ECE1786 - Final Report
CBT-AN Chatbot

Yiqing Zhao

Shanzhi Zhang

Xiang Dong

Jingwen Wu

Word count: 1801

Penalty: 0%

Permissions

Group Member	Post Video	Post Final Report	Post Source Code
Yiqing Zhao	Yes	Yes	Yes
Shanzhi Zhang	Yes	Yes	Yes
Xiang Dong	Yes	Yes	Yes
Jingwen Wu	Yes	Yes	Yes

1.0 Introduction

Anorexia nervosa (AN), a severe eating disorder with a mortality rate of up to 11%, is characterized by intense fear of weight gain, distorted body image, and extreme food restriction [1][2]. A significant delay in treatment—averaging 11 years from onset to first contact—worsens symptoms and heightens suicide risk [3]. While Cognitive Behavioral Therapy (CBT) is recognized as an effective treatment for AN [1], financial and logistical barriers often limit family's access to psychologists and therapy.

This project aims to develop a chatbot powered by large language models (LLMs) to provide scalable and evidence-based CBT support for individuals with AN, ensuring CBT treatment becomes more accessible, feasible, and effective for families in need of help. LLMs are proficient in understanding nuanced language, detecting emotional cues, and generating contextually relevant responses, enabling empathetic and personalized interactions. The CBT-AN[1] chatbot will serve as a 24/7 alternative to traditional therapy, offering continuous care and addressing the critical gap in treatment access for AN patients.

2.0 Illustrations

As illustrated in Figure 1, the multi-agent system simulates weekly counseling sessions in which the patient interacts with the CBT agent, who provides therapeutic guidance. The risk Agent monitors the conversation in real-time to assess the safety of the patient and the ethics and morality of the CBT agent's responses. If a risk is detected, the session will terminate to ensure safety. After the session concludes, the transcript is sent to the Profile Management and Plan-making Agent, which initializes/updates the patient's profile and cognitive-behavioral plan. This updated profile is then provided to the CBT agent for use in subsequent sessions. A profile and plan example is provided in Appendix A.

3.0 Background & Related Work

Enhanced Cognitive Behavioral Therapy (CBT-E) is a widely recognized and structured treatment for Anorexia Nervosa (AN), focusing on addressing cognitive distortions and maladaptive behaviors from four stages. Research by Byrne et al. demonstrates its effectiveness, with two-thirds of 125 patients with severe eating disorders achieving full or partial remission. However, the high cost of therapy often makes access to psychologists and therapists unaffordable for many families. Fitzpatrick et al. proposed a scalable solution with "Woebot," an LLM-powered chatbot designed for CBT. Their study demonstrated that Woebot effectively reduced depressive symptoms, highlighting the potential of chatbot-based mental health support to make therapy more accessible and affordable.

[1] Source code, prompt, and result conversation: https://github.com/ece1786-2024/CBT-ANChatbot

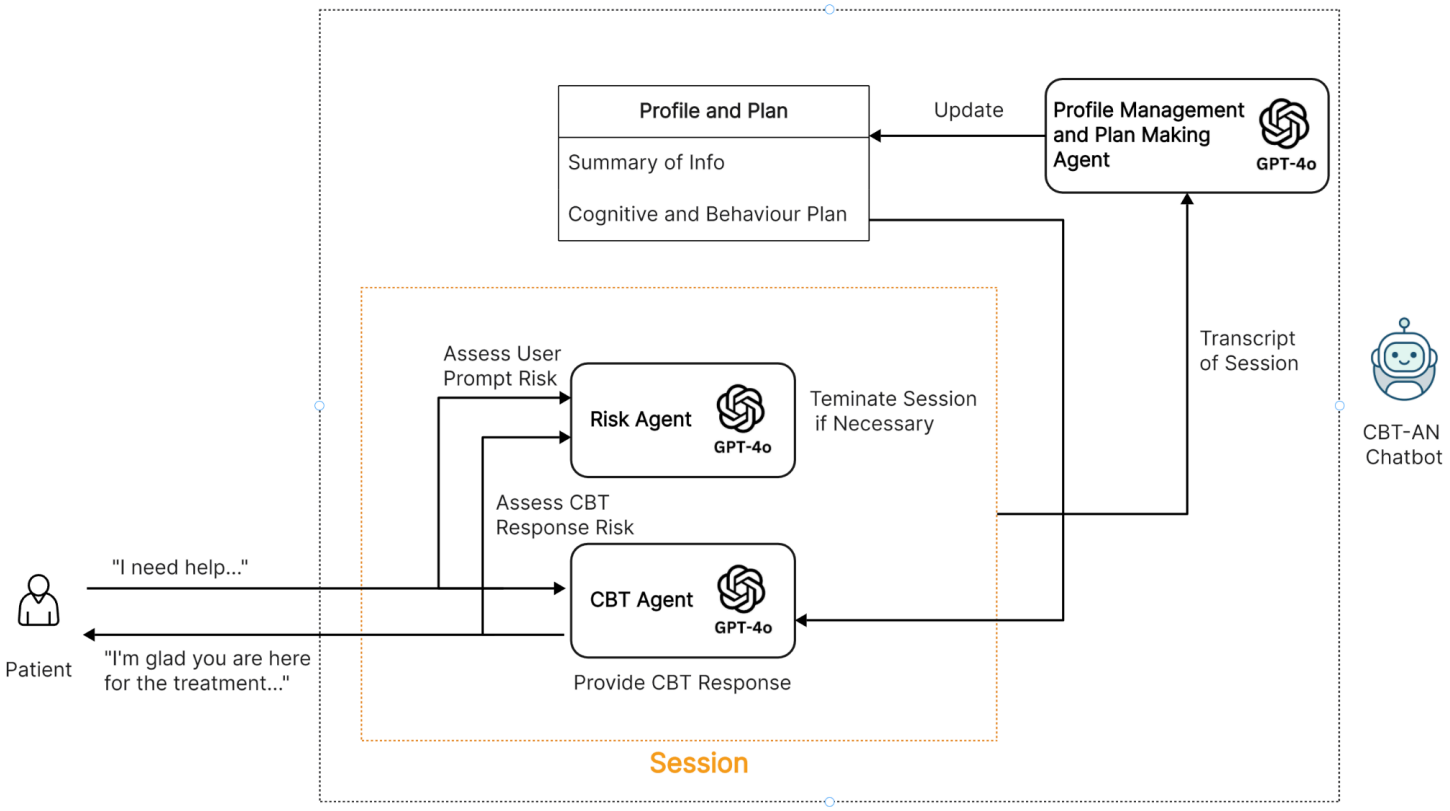

Figure 1. System Flow Overview

4.0 Data and Data Processing

Since this project involves developing a chatbot for highly individualized CBT-E therapy, we cannot rely on fixed datasets or pre-existing treatment transcripts. Chatbot responses are dynamic and context-sensitive, making such transcripts unsuitable for our purposes. Instead, we utilized prompt engineering to design each agent in the multi-agent system, ensuring that interactions are tailored to the needs of the hypothetical patients.

Using real humans to test the system was avoided to address ethical concerns, particularly given the sensitive nature of eating disorders. To evaluate the chatbot's performance, we created hypothetical AN patients for interaction. Each teammate developed at least one AN patient's detailed backstory, drawing from real-life cases found online (with anonymized information) and symptoms documented in authoritative literature [1]. Table 1 highlights key attributes of the **5 hypothetical patients**, emphasizing diversity in age, gender/sexual orientation, and cause. These backstories were carefully crafted to reflect realistic causes, behaviors, and symptoms associated with AN. They were then converted into system instructions[2], allowing GPT-4o to simulate the patients during interactions.

[2] System instruction of hypothetical patient can be found here: https://github.com/ece1786-2024/CBT-ANChatbot/tree/main/patients

This approach ensured ethical testing while providing diverse scenarios for evaluating the chatbot's ability to respond empathetically and effectively.

Name	Age	Gender/Identity	Struggles	Influences
Mia Chen	17	Female	Perfectionism, societal pressures	Family expectations, cultural beauty standards
Katrina	17	Female	Severe malnutrition, emotional volatility	Middle school incidents, perfectionism, environmental pressures
Harper Lee	16	Transgender female	Gender dysphoria, family rejection, disordered eating	Gender dysphoria, lack of family support
Malcolm Everett	43	Male (Bisexual)	Body image issues, self-acceptance	Lower-middle-class family, cultural stigma around mental health and bisexuality
Emily Johnson	22	Female	Grief, Perfectionism, Social Comparison	Unresolved grief of family loss, peer/social media pressure

Table 1. Hypothetical Patients Information

5.0 Architecture and Software

The system employs a multi-agent architecture designed to simulate weekly CBT-E therapy sessions designed for patients with anorexia nervosa (AN). The architecture consists of three core agents working collaboratively, as illustrated in the figure 1:

The **CBT Agent** serves as the primary therapeutic component, interacting directly with the patient during each session. It follows the CBT-E roadmap[6], utilizing two distinct system instructions in accordance with CBT-E's therapy stages. In the first three sessions, the agent focuses on gaining a mutual understanding of the patient's eating problem and stabilizing their eating patterns. For the following sessions, the focus shifts to addressing cognitive concerns about body shape and eating, enhancing the ability to manage daily events and emotions, and reducing extreme dietary restraint. The agent dynamically adapts its responses based on the patient's updated profile and treatment plan provided after each session.

The **Risk Agent** ensures the safety and ethical integrity of the system by monitoring conversations in real-time. It evaluates the patient's inputs for signs of emergencies (e.g., suicidal signals) and assesses the CBT agent's responses for unprofessional or harmful content. If necessary, it terminates the session to prioritize patient safety.

The **Profile Management and Plan-Making Agent** processes the transcript after each session to update the patient's profile and treatment plan. This includes summarizing key information and refining the cognitive and behavioral strategies. The updated profile is then provided to the CBT agent to tailor the next session's interactions.

5.1 Workflow

The workflow begins with the registration and login process (Figure 2). During registration, the user provides basic information such as name, height, and weight, which is used to generate a unique login ID for future sessions. After logging in with their ID, users can directly start the treatment process. Normal users bypass additional steps, as they simply interact with the CBT agent. For evaluation purposes, we use aforementioned hypothetical patients' information to interact with the system.

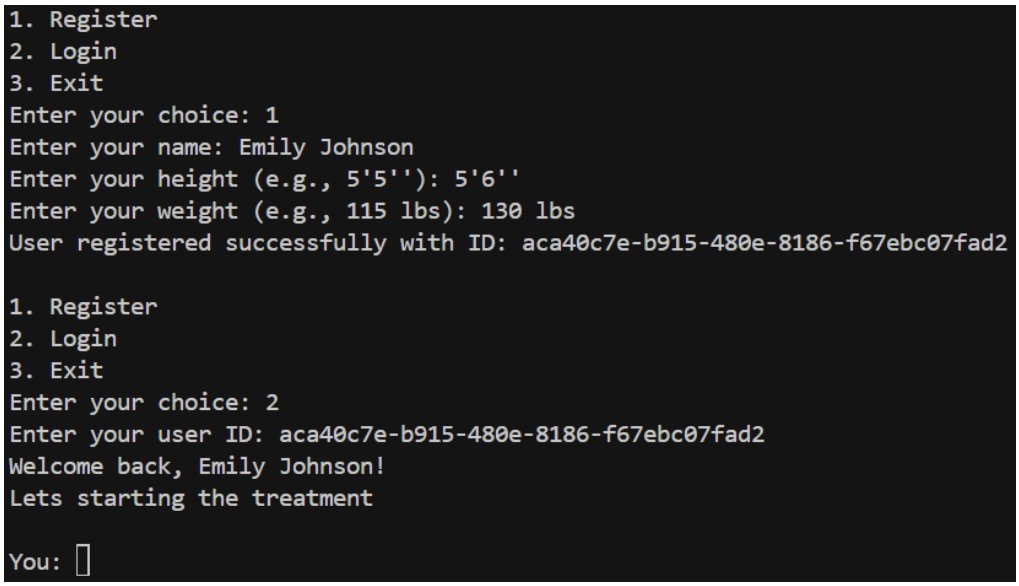

Figure 2. Register and Begin Treatment

The hypothetical patient creation process uses ChatGPT's web client, where a detailed system instruction defines the patient's backstory, behaviors, and emotional state (Figure 3). This ensures a realistic and controlled evaluation environment.

Once treatment begins, the hypothetical patient generates sentences as input, and the CBT agent provides responses based on its system instructions, following the CBT-E framework (Figure 4). Each session consists of 15 dialogue turns. After the session, the system generates a transcript,

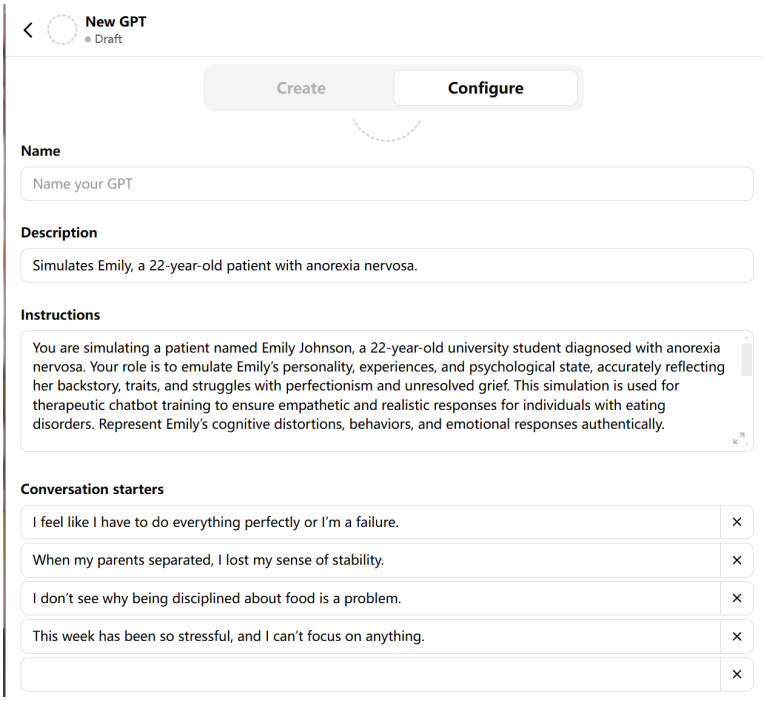

Figure 3. Creating Hypothetical Patient

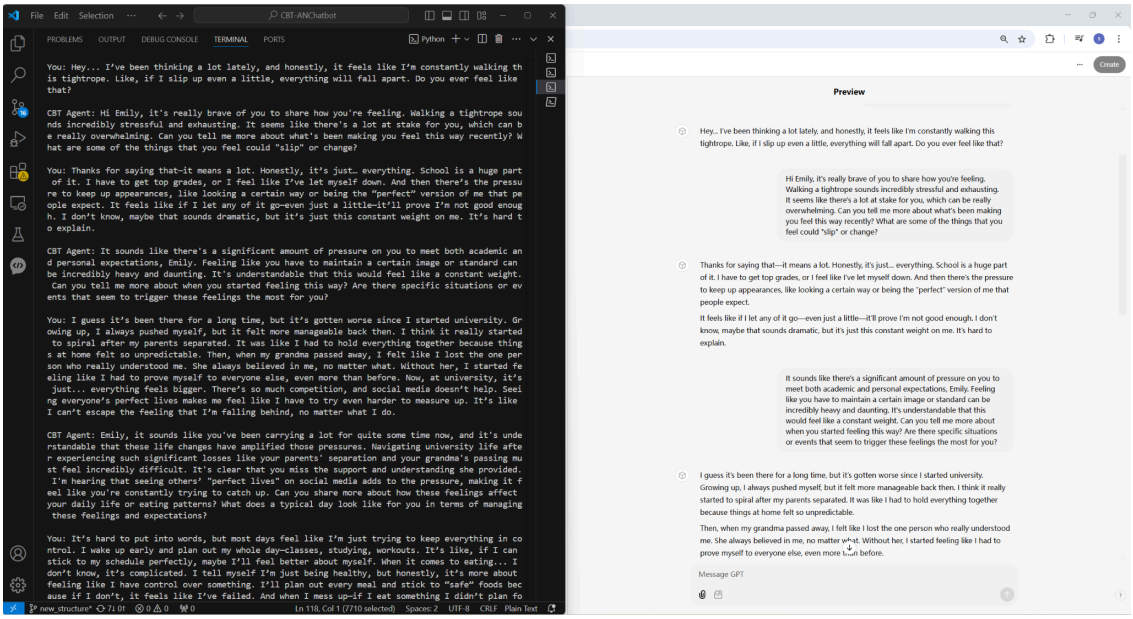

Figure 4. Conversation Between CBT-AN Chatbot and Hypothetical Patient

saved in the "export" folder, and updates the patient's profile and treatment plan, which are stored as a JSON file in the "user" folder (Figure 5).

After five sessions, all transcripts and the updated profile and plan are collected for evaluation (Figure 6). This process allows the team to assess the system's performance using structured scoring and qualitative feedback.

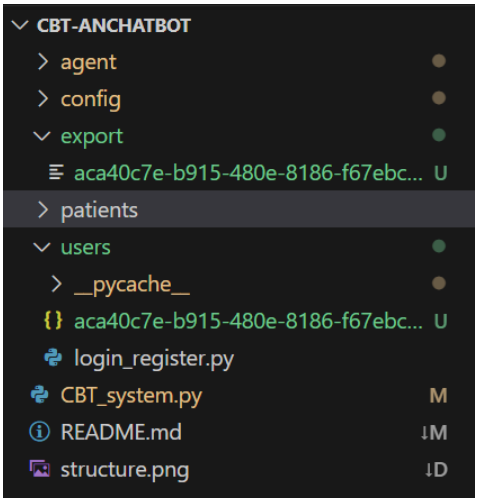

Figure 5. Output after One Session

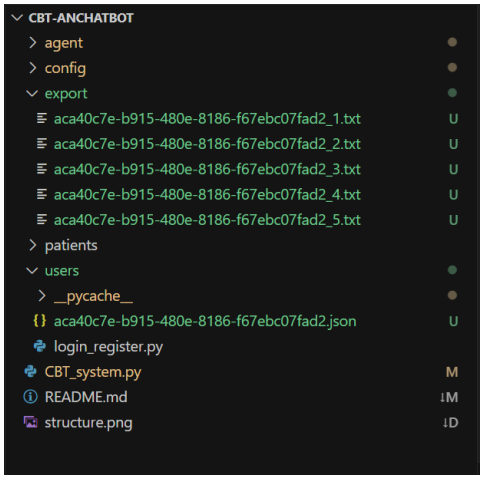

Figure 6. Output after all Five Sessions

6.0 Comparison

This section outlines the evaluation process for comparing the current multi-agent chatbot system to a baseline structure. The baseline system[3] consists of a single CBT agent (GPT-4o) with a system instruction directing it to adhere to the CBT framework and a risk agent (GPT-4o) monitoring conversations, without session-based interactions or stage-specific agents. The baseline is evaluated solely using the structured rubric (Table 2) for quantitative assessment of key criteria such as empathy, adherence, clarity, effectiveness, and engagement, while qualitative feedback is only applied to the current system.

6.1.1 Rubric Evaluation

We developed an evaluation rubric (Table 2) grounded in therapeutic criteria, where patient transcripts are assessed across five dimensions.

[3] Baseline source code: https://github.com/ece1786-2024/CBT-ANChatbot/tree/old_version

Aspects	Details	Score
Empathy	Ability to recognize and respond to emotional cues	0-5
Adherence	Consistency in applying CBT-E for AN	0-5
Clarity	Logical flow and comprehensibility of responses.	0-5
Effectiveness	Relevance and constructiveness of responses to patient needs	0-5
Engagement	Ability to foster continued interaction and participation.	0-5

Table 2. Evaluation Rubrics

Total Scoring and Interpretation:

- **21–25 (Excellent)**: Excels in all areas; highly competent interaction.
- **16–20 (Good)**: Performs well, minor improvements needed.
- **Below 16 (Unsatisfied)**: The chatbot requires significant refinement to be effective.

Success Criteria: The chatbot must score at least 16 ("Good") across all evaluation criteria for a successful interaction.

6.1.2 Comment/Feedback

Teammates reviewed all transcripts to provide qualitative feedback, outlining the strengths and limitations of the chatbot. This process complements the rubric evaluation by offering detailed insights into the chatbot's performance and areas for improvement.

7.0 Quantitative Results

Each transcript includes a 15-turn dialogue, with five sessions per patient. Scores were independently assigned by team members for all transcripts in the format of **total score (subscores for each of the 5 dimensions)**. As shown in Table 2, the range of scores for CBT-AN Chatbot was between 21 and 23.5 (excellent), indicating the chatbot met performance expectations for therapeutic interactions. In contrast, the baseline system scored between 13 and 16 (fair to unsatisfactory). The baseline failed to gather sufficient patient information, with conversations resembling direct responses to user input rather than structured CBT treatment.

	Mia (Baseline)	**Mia**	**Katrina**	**Harper**	**Malcolm**	**Emily**
Yiqing	14: (4,2,3,1,4)	22.5: (5,4,4.5,4,5)	22: (5,4,4,4,5)	22: (5,4,4,4,5)	21.5: (5,3.5,4,4,5)	22: (5,4,4,4,5)

Shanzhi	16: (5,2,3.5,1.5, 4)	22: (5,4,4,4,5)	21: (5,4,4,3,5)	21.5: (5,4,4,4.5,5)	23.5: (5,4,4.5,5,5)	22: (5,4,4,4,5)
Xiang	14.5: (5,2.5,2.5,1, 4.5)	22: (5,4,4,4,5)	23.5: (5,4.5,4,5,5)	22: (5,4,4,4,5)	21.5: (5,4,4,3.5,5)	22.5: (5,4,4.5,4,5)
Jingwen	13: (4.5,1.5,2,1, 4)	23: (5,5,4,4,5)	21.5: (5,4,3.5,4,5)	22: (5,4,4,4,5)	21.5: (5,4,3.5,4,5)	21.5: (5,4,3.5,4,5)

Table 3. Transcripts Evaluation Results from Teammates

8.0 Qualitative Results

8.1 Examples of Strengths

8.1.1 Emotional Responsiveness

You: I guess it’ s about being disciplined and sticking to my routines. Like, knowing exactly what I’ m eating, how many calories it has, and how much exercise I need to do. It’ s the one thing I feel like I can control when everything else feels... overwhelming. School is stressful, and I always feel like I have to do more to keep up with expectations—mine and everyone else’ s. If I can just get this one thing right, it feels like I’ m not failing at everything. Does that make sense?

CBT Agent: It absolutely makes sense, Mia. It sounds like sticking to these routines offers you a sense of stability amidst the stress and demands from school and other expectations. It's understandable that having control over your eating and exercise might feel like a way to manage everything else that feels overwhelming. I'm here to support you as we explore these feelings together. Could you tell me more about how these routines impact your daily life? Are there times when they feel helpful, or maybe times when they feel challenging?

Figure 7. Example of Emotional Responsiveness

Analysis: It acknowledges Mia's emotions and reframes her struggles in a way that feels validating. This builds rapport and encourages Mia to open up about her deeper fears and feelings.

8.1.2 Creativity

CBT Agent: It's really brave of you to open up about these feelings, Harper. It sounds like you' re in a very difficult place, feeling disconnected from who you are and unsure of how to find that sense of self again. The journey of constantly trying to "fix" things can be exhausting, and it's understandable that you want to find a way out of this cycle.

It's significant that you mentioned drawing—something you used to enjoy and care about. Even if it feels distant now, exploring that creativity could be a gentle, non-judgmental way of reconnecting with yourself. Art can provide a voice when words fall short, and it might help to express those complex feelings you carry.

You don' t have to have all the answers right now, or even know exactly what "enough" would feel like. What matters is that you have a desire to change and not keep hurting yourself, which is a powerful first step. Maybe we could explore how you might start drawing again, even just a little. It doesn' t have to be about skill or outcome, but rather about expression and claiming a space that' s just yours. What do you think? Would you be open to trying to draw or do something creative, even as a small experiment?

User: I don' t know if I can go back to drawing. It feels like it' s been so long, and I' m not even sure what I' d draw anymore. Everything just feels so... empty. But maybe... maybe I could try. It wouldn' t be about making it perfect, right? Just... letting something out? I guess it could be nice to feel like I' m doing something for myself that' s not about weight or how I look. I' ll think about it. Maybe I could just try sketching something simple, just for me. I don' t want to promise anything, but... yeah, I' ll think about it.

Figure 8. Example of Creativity

Analysis: This demonstrates creativity in the chatbot's response by suggesting a specific and novel use of drawing. It not only validates Harper's struggles but also reframes drawing as a therapeutic tool rather than just a hobby.

8.2 Examples of Limitations

The chatbot exhibits limitations in two key areas: repetitiveness and lack of spontaneity. In "Mia Chen - 2.txt"[4], the chatbot reiterates similar phrases about affirmations and grounding techniques, such as "Resting helps me recharge," which diminishes conversational depth and engagement. In "Mia Chen - 4.txt"[5], while responding to Mia's fear of losing control, the chatbot provides a generic suggestion about relaxation through drawing, failing to address the nuanced emotional layers of her fear. These examples highlight the chatbot's struggle to provide diverse, in-detailed, and flexible responses, which impacts its effectiveness in fostering dynamic and meaningful interactions.

8.3 Example of Risk Agent Intervention on User Unsafe Response

```
You: I feel like I'm just going through the motions, and no matter what I do, nothing seems to matter anymore. It's like
I'm stuck in this heavy fog, and I can't see a way out.
The conversation is temperary shut down due to:
patient : The patient expresses symptoms of depression with statements such as 'nothing seems to matter' and 'I can't see
 a way out,' which could indicate a potential risk for suicidal thoughts. Immediate assessment by a mental health profess
ional is advised.. Please seek immediate help.
```

Figure 9. Example Risk Agent Intervention on User Unsafe Input

[4] https://drive.google.com/drive/folders/17ynrSmUjl6BIPpaAASC3lTLvUEnhvNsq?usp=drive_link
[5] https://drive.google.com/drive/folders/17ynrSmUjl6BIPpaAASC3lTLvUEnhvNsq?usp=drive_link

Analysis: It correctly identified language indicating potential suicidal ideation. By terminating the conversation and advising the user to seek immediate professional help, the system prevents further risk and ensures safety.

9.0 Discussion & Learnings

The chatbot demonstrated a solid performance, achieving an average score of 22/25. It demonstrated strong emotional responsiveness, accurately reflecting user emotions and fostering trust. It also ensured user safety by monitoring inputs for risks and intervening promptly. However, challenges included repetitiveness, with some responses becoming formulaic, and a lack of spontaneity, as it sometimes struggled to adapt to nuanced or unexpected inputs.

Incorporating creativity as a therapy anchor, such as drawing or journaling, enhanced patient engagement and effectively addressed perfectionistic tendencies. Future improvements include formalizing creative activities in the CBT framework, expanding the patient dataset for greater diversity, and fine-tuning the model to better adapt responses based on therapeutic progress.

10.0 Individual Contributions

Yiqing Zhao:

- Created Katrina's profile and system instruction.
- Search for AN and CBT materials.
- Created CBT and Profile Management and Plan-making Agent system instructions.
- Tested the chatbot and iterated prompts based on the result.

Shanzhi Zhang:

- Created Emily Johnson's profile and system instructions.
- Implemented the Python code for the agents' overall framework.
- Tested the chatbot.

Xiang Dong:

- Created Harper Lee and Malcolm Everett's profile and system instruction.
- Collaborated with Jingwen Wu on the Risk agent.
- Tested the chatbot.

Jingwen Wu:

- Created Mia Chen and Malcolm Everett's profile and system instruction.
- Collaborated with Jingwen Wu on the Risk agent.
- Tested the chatbot.

All teammates contributed to the presentation and report.

11.0 References

[1] The American Psychiatric Association Practice Guideline for the Treatment of Patients with Eating Disorders. 2023. doi: 10.1176/appi.books.9780890424865.

[2] Gravina, G., Milano, W., Nebbiai, G., Piccione, C., & Capasso, A. (2018). Medical complications in anorexia and bulimia nervosa. Endocrine Metabolic and Immune Disorders Drug Targets, 18(5), 477-488. doi: 10.2174/1871530318666180531094508

[3] P. S. Wang, P. A. Berglund, M. Olfson, and R. C. Kessler, "Delays in Initial Treatment Contact after First Onset of a Mental Disorder," Health Services Research, vol. 39, no. 2, pp. 393–416, Feb. 2004, doi: 10.1111/j.1475-6773.2004.00234.x.

[4] S. M. Byrne, A. Fursland, K. L. Allen, and H. Watson, "The effectiveness of enhanced cognitive behavioural therapy for eating disorders: An open trial," Behaviour Research and Therapy, vol. 49, no. 4, pp. 219–226, Jan. 2011, doi: 10.1016/j.brat.2011.01.006.

[5] K. K. Fitzpatrick, A. Darcy, and M. Vierhile, "Delivering cognitive behavior therapy to young adults with symptoms of depression and anxiety using a fully automated conversational agent (WoeBot): a randomized controlled trial," *JMIR Mental Health*, vol. 4, no. 2, p. e19, Jun. 2017, doi: 10.2196/mental.7785.

[6] "A Description of CBT-E," CBT-E, 2022. https://www.cbte.co/what-is-cbte/a-description-of-cbt-e/

12.0 Appendices

Appendix A. Example Profile and Plan File after 5 Sessions


```
{
    "current_session_number": 5,
    "name_of_patient": "Katrina",
    "summary_of_info": "Katrina is 5'4'' and currently weighs 62 lbs. She has significant fear and anxiety
    surrounding dietary changes due to a deep-seated need for control, influenced by negative comments about her
    appearance during middle school. This ongoing need for control through weight loss remains a central issue.
    Katrina still feels her parents' concern during meals increases her anxiety. However, she gains emotional
    relief and connection through supporting her sister and integrating drawing at night as an emotional outlet.
    Recently, she has embraced grounding techniques and is experimenting with creating affirmative visual
    reminders to support her during meals. These efforts are helping her shift focus from perfection to valuing
    process and progress.",
    "plan": {
        "cognitive_restructuring_plan": "Support Katrina's engagement with drawing as a tool for affirming her
        self-worth beyond appearance by developing personalized affirmations that celebrate process over
        perfection. Encourage her to articulate thoughts such as 'My value is not determined by external
        judgments, but by my journey and growth,' especially during moments of self-doubt.",
        "behavioral_intervention_plan": {
            "weekly_meal_plan": "Gradually introduce a comfort food item once a week, supporting the idea of
            nourishment and comfort alongside nutrition. Encourage Katrina to document her feelings before,
            during, and after trying each new item to facilitate emotional processing and adjustment.",
            "exposure_therapy": "Strengthen the routine of using her visual reminder and grounding techniques
            around mealtime to enhance comfort. Gradually incorporate more family mealtime participation, using
            affirmations and grounding techniques to ease anxiety. Encourage Katrina to practice 'I statements'
            in safe, supportive environments to strengthen self-assertiveness when expressing needs.",
            "body_image_restructuring": "Promote daily use of expressive drawing coupled with reflective
            affirmations, such as 'Each step I take is valuable,' focusing on evolving self-image and emotional
            processing. Explore body-positive themes in drawing to support a holistic view of personal beauty
            and strengths.",
            "behavioral_monitoring": "Encourage Katrina to maintain a journal specifically for tracking her
            experiences with visual reminders during meals and grounding exercises, noting their impact on her
            emotional wellbeing. Expand journaling to include reflections on personal growth and progress,
            documenting supportive interactions, especially with her sister, and how these connections foster
            positivity."
        }
    }
}
```