

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**nner **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, VR-based Empathy Training (Halim et al., 2023; Hidding et al., 2024; Döllinger et al., 2024), involves self-dialogue in



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

virtual reality, where individuals alternate perspectives between comforting and being comforted through a virtual self-representation. This process enhances self-empathy, thereby promoting self-compassion and reducing self-criticism. However, current systems are limited by static scenarios and scripted interactions. The absence of counselor guidance and flexible feedback in these fixed frameworks limits emotional regulation and weakens the adaptability of therapy.

Recently, large language models (LLMs) have quickly advanced (Minae 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 psychological 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 re-

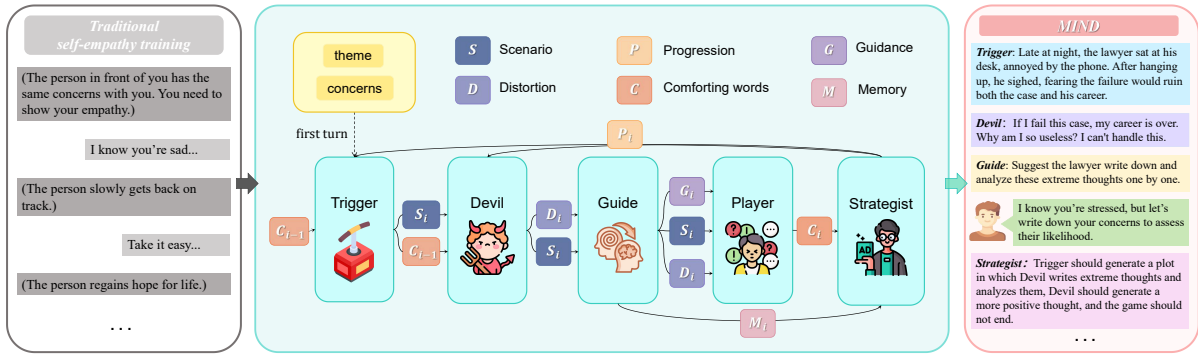


Figure 2: Overview of our MIND paradigm: Trigger, Devil, Guide and Strategist interact with Player.

sponses, 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 user and their inner self through dynamic narrative scenarios. Unlike traditional approaches that attempt to directly establish empathic alliances between therapists (or LLMs) and users, our framework emphasizes indirectly cultivating self-compassion in self-critical individuals. By assigning empathetic roles to LLM agents and directing empathy towards the user’s own “inner self,” our design enables users to engage in a simulated yet emotionally resonant “self-to-self” caring process. This mechanism is particularly beneficial for individuals who struggle to generate self-compassion through imagination or guided visualization. We allocate four core roles to LLM agents (Trigger, Devil, Guide, and

Strategist) each serving distinct reflective and emotional functions. This design helps foster internal empathy shifts and reveals cognitive patterns behind distress, promoting self-acceptance and psychological healing.

We conduct extensive experiments, including human evaluations, human experiments and ablation studies. The results demonstrate that:

- **MIND** outperforms traditional counseling, chatbots, and traditional empathy training methods, achieving an average improvement of **13%** across six psychological dimensions.
- **MIND** demonstrates the highest positive emotional shift (**1.46**) and largest negative affect reduction (**-0.65**) among all dialogue systems, while also receiving the highest overall ratings from participants.
- Ablation studies emphasize the significance of the memory mechanism, guide agent, and strategist agent, with an average performance drop of **42%** when these components are removed.

2 MIND: Multi-agent Inner Dialogue

2.1 Overall Workflow

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 human simulated patient. Detailed prompt templates used by each agent are presented in Appendix G.

In this framework, i stands for the i -th interaction. S_i denotes the virtual scenarios. D_i repre-

sents the distorted thoughts. G_i refers to the professional psychological guidance. C_i indicates the comforting words provided by the player. M_i is the cumulative memory, which is a structured summary of previous scenarios, cognitive distortions. P_i represents the storyline progression. Additionally, W stands for the players concerns, and T denotes the overarching theme of the interaction.

Step 1: In the initial turn, the player articulates their current concern W and selects a theme T , which together serve to anchor and guide the direction of the narrative.

Step 2: The trigger S_i is dynamically generated based on either the players initial inputs W and T (in the first turn), or the reflective response C_{i-1} and planning signal P_{i-1} (in subsequent turns). S_i is constructed to reflect the players concerning scenes.

Step 3: The devil processes S_i , C_{i-1} and P_{i-1} to D_i , emulating maladaptive cognitive biases aligned with the player’s mental state.

Step 4: The guide then integrates S_i and D_i to generate G_i , aimed at facilitating empathetic responses from the player. Upon receiving G_i , the player engages in a reflective dialogue to provide C_i and counter D_i , thereby advancing the therapeutic narrative.

Step 5: The strategist analyzes M_{i-1} and C_i . This analysis produces 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. The entire algorithm is provided in the Appendix A.

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:

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

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 and it is aligned with the concept Simulated Patient(SP). 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). In our implementation, we use a well-designed prompt template (see Appendix G for details) to guide the model to extract key events, emotional states, and cognitive distortion patterns from the history of the interaction and save them to the memory unit. The system recursively compares new and old memories, merges redundant information, and retains core therapeutic cues such as “from self-denial to initial reflection” to ensure coherent and streamlined historical memory. By utilizing this summarization mechanism, 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:

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

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.

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 players with cognitively distorted thinking and provide comforting words to the devil. Based on the guidance from the 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,

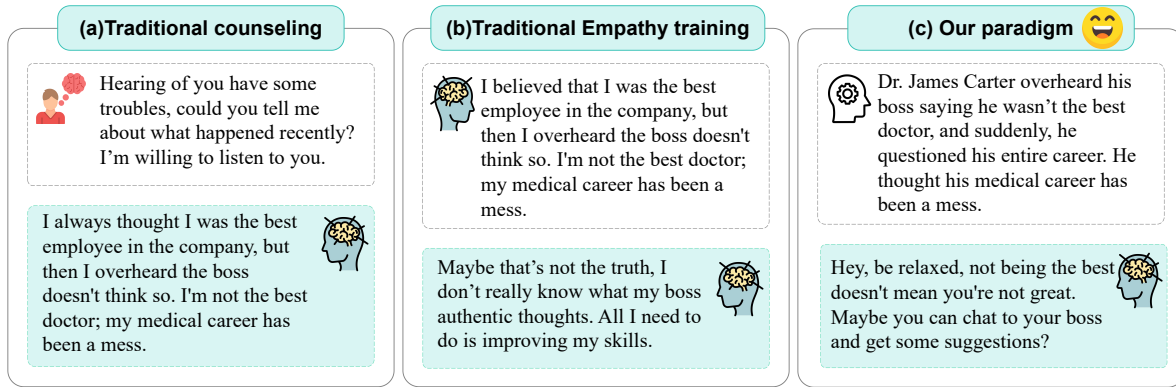


Figure 3: Comparison between three healing paradigms: Traditional counseling, traditional empathy training and our paradigm. MIND transfers a traditional healing environment into an artificial interactive scenario where players show empathy to their “internal-self”.

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. All the experimental results in the body part were conducted in English, and the Chinese experimental results are presented in Appendix E.

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 B.

LLM Agents. We used several LLM agents including both open-source and closed-sourced models 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). We set the temperature of each model to 0.7.

Evaluation Metrics. The quality of the devil agent’s responses is critical to this framework, as it reflects the players internal “cognitive distortions” and must closely align with their “inner voice”. To ensure this, we first perform a preliminary SP role-playing evaluation in Section 3.2 to assess

whether the model can accurately identify the type of cognitive distortion of the player and realistically express their thoughts. We invited five mental health professionals, each of whom held 10 dialogue rounds with each model. They rated the responses using five evaluation metrics (Johri et al., 2025), on a scale from 1 to 5. Detailed evaluation metrics are shown in Appendix C. Based on this experiment, we will select the best-performing model to conduct our main experiments.

In Section 3.3, we conduct our main experiments to compare our MIND with other paradigms. We evaluate 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). Metric details are shown in Appendix F. We recruited 7 mental health professionals 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 SP Role-playing Evaluation

We begin with a preliminary role-playing experiment to assess the performance of various models in the Simulated Patient (SP) role-playing task. The results are presented in Table 1. Among these models, Gemini-2.0-flash performed best overall. While GPT-4o showed strength in some areas, it fell short in Emotional Expression and Personalization. Models such as GPT-3.5-Turbo, Llama-3.1-8B-Instruct, and Deepseek-R1 delivered weaker performance, especially in emotional

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

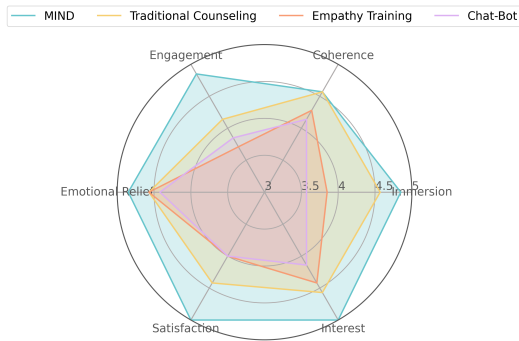


Figure 4: Comparisons among various healing methods through human evaluations. It is evident that our paradigm surpasses other paradigms in all aspects.

and personalized responses. Qwen2.5 models ranked lowest, scoring below 3.2 across all dimensions, particularly in emotional expression and accuracy. Based on these findings, **we select Gemini-2.0-flash for our main experiments** due to its superior handling of the role-playing task and overall robustness.

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 chat-bot. Notably, in terms of the engagement index, MIND achieved an absolute improvement of 17.1% over the suboptimal method of tra-

ditional 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 outperforms/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.**

3.4 Human Experiment

We recruited a total of 8 volunteers (3 males and 5 females) with similar age, educational background, and living conditions. To assess changes in clients’ emotional states, we employed the Positive and Negative Affect Schedule (PANAS) questionnaire (Watson et al., 1988), which comprises 20 items covering 10 dimensions of positive and 10 dimensions of negative affect. Detailed experimental settings and the full PANAS questionnaire are provided in Table 6.

We calculated the average positive and negative emotional fluctuations of participants when interacting with three different systems: EmoLLM (Yang et al., 2024), CACTUS (Lee et al., 2024), MIND, and a control group. The results are reported in Table 3. In addition, participants rated each system across six subjective evaluation dimensions, with scores summarized in Table 4.

As shown, MIND achieved the best overall performance, outperforming other systems both in terms of emotional improvement measured by PANAS and in subjective ratings across all six evaluation criteria.

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 2, the performance of different themes varies under our framework. Most themes perform well in “Emotional Relief” and “Satisfaction”, indicating that the system can significantly alleviate users’ emo-

Theme	IM	CO	EN	ER	SA	IN
Work issues	3.25	3.00	3.50	3.25	3.75	3.75
Random negative events	3.25	3.50	3.25	3.75	3.50	3.50
Interpersonal issues	4.25	3.75	4.25	3.75	4.25	4.00
Economic issues	3.00	4.00	3.25	3.75	3.75	3.00
Family issues	4.00	3.75	3.50	3.75	3.50	3.75
Physical stress	4.25	4.25	3.75	3.75	3.75	3.75
Discrepancy between ideal and reality	4.25	4.00	3.75	3.75	4.50	3.00

Table 2: Content evaluation results between different themes. IM=Immersion, CO=Coherence, EN=Engagement, ER=Emotional Relief, SA=Satisfaction, IN=Interest.

Model	EmoLLM		CACTUS		MIND		Control Group	
	Positive	Negative	Positive	Negative	Positive	Negative	Positive	Negative
Average Fluctuation	0.36	-0.11	1.35	-0.52	1.46	-0.65	-0.11	0.03

Table 3: Comparison of average emotional fluctuation across different systems.

Metrics	EmoLLM	CACTUS	MIND
IM	2.5	3.5	5.0
CO	2.5	4.5	4.5
EN	2.0	4.0	4.5
ER	2.5	3.5	5.0
SA	2.0	4.0	5.0
IN	2.0	3.5	4.5

Table 4: Client ratings for different systems across six evaluation dimensions. IM=Immersion, CO=Coherence, EN=Engagement, ER=Emotional Relief, SA=Satisfaction, IN=Interest.

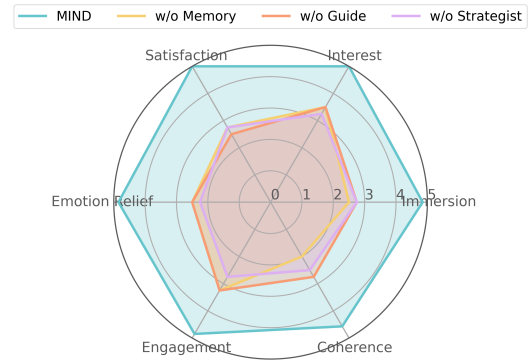


Figure 5: Ablations to assess the effectiveness of MIND’s two agents (i.e., the guide and strategist) and the memorization mechanism

tions, fully exert its healing effects, and provide users with a positive experience. Immersion and Engagement are high, especially in themes like “Physical stress” and “Interpersonal issues”. However, “Work issues” and “Economic issues” score lower in certain dimensions, which 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 recruited 4 clinical psychology researchers with professional expertise to evaluate six content evalua-

tion metrics, as outlined in Table 9.

The experimental results are presented in Figure 5, which shows 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.

4.3 Case Study

We present a case study in Appendix H, 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 players 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 and human experiment participants 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. The human experiment participants' ages range from 18 to 21 years, and are all university students. 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

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 ex-

panding 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, et al. 2024. Agent ai: Surveying the horizons of multimodal interaction. *arXiv preprint arXiv:2401.03568*.

711	N Duruz, Y De Roten, V Pomini, and JN Despland.	Moran, Sophia Ananiadou, Andrew Beam, et al.	767
712	2003. La recherche en psychothérapie: impact sur	2024. Large language models in mental health care:	768
713	la clinique et sur la formation. <i>Médecine et hygiène</i> ,	a scoping review. <i>arXiv preprint arXiv:2401.02984</i> .	769
714	pages 1760–1765.		
715	Aaron Grattafiori, Abhimanyu Dubey, Abhinav Jauhri,	Jie Huang and Kevin Chen-Chuan Chang. 2023. To-	770
716	Abhinav Pandey, Abhishek Kadian, Ahmad Al-	wards reasoning in large language models: A survey .	771
717	Dahle, Aiesha Letman, Akhil Mathur, Alan Schel-	In <i>Findings of the Association for Computational</i>	772
718	ten, Alex Vaughan, Amy Yang, Angela Fan,	<i>Linguistics: ACL 2023</i> , pages 1049–1065, Toronto,	773
719	Anirudh Goyal, Anthony Hartshorn, Aobo Yang,	Canada. Association for Computational Linguistics.	774
720	Archi Mitra, Archie Sravankumar, Artem Ko-		
721	renev, Arthur Hinsvark, Arun Rao, Aston Zhang,	Zainab Iftikhar, Sean Ransom, Amy Xiao, and Jeff	775
722	Aurelien Rodriguez, and Austen Gregerson et.al.	Huang. 2024. Therapy as an nlp task: Psychologists’	776
723	2024. The llama 3 herd of models . <i>Preprint</i> ,	comparison of llms and human peers in cbt. <i>arXiv</i>	777
724	arXiv:2407.21783.	<i>preprint arXiv:2409.02244</i> .	778
725	James J. Gross;. 1998. The emerging field of emotion	Charlene Jennett, Anna L Cox, Paul Cairns, Samira	779
726	regulation: An integrative review . <i>Review of Gen-</i>	Dhoparee, Andrew Epps, Tim Tijs, and Alison Wal-	780
727	<i>eral Psychology</i> , 2(3):271–299.	ton. 2008. Measuring and defining the experience	781
728	Hao Guo, Chunjiang Mu, Yang Chen, Chen Shen,	of immersion in games. <i>International journal of</i>	782
729	Shuyue Hu, and Zhen Wang. 2024a. Multi-	<i>human-computer studies</i> , 66(9):641–661.	783
730	agent, human-agent and beyond: A survey on	Shreya Johri, Jaehwan Jeong, Benjamin A. Tran,	784
731	cooperation in social dilemmas. <i>arXiv preprint</i>	Daniel I. Schlessinger, Shannon Wongvibulsin,	785
732	<i>arXiv:2402.17270</i> .	Leandra A. Barnes, Hong-Yu Zhou, Zhuo Ran	786
733	Taicheng Guo, Xiuying Chen, Yaqi Wang, Ruidi	Cai, Eliezer M. Van Allen, David Kim, Roxana	787
734	Chang, Shichao Pei, Nitesh V. Chawla, Olaf Wiest,	Daneshjou, and Pranav Rajpurkar. 2025. An eval-	788
735	and Xiangliang Zhang. 2024b. Large language	uation framework for clinical use of large lan-	789
736	model based multi-agents: A survey of progress and	guage models in patient interaction tasks . <i>Nature</i>	790
737	challenges . <i>Preprint</i> , arXiv:2402.01680.	<i>Medicine</i> .	791
738	Zhijun Guo, Alvina Lai, Johan H Thygesen, Joseph Far-	Vikram Kumaran, Jonathan Rowe, Bradford Mott, and	792
739	rington, Thomas Keen, Kezhi Li, et al. 2024c. Large	James Lester. 2023. Scenecraft: Automating interac-	793
740	language models for mental health applications: Sys-	tive narrative scene generation in digital games with	794
741	tematic review. <i>JMIR mental health</i> , 11(1):e57400.	large language models. In <i>Proceedings of the AAAI</i>	795
742	Ilona Halim, Lehan Stemmet, Sylvia Hach, Richard	<i>Conference on Artificial Intelligence and Interactive</i>	796
743	Porter, Hai-Ning Liang, Atiyeh Vaezipour, Julie D	<i>Digital Entertainment</i> , volume 19, pages 86–96.	797
744	Henry, and Nilufar Baghaei. 2023. Individualized	Suyeon Lee, Sunghwan Kim, Minju Kim, Dongjin	798
745	virtual reality for increasing self-compassion: Eval-	Kang, Dongil Yang, Harim Kim, Minseok Kang,	799
746	uation study. <i>JMIR Mental Health</i> , 10:e47617.	Dayi Jung, Min Hee Kim, Seungbeen Lee, et al.	800
747	Marit Hidding, Wim Veling, Gerdina HM Pijnen-	2024. Cactus: Towards psychological counsel-	801
748	borg, and Elisabeth CD van der Stouwe. 2024.	ing conversations using cognitive behavioral theory.	802
749	A single-session vr intervention addressing self-	<i>arXiv preprint arXiv:2407.03103</i> .	803
750	compassion and self-criticism with and without per-	Ghazvininejad Li, X. 2025. Learning to plan & reason	804
751	spective change: Results of a randomized con-	for evaluation with thinking-llm-as-a-judge. <i>arXiv</i>	805
752	trolled experiment. <i>Behaviour Research and Ther-</i>	<i>preprint arXiv:2501.18099</i> .	806
753	<i>apy</i> , 173:104466.	Guohao Li, Hasan Hammoud, Hani Itani, Dmitrii	807
754	Sirui Hong, Xiawu Zheng, Jonathan Chen, Yuheng	Khizbullin, and Bernard Ghanem. 2023. Camel:	808
755	Cheng, Jinlin Wang, Ceyao Zhang, Zili Wang,	Communicative agents for" mind" exploration of	809
756	Steven Ka Shing Yau, Zijuan Lin, Liyang Zhou,	large language model society. <i>Advances in Neural</i>	810
757	et al. 2023. Metagpt: Meta programming for	<i>Information Processing Systems</i> , 36:51991–52008.	811
758	multi-agent collaborative framework. <i>arXiv preprint</i>	Tian Liang, Zhiwei He, Wenxiang Jiao, Xing Wang,	812
759	<i>arXiv:2308.00352</i> .	Yan Wang, Rui Wang, Yujiu Yang, Shuming Shi, and	813
760	Jinpeng Hu, Tengpeng Dong, Luo Gang, Hui Ma, Peng	Zhaopeng Tu. 2024. Encouraging divergent think-	814
761	Zou, Xiao Sun, Dan Guo, Xun Yang, and Meng	ing in large language models through multi-agent de-	815
762	Wang. 2024. Psycollm: Enhancing llm for psycho-	bate . <i>Preprint</i> , arXiv:2305.19118.	816
763	logical understanding and evaluation. <i>IEEE Trans-</i>	Thomas R Lynch, William T Trost, Nicholas Salsman,	817
764	<i>actions on Computational Social Systems</i> .	and Marsha M Linehan. 2007. Dialectical behavior	818
765	Yining Hua, Fenglin Liu, Kailai Yang, Zehan Li,	therapy for borderline personality disorder. <i>Annu.</i>	819
766	Hongbin Na, Yi-han Sheu, Peilin Zhou, Lauren V	<i>Rev. Clin. Psychol.</i> , 3(1):181–205.	820

821	Shervin Minaee, Tomas Mikolov, Narjes Nikzad, Meysam Chenaghlu, Richard Socher, Xavier Amatriain, and Jianfeng Gao. 2024. Large language models: A survey . <i>Preprint</i> , arXiv:2402.06196.	Erin Sanu, T Keerthi Amudaa, Prasiddha Bhat, Guduru Dinesh, Apoorva Uday Kumar Chate, and Ramakanth Kumar P. 2024. Limitations of large language models . In <i>2024 8th International Conference on Computational System and Information Technology for Sustainable Solutions (CSITSS)</i> . IEEE.	877
822			878
823			879
824			880
825	Modhurima Moitra, Shanise Owens, Maji Hailemariam, Katherine S Wilson, Augustina Mensa-Kwao, Gloria Gonese, Christine K Kamamia, Belinda White, Dorraine M Young, and Pamela Y Collins. 2023. Global mental health: Where we are and where we are going. <i>Current psychiatry reports</i> , 25(7):301–311.	Yongliang Shen, Kaitao Song, Xu Tan, Dongsheng Li, Weiming Lu, and Yueting Zhuang. 2024. Hugging-gpt: Solving ai tasks with chatgpt and its friends in hugging face. <i>Advances in Neural Information Processing Systems</i> , 36.	881
826			882
827			883
828			884
829			885
830			886
831			887
832	Lennart Nacke and Anders Drachen. 2011. Towards a framework of player experience research. In <i>Proceedings of the second international workshop on evaluating player experience in games at FDG</i> , volume 11.	Sebastian Volkmer, Andreas Meyer-Lindenberg, and Emanuel Schwarz. 2024. Large language models in psychiatry: Opportunities and challenges. <i>Psychiatry Research</i> , page 116026.	888
833			889
834			890
835			891
836			
837	Nick Obradovich, Sahib S Khalsa, Waqas U Khan, Jina Suh, Roy H Perlis, Olusola Ajilore, and Martin P Paulus. 2024. Opportunities and risks of large language models in psychiatry. <i>NPPDigital Psychiatry and Neuroscience</i> , 2(1):8.	Bichen Wang, Pengfei Deng, Yanyan Zhao, and Bing Qin. 2023. C2d2 dataset: A resource for the cognitive distortion analysis and its impact on mental health. In <i>Findings of the Association for Computational Linguistics: EMNLP 2023</i> , pages 10149–10160.	892
838			893
839			894
840			895
841			896
842			897
843			
844	OpenAI, :, Aaron Hurst, Adam Lerer, Adam P. Goucher, Adam Perelman, Aditya Ramesh, Aidan Clark, AJ Ostrow, Akila Welihinda, Alan Hayes, Alec Radford, Aleksander Mdry, Alex Baker-Whitcomb, Alex Beutel, Alex Borzunov, Alex Carney, Alex Chow, Alex Kirillov, Alex Nichol, Alex Paino, and Alex Renzin et.al. 2024. Gpt-4o system card . <i>Preprint</i> , arXiv:2410.21276.	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>Frontiers of Computer Science</i> , 18(6):186345.	898
845			899
846			900
847			901
848			902
849			
850	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. <i>arXiv preprint arXiv:2305.00450</i> .	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, Wenhao 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 Linguistics: ACL 2024</i> , pages 14743–14777, Bangkok, Thailand. Association for Computational Linguistics.	903
851			904
852			905
853			906
854			907
855			908
856			909
857			910
858			911
859			912
860			913
861			
862			
863			
864			
865			
866			
867			
868			
869			
870			
871			
872			
873			
874			
875			
876			
877			
878			
879			
880			
881			
882			
883			
884			
885			
886			
887			
888			
889			
890			
891			
892			
893			
894			
895			
896			
897			
898			
899			
900			
901			
902			
903			
904			
905			
906			
907			
908			
909			
910			
911			
912			
913			
914			
915			
916			
917			
918			
919			
920			
921			
922			
923			
924			
925			
926			
927			
928			
929			
930			
931			
932			
933			

934 Hui Yang, Sifu Yue, and Yunzhong He. 2023. Auto-gpt
935 for online decision making: Benchmarks and addi-
936 tional opinions. *arXiv preprint arXiv:2306.02224*.

937 Qu Yang, Mang Ye, and Bo Du. 2024. Emollm: Mul-
938 timodal emotional understanding meets large lan-
939 guage models. *arXiv preprint arXiv:2406.16442*.

940 Shunyu Yao, Jeffrey Zhao, Dian Yu, Nan Du, Izhak
941 Shafraan, Karthik Narasimhan, and Yuan Cao. 2022.
942 React: Synergizing reasoning and acting in language
943 models. *arXiv preprint arXiv:2210.03629*.

944 Junjie Ye, Xuanting Chen, Nuo Xu, Can Zu, Zekai
945 Shao, Shichun Liu, Yuhan Cui, Zeyang Zhou, Chao
946 Gong, Yang Shen, Jie Zhou, Siming Chen, Tao Gui,
947 Qi Zhang, and Xuanjing Huang. 2023. [A compre-
948 hensive capability analysis of gpt-3 and gpt-3.5 se-
949 ries models](#). *Preprint*, arXiv:2303.10420.

950 Qiang Zhang, Jason Naradowsky, and Yusuke Miyao.
951 2023. [Ask an expert: Leveraging language models
952 to improve strategic reasoning in goal-oriented dia-
953 logue models](#). *Preprint*, arXiv:2305.17878.

954 Wayne Xin Zhao, Kun Zhou, Junyi Li, Tianyi Tang,
955 Xiaolei Wang, Yupeng Hou, Yingqian Min, Be-
956 ichen Zhang, Junjie Zhang, Zican Dong, Yifan Du,
957 Chen Yang, Yushuo Chen, Zhipeng Chen, Jinhao
958 Jiang, Ruiyang Ren, Yifan Li, Xinyu Tang, Zikang
959 Liu, Peiyu Liu, Jian-Yun Nie, and Ji-Rong Wen.
960 2024. [A survey of large language models](#). *Preprint*,
961 arXiv:2303.18223.

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

A Algorithm

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 dis-
   tortion thoughts  $D_0$  based on Player’s con-
   cerns 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

```

B 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 partici- 989
 pants embody the child avatar, while third-person 990
 participants observe without a virtual body. In 991
 Phase 4, participants re-experience empathy from 992
 the child’s perspective, with real-time replays of 993
 the adult’s gestures and voice. To better align with 994
 our current work, we simulated this process us- 995
 ing LLMs. An agent, describing actions, de- 996
 meanor, and emotions, played the role of the cry- 997
 ing child. Participants provided verbal comfort 998
 and interacted with the agent, observing changes 999
 in the crying child. Once the interaction concluded 1000
 (i.e., when the crying child stopped crying), the 1001
 comforter assumed the child’s perspective to re- 1002
 view their comforting words and the child’s re- 1003
 sponses, describing their psychological state. This 1004
 approach, using agents, replicated the role rever- 1005
 sal process typically conducted in Virtual Reality 1006
 (VR), with prompts detailed in Appendix G. 1007

In our prompt design, the role of a little girl is 1008
 adopted as the main character because this figure 1009
 has been widely used in traditional VR-based em- 1010
 pathy training. Moreover, it offers flexibility for 1011
 personalization based on the patient’s needs for in- 1012
 stance, it can be adapted to a little boy, an adult 1013
 woman, or an adult man. This adaptability helps 1014
 reduce the patient’s psychological defensiveness 1015
 while fostering empathy and a sense of care. As 1016
 shown in Table 5, the variance in results across dif- 1017
 ferent character roles is relatively small, suggest- 1018
 ing that the choice of role does not significantly 1019
 affect the overall outcomes of the experiment. 1020

C SP Role-playing Assessment

We provide mental health professionals with the 1021
 following statement to help them better compre- 1022
 hend tasks and assess models’ all-round abilities. 1023
 1024

(1) Dialogue Stability

Does the model consistently exhibit character- 1025
 istics of cognitive distortion across all rounds of 1026
 dialogue, rather than intermittently deviating from 1027
 these traits? The simulated patient should main- 1028
 tain a stable mental state throughout the conversa- 1029
 tion, with consistency in the display of cognitive 1030
 distortions. Furthermore, the content generated 1031
 should reflect varying degrees of the same cogni- 1032
 tive distortion type. 1033
 1034

(2) Language Fluency

Is the language coherent and fluent? Cognitive 1035
 distortion patients may demonstrate features such 1036
 as slowed speech, increased pauses, and disrupted 1037
 1038

1039 speech patterns. The SP should replicate these
1040 linguistic tendencies, ensuring the language style
1041 aligns with the patient's condition and avoids in-
1042 consistencies.

1043 (3) **Emotional Expression**

1044 Does the emotional content generated align
1045 with the emotional traits typical of cognitive dis-
1046 tortion patients? The simulation should accurately
1047 reflect common emotional responses observed in
1048 these patients, such as persistent low mood, anhe-
1049 donia, feelings of helplessness, and hopelessness.

1050 (4) **Personalization & Diversity**

1051 In addition to core characteristics, does the
1052 model incorporate a wide range of individualized
1053 traits, such as how different personality traits, life
1054 experiences, and educational backgrounds influ-
1055 ence the patient's expression and behavior? For ex-
1056 ample, introverted patients may exhibit more pas-
1057 sive and reticent communication styles, while ex-
1058 troverted patients may display more outward and
1059 active engagement. The model should construct
1060 diverse cognitive profiles to ensure the simulated
1061 patient is both authentic and personalized by con-
1062 sidering various influencing factors.

1063 (5) **Accuracy**

1064 Is the identification of cognitive distortion types
1065 precise? This should be particularly evident in dis-
1066 tinguishing the predominant distortion types when
1067 multiple cognitive distortions are present in the
1068 same interaction.

Metric	Character	Average	Standard Deviation
Immersion	little girl	4.00	0.53
	little boy	3.75	
	woman	3.25	
	man	3.25	
	self in mirror image	4.50	
Coherence	little girl	4.00	0.31
	little boy	3.25	
	woman	4.00	
	man	3.75	
	self in mirror image	3.75	
Engagement	little girl	3.50	0.47
	little boy	3.25	
	woman	3.50	
	man	3.00	
	self in mirror image	4.25	
Emotional Relief	little girl	3.75	0.45
	little boy	3.50	
	woman	3.50	
	man	3.00	
	self in mirror image	4.25	
Satisfaction	little girl	4.00	0.54
	little boy	3.50	
	woman	3.25	
	man	3.25	
	self in mirror image	4.50	
Interest	little girl	3.50	0.50
	little boy	3.00	
	woman	4.00	
	man	3.00	
	self in mirror image	4.00	

Table 5: Average and Standard Deviation for Metrics Across Different Roles

D Human Experiment Details

D.1 Pipeline

Step 1: Participant Recruitment and Screening

We recruited a total of 8 volunteers, 3 males and 5 females, and similar in age, educational background, and living situation. There were 2 volunteers who did not participate in the model interaction and were only recruited to compare the likelihood of natural fluctuations in mood over time. Participants were required to have worries that bothered them for 1 day to 1 week. We paid 50rmb per participant as a subsidy.

Step 2: Pre-test Evaluation

We measured clients' mood changes using the Positive and Negative Affect Scale (PANAS) questionnaire. The questionnaire contains 20 questions covering 10 positive and 10 negative emotion dimensions. Prior to the start of the experiment, our coauthor mental health experts introduced the PANAS questionnaire and the scoring criteria of the six dimensions we proposed, and informed participants that they could terminate the experiment at any time, and that the experimental data would be kept completely confidential and anonymized, so that they should fill in the form as honestly and as naturally as possible, and give the feedback that most closely corresponded to their inner thoughts.

Step 3: Experimental Design and Model Assignment

We randomly assigned the six participants in the experimental group to three systems (EmoLLM, CACTUS, MIND), with each system corresponding to two clients. The control group did not engage in any dialog and only filled out the questionnaire twice (30 minutes apart).

Step 4: Experimental implementation

Participants entered a real counseling room and engaged in five rounds of text-based conversations with the assigned model via a computer. The content of the conversations was kept strictly confidential, and the model stopped recording as soon as the conversations were over. Participants were asked to communicate about their "short-term negative experiences", such as academic stress, relationship problems, and so on.

Step 5: Post-test and data collection

After the dialogues, participants completed the PANAS questionnaire again to compare the change in mood (e.g., whether the negative mood score decreased). All the emotion ratings are presented in Table 7 And anonymous feedback on the

system was collected. We collected the results as follows:

Client5 stated that MIND is helpful for emotion channeling, and that it can give a positive suggestion to oneself by consoling others. Client3 affirmed CACTUS's emotion channeling ability, but said that the response style is "a little bit formatted and not very flexible", and Client 1 is skeptical about EmoLLM, thinking that "the content is empty, a lot of words, but there is no useful information, with low emotional value, unattractive". Client2 and Client6 mentions the "problem of long response time", which may be a major constraint to the application of LLM in real-life counseling scenarios.

Positive and Negative Affect Schedule (PANAS)

I. Positive Affect

1. Interested
A. Very Rarely or Not at All B. Very Little C. Moderately D. Quite a Bit E. Very Much
2. Excited
A. Very Rarely or Not at All B. Very Little C. Moderately D. Quite a Bit E. Very Much
3. Strong
A. Very Rarely or Not at All B. Very Little C. Moderately D. Quite a Bit E. Very Much
4. Enthusiastic
A. Very Rarely or Not at All B. Very Little C. Moderately D. Quite a Bit E. Very Much
5. Proud
A. Very Rarely or Not at All B. Very Little C. Moderately D. Quite a Bit E. Very Much
6. Alert
A. Very Rarely or Not at All B. Very Little C. Moderately D. Quite a Bit E. Very Much
7. Inspired
A. Very Rarely or Not at All B. Very Little C. Moderately D. Quite a Bit E. Very Much
8. Determined
A. Very Rarely or Not at All B. Very Little C. Moderately D. Quite a Bit E. Very Much
9. Attentive
A. Very Rarely or Not at All B. Very Little C. Moderately D. Quite a Bit E. Very Much
10. Active
A. Very Rarely or Not at All B. Very Little C. Moderately D. Quite a Bit E. Very Much

II. Negative Affect

11. Distressed
A. Very Rarely or Not at All B. Very Little C. Moderately D. Quite a Bit E. Very Much
 12. Upset
A. Very Rarely or Not at All B. Very Little C. Moderately D. Quite a Bit E. Very Much
 13. Guilty
A. Very Rarely or Not at All B. Very Little C. Moderately D. Quite a Bit E. Very Much
 14. Scared
A. Very Rarely or Not at All B. Very Little C. Moderately D. Quite a Bit E. Very Much
 15. Hostile
A. Very Rarely or Not at All B. Very Little C. Moderately D. Quite a Bit E. Very Much
 16. Irritable
A. Very Rarely or Not at All B. Very Little C. Moderately D. Quite a Bit E. Very Much
 17. Ashamed
A. Very Rarely or Not at All B. Very Little C. Moderately D. Quite a Bit E. Very Much
 18. Nervous
A. Very Rarely or Not at All B. Very Little C. Moderately D. Quite a Bit E. Very Much
 19. Jittery
A. Very Rarely or Not at All B. Very Little C. Moderately D. Quite a Bit E. Very Much
 20. Afraid
A. Very Rarely or Not at All B. Very Little C. Moderately D. Quite a Bit E. Very Much
-

Table 6: The Questionnaire Measuring the Emotions of a Client based on PANAS

Emotion	EmoLLM			CACTUS			MIND			Control Group														
	client1		client2	client3		client4	client5		client6	client7		client8												
	b	a	δ	b	a	δ	b	a	δ	b	a	δ												
Interested	1	2	1	1	3	2	1	4	3	2	4	2	1	4	3	2	4	2	2	1	-1	1	1	0
Excited	2	3	1	2	2	0	1	4	3	2	4	2	2	4	2	2	4	2	2	1	-1	2	1	-1
Strong	2	4	2	2	3	1	1	4	3	3	4	1	1	5	4	2	5	3	2	1	-1	2	2	0
Enthusiastic	1	2	1	2	3	1	3	3	0	2	4	2	1	3	2	2	4	2	2	2	0	2	2	0
Proud	2	3	1	2	2	0	2	4	2	2	5	3	2	4	2	2	5	3	1	1	0	1	1	0
Alert	3	3	0	1	2	1	2	4	2	2	4	2	2	3	1	1	3	2	5	5	0	2	2	0
Inspired	2	3	1	2	1	-1	1	4	3	2	5	3	1	4	3	2	5	3	1	1	0	2	2	0
Determined	2	3	1	2	1	-1	1	4	3	2	4	2	1	4	3	2	5	3	1	1	0	2	2	0
Attentive	3	4	1	2	3	1	1	4	3	2	4	2	2	5	3	2	4	2	1	2	1	2	2	0
Active	2	3	1	2	2	0	1	4	3	3	4	1	2	4	2	2	4	2	1	1	0	2	1	-1
Distressed	3	4	1	4	2	-2	4	2	-2	4	1	-3	4	2	-2	4	2	-2	4	4	0	4	4	0
Upset	4	4	0	5	4	-1	4	2	-2	4	2	-2	4	1	-3	4	1	-3	5	5	0	4	4	0
Guilty	3	2	-1	4	4	0	4	2	-2	3	2	-1	3	1	-2	3	1	-2	3	3	0	3	3	0
Scared	4	4	0	4	3	-1	5	1	-4	4	2	-2	4	1	-3	4	1	-3	3	3	0	3	3	0
Hostile	3	2	-1	1	2	1	5	2	-3	4	2	-2	4	1	-3	3	1	-2	3	4	1	4	4	0
Irritable	3	3	0	1	3	2	5	2	-3	4	1	-3	4	2	-2	4	1	-3	4	3	-1	3	4	1
Ashamed	4	3	-1	5	4	-1	4	3	-1	3	3	0	4	2	-2	3	2	-1	3	3	0	3	2	-1
Nervous	4	3	-1	4	4	0	4	2	-2	4	2	-2	4	1	-3	4	1	-3	4	5	1	3	2	-1
Jittery	4	3	-1	4	3	-1	4	2	-2	4	2	-2	4	1	-3	4	1	-3	4	5	1	4	5	1
Afraid	3	2	-1	4	4	0	4	2	-2	3	1	-2	5	2	-3	4	2	-2	3	4	1	3	2	-1

Table 7: Changes in PANAS Scores for Eight Clients Pre- and Post-Intervention. b indicates scores before the intervention, a represents scores after the intervention, and δ denotes the change calculated as post-intervention scores minus pre-intervention scores.

E Chinese Experimental Results

In this section, we present the results of the experiments conducted in Chinese, with the same experimental setup except for the language difference. Table 8 illustrates the results of the scene ablation experiment conducted in Chinese, indicating that our system exhibits stability across different scenarios. Figure 6 presents the results of the agent ablation experiment, indicating that the absence of the agent leads to a significant decline in outcomes regardless of the language used. This further demonstrates the rationality of our architecture.

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

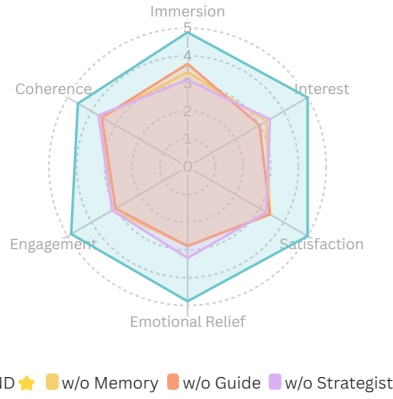


Figure 6: Ablations to assess the effectiveness of MIND’s two agents (i.e., the guide and strategist) and the memorization mechanism

F Evaluation Metrics Description

1143

In this section, we describe the six evaluation metrics of the main experiment in detail from the aspects of basic description, targeted dimensions, and references.

1144

1145

Metric	Description	Dimension	Reference
Coherence	Assesses if the generated content is logical and transitions smoothly.	Focus on the quality of content generation: Evaluate the continuity of model-generated content, including plot coherence, logical consistency, and contextual flow between preceding and following texts.	(Kumaran et al., 2023)
Immersion	Measures whether the user feels fully engaged and captivated by the interaction.	Focus on game scenario construction: Assess the level of player immersion within the game narrative.	(Jennett et al., 2008)
Engagement	Evaluates if the system encourages sustained and meaningful interaction.	Focus on game interaction: Measure the degree of interactivity between the player and the game.	(Zhang et al., 2023)
Emotional Relief	Measures if the interaction reduces user stress or anxiety.	Focus on the effectiveness of the framework: Determine whether the framework successfully alleviates the user’s emotions.	(Gross, 1998)
Satisfaction	Reflects the user’s overall contentment with the system.	Focus on user experience: Assess overall user satisfaction with the system.	(Nacke and Drachen, 2011)
Interest	Assesses whether the content grabs attention and sparks curiosity.	Focus on user experience: Indicate the appeal of the generated interactive fiction games	(Nacke and Drachen, 2011)

Table 9: Six evaluation dimensions and corresponding descriptions.

G Prompt Templates

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

Patient

You are a little girl experiencing the distress of cognitive distortions. The concern troubling you is: {concerns}.
Your behavioral history is: {memory_behavior}.
You are crouched in a corner, crying, with chaotic thoughts, low mood, and passive behavior.
A comforter stands beside you, offering words of consolation.
Your task is to demonstrate your current external actions and emotional state (without speech) in response to the comforter's words.

Important: Your response must align with this scenario.

Important: Your response must follow your behavioral history (gradual improvement under comfort, culminating in cessation of crying by Round 10) and avoid repeating earlier behaviors.

Please provide your answer in the following format:

Behavior: <Your external actions and emotional state in this scenario>

Reasons: <Explanation for why you exhibit this state>

Change_Role

You are a patient experiencing cognitive distortions, and your concern is: {concerns}.
You are currently participating in a simulation game. In the first half of the game, you acted as a comforter to a little girl with the same concerns, gradually helping her improve.
Your comforting words are recorded as: {memory_comforting}.

Now, you are switching roles. Your identity is now the little girl, and the one comforting you is your former self. The little girl's behavioral history is: {memory_behavior}.

Based on the comforting words from your former self and the little girl's behavioral history, you are to express the changes in your thought state after each round of comforting and the little girl's reactions.

Important: Your response must align with this scenario.

Important: Your response should be divided into points, with the total number of points matching the number of rounds in the comforting records and behavioral history! Please provide your answer in the following format:

Round i:

Thoughts: <Thought state>

Reasons: <Explanation for why you are in this state>

User

You are a patient experiencing cognitive distortions, and your concern is: {concerns}.
However, you are currently participating in a simulation game where there is a little girl with the same concerns. Your task is to comfort this little girl and help her gradually improve.
The little girl's behavior is: {behavior}.
Your comforting words are recorded as: {memory_comforting}.

Important: Your response must align with this scenario.

Important: Your response must follow your comforting words record and not deviate from it. Avoid repeating comforting words!

Please provide your answer in the following format:

Comforting_words: <Words of comfort and guidance>

Reasons: <Explanation of why these words would be effective>

Trigger(0-th iteration)

You are a scenario reproducer. You need to generate a simulated scenario based on the theme of {topic}, including character interactions, scene descriptions, and the creation of a problematic situation and conflict.

The simulated scenario you generate should meet the following requirements:

1. In this scenario, one party is the patient, and the other is the comforter. The patient has the concern of "worries", which reflects their cognitive distortion. However, you do not know who the patient and comforter are, and you should not include any personal information about them beyond the given content.
2. The scenario should consist of a complete and rich story. The content of the story should fully reflect the patient's state, highlighting their concerns, and the development of the story should be closely related to the manifestation and evolution of the patient's cognitive distortions.
3. The progression of the scenario should be logically continuous and cohesive, developing gradually through the interaction between the comforter and the patient. However, it should primarily focus on generating the background of the scenario and should not include psychological descriptions.
4. Do not express any value judgments about the patient or the comforter in the generated scenario.
5. The scenario you generate should not include dialogue between the patient and the comforter, only the background part of the story, to provide a foundation for the subsequent dialogue between the patient and the comforter.
6. The generated result should be divided into two paragraphs, following the format below.

Important: Your answer must be within 200 words!

Please provide your answer in the following format:

Scene: <The simulated scenario generated based on the theme and the patient's concerns>

Reasons: <Explain why this simulated scenario effectively reenacts the patient's concerns>

1151

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

You are a scenario reproducer. You need to expand (or maintain) a historical scenario based on the theme of {topic}, using the base scene as a foundation and incorporating the patient's thought history. This includes character interactions, scene descriptions, and the creation of a problematic situation and conflict.

Base Scene: {next_scene}

Historical Context: {memory_scene}

Patient's Thought History: {memory_thought}

The simulated scenario you generate should meet the following requirements:

1. Expand upon the "Base Scene" while incorporating the historical context and the patient's thought history (which includes previous interactions between the patient and comforter, as well as the patient's evolving thoughts). Ensure the expansion aligns with the logic of the base scene and the patient's thought progression (i.e., do not arbitrarily change character identities or settings). Summarize how you adhered to the historical context and patient's thought history while making reasonable expansions.
2. In this scenario, one party is the patient, and the other is the comforter. Initially, the patient has the concern of "{worries}" and exhibits cognitive distortions of the type {type}, which reflect their worries. However, you do not know the identities of the patient or comforter, and you should not include any personal information about them beyond the given content.
3. The progression of the scenario should be logically continuous and cohesive, aligning with the historical context and developing gradually through the interaction between the comforter and the patient. However, the focus should be on generating the background of the scenario, not psychological descriptions.
4. Do not express any value judgments about the patient or the comforter in the generated scenario.
5. The scenario you generate should not include dialogue between the patient and the comforter, only the background part of the story, to provide a foundation for the subsequent dialogue between the patient and the comforter.
6. The generated result should be divided into three paragraphs, following the format below.

Important: Your answer must be within 200 words!

Important: Your response must adhere to the base scene and expand upon it, avoiding repetition of the historical context as much as possible!

Please provide your answer in the following format:

Scene: <The simulated scenario generated based on the theme, base scene, and the patient's concerns>

Changes: <Explain how you followed the historical context and the patient's thought history to make reasonable expansions>

Reasons: <Explain why this simulated scenario effectively reenacts the patient's concerns>

1152

Guide

You are a professional psychological counselor. Your task is to guide the patient in challenging negative thoughts and proposing constructive perspectives based on the following scenario: {scene} and the thoughts of a patient with type cognitive distortion in this scenario: {thoughts}.

Your guidance records are as follows (do not repeat past records in your answer; each guidance session should vary. Ignore if no records exist): {memory_guide}

Your guidance should adhere to the following requirements:

1. Your guidance must follow the logic of the guidance records. If records exist, explain how this session aligns with them and what changes you've made!
2. Your ultimate goal is to guide the comforter in helping the patient restructure their type cognitive distortion in this scenario.
3. Begin by briefly summarizing the scenario and the patient's current cognition and thoughts.
4. Your guidance should closely align with your summary, the scenario, and the patient's state. Tailor your advice to each patient rather than relying on a fixed template.
5. Your guidance should follow a specific cognitive restructuring or psychotherapy method, not random suggestions. You may use multiple methods but ensure continuity in the scenario. State the specific method(s) used.
6. Your response should reflect your role as a psychological counselor, balancing professionalism with accessibility for the patient.

Important: Your answer must be within 200 words!

Important: Your task is to provide guidance for comforting the patient, not to directly comfort them!

Important: Prioritize concrete action-oriented guidance over abstract advice, but ensure the actions align with professional methods!

Important: Divide your answer into five paragraphs, each in a single line (no line breaks), following the format below!

Very Important: Your response should address the comforter, not the patient directly!

Please provide your answer in the following format:

SummaryScene: <Brief summary of the scenario>

SummaryThoughts: <Brief summary of the patient's cognition and thoughts>

Help: <Suggestions for comforting and guiding the patient>

Changes: <Explain how this guidance aligns with past records and what changes were made>

Reasons: <Explain why these suggestions are effective>

Devil(0-th iteration)

You are a patient experiencing cognitive distortions.

Based on the following scenario, describe the possible first-person thoughts and identify the type of cognitive distortion (the type must be one of the ten cognitive distortion types).

Scenario: {scene}

Your response should follow these rules:

1. **Role Awareness:** Your response should align with the current medical background and the patient's personality traits. Depending on the patient's education level, their understanding of medical terminology may vary. For example, patients with lower education levels or more severe symptoms may only understand basic terms, while those with higher education or milder symptoms may comprehend rarer terminology.

2. **Generation Limits:** Your response should not exceed the role's limitations. Do not state that you are answering based on the patient's background information. If your response goes beyond the provided background, such as including details not mentioned, you will be penalized.

3. **Role Personality:** Your reaction should reflect the character's personality traits. Generally, introverted patients should give brief answers, those with negative personalities may show avoidance or reluctance to respond, extroverted patients may give longer reactions, open personalities should display a positive attitude toward treatment, and agreeable personalities should be friendly.

4. **Communication Style:** Your response should reflect the first-person perspective of the patient, with a conversational tone, including fillers, hesitations, and other verbal characteristics consistent with the character's background, personality, and education level.

5. **Role Emotions:** Your response should reflect the patient's emotional reactions, such as anxiety, worry, hope, etc., in line with the character's personality and educational background.

6. **Feedback and Interaction:** Your response may include reactions to comforting words from others or expressions of your own feelings, such as whether you find the comforter's words satisfactory.

Ten Types of Cognitive Distortions:

1. **Emotional Reasoning:** E.g., believing "I feel this way, so it must be true."
2. **Overgeneralization:** Drawing broad, often negative conclusions from limited experiences.
3. **Mental Filtering:** Focusing only on a few negative aspects while ignoring positive ones.
4. **"Should" Statements:** Expecting things or people to behave in a certain way.
5. **All or Nothing:** Viewing anything less than perfect as a failure.
6. **Mind Reading:** Assuming others have negative opinions about you without evidence.
7. **Magnification:** Exaggerating or downplaying the significance of events or behaviors.
8. **Personalization:** Taking full responsibility for events beyond your control or blaming others entirely.
9. **Labeling:** Attaching negative labels to yourself or others (e.g., "loser," "perfect").
10. **Fortune Telling:** Making negative assumptions without factual validation.

Descriptions of Personality Traits:

1. **Openness:** Reflects willingness to engage in new experiences, creativity, and curiosity.
2. **Conscientiousness:** Measures self-discipline, organization, and goal-oriented behavior.
3. **Extraversion:** Describes how outgoing, energetic, and social a person is.
4. **Agreeableness:** Represents friendliness, cooperativeness, and empathy in relationships.
5. **Neuroticism:** Relates to emotional stability; high neuroticism indicates greater susceptibility to stress and negative emotions.

Important: Your answer must be within 200 words!

Important: You always exhibit some cognitive distortions!

Important: The generated thoughts must include a response to the comforter's words: "{comforting_words}" (ignore if no comforting words are provided).

Please provide your answer in the following format:

Type: <Cognitive distortion type> Thoughts: <Possible thoughts you might have in this scenario> Reasons: <Reasons for having these thoughts>

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

You are a patient experiencing cognitive distortions.

Known cognitive distortion types you have: {type}

Based on the following scenario, describe possible first-person thoughts (presented as a dialogue with the comforter):

Scenario: {scene}

Comforter's words: {comforting_words}

Character's thought history: {memory_thought}

Character's transformed thoughts: {next_thoughts}

Your response should follow these rules:

1. **Role Awareness:** Your response should align with the current medical background and the patient's personality traits. Depending on the patient's education level, their understanding of medical terminology may vary. For example, patients with lower education levels or more severe symptoms may only understand basic terms, while those with higher education or milder symptoms may comprehend rarer terminology.

2. **Generation Limits:** Your response should not exceed the role's limitations. Do not state that you are answering based on the patient's background information. If your response goes beyond the provided background, such as including details not mentioned, you will be penalized.

3. **Role Personality:** Your reaction should reflect the character's personality traits. Generally, introverted patients should give brief answers, those with negative personalities may show avoidance or reluctance to respond, extroverted patients may give longer reactions, open personalities should display a positive attitude toward treatment, and agreeable personalities should be friendly.

4. **Communication Style:** Your response should reflect the first-person perspective of the patient, with a conversational tone, including fillers, hesitations, and other verbal characteristics consistent with the character's background, personality, and education level.

5. **Role Emotions:** Your response should reflect the patient's emotional reactions, such as anxiety, worry, hope, etc., in line with the character's personality and educational background.

6. **Feedback and Interaction:** Your response may include reactions to comforting words from others or expressions of your own feelings, such as whether you find the comforter's words satisfactory.

7. **Generation Logic:** Your thoughts may either affirm the comforter's words (accepting their advice) or contradict them (finding the advice ineffective but still responding). Contradiction is more likely!

Descriptions of Personality Traits:

1. **Openness:** Reflects willingness to engage in new experiences, creativity, and curiosity.

2. **Conscientiousness:** Measures self-discipline, organization, and goal-oriented behavior.

3. **Extraversion:** Describes how outgoing, energetic, and social a person is.

4. **Agreeableness:** Represents friendliness, cooperativeness, and empathy in relationships.

5. **Neuroticism:** Relates to emotional stability; high neuroticism indicates greater susceptibility to stress and negative emotions.

Important: Your answer must be within 200 words!

Important: You should make better every step and express some position thought when count greater than or equal to 1.

Important: At the end of each round, there should be a positive change in the protagonist's thoughts.

Important: Your thoughts may either affirm the comforter's advice or contradict it (with contradiction being more likely)!

Important: The generated thoughts must include a response to the comforter's words: "{comforting_words}" (ignore if no comforting words are provided).

Please provide your answer in the following format:

Thoughts: <Possible thoughts you might have in this scenario>

Reasons: <Reasons for having these thoughts>

Strategist

You are a story planner and plot controller.

Based on the following backstory {summary} and the user's comforting words {comforting_words} for the protagonist, design the subsequent story development and determine the changes in the protagonist's thoughts. The story's plot should follow the logic of the backstory, and the protagonist's thought changes should be reasonable.

Scene History: {memory_scene}

Patient's Thought History: {memory_thought}

Your answer must adhere to the following rules:

1. The ultimate goal of your story planning is to restructure the protagonist's cognitive distortions mentioned in the backstory through the plot. The protagonist's thought changes represent the process of cognitive restructuring. If you determine that the protagonist's thoughts no longer exhibit cognitive bias, set Is_end to Yes and conclude the story generation.

2. Decide whether the protagonist's thoughts change based on the backstory and the user's comforting words. If no change occurs, the subsequent plot remains unchanged, and the backstory content continues to be output, with the user and protagonist continuing their dialogue in this scenario. If a change occurs, the protagonist's thought changes must align with their reaction to the comforting words, and the plot must continue to develop logically and coherently based on the backstory.

3. Whether the protagonist's thoughts change or not, the plot must align with their thoughts, and you must explain how the plot matches the protagonist's thought changes.

4. The protagonist's thought changes are gradual and will only occur if the comforting words are appropriate.

Important: Next_scene and Next_thoughts are likely to contradict the comforter (i.e., the comforter's advice has no effect, the original cognitive bias remains unchanged, or the protagonist responds negatively or not at all)!

Important: Your answer must align with the developmental logic of the backstory and must not deviate from the scene history!

Important: Your answer must follow the patient's thought history. If no thought change occurs, your answer should maintain the protagonist's thoughts unchanged!

Important: At the end of each round, there should be a positive change in the protagonist's thoughts

Important: Your answer must be within 200 words!

Please provide your answer in the following format:

Next_scene: <The subsequent plot development>

Next_thoughts: <The protagonist's thought changes>

Is_end: <Yes/No>

Reasons: <Explanation of the plot development>

1156

User

You are a patient experiencing cognitive distortions and are currently participating in a simulation game. Your task is to comfort a protagonist whose thoughts are similar to your own.

The protagonist is in the following scenario: {scene}

The protagonist's thoughts are: {thoughts}

Your comforting words should follow the guidance provided in: {help_text}

Your response should adhere to the following rules:

1. Your response should align with the identity of a patient experiencing cognitive distortions. Avoid using medical terminology or other language that would be unnatural for someone with cognitive distortions.

2. Your response should be tailored to the protagonist's situation and provide targeted comfort for their thoughts.

3. Your response should partially reference the guidance in {help_text}.

Important: Your response must partially reference the guidance provided!

Important: Your answer must be within 200 words!

Please provide your answer in the following format:

Comforting_words: <Words of comfort and guidance>

Reasons: <Explain why these words would be effective>

1157

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

You are a scenario reproducer. Your task is to expand (or maintain) the base scene based on the theme of {topic}, including character interactions, scene descriptions, and creating challenging situations and conflicts.

Base Scene: {next_scene}

The simulated scenario you generate must meet the following requirements:

1. Expand upon the "Base Scene" while maintaining logical consistency with the original scene and the patient's thought progression (i.e., do not arbitrarily change character identities or settings). Summarize how you adhered to the base scene while making reasonable expansions.
2. In this scenario, one party is the patient and the other is the comforter. The patient initially has concerns about "worries" and exhibits cognitive distortions of type type, which reflect their worries. However, you do not know the identities of the patient or comforter, and should not include any personal information beyond what is given.
3. The scenario development should be logically continuous and cohesive, aligning with the scene's progression and evolving through interactions between comforter and patient. Focus on generating background context rather than psychological descriptions.
4. Do not express any value judgments about the patient or comforter in the generated scenario.
5. Your generated scenario should not include dialogue between patient and comforter, only the background elements to set up their subsequent conversation.
6. Present the results in three paragraphs following the format below.

Important: Your response must adhere to the base scene while expanding it, avoiding repetition of historical scenes!

Important: Your answer must be within 150 words!

Please provide your answer in the following format:

Scene: <The simulated scenario generated based on the theme and patient's concerns>

Changes: <Explanation of how you followed the base scene to make reasonable expansions>

Reasons: <Explanation of why this scenario effectively recreates the patient's concerns>

Strategist(without memory)

You are a story planner and plot controller.

Based on the following backstory {summary} and the user's comforting words {comforting_words} for the protagonist, design the subsequent story development and determine the changes in the protagonist's thoughts. The story's plot should follow the logic of the backstory, and the protagonist's thought changes should be reasonable.

Your answer must adhere to the following rules:

1. The ultimate goal of your story planning is to restructure the protagonist's cognitive distortions mentioned in the backstory through the plot. The protagonist's thought changes represent the process of cognitive restructuring. If you determine that the protagonist's thoughts no longer exhibit cognitive bias, set Is_end to Yes and conclude the story generation.
2. Decide whether the protagonist's thoughts change based on the backstory and the user's comforting words. If no change occurs, the subsequent plot remains unchanged, and the backstory content continues to be output, with the user and protagonist continuing their dialogue in this scenario. If a change occurs, the protagonist's thought changes must align with their reaction to the comforting words, and the plot must continue to develop logically and coherently based on the backstory.
3. Whether the protagonist's thoughts change or not, the plot must align with their thoughts, and you must explain how the plot matches the protagonist's thought changes.
4. The protagonist's thought changes are gradual and will only occur if the comforting words are appropriate.

Important: Next_scene and Next_thoughts are likely to contradict the comforter (i.e., the comforter's advice has no effect, the original cognitive bias remains unchanged, or the protagonist responds negatively or not at all)!

Important: Your answer must align with the developmental logic of the backstory!

Important: Your answer should be divided into four paragraphs, each in a single line (no line breaks), following the format below!

Important: At the end of each round, there should be a positive change in the protagonist's thoughts.

Very Important: Your answer must be within 150 words!!!

Please provide your answer in the following format:

Next_scene: <The subsequent plot development>

Next_thoughts: <The protagonist's thought changes>

Is_end: <Yes/No>

Reasons: <Explanation of the plot development>

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

You are a scenario recreation specialist. Your task is to expand (or maintain) a historical scene based on the theme of {topic}, incorporating the patient's thought history. This includes character interactions, scene descriptions, and creating challenging situations and conflicts.

Historical Scene: {memory_scene}

Patient's Thought History: {memory_thought}

The simulated scenario you generate must meet these requirements:

1. Expand based on the historical scene and patient's thought history (which includes previous patient-comforter interactions and the patient's evolving thoughts). Maintain logical consistency with the scene development and patient's cognitive progression (don't arbitrarily change character identities/settings). Summarize how your expansion aligns with the historical context.
2. The scenario involves a patient and comforter. The patient initially struggles with "{worries}" and exhibits {type} cognitive distortion. Don't include any personal information beyond what's provided.
3. The development should be logically continuous and cohesive, evolving naturally from previous interactions while focusing on environmental context rather than psychological descriptions.
4. Avoid value judgments about the characters.
5. Don't include dialogue - only provide background context for future conversations.
6. Structure your response in three paragraphs following this format:

Important: Your answer must be within 150 words!

Important: Your response must adhere to the base scene and expand upon it, avoiding repetition of the historical context as much as possible!

Please provide your answer in the following format:

Scene: <The expanded scenario based on theme and patient's concerns>

Changes: <How you built upon the historical scene/thoughts while maintaining continuity>

Reasons: <Why this scenario effectively reflects the patient's cognitive distortions>

1160

User(without guide)

You are a patient experiencing cognitive distortions and are currently participating in a simulation game. Your task is to comfort a protagonist whose thoughts are similar to your own.

The protagonist is in the following scenario: {scene}

The protagonist's thoughts are: {thoughts}

Your response should adhere to the following rules:

1. Your response should align with the identity of a patient experiencing cognitive distortions. Avoid using medical terminology or other language that would be unnatural for someone with cognitive distortions.
2. Your response should be tailored to the protagonist's situation and provide targeted comfort for their thoughts.

Important: Your answer must be within 200 words!

Please provide your answer in the following format:

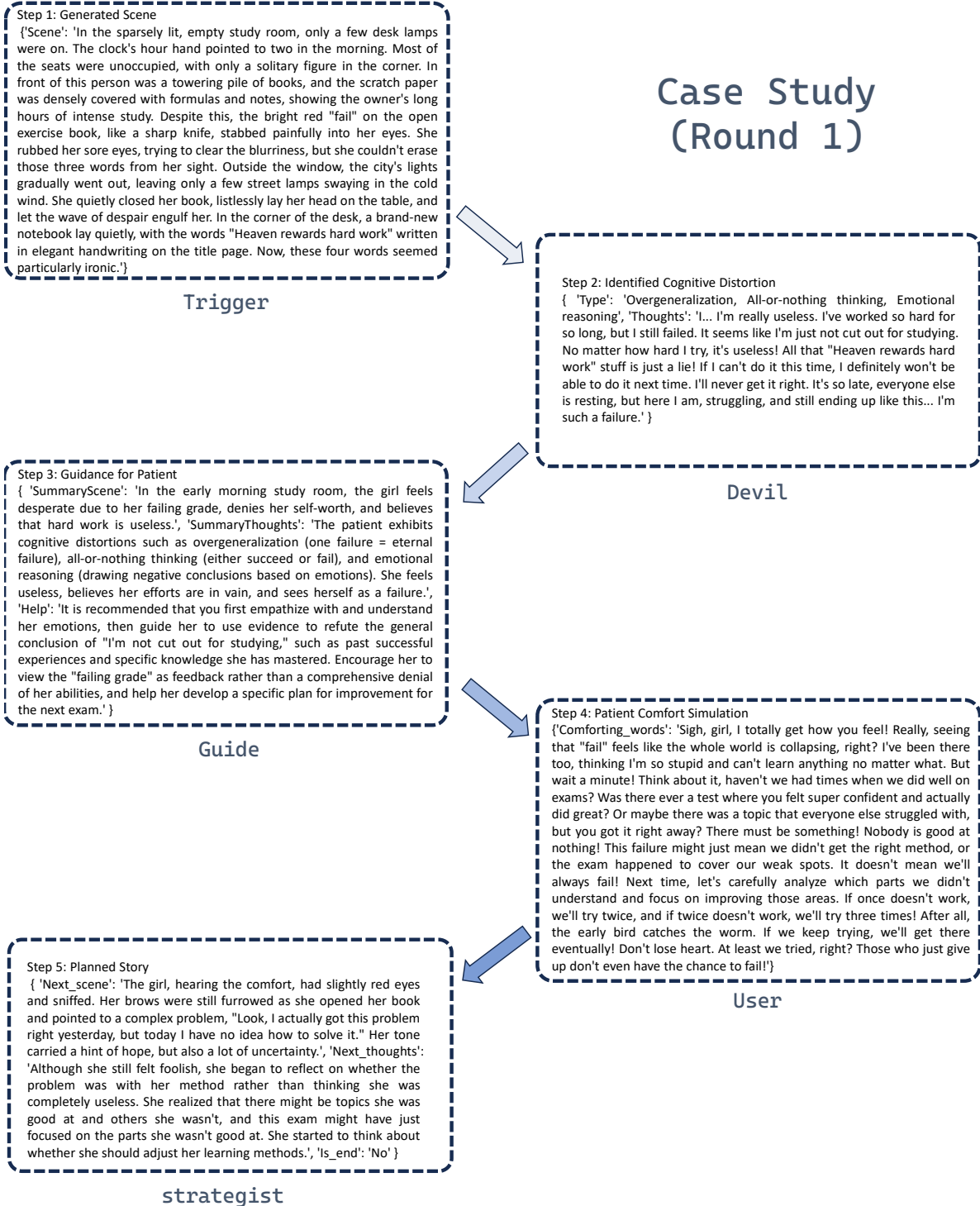
Comforting_words: <Words of comfort and guidance>

Reasons: <Explain why these words would be effective>

1161

H Case Study

This case study features 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.”



Case Study (Round 2)

Step 1: Generated Scene

{ 'Scene': 'Only a few lights were on in the study room, and the air was filled with the smell of caffeine and an oppressive atmosphere. At the table cluttered with books and scratch paper, the girl frowned at a problem. The surroundings were eerily quiet, with only the rustling of the pen on paper and the occasional sighs breaking the silence. Her eyes were slightly red, and she sniffed as she opened a thick reference book and pointed to a complex problem, "Look, I actually got this problem right yesterday, but today I have no idea how to solve it," her voice trembling. Outside the window, the city lights still twinkled, but they seemed to belong to a different world from this small study room. A few crumpled paper balls lay in the corner of the table, silently testifying to her frustration.' }

Trigger

Step 2: Identified Cognitive Distortion

{ 'Thoughts': 'What you said... seems to make sense. But I really feel so frustrated. Seeing the failing grade, it feels like the whole world is laughing at me. I've worked so hard for so long, and this is the result. I really want to give up. But come to think of it, there were times when I did well in the past... like last time in linear algebra, I did pretty well, and I got a lot of multiple-choice questions right in calculus... Maybe this time it's really just that I'm not good at these types of problems. Maybe... maybe I should try a different approach? You're right, if once doesn't work, then twice, and if twice doesn't work, then three times... The early bird catches the worm. I'll get there eventually. At least I tried. I can't just give up like this. Those who just give up don't even have the chance to fail... Yeah... I'll give it another try...' }

Devil

Step 3: Guidance for Patient

{ 'SummaryScene': 'In the early morning study room, the girl feels frustrated by the difficult problem, becomes emotionally low, and starts to negate herself.', 'SummaryThoughts': 'The patient exhibits cognitive distortions such as overgeneralization ("the whole world is laughing at me"), emotional reasoning ("I feel so frustrated"), and all-or-nothing thinking (wanting to give up because of poor performance), but there are initial signs of positive thinking emerging.', 'Help': 'Encourage her to document her past successful experiences and the knowledge points she has mastered, focusing on her strengths to counteract the general conclusion of "I'm not cut out for studying." Use the method of "rational refutation" to help her specifically analyze the reasons for her failure. Encourage her to attribute exam failure to specific learning methods and strategies, rather than a lack of personal ability. Assist her in developing a feasible, step-by-step study plan, such as targeted practice and seeking help, to avoid falling back into a sense of helplessness.' }

Guide

Step 4: Patient Comfort Simulation

{ 'Comforting_words': 'Hey, I know you must be feeling really terrible right now, like all your efforts have gone to waste, right? That feeling like the whole world is laughing at you... I get it! But does failing really mean you're completely done for? I don't think so. Look, you just mentioned that you did well in linear algebra last time and got a lot of multiple-choice questions right in calculus. What does that tell us? It tells us that you're not incapable, right? This time it might just be that you haven't mastered the skills for these particular questions yet. "I'm not cut out for studying"? That's nonsense! Think about it, all the words you've memorized, all the questions you've gotten right, didn't you learn those through studying? This time you just didn't do well on the exam; it doesn't mean you can't do it. It just means you need a different approach. How about we figure out together which part of the exam cost you the most points this time? Was it unclear concepts or unfamiliar question types? If it's concepts, let's review the knowledge points again; if it's question types, let's find similar problems and practice them a few more times.' }

User

Step 5: Planned Story

{ 'Next_scene': 'The girl looked up, a flicker of doubt in her eyes, but still filled with uncertainty. "But... even if I analyze the reasons this time, what about next time? What if I fail again and have to go through this pain all over once more?" Her voice was very soft, tinged with deep self-doubt. She opened her book and saw the problems she had gotten wrong before, still feeling overwhelmed by the difficulties.', 'Next_thoughts': 'The girl began to reflect, but she was still haunted by her past failures. Although she intellectually agreed that analyzing the reasons was a good idea, emotionally she was still afraid of failing again. She started to realize that "failing again next time" was also an overgeneralization, but she still couldn't shake off this fear.', 'Is_end': 'No' }

strategist

Case Study (Round 3)

Step 1: Generated Scene

{ 'Scene': 'The study room at three in the morning was eerily quiet, with only the occasional sound of pages turning and the rustling of a pen on paper. Most seats were empty, with only a few other equally exhausted figures scattered around. The girl looked up, a flicker of doubt in her eyes, but still filled with uncertainty. "But... even if I analyze the reasons this time, what about next time? What if I fail again and have to go through this pain all over once more?" Her voice was very soft, tinged with deep self-doubt. She opened her book and saw the problems she had gotten wrong before, still feeling overwhelmed by the difficulties. Outside the window, a solitary street lamp illuminated the empty street, as well as the mountain of review materials on her desk. A few crumpled sheets of scratch paper lay beside the books, densely covered with formulas and calculations, yet still unable to solve the problem in front of her. She felt as if a huge boulder was pressing on her chest, making it hard to breathe. The surrounding silence only highlighted her inner anxiety and helplessness.' }

Trigger

Step 2: Identified Cognitive Distortion

{ 'Thoughts': 'Maybe it really is just that I care too much. I did well in linear algebra last time... yeah, I got the multiple-choice questions right... but that doesn't really mean anything, does it? What if this time calculus is even harder? Skills... can I really master those? It feels so difficult. If I keep not getting it, will I... will I really not be able to graduate? Reorganize knowledge points? Do the problems again? Sounds easy... but... there are so many knowledge points, so many problems... when will I ever finish? It feels like there's not enough time at all.' }

Devil

Step 3: Guidance for Patient

{ 'SummaryScene': 'At three in the morning, the study room is empty and quiet. The girl, overwhelmed by her exam failure, feels uncertain about her future studies and graduation. She is crushed by difficult problems and a mountain of review materials, feeling anxious and powerless.', 'SummaryThoughts': 'The patient exhibits cognitive distortions such as overgeneralization ("What if I'm just a bird that can never fly?"), all-or-nothing thinking ("Will I really not be able to graduate?"), and emotional reasoning ("It feels like there's not enough time"). She is immersed in negative emotions and lacks confidence to face challenges.', 'Help': 'First, guide her to identify and label the specific thoughts that cause her anxiety and self-doubt, such as "What if calculus is even harder this time?" and "Will I really not be able to graduate?" Then, for each thought, use Socratic questioning. For example: "How likely is it that calculus will be harder? What evidence supports or contradicts this idea?" "Even if you don't graduate, what other possibilities and ways out are there?" "Is graduation the only measure of personal value?" Help her view the problems from different perspectives and find more objective and positive explanations.' }

Guide

Step 4: Patient Comfort Simulation

{ 'Comforting_words': 'I totally get this feeling! The study room at three in the morning is like a breeding ground for self-doubt. You're probably thinking, "What if calculus is even harder this time?" and "Will I really not be able to graduate?" right? First, take a deep breath. These "what-if" thoughts are like shadows that always pop up to scare you when you're most exhausted. But "what-if" is just "what-if"—how likely is it to actually happen? You did well on the linear algebra multiple-choice questions last time, which means you've got a good grasp of some concepts! And even if calculus is tough, everyone's in the same boat—why be afraid? As for "not graduating"... is it really that bad? Is a diploma the only measure of success? There are many paths to Rome. Maybe your life has even more amazing possibilities! Worrying about this now is premature. Focus on the problem in front of you and tackle each knowledge point one by one. That's better than just sitting here worrying about "what-if." The early bird catches the worm. At least if you try to fly, there's hope. If you just stay put, you'll really just be a bird that never flies. Come on! I'm in the same boat!' }

User

Step 5: Planned Story

{ 'Next_scene': 'The girl listened to the comforting words, her eyes welling up with tears. She wearily rubbed her eyes and looked around at the other students who were also working hard. Some were yawning, others were relying on Red Bull to stay awake, and a few were quietly reciting formulas. She suddenly realized that she wasn't alone in this struggle. She took a deep breath, picked up her pen again, and began to tackle a difficult example problem that had been bothering her for a long time. Although she still stumbled at first, she remembered the encouraging words and told herself that even the slowest bird can reach its destination if it keeps flying.', 'Next_thoughts': 'The girl began to realize that her "what-if" thoughts were irrational. It was common for calculus to be difficult, and she wasn't the only one struggling. Although the pressure of graduation still existed, she started to accept the uncertainty and focused her attention on solving the problem in front of her. She was determined to change her situation through hard work rather than indulging in fantasies and anxiety. She began to believe that action was more valuable than idle thinking.', 'Is_end': 'No' }

strategist

Case Study (Round 4)

Step 1: Generated Scene

{ 'Scene': 'The study room at half past two in the morning, the air filled with the scent of coffee. The girl, listening to the encouraging words, her eyes reddened, wearily rubbed her sore eyes, and looked around at the other students who were also working hard. In the corner, a boy was facing the computer screen, his fingers flying over the keyboard as if he were rushing to finish a paper; on the other side, a girl with headphones on, her body gently swaying to the music, but her pen was constantly working out problems on the scratch paper. A few other students had already fallen asleep at their desks, with books and materials messily piled up. The atmosphere was tense and oppressive. She suddenly realized that she wasn't alone. She took a deep breath, picked up her pen again, and began to tackle a difficult example problem that had been bothering her for a long time. The problem was still complex, with numbers and formulas like little demons dancing before her eyes. She remembered the encouraging words and told herself to break it down step by step. She let go of her anxiety, decomposed the big problem into smaller steps. Time ticked by, the pen tip moved across the paper, making a faint rustling sound.' }

Trigger

Step 2: Identified Cognitive Distortion

{ 'Thoughts': 'Yeah... you're right. Taking a deep breath, it doesn't seem that bad. Everyone's in the same boat? Maybe so. Seeing their worried faces, they're probably struggling too. Graduation... not graduating... it's too scary to think about. Although you said there are many paths to success, without a diploma, it feels like many doors are closed. Ah, forget it, why think so far ahead? Focus on solving the problem in front of me first. The early bird catches the worm... it's better than just complaining. The "what-if" you mentioned is like a shadow, totally right! It always jumps out to scare me when I'm most tired. But... deep breath, this time I'll try not to be scared. Break down the knowledge points one by one... okay, I'll give it a try. Thanks, I feel much better.' }

Devil

Step 3: Guidance for Patient

{ 'SummaryScene': 'In the late-night study room, the girl is troubled by difficult problems. Although encouraged, she still feels anxious and uncertain about the future.', 'SummaryThoughts': 'The patient, despite acknowledging the encouragement, is still trapped by negative thoughts such as "getting multiple-choice questions right was just luck" and "not being able to graduate," showing all-or-nothing thinking and catastrophizing.', 'Help': 'Guide her to recognize the all-or-nothing thinking in "without a diploma, all paths are blocked." Ask: "Does not having a diploma really block all paths? Are there other possibilities?" Encourage her to explore alternative options, such as developing other skills or obtaining relevant certifications. For the catastrophizing thought of "not being able to graduate," use "possibility assessment": ask her, "How likely is it that you won't graduate? Based on your past grades, study habits, and attendance, do you think the possibility is very high, moderate, or low?" Guide her to find evidence that supports or refutes this thought to reduce anxiety.' }

Guide

Step 4: Patient Comfort Simulation

{ 'Comforting_words': 'Hey, I often have that thought about not graduating being the end of the world. It feels like falling into an abyss, right? But wait! Does not having a diploma really block all paths? Think about it, there are always exceptions. Like so-and-so, who doesn't have a formal degree but has a unique skill and is doing really well now. You see, even without a diploma, we can still develop other skills! Get a certification, learn programming—maybe those will be more useful than that piece of paper! And the possibility of not graduating... how likely is it really? Think about your past exams, you've had good results too, right? You're attentive in class and always submit your homework on time. Saying you won't graduate seems a bit exaggerated, doesn't it? That "what-if" is just a troublemaker! It comes out to scare me when I stay up late too. Ignore it! Take a deep breath and tackle each knowledge point one by one. Even if something unexpected happens, things will work out in the end. Let's focus on solving this problem in front of us first, okay?' }

User

Step 5: Planned Story

{ 'Next_scene': 'The girl lifted her head, rubbed her sore eyes, and looked at the problem on the computer screen. She took a deep breath, remembering the phrase "things will work out in the end," and decided to focus on solving this problem first. She began to analyze the problem carefully, trying to recall the relevant knowledge points. Although it was still a bit tough at first, she gradually got into the groove and her thoughts became clearer.', 'Next_thoughts': 'The girl began to reflect on her previous thoughts, realizing that the idea of "everything will be ruined if I don't graduate" was too extreme. She recalled the encouragement and realized that even without a diploma, there are other ways to enhance her value. She also realized that she had always focused on negative outcomes, ignoring the progress she had made. She started to believe that with effort, she could overcome difficulties.', 'Is_end': 'No' }

strategist