



Exploration of an Innovative AI-Based Mental Health Education Model in Universities Based on Emotion-Oriented Psychotherapy

Huansheng Li

Zhejiang Institute of Economics and Trade, Hangzhou, Zhejiang, China

Abstract: Objective — There is a shortage of professional faculty in university mental health education settings, and the dissemination of Emotion-Oriented Therapy and Accelerated Experiential Dynamic Psychotherapy within universities is constrained. Methods — This study explores the integration of the theoretical core of the two emotion-oriented therapies with artificial intelligence technologies. The research is conducted using a mixed-methods approach. Bibliometric analysis and theoretical analysis are employed to examine the digital integration forms of Emotion-Oriented Therapy and Accelerated Experiential Dynamic Psychotherapy. Five universities in Zhejiang Province were selected as research subjects. A total of 217 valid questionnaires were collected, and 30 students exhibiting emotional distress were selected for interviews, forming corresponding emotional corpus data. Based on a large language model, instruction fine-tuning was conducted to develop an AI agent with capabilities of deep emotional recognition and transformation. Results — Survey data indicate that emotional expression among college students shows characteristics of concealment. Among the respondents, 44.6% exhibited anxiety, and 2.8% demonstrated active help-seeking behavior. Conclusion — The AI agent achieved an accuracy rate of 82.3% in distinguishing secondary emotions such as anger from core emotions such as fear.

Keywords: Emotion-Oriented Therapy; accelerated experiential dynamic psychotherapy; generative artificial intelligence; university mental health education; AI agent

1. Introduction

With the continuous expansion of higher education, social competitive pressures are increasingly transmitted to university students, and the proportion of emotional conditions such as anxiety and depression among college students is rising. The clinical application of Emotion-Oriented Therapy and Accelerated Experiential Dynamic Psychotherapy depends on practitioners' professional competence and the duration of the therapeutic relationship. Human resource constraints and faculty allocation limit the application scope of these two therapies in universities [1]. Emotion-Oriented Therapy is a therapeutic approach that emphasizes the key role of emotion in psychological change, and the development of generative artificial intelligence technology provides technical support for the digital transformation of psychological services [2-3]. This study analyzes the integration of the emotional acceptance theory of Emotion-Oriented Therapy and the accelerated transformation theory of Accelerated Experiential Dynamic Psychotherapy into AI systems, and systematically organizes the processes through which AI identifies and intervenes in college students' emotions, thereby forming a corresponding model for university mental health education.

2. Theoretical Framework: The Logic of Digital Integration of Emotion-Oriented Psychotherapy

Emotion-oriented psychotherapy includes various therapeutic approaches that take emotion as the core of intervention. Emotion-Oriented Therapy and Accelerated Experiential Dynamic Psychotherapy are the main components of the emotion-oriented therapeutic system. The integration and digital transformation of these two therapies constitute the core of this study [4].

2.1 Core Mechanisms and Encodability of Emotion-Oriented Therapy

When practitioners assist clients in distinguishing primary emotions from secondary emotions, they adopt a "scenario recall + layered questioning" approach, guiding clients to recall in detail the specific situations that trigger emotions, accurately capturing the initial emotional response (primary emotion), and then comparing it with subsequent derived emotions (secondary emotions). For anger, practitioners inquire about bodily sensations and thoughts preceding its emergence to determine whether it is a secondary emotion of fear. When encountering emotional numbness, concrete guidance (e.g., "If you were to describe your current mood with a color, what would it be?") is used to uncover the underlying avoided sadness.

After each emotional distinction, clients are asked to label the differences between the two emotions using simple terms, forming a personal emotional comparison table, thereby repeatedly reinforcing the logic of differentiation and avoiding confusion [5].

According to the theoretical background of Emotion-Oriented Therapy, emotions are categorized into five basic types and transformed into different emotional labels. Correspondingly, an AI text recognition model is constructed, marking the core textual features and tonal tendencies of each type of emotion. At the same time, three intensity levels — mild, moderate, and severe — are defined and quantitatively determined based on keyword frequency, use of modal particles, and sentence structure [6]. Accelerated Experiential Dynamic Psychotherapy (AEDP), developed by Dr. Diana Fosha in the 1990s on the basis of psychodynamic psychotherapy, integrates contemporary developments in attachment theory, transformation theory, and affective neuroscience, representing the latest direction in psychotherapy [7-8]. Its core lies in focusing on and facilitating transformation. The models of "worst self-functioning" (source of psychopathology) and "optimal self-functioning" (thriving growth) provide the theoretical basis for explaining the formation of the worst self and the best self [9-10]. The AI system incorporates a lexicon of avoidant expressions. For statements such as "I'm fine" or "it's okay," it combines contextual semantic analysis to capture details such as negative sentence patterns and vague responses, linking them to underlying emotions such as suppressed sadness and unmet attachment needs. It then generates a text analysis report including emotional type, intensity, and potential underlying emotions. Each report simultaneously provides three targeted follow-up questioning directions, offering precise textual support for subsequent practitioner interventions. At the same time, based on client feedback, the AI dynamically adjusts keyword weights and associative logic [11].

2.2 Core Mechanisms and Digital Translation of Accelerated Experiential Dynamic Psychotherapy

The AI agent implements specific interventions according to different therapeutic stages. In the intake stage, open-ended questions without preset answers are adopted, such as "What emotion do you most want to be seen right now?" and "What moment has touched you the most during this period?" The process involves no interruption or judgment, while accurately recording keywords related to attachment needs in the user's expressions, such as "desire to be valued" or "sadness with no response," and simultaneously capturing changes in tone and speech rate to establish a pressure-free communicative connection. In the resonance stage, mechanical empathy is abandoned in favor of empathic responses. When users express negative emotions, the AI directly mirrors and reflects their emotional details. For example, if a user says, "I feel like I'm always not needed," the response would be, "At this moment, you are being overwhelmed by the sense of 'not being needed.' This empty feeling has been with you for a long time." This approach accurately reaches surface emotions and lays the groundwork for deeper exploration.

In the emotional deepening stage, the AI agent actively follows up on defensive expressions while avoiding rigid preaching. For instance, when a user avoids sadness, it may ask, "When you say 'I don't care,' is there a trace of sadness that you don't dare to reveal?" This gradually breaks down defense mechanisms and guides the user to confront core emotions. In the transformation stage, the focus is on guiding self-awareness, conducting dialogues centered on the user's internal emotional needs. For example, if a user longs to be understood, the AI guides them to describe "what kind of response you hope to receive when you are understood," and further asks, "I hope you can feel that I am here accompanying you. I invite you to experience this longing and the emotions and needs behind it." Rather than instilling cognitive viewpoints, the AI enables users to independently organize their emotional logic. In the integration stage, the AI assists users in sorting out the trajectory of emotional changes and guides them to summarize concrete methods for coping with core emotions. For example, "The next time you experience this sense of grievance, I hope you can recall that I am here accompanying you to experience this emotion, and you can try to first express your longing, just like you are doing now." This makes the intervention effects actionable and reusable.

3. Research Methods and Construction of the Emotional Data Collection System

An empirical investigation was conducted from October to December 2025. A mixed-methods approach was adopted, combining quantitative questionnaires and qualitative interviews to record the characteristics of emotional expression and psychological needs among university students.

3.1 Research Subjects and Sampling Strategy

A stratified convenience sampling method was used to select research subjects. Five universities in Zhejiang Province were chosen, including comprehensive universities, science and engineering institutions, normal universities, and finance and economics colleges at both undergraduate and junior college levels. The investigation was initiated after obtaining approval from university student affairs departments and ethical review committees. The questionnaire survey was conducted through

a combination of online and offline distribution. A total of 235 questionnaires were distributed. After excluding those with abnormal response times, response bias, or missing information, 217 valid questionnaires remained, resulting in an effective response rate of 92.3%. The sample covered students from freshman to senior years, with a balanced distribution in terms of gender and academic disciplines. Based on questionnaire scale scores and responses to open-ended questions, 30 students exhibiting anxiety, social avoidance, and academic stress were selected for interviews. All participants signed informed consent forms. Interviews were conducted in independent rooms within university psychological counseling centers, with each session lasting no more than 90 minutes. The total interview duration reached 45 hours, and the transcribed interview texts amounted to approximately 120,000 words.

3.2 Data Processing and Emotional Coding System

All interviews were audio-recorded. With participants' informed consent, two graduate students majoring in psychology completed the transcription within 24 hours. The transcription process retained nonverbal expressions such as pauses, repetitions, and sighs. All interview texts were imported into NVivo 14 software for three-level coding. The coding criteria were based on a self-developed EFT-AEDP hybrid emotional annotation manual. The manual classifies emotional expressions into four dimensions: defensive emotions, secondary emotions, traumatic emotions, and core emotions. Traumatic emotions include ungrieved sadness, fear of abandonment, fear of being hurt, shame of worthlessness, and sadness lacking joy (lonely sadness). Secondary emotions include helplessness, shame, and anxiety. Defensive emotions include psychological manifestations such as avoidance, denial, and intellectualization. Core emotions (primary emotions) refer to internal emotions related to attachment needs, including anger, joy, shame, sadness, and fear. After completing training on the annotation manual, three coders conducted preliminary coding. Five texts were randomly selected for joint coding, yielding an inter-coder reliability Kappa value of 0.81. Upon completion of the coding process, a corpus of annotated emotional data was formed, containing thousands of labeled sentences.

4. Model Construction and Training Process of the AI Agent

Based on the theoretical framework and empirical corpus, this study developed an AI mental health agent named Xinyu. The development process includes architecture design, instruction fine-tuning, and function encapsulation.

4.1 Algorithm Architecture Design and Instruction Fine-Tuning

The model adopts a two-stage hybrid architecture. In the first stage, an emotion recognition layer is constructed. Based on a pre-trained BERT model, emotionally annotated corpora covering multiple scenarios such as daily conversations and psychological interactions are selected. The corpus is divided according to multi-label emotion classification standards, with each sentence annotated for emotional categories and corresponding features for model fine-tuning. During the fine-tuning process, for ironic and metaphorical language, a recognition algorithm combining semantic similarity calculation and contextual association analysis is designed. It extracts ironic signal words and metaphorical mapping relationships in the text, and integrates contextual semantic weights to achieve accurate recognition of irony and metaphor [12]. For defensive language, a module for repeated vocabulary statistics and semantic integrity detection is constructed. When users produce expressions such as "I'm fine, really fine," the system automatically counts the frequency of repeated words and calculates semantic vector similarity. If the repetition frequency is ≥ 2 and no new semantic information is added, it is directly identified as a secondary emotional defense expression. The semantic features of such expressions are simultaneously recorded for subsequent model iteration and optimization.

In the second stage, a dialogue generation layer is constructed based on the GPT-3.5 model. The study systematically organizes complete dialogue cases from the five stages of Accelerated Experiential Dynamic Psychotherapy and decomposes them into three core components: exploratory questioning, empathic response sentences, and reconstructive responses. These are compiled into more than 1,200 pairs of instructions. Exploratory questioning focuses on designing specific prompts targeting users' emotional triggers; empathic response sentences accurately mirror users' emotional expressions and enhance emotional resonance; reconstructive responses integrate users' past experiences to guide cognitive adjustment. This set of instruction pairs is used to fine-tune the GPT-3.5 model. A reinforcement learning mechanism based on human feedback is introduced. A panel of professional psychotherapists is invited to form a scoring group, using the alignment of Emotion-Oriented Therapy's validating responses and the effectiveness of Accelerated Experiential Dynamic Psychotherapy's transformational questioning as core evaluation dimensions. Model-generated dialogues are scored on a scale of 1 to 5, and the scoring results are converted into reward signals to adjust model parameters in reverse, thereby improving the specificity and therapeutic adaptability of dialogue generation [13].

4.2 Core Functional Modules and Hierarchical Intervention Strategies

The system extracts continuous user dialogue texts, capturing emotional keywords, tonal features, and semantic tendencies. Data statistics are conducted along daily and weekly temporal dimensions to generate a three-dimensional emotional profile. The emotional stability index is divided into five quantitative levels based on the frequency of daily emotional fluctuations and the range of extreme value differences, while peak fluctuation periods are recorded simultaneously. The defense intensity score annotates linguistic features corresponding to behaviors such as avoidance, intellectualization, and denial, and calculates weighted scores based on their frequency within a single dialogue, thereby clarifying behavioral manifestations associated with different score levels [14]. The core needs map is constructed using semantic clustering algorithms to extract attachment-related themes (such as companionship, recognition, and sense of security). A heat map is used to present the frequency of theme occurrence, associated vocabulary, and contextual scenarios, enabling precise identification of users' core demands.

Users' emotional states are categorized into three levels for intervention. For mild distress, Emotion-Oriented Therapy techniques are applied, guiding users to accurately label their current emotions through "emotion reflection + targeted feedback," encouraging acceptance of their emotions without avoiding negative feelings. For moderate distress, the approach of Accelerated Experiential Dynamic Psychotherapy is adopted, using open-ended questions to guide users in recalling emotional triggering scenarios, organizing emotional responses within those contexts, and promoting deeper expression of internal feelings. The system continuously monitors dialogue texts in real time. It predefines more than 20 core keywords and synonymous expressions related to self-harm, hopelessness, and suicide planning, and establishes a three-level trigger threshold. Once the highest threshold is reached, the AI dialogue is immediately terminated, and a pop-up window displays the 24-hour hotline of the university psychological counseling center, the address of psychiatric emergency services, and treatment procedures [15-16]. At the same time, an anonymous alert is sent to the student affairs system, including periods of emotional abnormalities and keyword excerpts, with all personal information removed [17].

5. Analysis of Application Effectiveness and Discussion of Model Advantages

After completing closed testing, an analysis of application effectiveness was conducted. The analysis focuses on the accuracy of complex emotion recognition and the specific characteristics of the model.

5.1 Model Performance Test Results

The testing process used 200 texts independent of the training set. These texts were derived from student interview excerpts that were not involved in model training, and were annotated with gold standards by experts. The test results show that the AI agent achieved an overall accuracy of 82.3% in distinguishing primary emotions from secondary emotions. The model's accuracy in identifying secondary anger reached 78%, while its accuracy in identifying core fear reached 85%. The model also demonstrates the ability to recognize defensive language. When users express statements such as "I'm fine" or "it doesn't matter," the model identifies the text as a defensive state and generates dialogue that bypasses defensive content. The conversation focuses on the user's internal emotional experiences without involving discussion of specific events.

5.2 Comprehensive Characteristics of the Innovative Model

Traditional university psychological counseling primarily adopts cognitive behavioral therapy and solution-focused brief counseling, with intervention content oriented toward cognitive restructuring and problem-solving. The model proposed in this study integrates the content of Emotion-Oriented Therapy and Accelerated Experiential Dynamic Psychotherapy, with AI interventions covering relational issues and the processing of core emotions. Emotion-Oriented Therapy and Accelerated Experiential Dynamic Psychotherapy require a high level of professional competence from practitioners. The AI system transforms the core logic of these two therapies into operational rules and trained models, forming an automated service tool capable of long-term response. The number of students served is not limited by human resource conditions. The AI agent can simultaneously conduct large-scale emotional screening of students and classify their emotional states based on emotional profiles. It then outputs corresponding levels of dialogue content for students in different states, covering types such as mild guidance, deep transformation, and crisis referral.

6. Conclusion

This study focuses on the current situation of professional service provision and students' emotional needs in university mental health education, and systematically outlines the integration pathways of Emotion-Oriented Therapy, Accelerated Experiential Dynamic Psychotherapy, and generative artificial intelligence technologies. The research completes theoretical analysis, empirical investigation, and model training, forming a model for university mental health education that integrates

AI with emotion-oriented psychotherapy. In psychological service scenarios, the AI agent functions as both a carrier and a tool, providing students with emotion-related dialogue services. Future research can expand the model's capability to recognize multimodal emotional content such as speech intonation and facial micro-expressions, and conduct long-term experiments in university campus environments to record the model's operational processes and data performance in real-world settings.

Acknowledgments

This paper is supported by the following fund project: General scientific research project of Zhejiang Province Education Department in 2025 "Innovative research on AI-enabled emotion-oriented psychotherapy in colleges and universities" (Project No.: Y202559523).

References

- [1] Gao Ge. The impact of AI emotion recognition technology on university psychological services from the perspective of digital-psychological integration [J]. *Heilongjiang Science*, 2025, 16(23): 82-84.
- [2] Sun Juncai, Gao Zengming, Zheng Xinjun. Understanding and treatment of pathological emotions in Emotion-Oriented Therapy [J]. *Psychological Research*, 2014, 7(4): 6.
- [3] Leslie S. Greenberg. *Emotion-Oriented Therapy* [M]. Chongqing University Press, 2015.
- [4] Zhu Kangtai, Xie Shuqun. A study on the relationship among employment pressure, emotion regulation strategies, and psychological capital of university students [J]. *Vocational Education*, 2025, 24(26): 75-80.
- [5] Tan Yaya, Liu Jun, Wang Shumin. An empirical study on embodied cognition-based teaching design in university mental health courses: A case study of the "emotion management" module [J]. *Media and Art Research*, 2025, (02): 173-179.
- [6] Elliott R. Emotion-Oriented therapy [J]. *Psychotherapy*, 2022(2): 59.
- [7] Yang Zhaoqian. *The Instinct to Heal* [M]. China Renmin University Press, 2023: 343.
- [8] Iwakabe Shigeru; Edlin Jennifer; Fosha Diana; Thoma Nathan C; Gretton Heather; Joseph Andrew J; Nakamura Kaori. The long-term outcome of accelerated experiential dynamic psychotherapy: 6- and 12-month follow-up results [J]. *Psychotherapy* (Chicago, Ill.), 2022.
- [9] Wang Juan, Danny Yeung, Guo Zengrui, et al. Current status and progress of accelerated experiential dynamic psychotherapy [J]. *Medicine and Philosophy*, 2024, 45(16): 30-34+44.
- [10] Iwakabe Shigeru; Edlin Jennifer; Fosha Diana; Gretton Heather; Joseph Andrew J; Nunnink Sarah E; Nakamura Kaori; Thoma Nathan C. The effectiveness of accelerated experiential dynamic psychotherapy (AEDP) in private practice settings: A transdiagnostic study conducted within the context of a practice-research network [J]. *Psychotherapy* (Chicago, Ill.), 2020.
- [11] Hu Chun. From emotion management to psychological counseling: Practical pathways for university mental health education [J]. *Pinwei Classics*, 2025, (15): 73-75.
- [12] Li Jialin, Su Jinyao, Zhong Weiwei, Hao Jinying, Li Yan. From the perspective of service-oriented education, evaluation of the intervention effects of group counseling on emotional intelligence and listening ability of university peer counselors [J]. *Society and Public Welfare*, 2025, (13): 240-243.
- [13] Zhou Li. The impact of stress events on the mental health of university counselors: A chain mediation effect study of cognitive emotion and meaning in life [J]. *Psychological Monthly*, 2024, 19(14): 76-79+86.
- [14] Wang Feifei. The relationship between emotional intelligence and psychological resilience among university students [J]. *Heilongjiang Science*, 2024, 15(05): 137-139.
- [15] Shi Huawei, Wang Fengxia, Ma Zhiqiang, Deng Rong. The mechanism of organizational support influencing university teachers' professional well-being through psychological capital and emotional exhaustion [J]. *Chinese Journal of Health Education*, 2023, 39(11): 1011-1014.
- [16] Zhang Jiawen, Ma Lin, Liang Yufeng, Ma Shichao. Emotional labor and emotional exhaustion among university psychological service personnel: The moderating effect of Zhongyong thinking [J]. *Psychological Monthly*, 2023, 18(18): 80-82.
- [17] Ji Yeshan, Wang Zhaodong, Li Yongqin. A study on the relationship between psychological resilience and core self-evaluation among university athletes: The mediating role of emotional intelligence [J]. *Contemporary Sports Technology*, 2023, 13(25): 151-156.