Empathy-Enhanced Chatbot for Psychological Support: A Retrieval-Augmented and Therapy-Informed Approach

Yong-Xiang Chen
Intelligent Computing and Big Data
Chung Yuan Christian University
Taoyuan, Taiwan
yxchen@cycu.edu.tw

Yu-Chiao Chang
Intelligent Computing and Big Data
Chung Yuan Christian University
Taoyuan, Taiwan
s11120403@cycu.edu.tw

Hsi-Wei Ho
Intelligent Computing and Big Data
Chung Yuan Christian University
Taoyuan, Taiwan
s11120420@cycu.edu.tw

Abstract*—In modern society, busy work schedules and environmental factors contribute to high levels of stress, negatively impacting mental well-being. This issue is especially critical for people who often experience prolonged stress with limited access to psychological support. This study proposes a chatbot system designed to provide real-time, empathetic, and effective psychological assistance. Unlike conventional chatbots that offer superficial responses, our system integrates multiple psychotherapy techniques to deliver deeper emotional resonance and sustained support. Built on a large language model (LLM), the chatbot is enhanced with Retrieval-Augmented Generation (RAG) and optimized for personalized, context-aware interactions. The system adopts RA-CoP, a novel framework that combines RAG with structured psychotherapeutic guidance, as its core mechanism for generating emotionally attuned and clinically informed responses. This study not only proposes the RA-CoP framework and designs a psychological support system, but also introduces an evaluation method that combines objective scoring with users' subjective experiences to assess the usability of such systems.

***Keywords*—Psychological Support, Large Language Model, Empathy, Chatbot**

I. INTRODUCTION

With the advancement of artificial intelligence and natural language processing technologies, chatbots have been widely applied in areas such as customer service, automated responses, and healthcare. However, current chatbots still face significant challenges in providing psychological support and often fail to meet users' emotional needs effectively. Most chatbots respond to users' negative emotions with superficial empathy, offering statements like "I understand how you feel" or "That sounds difficult," before immediately shifting to problem-solving. Psychological research [1, 2] suggests that when individuals experience stress or emotional distress, they need more than just problem-solving advice—they also require deep emotional understanding and resonance. As a result, many chatbot responses lack warmth and genuine care, making users feel unsupported.

The core objective of the study is to design an enhanced psychological support system by building upon and integrating existing chatbot solutions. The system specifically targets individuals who face challenges in accessing timely psychological relief due to time or environmental constraints, such as caregivers. While previous research [3, 4, 5] has developed various psychological support chatbots, some studies have explored using Retrieval-Augmented Generation (RAG) [5] to enhance response relevance, while others have incorporated psychotherapy theories during model training [3] to strengthen therapeutic alignment. However, no existing research has combined these two approaches. This study aims to design a novel framework that integrates both RAG and psychotherapy-informed training to develop a more effective and context-aware psychological support system.

By optimizing the chatbot's language model, this research aims to achieve deeper empathetic and comforting conversations. The study particularly focuses on enhancing the chatbot's capabilities in emotional recognition, empathetic expression, and sustained support. The outcomes of this research are expected to offer new directions for AI applications in mental health and enhance the feasibility and effectiveness of psychological support systems, especially in serving the needs of specific populations.

II. RELATED WORKS

2.1 Empathy Recognition

Empathy is one of the core elements of psychological support, helping users feel understood and cared for.[4] Recent studies have shown that many chatbot empathy recognition systems rely on surface-level language patterns and fail to fully consider the context of conversations.[6] Current empathy recognition systems often assess whether a response is empathetic based solely on semantic similarity, ignoring the relationship between preceding and following dialogue, which can lead to misclassification.[6]

2.2 Chain-of-Psychotherapies (CoP)

Many studies [3,4] have explored integrating psychotherapy techniques into chatbots to enhance their psychological support capabilities. One such approach is the CoP method, which combines various psychotherapy theories. CoP enables chatbots to assess users' psychological states by analyzing their input and selecting appropriate response strategies based on these evaluations. This method ensures that the chatbot's responses are tailored to the user's emotional needs, enhancing the interaction's effectiveness.

2.3 *Mixed Chain-of-Psychotherapies for Emotional Support Chatbot*

Chen et al. [3] expands upon the CoP method by introducing a more flexible and diverse range of psychotherapy techniques into the chatbot's response mechanism. They utilized CBT, PCT, and SFBT to create a dynamic response system that adapts to individual user needs. Given their consideration of the response mechanism, we anticipate that adding RAG would yield better results.

2.4 *Evaluation Methods for Chatbots*

Accurately evaluating the empathy performance of psychological support chatbots is crucial in research. Sharma, et al. [6] proposed an evaluation framework based on "Emotional Reactions," "Interpretations," and "Explorations." Li et al. [11] proposed an automated dialogue quality assessment metric called FlowScore. It primarily used to measure the contextual fluency and information dynamics of human-machine conversations. This technology is based on the DialoFlow model and evaluates the quality of responses by calculating the "semantic flow" of the conversation history. Traditional dialogue evaluation methods, such as BLEU or ROUGE, rely on the degree of match with reference answers [10], while FlowScore can dynamically assess the content of a chatbot's response based on the dialogue context without the need for a reference answer.

2.5 *Retrieval Augmented Generation Integrate with Large Language Models*

Zhang et al. [5] proposed an innovative system that integrates Retrieval-Augmented Generation (RAG) techniques into large language models, enabling the combination of both long-term and short-term memory. This approach leverages external knowledge bases to supplement the internal knowledge of the language model. The core of the RAG technique is its ability to combine retrieval mechanisms with generative models, not only improving the accuracy of generated text but also enhancing its personalization and coherence. This allows the model to provide more contextually relevant and coherent responses based on the user's historical interactions or background information.

III. ARCHITECTURE DESIGN

The system is built to overcome the current limitations of existing chatbots in understanding and responding to emotional distress, particularly in the context of psychological support. The architecture incorporates advanced AI technologies, such as Mistral Large 2, RAG [8], and CoP, to enhance empathy, context awareness, and response personalization. To achieve this, the design requirements (DRs) that guide our system development are as follows:

DR1: Provide users with deeper psychological support. This requirement is essential because emotional distress often involves complex and layered issues. To offer meaningful help, the chatbot must engage in conversations that delve into the underlying emotional states of users. By providing deeper psychological support, the chatbot can better understand and address the root causes of distress, creating a more effective therapeutic experience.

DR2: Offer more personalized and individualized supportive responses based on the user's input. Psychological support should not be one-size-fits-all. Each user's emotional state, background, and needs are unique. This requirement ensures that the chatbot tailors its responses to everyone, fostering a sense of personalization. Personalized responses build rapport and trust, making users more likely to feel heard and understood, which is critical for therapeutic effectiveness.

DR3: Provide specific, detailed, and actionable suggestions based on the user's environment and situation. General advice is often insufficient when dealing with psychological distress. To be truly helpful, the chatbot must offer practical, context-sensitive solutions that users can act upon. By understanding the user's environment and specific circumstances, the system can provide suggestions that are not only relevant but also feasible, helping users to feel empowered and supported in taking concrete steps toward recovery.

3.1 System Architecture Components

Based on the DRs outlined, the system is designed to provide deeper psychological support, personalized responses, and contextually relevant suggestions. To achieve these goals, the architecture is structured around several interconnected modules that work together to address the emotional distress of users in a comprehensive and responsive manner. These modules are crafted to ensure that the chatbot can provide meaningful, individualized support based on the user's emotional state, background, and environment. The system consists of two main core architectures: the training model architecture and the response generation architecture.

3.1.1 **Model Training Components**

- Psychotherapy Theory Integration: During the model training phase, multiple psychotherapy approaches were systematically integrated to inform. This integration ensures that the model learns from diverse therapeutic perspectives, enhancing its ability to provide deeper and more professional psychological support responses.
- Fine-Tuning: Fine-tuning the model based on psychiatrist feedback (ground truth), performance evaluations, and CoP ensures the chatbot's responses are empathetic, comforting, and contextually coherent. This iterative optimization process ensures that responses improve over time in terms of empathy, relevance, and practicality.

3.1.2 **Chatbot Response Generation Components**

- User Input: This module captures user's input, including emotional cues and context from previous interactions.
- Knowledge Base: This module stores a structured repository of psychological support materials, including

therapeutic dialogues and frameworks from established psychotherapy methods. This knowledge base serves as the foundation for generating accurate, relevant, and context-sensitive responses, as well as actionable suggestions tailored to the user's situation.

- Long-term Store & Short-term History: These modules enhance the system's contextual awareness by maintaining two complementary memory structures. The Long-term Store preserves essential user information across sessions (key individuals, events, emotional patterns), while the Short-term History captures recent conversational context.

- Retriever Module: A dense retrieval mechanism is employed to extract pertinent psychological content from the knowledge base, tailored to the user's input. By enhancing the relevance of the retrieved materials, the system is better to generate personalized and contextually appropriate responses.
- Response Generation Module: Leveraging Mistral Large 2 with RAG [5], this module generates contextually relevant and psychologically supportive responses. This approach enables the chatbot to deliver in-depth, context-sensitive support based on professional expertise.

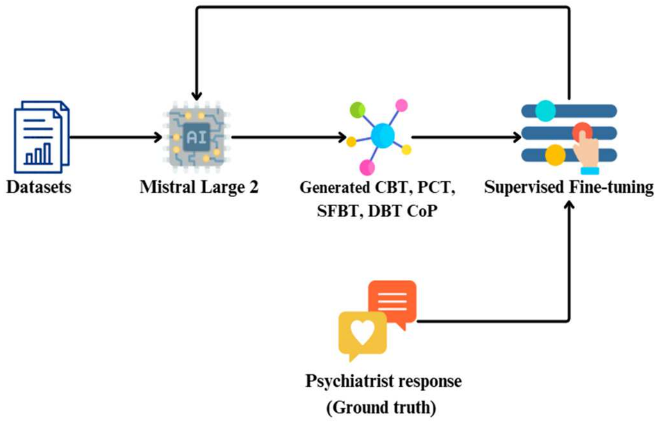

Figure 1. This diagram illustrates the training pipeline for the chatbot system. The Mistral Large 2 model undergoes supervised fine-tuning using specialized datasets and the CoP framework. Generated responses are evaluated against psychiatrist-provided ground truth to enhance the model's empathetic capabilities and psychological support language.

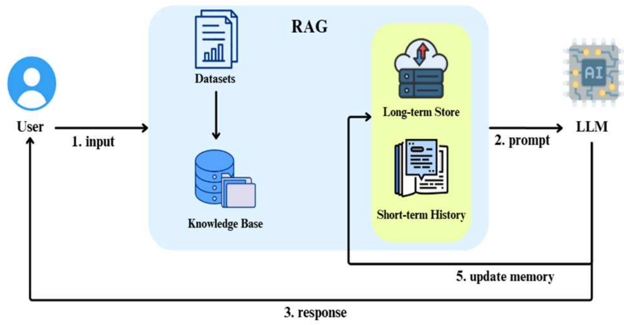

Figure 2. This figure demonstrates the RAG-enhanced deployment architecture of the system. Building upon the trained Mistral Large 2 model, this implementation incorporates knowledge retrieval mechanisms and both long-term and short-term memory components to deliver contextually relevant psychological support. The integration of previously learned therapeutic techniques enables dynamic response generation tailored to users' emotional states and specific situations.

The training model architecture (as shown in Figure 1.) utilizes Mistral Large 2 as the base model and incorporates various psychotherapy techniques to generate responses. The system fine-tunes the model using supervised learning by comparing generated responses with psychiatrist-provided responses. This process enhances the model's psychological support language, ensuring professional and empathetic responses.

The response generation architecture (as shown in Figure 2.) is based on RAG technology. By integrating a knowledge base and both long-term and short-term memory storage, the system dynamically adapts to the user's conversation history, providing deeper, more personalized, and actionable psychological support responses.

This dual-architecture design ensures that the system not only generates high-quality psychological support responses but also dynamically adjusts to users' emotions and needs during real-time conversations. This enables the chatbot to offer more professional and empathetic psychological support services.

3.2 Design Features

The core design of this system integrates RAG and CoP technologies into a unified framework, creating a uniquely therapy-oriented interaction mechanism. This innovative integration leverages the knowledge retrieval capabilities of RAG with the structured therapeutic guidance of CoP, ensuring that the chatbot delivers responses that are both informative and psychologically therapeutic.

3.2.1 **The Design of Integrated Framework: RA-CoP**

The feature design proposed in this research lies in the first-ever combination of RAG and CoP within a single system, addressing two key limitations of existing psychological support chatbots: insufficient contextual relevance and weak therapeutic foundations. By integrating these technologies, the system can dynamically retrieve professional mental health knowledge relevant to the user's emotional state, apply structured therapeutic frameworks to guide response generation, and provide personalized support based on the user's unique situation and interaction history. This dual enhancement mechanism ensures that the chatbot delivers responses that are not only informationally accurate but also aligned with professional psychotherapeutic principles.

3.2.2 **Extended Psychotherapy Integration**

Building upon the CoP framework of Chen et al. [3], this research extends the integration of psychotherapeutic theories, particularly by incorporating Dialectical Behavior Therapy to enhance the system's capability in supporting users experiencing high emotional distress or difficulties in emotion regulation. The system flexibly applies four major therapeutic approaches:

- Cognitive Behavioral Therapy (CBT): For identifying and challenging distorted thought patterns
- Person-Centered Therapy (PCT): Providing non-directive, empathetic validation
- Solution-Focused Brief Therapy (SFBT): Guiding users toward developing practical solutions
- Dialectical Behavior Therapy (DBT): Newly added therapy approach for emotion regulation and distress tolerance strategies

The system selects and applies these therapeutic approaches flexibly based on the user's psychological needs. For instance, when a user exhibits self-doubt, the system may employ PCT principles to provide acceptance and understanding; for users needing problem-solving, it may utilize SFBT techniques to guide toward viable solutions; and for users with high emotional volatility, it may integrate DBT mindfulness practices and emotion regulation techniques.

3.2.3 **Fine-Tuning for Empathy and Comfort**

The Fine-Tuning module plays a crucial role in refining the chatbot's ability to engage in emotionally resonant conversations. During training, researchers first generate CoP [3] based analyses derived from user dialogues.

The generated CoP insights are then integrated into the training dataset, serving as additional guidance for Fine-Tuning of the large language model. By exposing the model to these diverse therapeutic viewpoints during training, the chatbot gains a deeper understanding of user emotions and psychological needs, allowing it to adapt dynamically to different emotional contexts in real-world interactions. Furthermore, the fine-tuning process is iteratively refined with the assistance of CoP analysis and ground truth from psychiatrists, ensuring that the chatbot's responses align with professional therapeutic standards and provide effective psychological support.

IV. METHODOLOGY

The study aims to design a chatbot system for psychological support, optimized using fine-tuning and CoP design [3]. The methodology follows a structured process consisting of three key phases:

1) **Model Training (Supervised Fine-Tuning Phase)**

Training the chatbot using CoP-based psychotherapy insights and expert-labeled counseling dialogues to impr ove its empathetic and therapeutic capabilities.

2) **Response Generation (Deployment Phase using RAG)**

Implementing RAG to retrieve relevant psychological knowledge and generate responses dynamically.

3) **Test and Optimization**

Conducting both automated (Flow Score) and human-centered (Likert Scale) evaluations, refining the chatbot's responses through an iterative improvement process.

4.1 Model Training with Mistral Large 2 (Supervised Fine-Tuning Phase)

We utilize the Mistral Large 2 model, a 123-billion-parameter Mixture of Experts (MoE) model optimized for psychological dialogues through supervised fine-tuning. The training process begins with structured datasets, including Counsel Chat, Xinling [3], Empathetic Dialogues[5], forming the foundation of the chatbot's knowledge. During training, CoP analyses integrate multiple psychotherapy approaches to enhance therapeutic response generation. Psychiatrists iteratively review and refine chatbot responses, ensuring alignment with psychological principles by providing expert feedback on therapeutic accuracy. This supervised fine-tuning process enables the chatbot to continuously improve in delivering empathetic and psychologically sound support. In addition, since most existing systems rely on ChatGPT, we chose to explore the use of a different model with the expectation that it may offer the potential for improved results. By leveraging Mistral Large 2, we aim to evaluate whether alternative large language models can provide enhanced performance in delivering psychological support.

4.2 Response Generation Using RAG (Deployment Phase)

Once fine-tuning is complete, the chatbot is deployed with RAG [5] to dynamically generate responses during real-time interactions.

4.2.1 RAG-Based Response Generation Process

1. Knowledge Retrieval

When a user provides input, the chatbot retrieves expert-verified therapeutic materials from a structured knowledge base. Using a dense retrieval mechanism [9], it selects psychological content based on semantic similarity.

2. Context-Aware Response Generation

The retrieved knowledge is combined with the user's query and processed by Mistral Large 2. The chatbot then dynamically applies information from the Long-term Store and Short-term History to generate structured and personalized responses tailored to the user's needs. By incorporating both past interactions and recent context, the chatbot ensures that the responses are relevant, empathetic, and aligned with the user's current emotional state and situation.

By integrating retrieved expert knowledge with generative modeling, this approach ensures the delivery of psychologically informed and contextually appropriate responses. The system focus on semantic-based knowledge retrieval and dynamic context integration enables it to generate responses that are not only therapeutically aligned but also personalized to the user's emotional state, situational context, and interaction history. The comprehensive methodology enhances response quality across multiple dimensions, including psychological relevance, contextual coherence, and therapeutic effectiveness.

4.3 Evaluation

The chatbot is evaluated at two key stages:

4.3.1 FlowScore Assessment

FlowScore [11] is used as an automated evaluation metric after training to assess the chatbot's fluency, coherence, and emotional resonance. By ensuring that responses maintain both logical and emotional continuity, it helps validate the chatbot's effectiveness before deployment.

4.3.2 Likert Scale Assessment

System users rate chatbot responses on a Likert Scale, comparing outputs from:

- Psychiatrists (ideal standard).
- ChatGPT (baseline model).
- The Developed Chatbot (CoP-RAG optimized version).

Evaluators assess empathy, comfortability, and contextual coherence.

V. USE CASE

The conversation illustrated in the Figure 3. demonstrates a chatbot engaging in psychological support by responding to a service user struggling with chronic lateness. The chatbot applies different therapeutic techniques, aligning with principles from CBT and PCT. Initially, the service user exhibits overly negative thinking and self-frustration, which the chatbot addresses with a mix of humor and a challenge to negative self-talk, encouraging a shift in perspective. As the conversation progresses, the user expresses a deeper emotional struggle, revealing feelings of helplessness and externalizing blame. In response, the chatbot employs empathy and proactive guidance, suggesting practical strategies to regain control over time management. The interaction demonstrates how a chatbot can integrate psychological techniques to provide emotional validation while offering constructive behavioral interventions. The dialogue structure is adapted from an example in a psychological support reference book [14].

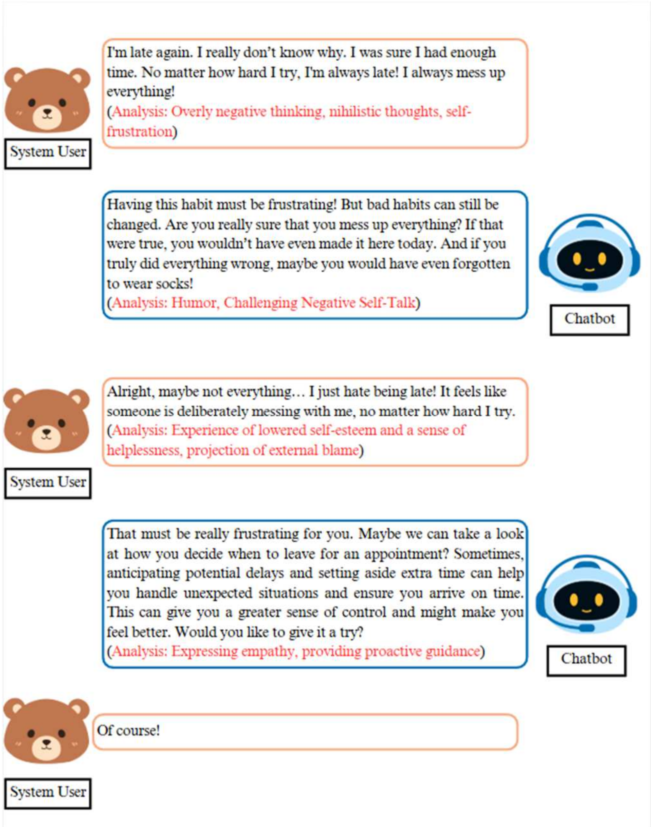

Figure 3. Example of a chatbot providing psychological support for a user struggling with chronic lateness. The chatbot utilizes cognitive-behavioral and person-centered therapy techniques, such as humor, challenging negative self-talk, empathy, and proactive guidance. Through this interaction, the chatbot helps the user shift their perspective, validate emotions, and adopt practical time management strategies.

VI. ETHICAL CONSIDERATION

The chatbot system and RA-CoP we designed offers numerous benefits, such as contextual relevance and therapeutic grounding in psychological support conversations. However, the development of psychological support chatbots raises several important ethical concerns. Data privacy and security is a primary concern, as psychological support conversations contain sensitive personal information that requires protection. Scope limitations are also significant, as chatbots have inherent constraints in their ability to provide comprehensive mental healthcare. Misinformation risks emerge because AI models may generate misleading or inaccurate responses due to limitations in training data and contextual understanding. Dependency concerns arise when users develop excessive reliance on chatbots, potentially delaying seeking professional care. Limited emotional intelligence remains a critical limitation, as AI cannot fully emulate human intuition in recognizing subtle psychological cues. These ethical considerations highlight that while chatbots can provide accessible emotional support tools, they cannot substitute the nuanced judgment of trained mental health professionals and should be viewed as complementary technologies within the broader healthcare ecosystem.

VII. CONCLUSION AND FUTURE WORK

Conclusion

In this study, we propose chatbot system with RA-CoP, a psychological support chatbot system designed to assist individuals who experience prolonged stress and have limited access to timely psychological support. RA-CoP addresses existing challenges in generating psychologically supportive language in chatbots by integrating an optimized CoP approach with RAG technology. This integration enables the chatbot to generate more professional and empathetic responses. Additionally, the chatbot can adapt its replies based on past conversations to ensure contextual relevance and responsiveness to users' current needs. We anticipate that our proposed framework will provide a new direction for the future development of AI in psychological support.

Future Work

In future research, we plan to implement strategies to address the ethical concerns identified. We will develop enhanced data protection through improved encryption methods, anonymization protocols, and secure storage infrastructures with clear informed consent procedures. Clear boundary communication will be implemented through better disclaimers and conversational reminders about the chatbot's limitations. We will incorporate improved safety mechanisms using expert-verified therapeutic data and safety filtering mechanisms supported by continuous user feedback. Professional care integration will be enhanced by designing systems that periodically encourage users to seek support from qualified professionals. We plan to develop better emotional state classification models to recognize subtle psychological cues with appropriate human oversight to identify high-risk patterns. Furthermore, we plan to further refine and implement our framework in real-world applications, with robust escalation protocols for complex scenarios that require human intervention.

VIII. REFERENCE

[1] H. Qiu, H. He, S. Zhang, A. Li, and Z. Lan, "SMILE: Single-turn to Multi-turn Inclusive Language Expansion via ChatGPT for Mental Health Support," Findings of the Association for Computational Linguistics: EMNLP 2024, Miami, FL, USA, pp. 615–636, 2024. [Online]. Available: https://doi.org/10.18653/v1/2024.findings-emnlp.34.

[2] S. B. Loh and A. S. Raamkumar, "Harnessing Large Language Models' Empathetic Response Generation Capabilities for Online Mental Health Counselling Support," Oct. 2023. [Online]. Available: https://doi.org/10.48550/arxiv.2310.08017.

[3] S. Chen, C. Ming, Z. Zhang, Y. Chen, K. Zhu, Q. Kenny, and M. Wu, "Mixed Chain-of-Psychotherapies for Emotional Support Chatbot,"Jan. 2024. [Online]. Available: https://doi.org/10.48550/arxiv.2409.19533.

[4] Y. K. Lee, I. Lee, M. Shin, S. Bae, and S. Hahn, "Chain of Empathy: Enhancing Empathetic Response of Large Language Models Based on Psychotherapy Models," Nov. 2023. [Online]. Available: https://doi.org/10.48550/arxiv.2311.04915.

[5] X. Zhang and Z. Luo, "Advancing Conversational Psychotherapy: Integrating Privacy, Dual-Memory, and Domain Expertise with Large Language Models," Dec. 2024. [Online]. Available: https://doi.org/10.48550/arxiv.2412.02987.

[6] A. Sharma, W. I. Lin, A. S. Miner, D. C. Atkins, and T. Althoff, "Towards Facilitating Empathic Conversations in Online Mental Health Support: A Reinforcement Learning Approach," in Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems (CHI), 2021. [Online]. Available: https://doi.org/10.1145/3442381.3450097.

[7] A. Li, L. Ma, Y. Mei, H. He, S. Zheng, H. Qiu, and Z. Lan, "Understanding Client Reactions in Online Mental Health Counseling," in *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (ACL)*, 2023. [Online]. Available: https://doi.org/10.18653/v1/2023.acl-long.577.

[8] L. Patrick et al., "Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks," May 2020. [Online]. Available: https://doi.org/10.48550/arxiv.2005.11401.

[9] A. M. Agrawal et al., "Conversation AI Dialog for Medicare powered by Fine-tuning and Retrieval Augmented Generation," Feb. 2025. [Online]. Available: https://arxiv.org/pdf/2502.02249.

[10] K. Papineni, S. Roukos, T. Ward, and W. Zhu, "BLEU: A Method for Automatic Evaluation of Machine Translation," in *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics (ACL)*, 2002, pp. 311–318. [Online]. Available: https://dl.acm.org/doi/10.3115/1073083.1073135.

[11] Z. Li, J. Zheng, Z. Fei, Y. Feng, and J. Zhou, "Conversations Are Not Flat: Modeling the Dynamic Information Flow across Dialogue Utterances," in *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics (ACL)*, 2021. [Online]. Available: https://doi.org/10.18653/v1/2021.acl-long.11.

[12] N. Chi et al., "Usability Testing of the Pace App to Support Family Caregivers in Managing Pain for People with Dementia," in *Innovation in Aging*, vol. 7, no. S1, 2023. [Online]. Available: https://doi.org/10.1093/geroni/igad104.2763.

[13] H. Rashkin, E. M. Smith, M. Li, and Y. Boureau, "Towards Empathetic Open-domain Conversation Models: a New Benchmark and Dataset," Nov. 2019. [Online]. Available: https://doi.org/10.48550/arXiv.1811.00207.

[14] W. Arnold, R. Richard, and H. Pinsker, "*Introduction to Supportive Psychotherapy*", translated by L.-X. Zhou and D.-J. Tsai, 1st ed. Amer Psychiatric Pub Inc, 2005.

[15] O. Vesga-López, C. Blanco, K. Keyes, M. Olfson, B. F. Grant, and D. S. Hasin, "Psychiatric disorders in pregnant and postpartum women in the United States," *Archives of General Psychiatry*, vol. 65, no. 7, pp. 805–815, Jul. 2008. [Online]. Available: https://doi.org/10.1001/archpsyc.65.7.805.

[16] S. Liu, C. Zheng, O. Demasi, S. Sabour, Y. Li, Z. Yu, Y. Jiang, and M. Huang, "Towards Emotional Support Dialog Systems," in Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics (ACL), 2021. [Online]. Available: https://doi.org/10.18653/v1/2021.acl-long.269.