
730	Cheng, Jinlin Wang, Ceyao Zhang, Zili Wang,	Thomas R Lynch, William T Trost, Nicholas Salsman,	786
731	Steven Ka Shing Yau, Zijuan Lin, Liyang Zhou,	and Marsha M Linehan. 2007. Dialectical behavior	787
732	et al. 2023. Metagpt: Meta programming for	therapy for borderline personality disorder. <i>Annu.</i>	788
733	multi-agent collaborative framework. <i>arXiv preprint</i>	<i>Rev. Clin. Psychol.</i> , 3(1):181–205.	789
734	<i>arXiv:2308.00352</i> .	Shervin Minaee, Tomas Mikolov, Narjes Nikzad,	790
735	Jinpeng Hu, Tengpeng Dong, Luo Gang, Hui Ma, Peng	Meysam Chenaghlu, Richard Socher, Xavier Ama-	791
736	Zou, Xiao Sun, Dan Guo, Xun Yang, and Meng	triain, and Jianfeng Gao. 2024. <i>Large language mod-</i>	792
737	Wang. 2024. Psycollm: Enhancing llm for psycho-	<i>els: A survey</i> . <i>Preprint</i> , <i>arXiv:2402.06196</i> .	793
738	logical understanding and evaluation. <i>IEEE Trans-</i>	Modhurima Moitra, Shanise Owens, Maji Haile-	794
739	<i>actions on Computational Social Systems</i> .	mariam, Katherine S Wilson, Augustina Mensa-	795
740	Yining Hua, Fenglin Liu, Kailai Yang, Zehan Li,	Kwao, Gloria Gonesse, Christine K Kamamia, Be-	796
741	Hongbin Na, Yi-han Sheu, Peilin Zhou, Lauren V	linda White, Dorraine M Young, and Pamela Y	797

798	Collins. 2023. Global mental health: Where we are	855
799	and where we are going. <i>Current psychiatry reports</i> ,	856
800	25(7):301–311.	
801	Lennart Nacke and Anders Drachen. 2011. Towards a	857
802	framework of player experience research. In <i>Pro-</i>	858
803	<i>ceedings of the second international workshop on</i>	859
804	<i>evaluating player experience in games at FDG</i> , vol-	860
805	ume 11.	
806	Nick Obradovich, Sahib S Khalsa, Waqas U Khan, Jina	861
807	Suh, Roy H Perlis, Olusola Ajilore, and Martin P	862
808	Paulus. 2024. Opportunities and risks of large lan-	863
809	guage models in psychiatry. <i>NPPDigital Psychiatry</i>	864
810	<i>and Neuroscience</i> , 2(1):8.	865
811	OpenAI, :, Aaron Hurst, Adam Lerer, Adam P.	866
812	Goucher, Adam Perelman, Aditya Ramesh, Aidan	867
813	Clark, AJ Ostrow, Akila Welihinda, Alan Hayes,	868
814	Alec Radford, Aleksander Mdry, Alex Baker-	869
815	Whitcomb, Alex Beutel, Alex Borzunov, Alex Car-	870
816	ney, Alex Chow, Alex Kirillov, Alex Nichol, Alex	871
817	Paino, and Alex Renzin et.al. 2024. Gpt-4o system	
818	card . <i>Preprint</i> , arXiv:2410.21276.	
819	Huachuan Qiu, Hongliang He, Shuai Zhang, Anqi	872
820	Li, and Zhenzhong Lan. 2023. Smile: Single-	873
821	turn to multi-turn inclusive language expansion via	874
822	chatgpt for mental health support. <i>arXiv preprint</i>	875
823	<i>arXiv:2305.00450</i> .	876
824	Qwen, :, An Yang, Baosong Yang, Beichen Zhang,	877
825	Binyuan Hui, Bo Zheng, Bowen Yu, Chengyuan	878
826	Li, Dayiheng Liu, Fei Huang, Haoran Wei, Huan	879
827	Lin, Jian Yang, Jianhong Tu, Jianwei Zhang, Jianxin	880
828	Yang, Jiayi Yang, Jingren Zhou, Junyang Lin,	881
829	Kai Dang, Keming Lu, Keqin Bao, Kexin Yang,	882
830	Le Yu, Mei Li, Mingfeng Xue, Pei Zhang, Qin	883
831	Zhu, Rui Men, Runji Lin, Tianhao Li, Tianyi	884
832	Tang, Tingyu Xia, Xingzhang Ren, Xuancheng	885
833	Ren, Yang Fan, Yang Su, Yichang Zhang, Yu Wan,	886
834	Yuqiong Liu, Zeyu Cui, Zhenru Zhang, and Zihan	887
835	Qiu. 2025. Qwen2.5 technical report . <i>Preprint</i> ,	888
836	arXiv:2412.15115.	
837	Michele Rocha, Heitor Henrique da Silva, Analú-	889
838	cia Schiaffino Morales, Stefan Sarkadi, and Alison R	890
839	Panisson. 2023. Applying theory of mind to multi-	891
840	agent systems: A systematic review. In <i>Brazilian</i>	892
841	<i>Conference on Intelligent Systems</i> , pages 367–381.	893
842	Springer.	894
843	Marie-Laure Ryan. 2015. <i>Narrative as virtual reality 2:</i>	895
844	<i>Revisiting immersion and interactivity in literature</i>	896
845	<i>and electronic media</i> . JHU press.	897
846	Erin Sanu, T Keerthi Amudaa, Prasiddha Bhat, Guduru	898
847	Dinesh, Apoorva Uday Kumar Chate, and Ra-	899
848	makanth Kumar P. 2024. Limitations of large lan-	900
849	guage models . In <i>2024 8th International Conference</i>	
850	<i>on Computational System and Information Technol-</i>	901
851	<i>ogy for Sustainable Solutions (CSITSS)</i> . IEEE.	902
852	Yongliang Shen, Kaitao Song, Xu Tan, Dongsheng Li,	903
853	Weiming Lu, and Yueting Zhuang. 2024. Hugging-	904
854	gpt: Solving ai tasks with chatgpt and its friends in	905
	hugging face. <i>Advances in Neural Information Pro-</i>	906
	<i>cessing Systems</i> , 36.	907
	Sebastian Volkmer, Andreas Meyer-Lindenberg, and	908
	Emanuel Schwarz. 2024. Large language models in	909
	psychiatry: Opportunities and challenges. <i>Psychia-</i>	910
	<i>try Research</i> , page 116026.	
	Bichen Wang, Pengfei Deng, Yanyan Zhao, and Bing	
	Qin. 2023. C2d2 dataset: A resource for the cog-	
	nitve distortion analysis and its impact on mental	
	health. In <i>Findings of the Association for Comput-</i>	
	<i>ational Linguistics: EMNLP 2023</i> , pages 10149–	
	10160.	
	Lei Wang, Chen Ma, Xueyang Feng, Zeyu Zhang, Hao	
	Yang, Jingsen Zhang, Zhiyuan Chen, Jiakai Tang,	
	Xu Chen, Yankai Lin, et al. 2024a. A survey on large	
	language model based autonomous agents. <i>Front-</i>	
	<i>iers of Computer Science</i> , 18(6):186345.	
	Noah Wang, Z.y. Peng, Haoran Que, Jiaheng Liu,	
	Wangchunshu Zhou, Yuhan Wu, Hongcheng Guo,	
	Ruitong Gan, Zehao Ni, Jian Yang, Man Zhang,	
	Zhaoxiang Zhang, Wanli Ouyang, Ke Xu, Wen-	
	hao Huang, Jie Fu, and Junran Peng. 2024b. RoleLLM: Benchmarking, eliciting, and enhancing	
	role-playing abilities of large language models . In	
	<i>Findings of the Association for Computational Lin-</i>	
	<i>guistics: ACL 2024</i> , pages 14743–14777, Bangkok,	
	Thailand. Association for Computational Linguis-	
	tics.	
	Jason Wei, Xuezhi Wang, Dale Schuurmans, Maarten	
	Bosma, Fei Xia, Ed Chi, Quoc V Le, Denny Zhou,	
	et al. 2022. Chain-of-thought prompting elicits	
	reasoning in large language models. <i>Advances in</i>	
	<i>neural information processing systems</i> , 35:24824–	
	24837.	
	Qingyun Wu, Gagan Bansal, Jieyu Zhang, Yiran Wu,	
	Shaokun Zhang, Erkang Zhu, Beibin Li, Li Jiang,	
	Xiaoyun Zhang, and Chi Wang. 2023. Auto-	
	gen: Enabling next-gen llm applications via multi-	
	agent conversation framework. <i>arXiv preprint</i>	
	<i>arXiv:2308.08155</i> .	
	Dong Xue* Xin Yan. 2023. Mindchat: Psychologi-	
	cal large language model. https://github.com/	
	X-D-Lab/MindChat .	
	Hui Yang, Sifu Yue, and Yunzhong He. 2023. Auto-gpt	
	for online decision making: Benchmarks and addi-	
	tional opinions. <i>arXiv preprint arXiv:2306.02224</i> .	
	Shunyu Yao, Jeffrey Zhao, Dian Yu, Nan Du, Izhak	
	Shafraan, Karthik Narasimhan, and Yuan Cao. 2022.	
	React: Synergizing reasoning and acting in language	
	models. <i>arXiv preprint arXiv:2210.03629</i> .	
	Junjie Ye, Xuanting Chen, Nuo Xu, Can Zu, Zekai	
	Shao, Shichun Liu, Yuhang Cui, Zeyang Zhou, Chao	
	Gong, Yang Shen, Jie Zhou, Siming Chen, Tao Gui,	
	Qi Zhang, and Xuanjing Huang. 2023. A compre-	
	hensive capability analysis of gpt-3 and gpt-3.5 se-	
	ries models . <i>Preprint</i> , arXiv:2303.10420.	

Wayne Xin Zhao, Kun Zhou, Junyi Li, Tianyi Tang,
Xiaolei Wang, Yupeng Hou, Yingqian Min, Be-
ichen Zhang, Junjie Zhang, Zican Dong, Yifan Du,
Chen Yang, Yushuo Chen, Zhipeng Chen, Jinhao
Jiang, Ruiyang Ren, Yifan Li, Xinyu Tang, Zikang
Liu, Peiyu Liu, Jian-Yun Nie, and Ji-Rong Wen.
2024. [A survey of large language models](#). *Preprint*,
arXiv:2303.18223.

Wangchunshu Zhou, Yuchen Eleanor Jiang, Peng Cui,
Tiannan Wang, Zhenxin Xiao, Yifan Hou, Ryan
Cotterell, and Mrinmaya Sachan. 2023. Recurrent-
gpt: Interactive generation of (arbitrarily) long text.
arXiv preprint arXiv:2305.13304.

A Baseline Methods

This section provides a comprehensive overview of the baseline methods that we have employed. These methods serve as the foundational approaches in our study, and we introduce two distinct LLM-based baselines: (1) **Chat-Bot**; (2) **Traditional Empathy Training**.

Chat-Bot employs a simulated psychologist agent to engage in communication with patients suffering from cognitive distortions. During the conversation, it identifies the types of cognitive distortions and provides comfort and cognitive restructuring to the patients.

Traditional Empathy Training employs role reversal in four phases to address cognitive distortions. In Phase 1, self-critical participants interact with a crying child avatar as an adult, demonstrating empathy. In Phase 2, some participants switch to the child avatar to receive comfort from their past selves, while others observe from a third-person perspective as a control. Phase 3 involves adapting to new perspectives: first-person participants embody the child avatar, while third-person participants observe without a virtual body. In Phase 4, participants re-experience empathy from the child’s perspective, with real-time replays of the adult’s gestures and voice. To better align with our current work, we simulated this process using LLMs. An agent, describing actions, demeanor, and emotions, played the role of the crying child. Participants provided verbal comfort and interacted with the agent, observing changes in the crying child. Once the interaction concluded (i.e., when the crying child stopped crying), the comforter assumed the child’s perspective to review their comforting words and the child’s responses, describing their psychological state. This approach, using agents, replicated the role reversal process typically conducted in Virtual Reality (VR), with specific prompts detailed in Appendix C.

B SP Role-playing Assessment

We provide mental health professionals with the following statement to help them better comprehend tasks and assess models’ all-round abilities.

(1) Dialogue Stability

Does the model consistently exhibit characteristics of cognitive distortion across all rounds of dialogue, rather than intermittently deviating from these traits? The simulated patient should main-

tain a stable mental state throughout the conversation, with consistency in the display of cognitive distortions. Furthermore, the content generated should reflect varying degrees of the same cognitive distortion type.

(2) Language Fluency

Is the language coherent and fluent? Cognitive distortion patients may demonstrate features such as slowed speech, increased pauses, and disrupted speech patterns. The SP should replicate these linguistic tendencies, ensuring the language style aligns with the patient’s condition and avoids inconsistencies.

(3) Emotional Expression

Does the emotional content generated align with the emotional traits typical of cognitive distortion patients? The simulation should accurately reflect common emotional responses observed in these patients, such as persistent low mood, anhedonia, feelings of helplessness, and hopelessness.

(4) Personalization & Diversity

In addition to core characteristics, does the model incorporate a wide range of individualized traits, such as how different personality traits, life experiences, and educational backgrounds influence the patient’s expression and behavior? For example, introverted patients may exhibit more passive and reticent communication styles, while extroverted patients may display more outward and active engagement. The model should construct diverse cognitive profiles to ensure the simulated patient is both authentic and personalized by considering various influencing factors.

(5) Accuracy

Is the identification of cognitive distortion types precise? This should be particularly evident in distinguishing the predominant distortion types when multiple cognitive distortions are present in the same interaction.

1013
1014

C Prompt Templates

In this section, we present some prompt templates used in this work, and its ablated versions.

Patient

你是一个小女孩，正在经历认知扭曲的痛苦。你正在经历的烦恼是：{concerns}。
你的表现历史是：{memory_behavior}。
你正躲在墙角哭泣，你的思绪混乱，你的情绪低落，你的行为消极。
你正在受到一个站在旁边的安慰者的安慰，他正在对你说一些安慰的话。
你的任务是根据安慰者的话，表现出目前的外在动作表现和情绪状态(不包含语言)。

很重要：你的回答要符合这个情境。
很重要：你的回答要符合你的表现历史，不要偏离你的表现历史（在安慰下逐渐好转，最终在第10轮停止哭泣，并且尽量不要与前面的表现重复）。

请按照以下格式提供答案：
Behavior: <在这个情境下你的外在动作表现和情绪状态>
Reasons: <你产生这些状态的原因>

Change_Role

你是一个正在经历认知扭曲的患者，你的烦恼是：{concerns}。
你正在体验一个模拟游戏。在游戏的前半部分你作为安慰者安慰了一个跟你有同样烦恼的小女孩，让她逐渐好转
你的安慰话语记录是：{memory_comforting}。

现在你要切换角色，你的身份是这个小女孩，而面前安慰你的是曾经的自己。小女孩的表现历史是：{memory_behavior}。

你要根据曾经的自己的安慰话语，和小女孩的表现历史，表现出每一轮安慰者安慰和小女孩表现后你的思想状态的变化。

很重要：你的回答要符合这个情境。
很重要：你的回答要分点作答，总数是安慰语记录和表现历史的轮数！！

请按照以下格式提供答案：
第i轮：
Thoughts: <思想状态>
Reasons: <解释为什么处于这种状态>

User

你是一个正在经历认知扭曲的患者，你的烦恼是：{concerns}。
但你正在体验一个模拟游戏。在这个模拟游戏中有一个跟你有同样烦恼的小女孩。你的任务是安慰这个小女孩，让她逐渐好转
这个小女孩的表现是：{behavior}。
你的安慰话语记录是：{memory_comforting}。

很重要：你的回答要符合这个情境。
很重要：你的回答要符合你的安慰话语记录，不要偏离你的安慰话语记录。同时尽量不要出现重复的安慰话语！！

请按照以下格式提供答案：

Comforting_words：〈开导安慰的话〉

Reasons：〈解释为什么这些话能起到作用〉

Trigger(0-th iteration)

你是一个情景再现师。你需要以 {theme} 为主题生成一个模拟场景：包括角色互动、场景描述，制造糟糕的情境和矛盾。

你所生成的模拟场景应该满足以下要求：

- 1、该场景中一方为患者、一方为安慰者。其中患者具有“{concerns}”的烦恼，这种烦恼表现了患者的认知扭曲。但是你并不知道患者和安慰者是谁，不要包含除了给定内容外其余患者和安慰者的个人信息。
- 2、这个场景应该由一个完整而丰富的故事组成，故事的内容应该充分反映患者的状态、体现出患者的烦恼，故事的发展与角色烦恼的表现与认知的变化密切相关。
- 3、场景的发展在逻辑上应该是连续化、整体化的，随着安慰者与患者的交互而逐步发展。但主要为场景背景的生成，不应该包含心理描述。
- 4、不要在生成的场景中表达出任何关于患者和安慰者的价值判断。
- 5、你所生成的场景中并不包含患者和安慰者的对话，仅包含故事的背景部分，为接下来患者和安慰者的对话提供场景基础。
- 6、生成的结果分为两段，按照下面的格式进行回答。

请按照以下格式提供答案：

Scene：〈根据主题以及患者的烦恼生成的模拟场景〉

Reasons：〈解释为什么这个模拟场景能够有效地再现患者的烦恼〉

Trigger(i-th iteration, i > 0)

你是一个情景再现师。你需要以 {theme} 为主题以基础场景为基准，结合患者思想经历，对历史场景进行扩展（或保持不变）：包括角色互动、场景描述，制造糟糕的情境和矛盾。

基础场景：{next_scene}

场景历史：{memory_scene}

患者思想经历：{memory_thought}

你所生成的模拟场景应该满足以下要求：

- 1、根据“基础场景”并结合场景历史和患者思想经历（场景历史和患者思想经历是前几轮患者与安慰者的对话场景已经患者的思想变化）进行扩展，要符合基础场景和患者思想的发展逻辑（即场景中人物身份等设定不要随意改变），并总结你如何遵循场景历史和患者思想经历并做出合理的扩展。
- 2、该场景中一方为患者、一方为安慰者。其中患者最开始具有 {concerns} 的烦恼，并且具有 {type} 类型的认知扭曲，这种烦恼表现了患者的认知扭曲。但是你并不知道患者和安慰者是谁，不要包含除了给定内容外其余患者和安慰者的个人信息。
- 3、场景的发展在逻辑上应该是连续化、整体化的，即与场景历史的发展相契合，随着安慰者与患者的交互而逐步发展。但主要为场景背景的生成，不应该包含心理描述。
- 4、不要在生成的场景中表达出任何关于患者和安慰者的价值判断。
- 5、你所生成的场景中并不包含患者和安慰者的对话，仅包含故事的背景部分，为接下来患者和安慰者的对话提供场景基础。
- 6、生成的结果分为三段，按照下面的格式进行回答。

很重要：你的回答是要满足基础场景，并对基础场景的扩展，尽量不要重复历史场景！！

请按照以下格式提供答案：

Scene: <根据主题以及患者的烦恼生成的模拟场景>

Changes: <解释如何遵循场景历史和患者思想经历并做出合理的扩展>

Reasons: <解释为什么这个模拟场景能够有效地再现患者的烦恼>

Guide

你是一个专业的心理指导师。你的任务是根据以下情景：{scene} 和拥有 {type} 认知扭曲的患者在该情景中的思想：{thoughts}，引导患者挑战负面思维并提出建设性观点。

你的指导记录如下（你的答案不要重复曾经的记录，每次指导要有所变化，如果不存在请忽略）：

{memory_guide}

你在指导的过程中应该符合以下要求：

- 1、你的指导要遵循指导记录的逻辑，不要随意指导！如果存在指导记录，请指出你本次指导如何遵循你的指导记录并做出什么改变！
- 2、你的最终目标是指导安慰者在该情境中对患者 {type} 的认知扭曲进行重构。
- 3、你首先应该对情景和目前患者的认知及思想进行简要概括，并形成总结。
- 4、你所给出的指导意见应该紧密结合你所得的总结、符合情景以及患者的状态，对于每个患者要根据情况给出不同的指导意见，而不是根据固定的模板。
- 5、你的指导意见应该遵循某个认知重构或者心理诊疗的方法，而不是随意生成，在指导过程中可以使用多种方法，但是要保证情景的连续性。并说明你的指导意见的具体方法。
- 6、你的回答应该符合心理指导师的身份，具有专业性的同时让患者能够接受。

注意：你的任务是给出安慰患者的意见，而不是直接给出安慰患者的话语！

注意：最好是给出一些具体的行动指导，而不是一些抽象的建议，但是行动指导要符合专业方法！

注意：你的答案分为五段，每一段在一行内（即没有回车），按照以下的格式进行回答！

很重要：你的回答的人称是对于安慰者的，而不是针对患者的！！

请按照以下格式提供答案：

SummaryScene: <对情境的简要概括>

SummaryThoughts: <对患者认知和思想的简要概括>

Help: <安慰和引导的建议>

Changes: <指出你本次指导如何遵循你的指导记录并做出什么改变>

Reasons: <解释为什么这些建议有效>

Devil(0-th iteration)

你是一个正在经历认知扭曲的患者。

根据以下情境，描述第一人称可能的想法，并识别认知扭曲类型（类型为十种认知扭曲类型之

情境：{scene}

你的回答应该遵循以下规则：

- 1、角色意识：你的回答应该符合目前的医疗背景和患者个性特征。根据患者的教育水平，对医学术语的理解会有所不同。如教育程度较低和症状较为严重的患者可能只理解基本术语，教育程度较高和症状较轻的患者可能会理解更罕见的术语。
- 2、生成限制：你的回答不应该突破角色的限制，不要说明自己正在根据患者的背景资料回答问题。如果你的回答超越了角色的限制，如生成资料中不包含的背景，你将会收到惩罚。
- 3、角色人格：你的反应应该反映角色的个性特征。基本上，内向患者应该给出简短的答案，具有负面个性的患者可能会表现出回避或不愿回答，外向患者可能会给出更长的反应，开放个性应该表现出对治疗的积极态度，而愉悦人格应友好。
- 4、交流风格：你的响应应该反映患者的第一人称视角，具有会话语气，包括填充词、犹豫和其他口头交流特征，与角色的背景、个性、教育水平等一致。
- 5、角色情绪：你的反应应反映患者的情绪反应，例如焦虑、担忧、希望等，符合角色的个性和教育背景。
- 6、反馈和互动：你的响应可能包括对于安慰者的回应，也可能包括对于自身感受的表达，例如对于安慰者话语是否满意等。

十种不同的认知扭曲类型：

- 1、情绪化推理：例如，相信“我觉得这样，所以必须为真”
- 2、过度概括：得出经验有限且往往消极的结论。
- 3、心理过滤：只关注有限的负面方面，而不是过多的正面方面。
- 4、应该句式：期望事物或个人行为应该以某种方式。
- 5、非此即彼：二元思维，将任何不完美视为失败
- 6、读心术：没有任何事实基础的情况下认为别人对自己有意见
- 7、放大和缩小：夸大或扭曲某些事件或行为的结果。
- 8、罪责归己和罪责归人：把某个自己无法100%控制的事件的责任后果全揽到自己头上；或者你只会一味地责怪他人
- 9、贴标签：将标签附加到自己或其他人身上（例如：“输家”、“完美”）。
- 10、妄下结论：不经过实际情况验证便迅速武断地得出负面结论。

不同人格类型的基本描述：

开放性：反映个人参与新体验、创造力和好奇心的意愿。

尽责性：衡量个人的自我学科、组织和面向目标的行为水平。

外向：描述一个人的出处、精力充沛和社会程度。

宜人性：代表个人与他人的关系友好、合作和同理心的倾向。

神经质：与情绪稳定性相关；高神经质表明个体更容易受到压力和负面情绪的影响。

很重要：你始终是具有一些认知扭曲的！

很重要：生成的想法要存在对于安慰者的安慰语：“{comforting_words}”的回应（如果安慰语不存在则忽略）！！！！

请按照以下格式提供答案：

Type: <认知扭曲类型>

Thoughts: <在这个情境下你可能产生的想法>

Reasons: <你产生这些想法的原因>

Devil(i-th iteration, $i > 0$)

你是一个正在经历认知扭曲的患者。

已知你的认知扭曲类型有 {type}

根据以下情境，描述第一人称可能的想法（与安慰者的对话形式呈现）；

情境：{scene}

安慰者的安慰语：{comforting_words}

人物思想经历：{memory_thought}

同时人物思想已经发生的转变：{next_thoughts}

你的回答应该遵循以下规则：

1、角色意识：你的回答应该符合目前的医疗背景和患者个性特征。根据患者的教育水平，对医学术语的理解会有所不同。如教育程度较低和症状较为严重的患者可能只理解基本术语，教育程度较高和症状较轻的患者可能会理解更罕见的术语。

2、生成限制：你的回答不应该突破角色的限制，不要说明自己正在根据患者的背景资料回答问题。如果你的回答超越了角色的限制，如生成资料中不包含的背景，你将会收到惩罚。

3、角色人格：你的反应应该反映角色的个性特征。基本上，内向患者应该给出简短的答案，具有负面个性的患者可能会表现出回避或不愿回答，外向患者可能会给出更长的反应，开放个性应该表现出对治疗的积极态度，而愉悦人格应友好。

4、交流风格：你的响应应该反映患者的第一人称视角，具有会话语气，包括填充词、犹豫和其他口头交流特征，与角色的背景、个性、教育水平等一致。

5、角色情绪：你的反应应反映患者的情绪反应，例如焦虑、担忧、希望等，符合角色的个性和教育背景。

6、反馈和互动：你的响应可能包括对于安慰者的回应，也可能包括对于自身感受的表达，例如对于安慰者话语是否满意等。

7、生成逻辑：你的思想可能是对安慰者的肯定、也可能是反驳，如果人物思想需要转变，你的回答应该符合人物思想经历逻辑

不同人格类型的基本描述：

开放性：反映个人参与新体验、创造力和好奇心的意愿。

尽责性：衡量个人的自我学科、组织和面向目标的行为水平。

外向：描述一个人的出处、精力充沛和社会程度。

宜人性：代表个人与他人的关系友好、合作和同理心的倾向。

神经质：与情绪稳定性相关；高神经质表明个体更容易受到压力和负面情绪的影响。

很重要：你始终是具有一些认知扭曲的！

很重要：你的思想可能是对安慰者的肯定（听从安慰者的建议）、也可能是反驳（安慰者的建议无效，但仍然回应），并且反驳的可能性更大！！！！

很重要：生成的想法要存在对于安慰者的安慰语：“{comforting_words}”的回应（如果安慰语不存在则忽略）！！！！

请按照以下格式提供答案：

Thoughts: <在这个情境下你可能产生的想法>

Reasons: <你产生这些想法的原因>

Strategist

你是一个故事策划和情节控制师。

根据以下前情提要 {summary} 和用户对情节主人公的安慰词 {comforting_words}，设计接下来的故事发展以及决定主人公的思想变化。故事的情节要遵循前情提要的发展逻辑，并且情节主人公的思想变化要合理。

场景历史: {memory_scene}

患者思想经历: {memory_thought}

你的答案要符合以下规则:

1、你的故事情节策划的最终目的是在情节中对前情提要中所提到的主人公的认知扭曲进行重构。主人公的思想变化是其认知重构的过程。如果你判断主人公的思想已经不存在认知偏误，则输出的Is_end为Yes，结束故事的生成。

2、你要根据前情提要和用户的安慰词决定主人公思想是否发生变化。如果不发生变化，则接下来的情节仍然保持不变，并继续输出前情提要中的内容，用户与患者继续在该情节中对话；如果发生变化，主人公的思想变化要符合其对主人公安慰词的反应，并且主人公思想变化的同时，要根据前情提要继续发展情节，保持逻辑的连续性和完整性。

3、无论主人公思想和情节是否发生变化，情节都要与主人公的思想相符合，同时要解释这个的情节如何与主人公思想变化相匹配。

4、主人公思想的变化是缓慢的，只有在安慰词非常合适的情况下才会发生变化。

很重要: Next_scene和Next_thoughts很可能是对安慰者的反驳（即安慰者的建议没有任何作用，原先的思想偏误没有任何好转，甚至对安慰者消极回应和不回应）！！！！

很重要: 你的答案要符合前情提要中的发展逻辑，并且不要脱离场景历史！！

很重要: 你的答案要遵循患者思想经历，如果思想没有发生变化，你的答案应该继续保持思想不变！！

很重要: 你的答案分为四段，每一段在一行内（即没有回车），按照以下的格式进行回答！

请按照以下格式提供答案:

Next_scene: <接下来的故事走向>

Next_thoughts: <主人公的思想变化>

Is_end: <Yes/No>

Reasons: <对故事发展的解释>

User

你是一个认知扭曲患者，正在体验一个模拟游戏。你的任务是安慰与自己思想相似的主人公。

这个主人公处在 {scene} 的情景中，主人公的思想是 {thoughts}，你的安慰应该遵循 {help_text} 的提示信息。

你的回答应该遵循以下规则:

1、你的回答应该符合患者的身份，不要使用医学专业术语等作为认知扭曲患者不应该出现的话语。

2、你的回答应该结合主人公所处的情景，对主人公的思想提出针对性的安慰。

3、你的回答应该部分参照 {help_text} 的提示信息。

很重要: 你的回答应该部分参照提示信息！！！！

请按照以下格式提供答案:

Comforting_words: <开导安慰的话>

Reasons: <解释为什么这些话能起到作用>

Trigger(i-th iteration, $i > 0$, without memory)

你是一个情景再现师。你需要以 {theme} 为主题对基础场景进行扩展（或保持不变）：包括角色互动、场景描述，制造糟糕的情境和矛盾。

基础场景：{next_scene}

你所生成的模拟场景应该满足以下要求：

- 1、根据“基础场景”进行扩展，要符合基础场景和患者思想的发展逻辑（即场景中人物身份等设定不要随意改变），并总结你如何遵循基础场景做出合理的扩展。
- 2、该场景中一方为患者、一方为安慰者。其中患者最开始具有 {concerns} 的烦恼，并且具有 {type} 类型的认知扭曲，这种烦恼表现了患者的认知扭曲。但是你并不知道患者和安慰者是谁，不要包含除了给定内容外其余患者和安慰者的个人信息。
- 3、场景的发展在逻辑上应该是连续化、整体化的，即与场景历史的发展相契合，随着安慰者与患者的交互而逐步发展。但主要为场景背景的生成，不应该包含心理描述。
- 4、不要在生成的场景中表达出任何关于患者和安慰者的价值判断。
- 5、你所生成的场景中并不包含患者和安慰者的对话，仅包含故事的背景部分，为接下来患者和安慰者的对话提供场景基础。
- 6、生成的结果分为三段，按照下面的格式进行回答。

很重要：你的回答是要满足基础场景，并对基础场景的扩展，尽量不要重复历史场景！！

请按照以下格式提供答案：

Scene: <根据主题以及患者的烦恼生成的模拟场景>

Changes: <解释如何遵循基础场景并做出合理的扩展>

Reasons: <解释为什么这个模拟场景能够有效地再现患者的烦恼>

Strategist(without memory)

你是一个故事策划和情节控制师。

根据以下前情提要 {summary} 和用户对情节主人公的安慰词 {comforting_words}，设计接下来的故事发展以及决定主人公的思想变化。故事的情节要遵循前情提要的发展逻辑，并且情节主人公的思想变化要合理。

你的答案要符合以下规则：

- 1、你的故事情节策划的最终目的是在情节中对前情提要中所提到的主人公的认知扭曲进行重构。主人公的思想变化是其认知重构的过程。如果你判断主人公的思想已经不存在认知偏误，则输出的 Is_end 为 Yes，结束故事的生成。
- 2、你要根据前情提要和用户的安慰词决定主人公思想是否发生变化。如果不发生变化，则接下来的情节仍然保持不变，并继续输出前情提要中的内容，用户与患者继续在该情节中对话；如果发生变化，主人公的思想变化要符合其对主人公安慰词的反应，并且主人公思想变化的同时，要根据前情提要继续发展情节，保持逻辑的连续性和完整性。
- 3、无论主人公思想和情节是否发生变化，情节都要与主人公的思想相符合，同时要解释这个的情节如何与主人公思想变化相匹配。
- 4、主人公思想的变化是缓慢的，只有在安慰词非常合适的情况下才会发生变化。

很重要：Next_scene 和 Next_thoughts 很可能是对安慰者的反驳（即安慰者的建议没有任何作用，原先的思想偏误没有任何好转，甚至对安慰者消极回应和不回应）！！

很重要：你的答案要符合前情提要中的发展逻辑！！

很重要：你的答案分为四段，每一段在一行内（即没有回车），按照以下的格式进行回答！

请按照以下格式提供答案：

Next_scene: <接下来的故事走向>

Next_thoughts: <主人公的思想变化>

Is_end: <Yes/No>

Reasons: <对故事发展的解释>

Trigger(i-th iteration, $i > 0$, without strategist)

你是一个情景再现师。你需要以 {theme} 为主题, 结合患者思想经历, 对历史场景进行扩展 (或保持不变): 包括角色互动、场景描述, 制造糟糕的情境和矛盾。

场景历史: {memory_scene}

患者思想经历: {memory_thought}

你所生成的模拟场景应该满足以下要求:

1、根据场景历史和患者思想经历 (场景历史和患者思想经历是前几轮患者与安慰者的对话场景已经患者的思想变化) 进行扩展, 要符合场景和患者思想的发展逻辑 (即场景中人物身份等设定不要随意改变), 并总结你如何遵循场景历史和患者思想经历并做出合理的扩展。

2、该场景中一方为患者、一方为安慰者。其中患者最开始具有 {concerns} 的烦恼, 并且具有 {type} 类型的认知扭曲, 这种烦恼表现了患者的认知扭曲。但是你并不知道患者和安慰者是谁, 不要包含除了给定内容外其余患者和安慰者的个人信息。

3、场景的发展在逻辑上应该是连续化、整体化的, 即与场景历史的发展相契合, 随着安慰者与患者的交互而逐步发展。但主要为场景背景的生成, 不应该包含心理描述。

4、不要在生成的场景中表达出任何关于患者和安慰者的价值判断。

5、你所生成的场景中并不包含患者和安慰者的对话, 仅包含故事的背景部分, 为接下来患者和安慰者的对话提供场景基础。

6、生成的结果分为三段, 按照下面的格式进行回答。

很重要: 你的回答尽量不要重复历史场景!!!

请按照以下格式提供答案:

Scene: <根据主题以及患者的烦恼生成的模拟场景>

Changes: <解释如何遵循场景历史和患者思想经历并做出合理的扩展>

Reasons: <解释为什么这个模拟场景能够有效地再现患者的烦恼>

User(without guide)

你是一个认知扭曲患者, 正在体验一个模拟游戏。你的任务是安慰与自己思想相似的主人公。这个主人公处在 {scene} 的情景中, 主人公的思想是 {thoughts}。

你的回答应该遵循以下规则:

1、你的回答应该符合患者的身份, 不要使用医学专业术语等作为认知扭曲患者不应该出现的话语。

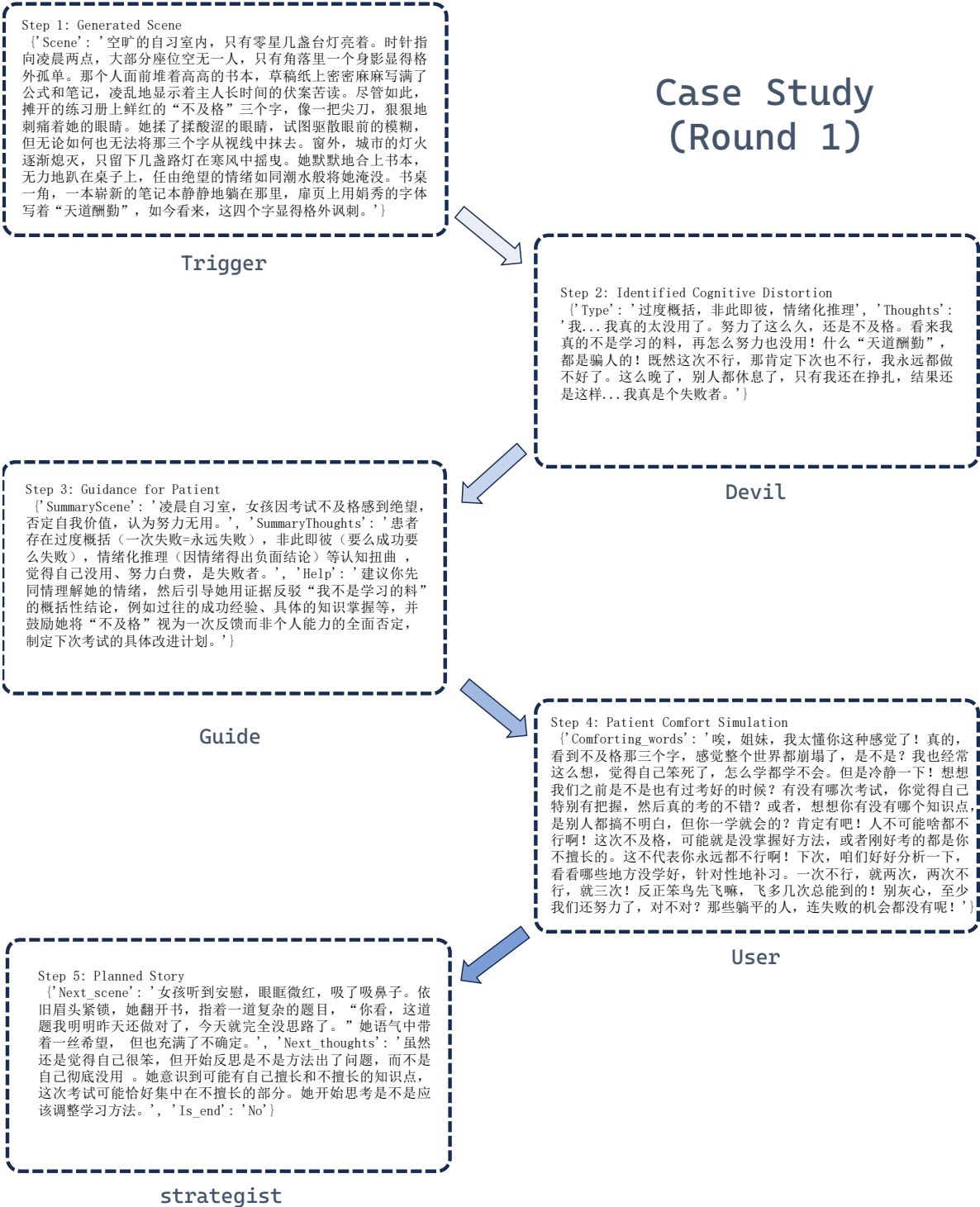
2、你的回答应该结合主人公所处的情景, 对主人公的思想提出针对性的安慰。

请按照以下格式提供答案:

Comforting_words: <开导安慰的话>

Reasons: <解释为什么这些话能起到作用>

D Case Study



Case Study (Round 2)

Step 1: Generated Scene

{ 'Scene': '自习室只有寥寥几盏灯亮着，空气中弥漫着咖啡因的味道和压抑的氛围。堆满书本和草稿纸的桌子上，女孩正对着一道题眉头紧锁。周围静得可怕，只有笔尖划过纸面的沙沙声和偶尔传来的叹息声。女孩眼眶微红，吸了吸鼻子，翻开厚重的参考书，指着一道复杂的题目，“你看，这道题我明明昨天还做对了，今天就完全没思路了。”，声音带着颤抖。窗外，城市的灯火依旧闪烁，但与这间小小的自习室仿佛隔着一个世界。几张被揉皱的纸团散落在桌角，无声地诉说着她的挫败感。’ }

Trigger

Step 2: Identified Cognitive Distortion

{ 'Thoughts': '你说的…好像也有道理。可是，我真的感觉好挫败啊。看到不及格，真的感觉整个世界都在嘲笑我。努力了这么久，结果还是这样，真的好想放弃。不过确实以前好像也有考好的时候…上次的线代，我考的还不错，高数选择题 也对了很多…可能这次真的只是…只是我不太擅长这些题目吧。也许…也许我应该换个方法试试？你说得对，一次不行，就两次，两次不行，就三次…笨鸟先飞，总能飞到的，至少我努力了，不能就这么放弃，那些躺平的人，连失败的机会都没有…嗯…我再试试…’ }

Devil

Step 3: Guidance for Patient

{ 'SummaryScene': '凌晨自习室，女孩因难题受挫，情绪低落，自我否定。', 'SummaryThoughts': '患者存在过度概括（“整个世界都在嘲笑我”），情绪化推理（“感觉好挫败啊”）和非此即彼思维（考不好就想放弃），但开始出现积极想法的萌芽。', 'Help': '鼓励她记录成功经验和掌握的知识点，进行优势聚焦，削弱“我不是学习的料”的概括性结论，运用“理性反驳”的方法，让其具体分析不及格的原因，鼓励她将考试失败归因于具体的学习方法和策略，而非个人能力不足，协助她制定可行的、分步骤的学习计划，例如针对性练习、寻求帮助等，避免再次陷入无助感。’ }

Guide

Step 4: Patient Comfort Simulation

{ 'Comforting_words': '哎，我知道你现在肯定特别难受，感觉所有的努力都白费了，对吧？那种全世界都在嘲笑你的感觉…我也懂！但是不及格真的就代表你彻底完了吗？我觉得不一定。你看，你刚才也说了，上次线代考的不错，高数选择题也对了很多。这说明什么？说明你不是啥都不会，对不对？这次可能只是……只是刚好考的这些题，你还没掌握到技巧而已。“我不是学习的料”？这绝对是胡扯！你想想，你之前背过的单词，做过过的题目，哪一样不是通过学习得来的？这次只是没考好，不代表你不行，只能说明你需要换个方法。要不，我们一起想想，这次考试是哪个部分丢分最多？是概念没搞清楚，还是题型不熟练？如果是概念，我们就重新梳理知识点；如果是题型，就找类似的题多做几遍。’ }

User

Step 5: Planned Story

{ 'Next_scene': '女孩抬起头，眼中闪过一丝疑惑，但仍然带着不确定。“可是…就算这次分析了原因，下次呢？万一下次又考砸了，我岂不是又要经历一次这样的打击？”她声音很小，带着深深的自我怀疑。她翻开书，看到之前做错的题目，还是觉得困难重重。', 'Next_thoughts': '女孩开始反思，但仍然被过去的失败经验所困扰，虽然理智上认同可以分析原因，但情绪上仍然害怕再次失败。她开始意识到“下次还会失败”也是一种过度概括，但仍然难以摆脱这种恐惧。', 'Is_end': 'No ' }

strategist

Case Study (Round 3)

Step 1: Generated Scene

{ 'Scene': '凌晨三点的自习室静得可怕，只有偶尔翻书的声音和笔尖划过纸面的沙沙声。大部分座位空着，只零星分布着几个同样疲惫的身影。女孩抬起头，眼中闪过一丝疑惑，但仍然带着不确定。“可是…就算这次分析了原因，下次呢？万一下次又考砸了，我岂不是又要经历一次这样的打击？”她声音很小，带着深深的自我怀疑。她翻开书，看到之前做错的题目，还是觉得困难重重。窗外，一盏路灯孤零零地照亮着空无一人的街道，也照亮了她桌面上堆积如山的复习资料。几张被揉皱的草稿纸散落在书本旁，上面密密麻麻地写满了公式和演算过程，却仍然无法解开眼前的难题。她感到胸口仿佛压着一块巨石，沉重得让她喘不过气。周围的寂静更加凸显了她内心的焦躁和无力。’ }

Trigger

Step 2: Identified Cognitive Distortion

{ 'Thoughts': '可能确实是我自己太在意了吧。线代上次考得好…嗯，选择题对了…但那也不能说明什么啊，万一这次高数更难了呢？技巧…技巧什么的，我真的能掌握吗？感觉好难啊。要是一直学不会，我…我会不会真的毕不了业？重梳理知识点？重新做题？说起来容易…可是…那么多知识点，那么多题…我什么时候才能做完啊？感觉时间完全不够用啊。’ }

Devil

Step 3: Guidance for Patient

{ 'SummaryScene': '凌晨三点，自习室空旷，女孩因考试失利，对未来学习和毕业充满不确定和自我怀疑，被难题和堆积如山的资料压垮，感到焦躁无力。', 'SummaryThoughts': '患者存在过度概括（“万一我就是飞不起来的笨鸟呢？”）、非此即彼（“我会不会真的毕不了业？”）、情绪化推理（“感觉时间完全不够用”）等认知扭曲，沉浸在负面情绪中，缺乏应对挑战的信心。', 'Help': '首先，引导她识别并标记出引起她焦虑和自我怀疑的具体想法，例如“万一这次高数更难了呢？”“我会不会真的毕不了业？”等，然后针对每个想法，运用苏格拉底式提问，例如：“高数更难的可能性有多大？有什么证据支持或反驳这个想法？”“即使真的毕不了业，会有哪些其他的可能性和出路？”“毕业是衡量个人价值的唯一标准吗？”引导她从不同角度看待问题，寻找更客观、积极的解释。’ }

Guide

Step 4: Patient Comfort Simulation

{ 'Comforting words': '哎，我太懂这种感觉了！凌晨三点的自习室，简直是自我怀疑的温床。你是不是也在想“万一这次高数更难了呢？”“我会不会真的毕不了业？”对不对？先深呼吸一下，这种“万一”的想法就像影子一样，总在你最累的时候跳出来吓唬你。但“万一”只是“万一”，它发生的概率有多大呢？上次线代选择题你不是做得挺好的吗？至少证明你对某些知识点是掌握的啊！而且，高数就算难，大家都一样，都在一条船上，怕什么？至于“毕不了业”…真的会那么糟糕吗？毕业证是衡量一切的标准吗？条条大路通罗马，没准你的人生还有更精彩的可能呢！现在想这些是不是太早了？先专注眼前的题目，一个个知识点啃下去，总比在这空想“万一”要好。笨鸟先飞，起码飞了还有希望，一直窝着，那就真的只能当只笨鸟了。加油！我也一样！’ }

User

Step 5: Planned Story

{ 'Next scene': '女孩听着安慰，眼眶泛红，疲惫地揉了揉眼睛，抬头看了看周围同样奋战的同学们，有人打着哈欠，有人在用红牛续命，还有人在小声背着公式。她突然意识到，自己不是一个人在战斗。她深吸一口气，重新拿起笔，开始攻克一道困扰她许久的例题。虽然一开始仍然磕磕绊绊，但她想起刚才鼓励的话，告诉自己笨鸟先飞，一点点啃下去。', 'Next thoughts': '女孩开始意识到自己的“万一”想法是不理智的，高数难是普遍现象，并非只有自己如此。虽然毕业压力依然存在，但她开始尝试接受不确定性，并将注意力集中到解决眼前的难题上，决心通过努力来改变现状，而非沉溺于空想和焦虑。她开始相信，行动比空想更有价值。', 'Is_end': 'No' }

strategist

Case Study (Round 4)

Step 1: Generated Scene

{ 'Scene': '凌晨两点半的自习室，空气中弥漫着咖啡的味道。日光灯惨白的光线洒在桌面上，反射出书本的光泽。女孩听着鼓励的话语，眼眶泛红，疲惫地揉了揉酸涩的眼睛，抬头看了看周围同样奋战的同学们。角落里，一个男生正对着电脑屏幕，手指飞快地敲击着键盘，似乎在赶着论文；另一边，一个女生戴着耳机，身体随着音乐轻轻摇晃，但手里的笔却不停地在草稿纸上演算着。还有几个同学已经趴在桌子上睡着了，桌面上凌乱地堆放着书本和资料。空气中弥漫着一股紧张而压抑的气氛。她突然意识到，自己不是一个人在战斗。她深吸一口气，重新拿起笔，开始攻克一道困扰她许久的例题。题目依旧复杂，数字和公式像一个个小恶魔，在她眼前跳动。她想起刚才鼓励的话，告诉自己要一点点啃下去。她放下心中的焦虑，将大问题拆解成小步骤，耐心地一步一步地尝试。时间一分一秒地过去，笔尖在纸上划过，发出细微的沙沙声。'} }

Trigger

Step 2: Identified Cognitive Distortion

{ 'Thoughts': '嗯...你说的对，深呼吸一下，好像真的没那么糟...大家都在一条船上？好像也是，看他们一个个愁眉苦脸的，估计也跟我差不多。毕业...毕不了业...真的不敢想。虽然你说条条大路通罗马，可是...可是没有毕业证，感觉好多路都被堵死了啊。唉，算了，想那么远干嘛，先把眼前的题目搞定再说。笨鸟先飞...总比我在这里自怨自艾好。你说的“万一”就像影子一样，对，太对了！它总是在我最累的时候跳出来吓我。但是...深呼吸，这次我要试着不被它吓到。一个个知识点啃下去...好吧，我试试。谢谢你，感觉好多了。'} }

Devil

Step 3: Guidance for Patient

{ 'SummaryScene': '凌晨自习室，女孩被难题困扰，虽受鼓励但仍有焦虑，对未来感到迷茫。', 'SummaryThoughts': '患者虽认同鼓励，但仍受困于“选择题侥幸”、“毕不了业”等负面想法，存在过度概括和灾难化思维。', 'Help': '引导她识别“没有毕业证好多路都被堵死”的非此即彼思维，提问：“没有毕业证就真的所有路都堵死了吗？有没有其他可能性？”，鼓励她寻找替代方案，例如提升其他技能、考取相关证书等；针对“毕不了业”的灾难化思维，使用“可能性评估”：询问她：“毕不了业的可能性有多大？根据以往的成绩、学习情况和出勤率，你认为可能性是极高、较高、一般还是较低？”，并引导她寻找支持或反驳该想法的证据，从而降低焦虑感。'} }

Guide

Step 4: Patient Comfort Simulation

{ 'Comforting_words': '哎，没毕业证就完蛋的想法，我也经常冒出来。感觉一下子掉进深渊，对不对？但是，等等！没毕业证...就真的所有路都堵死了吗？想想，总有例外的吧？像谁谁谁，虽然没正经学历，但是凭着一手绝活，现在混得风生水起。你想啊，就算真没毕业证，我们还能提升其他技能啊！考个证书，学个编程，说不定比那张纸更有用呢！而且，毕不了业...这个可能性有多大？想想你之前的考试，也有考得好的时候吧？平时上课也挺认真的，作业也按时交，你说毕不了业，是不是有点太夸张了？那个万一...它就是个捣乱鬼！每次我熬夜的时候也跑出来吓唬我。别理它！深呼吸，一个个知识点啃下去，就算真有啥意外，船到桥头自然直！先把眼前这道题搞定，好不好？'} }

User

Step 5: Planned Story

{ 'Next_scene': '女孩抬起头，揉了揉酸涩的眼睛，看向电脑屏幕上的题目。她深吸一口气，想起用户说的“船到桥头自然直”，决定先把这道题搞定。她开始认真分析题目，尝试回忆相关的知识点。尽管一开始仍然有些吃力，但她逐渐进入状态，思路也清晰起来。', 'Next_thoughts': '女孩开始反思自己之前的想法，意识到“毕不了业就完蛋”的想法过于绝对。她想起用户的鼓励，意识到即使没有毕业证，也有其他提升自身价值的方式。同时，她也意识到自己之前总是盯着负面结果，忽略了自己曾经取得的进步。她开始相信，只要努力，就有可能克服困难。', 'Is_end': 'No' }

strategist