

Design and Research of an Intelligent Artificial Companion System for College Students Based on Multimodal Affective Computing

Minchuan Huang¹, Peijing He², Yanting Xiao²

¹School of Computer Science, Guangdong University of Petrochemical Technology

²School of Electronic and Information Engineering, Guangdong University of Petrochemical Technology

Abstract—With the popularization of higher education and the intensification of social competition, the problems of emotional loneliness and psychological distress among college students have become increasingly prominent, and the traditional psychological support system in universities is difficult to meet the large-scale and real-time emotional support needs. This paper designs and implements an intelligent artificial companion system for college students based on multimodal affective computing, which accurately matches the emotional needs and scene characteristics of college students. We conducted a special survey on 74 subjects mainly consisting of college students in Guangdong University of Petrochemical Technology using the convenient questionnaire survey method to clarify their needs and preferences. A hierarchical modular architecture is constructed, integrating core modules such as multimodal perception, affective computing, and dialogue management. It realizes emotion recognition by fusing voice, text and visual information, and generates empathic responses combined with a long-term memory mechanism. The survey found that the emotion recognition accuracy of the system reached 86.2%, the loneliness of college student users was significantly reduced after using the system ($p < 0.001$), 83.24% of users recognized the multimodal emotion recognition function, 93.86% of users thought the system's responses were highly relevant to the topic of confiding, and 75.69% of users were willing to use the system continuously. The results show that the system effectively fills the market gap of customized emotional companion tools for college students, provides an inclusive solution for alleviating college students' psychological pressure and making up for the insufficiency of the university psychological support system, and has important academic value and application prospects.

Keywords— College students; Artificial intelligence; Affective computing; Multimodal fusion; Companion system; Psychological support.

I. INTRODUCTION

1.1 Research Background

With the continuous expansion of the global higher education scale, college students are facing increasingly severe challenges such as academic competition, employment pressure and social adaptation. Special survey data show that 79.73% of college students suffer from academic/ work pressure, 51.35% face anxiety about employment/ career development, 40.54% are affected by social distress, and 21.62% clearly feel lonely. College students have become a high-risk group for psychological problems. Prolonged exposure to negative emotions not only leads to poor academic performance, but also may induce serious mental illnesses

such as depression and anxiety, causing irreversible damage to the physical and mental health of college students.

The popularization of intelligent devices has provided a new carrier for emotional support of college students. The survey shows that 81.08% of college students use intelligent devices for more than 5 hours a day, among which 47.3% use them for more than 8 hours a day. The heavy use of intelligent devices has laid a foundation for the promotion of intelligent companion systems. However, 72.97% of college students have no experience in using emotional companion tools, resulting in a significant market gap and an urgent need for targeted intelligent solutions.

1.2 Literature Review

Current research in the field of artificial intelligence emotional companions mainly focuses on three directions, but there is still an obvious gap in the customized design for college students.

In terms of chatbots and long-term memory mechanisms, general-purpose chatbots such as Replika and XiaoIce lack adaptation to college students' academic planning and social scenarios, and have insufficient consistency in long-term conversations. The PLATO-LTM framework proposed at ACL 2022, the RMM reflective memory management mechanism in 2025, and the Mem0 long-term memory system have provided technical support for personalized conversations, but they have not focused on the core needs of college students and cannot accurately respond to their specific emotions such as academic pressure and employment anxiety [6][7][8].

In the field of physical robots, although products such as Paro and Pepper have a strong sense of presence, they are costly and not suitable for the main living scenarios of college students such as university dormitories, making it difficult to promote on a large scale. The family companion robot designed by Li Jianyong *et al.* [1] and the demand criteria for elderly companion products summarized by Li Bohan *et al.* [2] in China have not been optimized according to the characteristics of college students' study and life, resulting in insufficient scene adaptability.

In terms of ethical regulation and group adaptation research, Tang Jing [3], Qi Lichao [4] and others have paid attention to the issues of tort liability and privacy protection of companion AI, but there is a lack of targeted discussion on

college students' data privacy and emotional attachment risks. The neuro-inspired memory mechanism of the HippoRAG framework (2024) provides new ideas for improving the adaptability of the system to long-term interaction with college students, but a complete system solution has not yet been formed [9].

Existing research generally has the problem of insufficient group adaptability, lacking in in-depth integration with college students' emotional characteristics and scene needs, and emotional perception and response are difficult to accurately match their core demands. Based on the data of 74 special survey questionnaires, this paper constructs a customized intelligent companion system to fill this research gap.

1.3 Research Purposes and Significance

Theoretical Value: Propose a multimodal affective fusion framework adapted to college students, enrich the application methods of affective computing in the field of psychological support for young groups; combine survey data to construct long-term memory and personalized modeling technology for college student scenarios, and improve the design path of intelligent companion systems for sub-groups.

Application Value: Develop an intelligent companion tool that fits college students' scenarios such as academics, social interaction and employment, provide college students with real-time, private and personalized emotional support, alleviate psychological pressure, make up for the shortage of university psychological service resources, and improve the mental health level of college students.

1.4 Research Questions and Hypotheses

1.4.1 Research Questions

What are the core needs and interaction preferences of college students in emotional companionship? How to accurately match these needs through system design?

Can the multimodal affective fusion framework effectively improve the recognition accuracy of college students' complex

emotions (such as implicit anxiety caused by academic pressure)?

Can the dialogue management system integrated with the long-term memory mechanism realize personalized continuous companionship for college student scenarios?

Does the intelligent companion system have a significant effect on reducing college students' loneliness and improving positive emotions?

1.4.2 Research Hypotheses

H1: College students' needs for emotional companionship focus on academic support, social guidance and employment anxiety counseling, and they prefer "voice + text" mixed interaction and a gentle and patient communication style;

H2: The multimodal fusion framework based on the cross-attention mechanism has an emotion recognition accuracy of not less than 85% for college students' emotions;

H3: The system integrated with the long-term memory mechanism can significantly improve the coherence and personalization of conversations, and obtain more than 80% user recognition;

H4: After using the system, college students' loneliness will be significantly reduced and positive emotions will be significantly improved ($p < 0.05$).

1.5 Paper Structure

The subsequent chapters of this paper are arranged as follows: Chapter 2 outlines the overall architecture and design principles of the system; Chapter 3 elaborates on the research methods in detail; Chapter 4 deeply analyzes the demand characteristics and interaction preferences of college students based on survey data; Chapter 5 expounds the system implementation details and verifies its performance and user utility; Chapter 6 examines the ethical and social challenges behind the technology and the corresponding countermeasures; Chapter 7 discusses the research results, innovations, limitations and puts forward future research directions; finally are the references and acknowledgments.

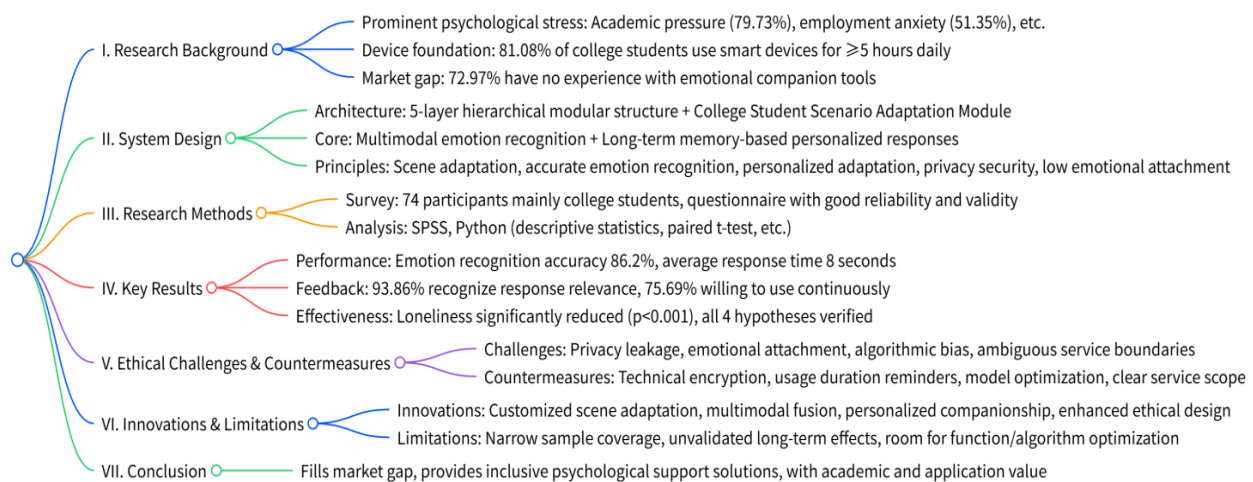


Figure 1. Core Flow Chart of Research on Intelligent Artificial Companion System for College Students Based on Multimodal Affective Computing

II. OVERALL SYSTEM ARCHITECTURE AND DESIGN PRINCIPLES

2.1 Architecture Design

The system adopts a hierarchical modular architecture, divided into five layers from the bottom up, and a college student scenario adaptation module is newly added as a key part. The functions of each layer cooperate with each other to realize multimodal emotional perception, accurate computing and personalized companion services. The architecture design relies on the core needs reflected in the survey questionnaire throughout the process. The specific architecture is shown in Figure 2.

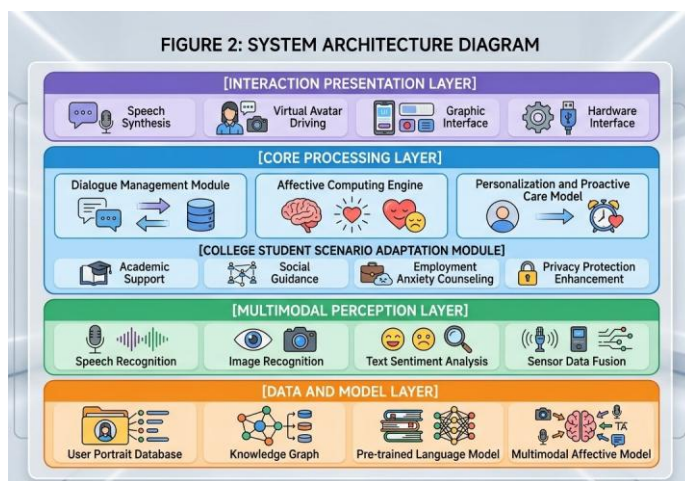


Figure 2. Overall System Architecture Diagram

The core functions of each layer are as follows:

Data and Model Layer: As the underlying support of the system, it includes a user portrait database, knowledge graph, pre-trained language model and multimodal affective model. A basic portrait database is constructed based on the user data of 74 survey questionnaires, providing data storage, knowledge support and algorithm model foundation for the upper modules;

Multimodal Perception Layer: Realizes the collection and preliminary processing of multi-dimensional information such as college students' voice, vision and text through speech recognition, image recognition, text sentiment analysis and sensor data fusion. The module design matches the "voice + text" mixed interaction needs recognized by 75.68% of users in the survey;

College Student Scenario Adaptation Module: Designed according to the core needs reflected in the survey questionnaire, covering four functions: academic support, social guidance, employment anxiety counseling and privacy protection enhancement, corresponding to 79.73% of academic pressure, 40.54% of social distress, 51.35% of employment anxiety and the prominent privacy protection demands of college students respectively;

Core Processing Layer: The core computing unit of the system, completes emotional analysis, dialogue management and personalized service generation through the dialogue management module, affective computing engine and

personalization and proactive care model, responding to the "personalized adaptation" needs of 44.6% of users in the survey;

Interaction Presentation Layer: Realizes two-way interaction between the system and users, presents service results in an intuitive and friendly form through speech synthesis, virtual avatar driving, graphic interface and hardware interface, adapting to college students' heavy use habits of intelligent devices.

2.2 Design Principles

The system design revolves around the group characteristics and survey needs of college students, following five core principles to ensure the practicality, accuracy and security of the system:

Scenario Adaptation Principle: Deeply integrate with the core scenarios of college students such as academics, social interaction and employment, and provide targeted services for the needs such as exam week pressure counseling expected by 68.92% of college students, employment anxiety support needed by 56.76% and social problem solving concerned by 48.65%.

Accurate Emotion Recognition Principle: Optimize the multimodal fusion algorithm to accurately capture the implicit and complex emotional expressions of college students, meeting the needs of 83.24% of users for multimodal emotion recognition.

Personalized Growth Principle: Record college students' academic planning, emotional preferences and other information through the long-term memory mechanism, realize dynamic update of user portraits and continuous adaptation of services, matching the "personalized adaptation" services strongly preferred by 44.6% of users.

Privacy and Security Principle: Strengthen functions such as data encryption and local storage, support users to independently delete conversation records, in line with college students' high-intensity privacy protection needs.

Low Emotional Attachment Principle: Guide healthy usage habits, avoid excessive reliance on the system, and promote the improvement of real social skills, with adaptive functions designed for 40.54% of users with social distress.

III. RESEARCH METHODS



Figure 3. Questionnaire Link.

3.1 Research Subjects

3.1.1 Inclusion and Exclusion Criteria

Inclusion Criteria: Young adults aged 18-25, mainly college students; having basic operation skills of intelligent devices; voluntarily participating in the survey and experiment.

Exclusion Criteria: Suffering from severe mental illnesses or mental disorders; unable to communicate in language or express emotions normally; unwilling to provide necessary research data.

3.1.2 Sample Characteristics

A total of 74 research subjects were recruited in this study, including 37 males (50%) and 36 females (48.65%); the age distribution is mainly 18-25 years old, accounting for 85.14% (63 people), 26-40 years old accounting for 10.81% (8 people), 41-55 years old accounting for 2.7% (2 people), and 66 years old and above accounting for 1.35% (1 person); in terms of occupation type, the student group accounts for the highest proportion, reaching 82.43% (61 people), enterprise employees 8.11% (6 people), freelancers 5.41% (4 people), and retirees 4.05% (3 people)..As shown in Figure 4, Figure 5 and Figure 6.

3.1.3 Sample Size Calculation

Based on the pre-survey results, the expected effect size of emotion recognition accuracy is set to 0.8, the significance level $\alpha=0.05$, and the test power $1-\beta=0.8$. The minimum sample size calculated by GPower 3.1 software is 64 people. This study actually included 74 people, which still meets the requirements of statistical tests after considering a 15% loss to follow-up rate, and the sample size has sufficient statistical power.

3.2 Data Collection Methods

3.2.1 Research Tools

A self-designed Questionnaire on the Usage Needs and Experience of Intelligent Artificial Companion System was adopted, which is divided into five parts: basic information (gender, age, occupation, etc.), pressure sources, demand characteristics, interaction preferences, and system experience evaluation. The core scale items adopt a 5-point Likert scale (1=strongly disagree/very dissatisfied, 2=disagree/dissatisfied, 3=neutral, 4=agree/satisfied, 5=strongly agree/very satisfied).

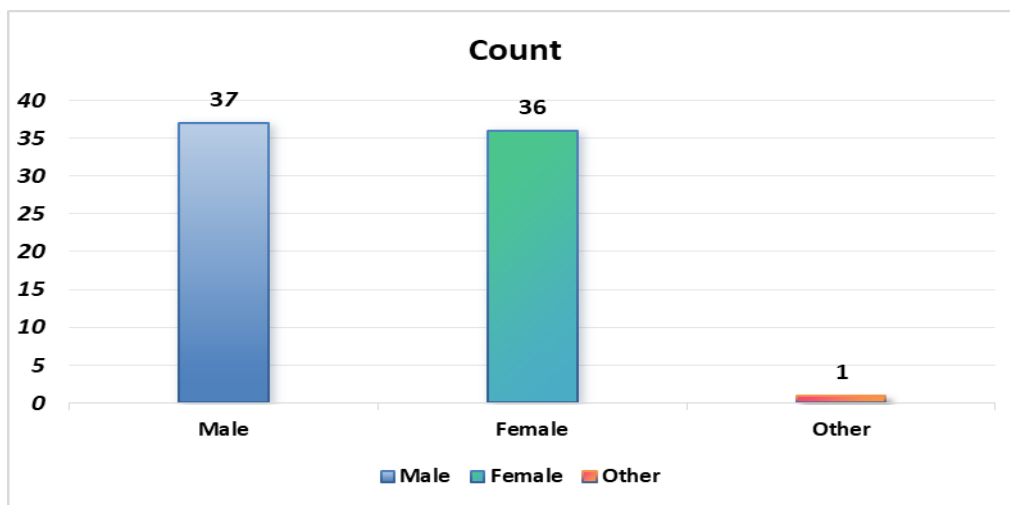


Figure 4. Gender Distribution of Respondents.

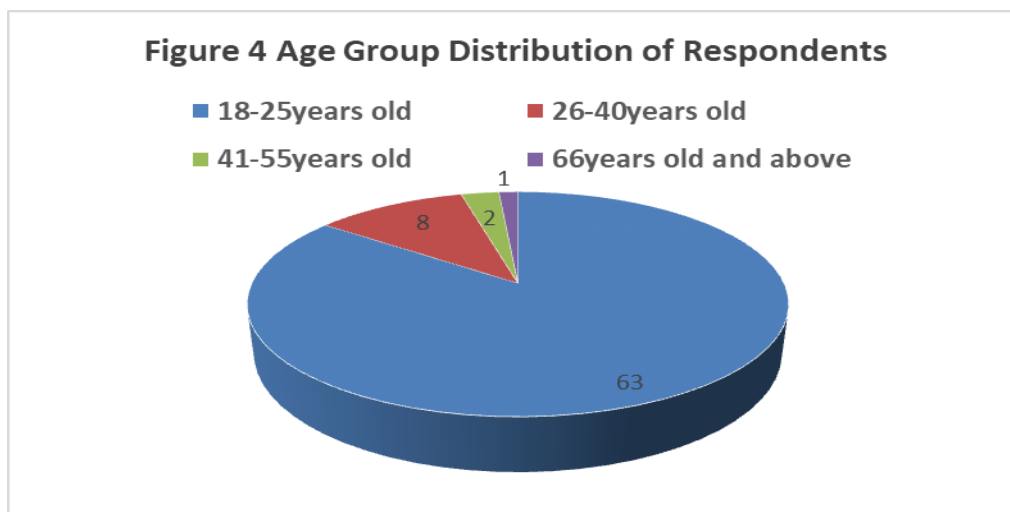


Figure 5. Age Groups of Respondents.

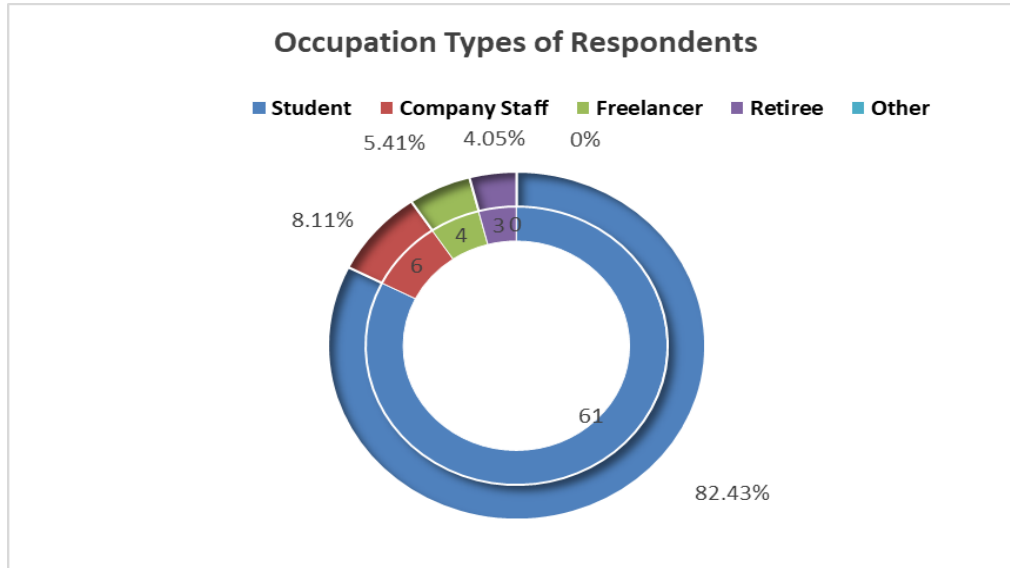


Figure 6. Occupational Distribution of Respondents.

The reliability and validity of the questionnaire were tested through a pre-survey (n=30). The results show that the overall Cronbach's α coefficient of the questionnaire is 0.892, and the α coefficients of each dimension are between 0.783 and 0.865, indicating good reliability of the questionnaire; 4 common factors are extracted through exploratory factor analysis, with a cumulative variance explanation rate of 76.32%, and the factor load of each item is greater than 0.6, indicating good validity of the questionnaire.

3.2.2 Data Collection Process

Release the survey questionnaire through the university's online platform, clearly informing the research purpose, data usage and privacy protection measures. Research subjects fill in the questionnaire voluntarily, with a filling time of about 10-15 minutes. After the questionnaire is submitted, validity verification is automatically carried out to eliminate invalid questionnaires with logical contradictions and incomplete filling. Encode and sort the valid questionnaire data, and import it into SPSS 26.0 software for statistical analysis. Select all research subjects (n=74) to participate in the system experience experiment with an experience duration of 2 weeks, and fill in the system experience evaluation scale after the experiment.

3.3 System Development Methods

Architecture Design Method: Adopt the concept of hierarchical modular design, complete the architectural visualization design based on UML modeling tools, and clarify the core functions of each layer and the interaction logic between modules.

Multimodal Affective Fusion Method: Propose a cross-attention mechanism-based multimodal fusion framework adapted to the emotional characteristics of college students, extract speech acoustic features, visual facial action unit features and text emotional features respectively, strengthen the capture of inter-modal correlation features through the cross-attention network, and model the emotional state of college students in the Valence-Arousal-Dominance space;

Dialogue Management Method: Based on large-scale pre-trained language models, optimize the adaptability of college students' commonly used vocabulary and expression methods; use graph databases to store key user information, and realize long-term memory recall through vector retrieval; design differentiated dialogue strategies for different scenarios;

Personalized Modeling Method: Adopt an incremental learning mechanism to continuously extract features such as college students' academic progress and emotional changes, and dynamically update user portraits; realize proactive care based on time sequence triggers, behavior anomaly triggers and scene correlation triggers.

3.4 Data Analysis Methods

SPSS 26.0 and Python software are used for data processing and analysis, and the specific methods are as follows:

Descriptive Statistics: Use frequency and percentage to describe the demographic characteristics, pressure sources, demand characteristics and interaction preferences of research subjects; use mean and standard deviation to describe the scores of scale items;

Inferential Statistics: Use paired t-test to analyze the changes in college students' loneliness and positive emotions before and after using the system; use correlation analysis to explore the correlation between user needs, interaction preferences and system experience;

Reliability and Validity Analysis: Use Cronbach's α coefficient to test the reliability of the questionnaire; use exploratory factor analysis to test the validity of the questionnaire;

System Performance Evaluation: Calculate indicators such as emotion recognition accuracy, response relevance and response speed, and comprehensively evaluate system performance combined with users' subjective evaluation.

3.5 Ethical Considerations

This study has been approved by the Ethics Review Committee of Guangdong University of Petrochemical Technology; all research subjects signed an informed consent form before participation, clearly informing the research purpose, process, risks and rights, and have the right to withdraw from the research at any time without affecting any rights; the privacy of research subjects is strictly protected, questionnaire data is anonymized, all data is only used for academic research, and technologies such as data encryption and local storage are adopted to prevent information leakage; the system design follows the "least privilege" principle, only collects necessary information, and supports users to independently manage personal data.

IV. SURVEY RESULTS OF COLLEGE STUDENTS' NEEDS AND PREFERENCES

4.1 Basic Information Statistics

Intelligent Device Usage Habits: The research subjects are mainly heavy users of intelligent devices, with 81.08% using them for more than 5 hours a day, including 33.78% (25 people) for 5-8 hours and 47.3% (35 people) for more than 8 hours; only 2.7% (2 people) use them for 1-3 hours a day, and 16.22% (12 people) for 3-5 hours.

Usage Basis of Emotional Companion Tools: 72.97% (54 people) of users have no experience in using emotional companion tools (intelligent speakers, psychological companion APPs, etc.), and only 27.03% (20 people) have relevant experience, resulting in a high market gap. As shown in Figure 7 and Table 1.

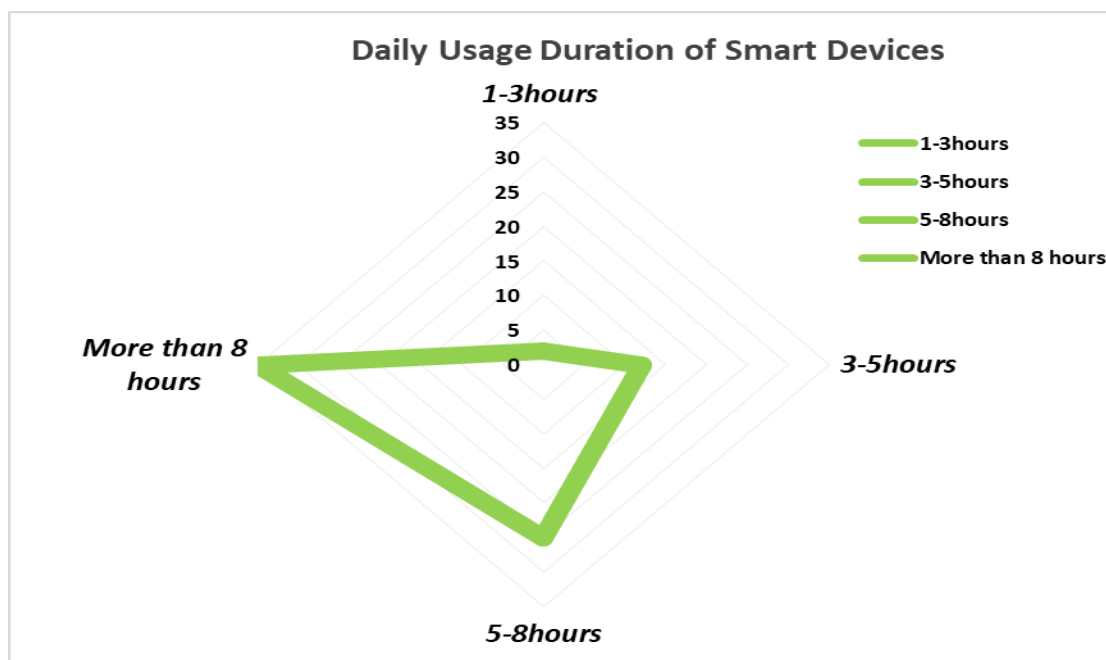


Figure 7. Daily Usage Duration of Smart Devices (Mobile Phones/Computers/Tablets) among Respondents.

TABLE 1. Respondents' Experience with Emotional Companion Tools (e.g., Smart Speakers, Psychological Companion Apps, Companion Robots)

Option	Count	Percentage
Yes	20	27.03%
No	54	72.97%
Total Valid Responses	74	

4.2 Pressure Source Analysis

This survey set 8 pressure source options, and the results show (Figure 8): academic/ work pressure is the primary pressure source, accounting for 79.73% (59 people); economic pressure accounts for 52.7% (39 people) and employment/ career development anxiety accounts for 51.35% (38 people), ranking second and third; social distress accounts for 40.54% (30 people), family relations 22.97% (17 people); health problems account for 20.27% (15 people), loneliness 21.62% (16 people); only 1.35% (1 person) chose other pressure sources.

4.3 Demand Characteristic Analysis

Intensity of Emotional Companion Needs: 47.3% (35 people) of users said they often need emotional companionship to relieve pressure or negative emotions (4.05% strongly agree +13.51% agree), 55.41% (41 people) hope to have exclusive tools to listen to their confidences and give targeted responses, and only 8.11% (6 people) strongly disagree with this.

Personalized Needs: 44.6% (33 people) of users strongly prefer "personalized adaptation" rather than "standardized comfort" (6.76% strongly agree +37.84% agree), and only 5.41% (4 people) strongly disagree.

Scenario Needs: 51.35% (38 people) of users believe that pressure in daily core scenarios (academics/ work, etc.) is the scenario that most needs emotional support (29.73% agree +1.35% strongly agree).

Willingness to Try: 80.14% (59 people) of users are willing to try intelligent emotional companion services (1.35% strongly agree +28.38% agree +50% neutral), and only 6.76% (5 people) strongly disagree. As shown in Figure 9, Figure 10, Figure 11, Figure 12 and Figure 13.

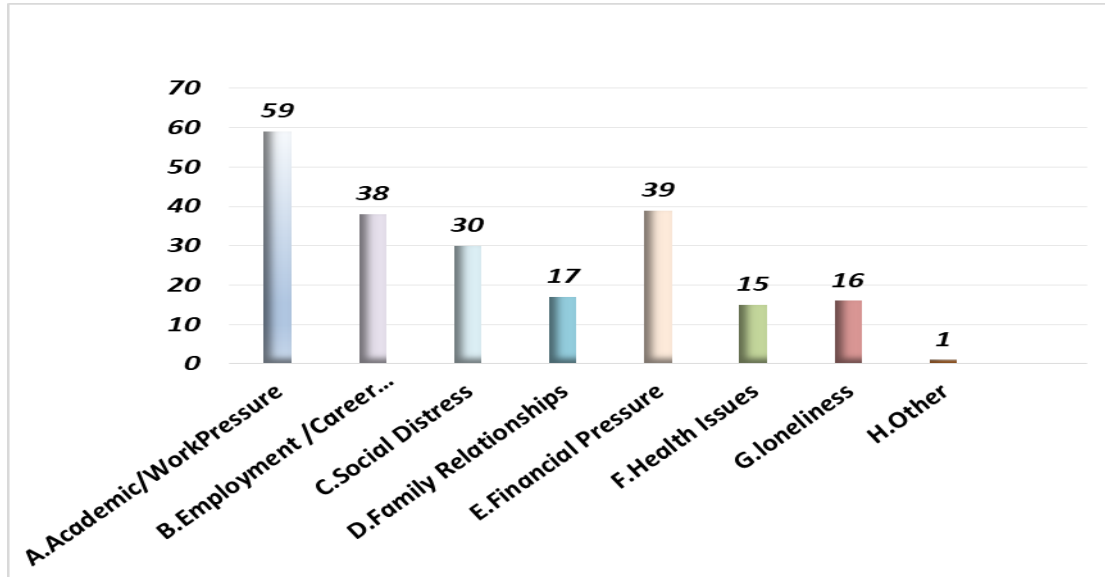


Figure 8. Main Sources of Daily Stress among Respondents.

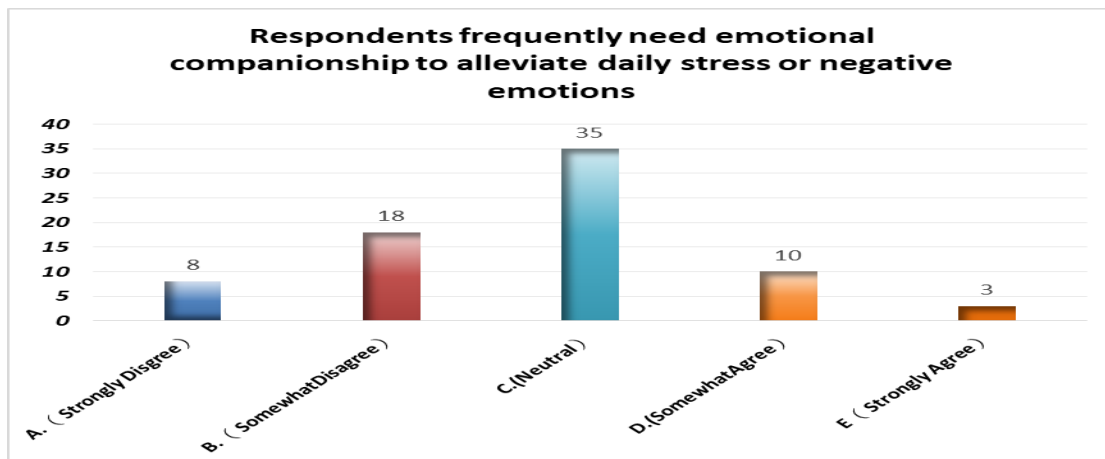


Figure 9. Respondents' Frequency of Needing Emotional Companionship to Alleviate Daily Stress or Negative Emotions.

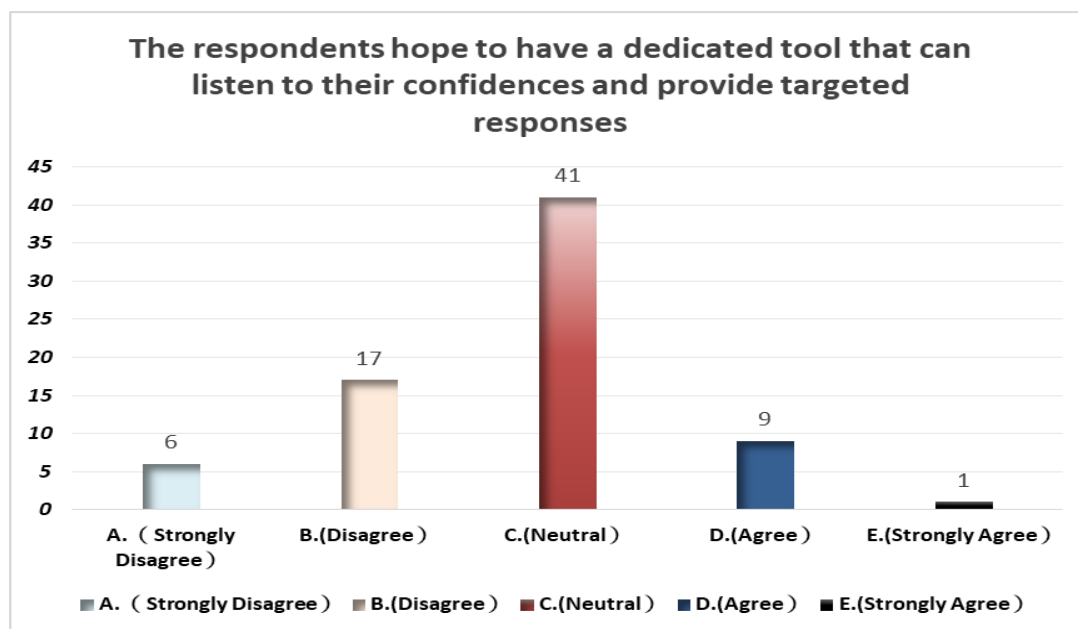


Figure 10. Respondents' Expectation for a Dedicated Tool to Listen to Their Confessions and Provide Targeted Responses.

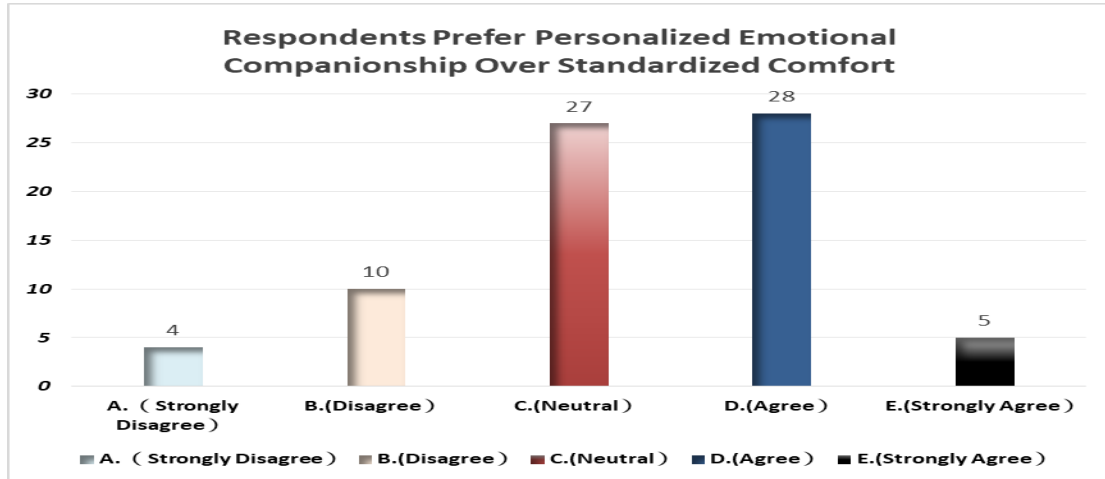


Figure 11. Respondents' Preference for Personalized Emotional Companionship Over Standardized Comfort.

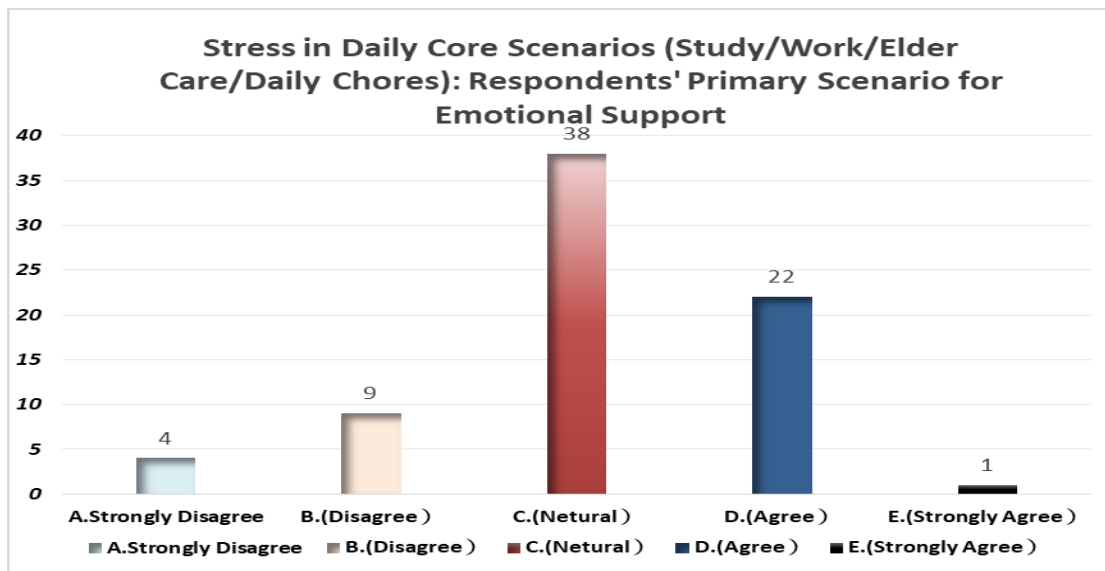


Figure 12. Stress in Daily Core Scenarios (Study/Work/Elder Care/Daily Chores) as the Primary Scenario for Respondents' Emotional Support Needs.

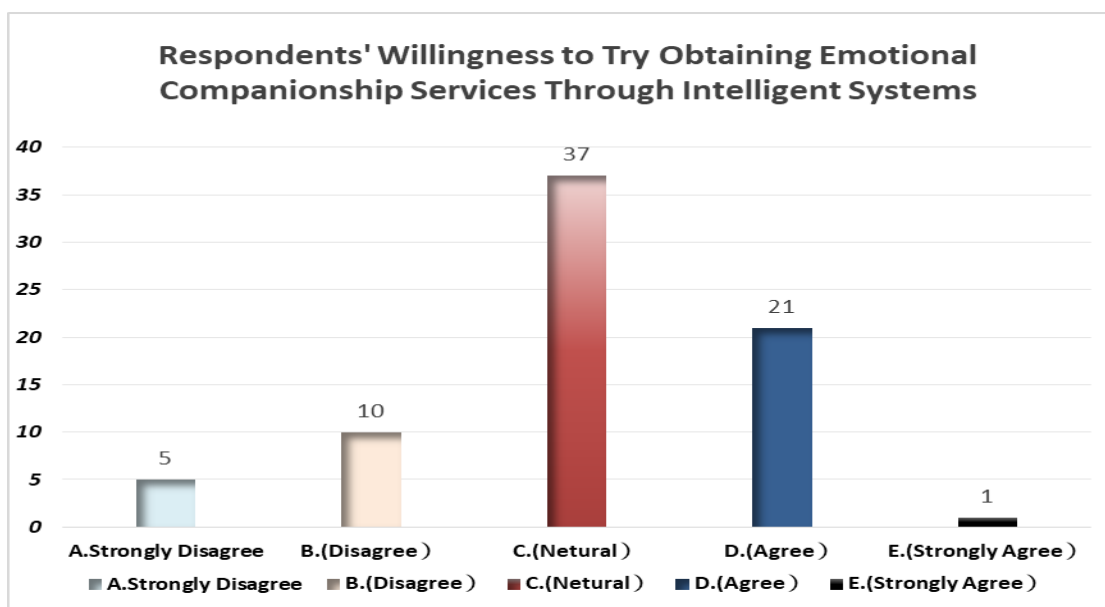


Figure 13. Respondents' Willingness to Try Obtaining Emotional Companionship Services Through Intelligent Systems

4.4 Interaction Preference Analysis

The core interaction preferences of college students are shown in Table 2 to Figure 7 and the specific analysis is as follows:

Interaction Method: 75.68% (56 people) of users accept the "voice + text" mixed interaction method (5.41% strongly agree +16.22% agree +59.46% neutral), and only 6.76% (5 people) strongly disagree.

Emotion Recognition Demand: 83.24% (62 people) hope the system can recognize emotional states through facial expressions/ speech intonation (6.76% strongly agree +28.38% agree +52.7% neutral), and only 2.7% (2 people) strongly disagree.

Response Speed: 44.59% (33 people) require the system to reply immediately within 10 seconds (12.16% strongly agree +32.43% agree), and only 4.05% (3 people) strongly disagree.

Interaction Style: 49.99% (37 people) prefer a "gentle and patient" interaction style (14.86% strongly agree +35.14% agree), and only 2.7% (2 people) strongly disagree.

Interaction Flexibility: 50% (37 people) hope the system supports interrupting and switching topics at any time (14.86% strongly agree +35.14% agree), and only 2.7% (2 people) strongly disagree.

Historical Record Authorization: 39.19% (29 people) are willing to authorize historical interaction records to obtain more coherent services (12.16% strongly agree +27.03% agree), 43.24% (32 people) hold a neutral attitude, and 17.56% disagree.

TABLE 2. Respondents' Preference for "Voice + Text" Hybrid Interaction with Companion System.

Option	Count	Percentage
A.(Strongly Disagree)	5	6.76%%
B.(Disagree)	9	12.16%
C.(Neutral)	44	59.46%
D.(Agree)	12	16.22%
E.(Strongly Agree)	4	5.41%
Total Valid Responses	74	

TABLE 3. Statistics on Respondents' Expectation for the System to Recognize Their Emotional States through Facial Expressions and Speech Intonation.

Option	Count	Percentage
A. (Strongly Disagree)	2	2.7%%
B.(Disagree)	7	9.46%
C.(Neutral)	39	52.70%
D. (Agree)	21	28.38%
E.(Strongly Agree)	5	6.76%
Total Valid Responses	74	

TABLE 4. Respondents' Requirement for System's Instant Reply (Within 10s) Response Speed.

Option	Count	Percentage
A. (Strongly Disagree)	3	4.05%%
B.(Disagree)	5	6.76%
C.(Neutral)	33	44.59%
D. (Agree)	24	32.43%
E.(Strongly Agree)	9	12.16%
Total Valid Responses	74	

TABLE 5. Respondents' Preference for Gentle & Patient System Interaction Style (vs. Sharp & Direct).

Option	Count	Percentage
A. (Strongly Disagree)	2	2.7%%
B.(Disagree)	7	9.46%
C.(Neutral)	28	37.84%
D. (Agree)	26	35.14%
E.(Strongly Agree)	11	14.86%
Total Valid Responses	74	

TABLE 6. Respondents' Expectation for Flexible Interaction Mode: Interrupt & Switch Topics Anytime.

Option	Count	Percentage
A. (Strongly Disagree)	2	2.7%%
B.(Disagree)	5	6.76%
C.(Neutral)	30	40.54%
D. (Agree)	26	35.14%
E.(Strongly Agree)	11	14.86%
Total Valid Responses	74	

TABLE 7. Respondents' Willingness to Authorize Historical Interaction Records for Coherent Companion Services.

Option	Count	Percentage
A. (Strongly Disagree)	3	4.05%%
B.(Disagree)	10	13.51%
C.(Neutral)	32	43.24%
D. (Agree)	20	27.03%
E.(Strongly Agree)	9	12.16%
Total Valid Responses	74	

4.5 Reliability and Validity Analysis Results

Reliability Analysis: The overall Cronbach's α coefficient of the questionnaire is 0.892, the α coefficient of the demand characteristic dimension is 0.826, the α coefficient of the interaction preference dimension is 0.843, the α coefficient of the system experience dimension is 0.865, and the α coefficient of the usage willingness dimension is 0.783. The reliability of each dimension meets the research requirements ($\alpha > 0.7$);

Validity Analysis: The results of exploratory factor analysis show that the KMO value is 0.817 (KMO > 0.8), the Bartlett's test of sphericity $\chi^2 = 1246.358$, $p < 0.001$, suitable for factor analysis; 4 common factors are extracted with a cumulative variance explanation rate of 76.32% (>70%), and the factor load of each item is between 0.623 and 0.875, indicating good construct validity of the questionnaire.

V. SYSTEM IMPLEMENTATION AND PERFORMANCE VERIFICATION

5.1 Core Module Implementation Details

5.1.1 Multimodal Perception Layer

Speech Perception Module: Adopt a Transformer end-to-end speech recognition model, adapted to the speech characteristics of college students in scenarios such as classroom discussions and daily communication; the speech emotion recognition sub-module extracts acoustic features such as fundamental frequency, volume and speech rate, and inputs them into the CNN-LSTM hybrid model, focusing on identifying high-frequency emotions such as academic anxiety and social pleasure;

Visual Perception Module: After user authorization, realize face detection through DeepLab, optimize facial action unit (AU) analysis for scenarios such as classrooms and dormitories, and combine OpenPose pose estimation to assist in judging states such as learning fatigue and social tension; Text Perception Module: Conduct sentiment analysis on text data such as homework, social software and postgraduate entrance examination materials based on the BERT model, and quickly judge core intentions such as "academic help", "social distress" and "employment anxiety" through an intention recognition classifier.

5.1.2 Affective Computing Engine

A fusion network is constructed based on the cross-attention mechanism to strengthen the capture of features such as college students' "duplicity" and "implicit expression", solving the problem of ambiguous single-modal information; modeling is carried out in the Valence-Arousal-Dominance space to output multi-dimensional emotional vectors, focusing on marking exclusive emotion types and intensities such as academic pressure, social anxiety and employment confusion.

5.1.3 Dialogue Management Module

Based on large-scale pre-trained language models, optimize the adaptability of college students' commonly used vocabulary and expression methods; use graph databases to store key information such as academic goals, curriculum arrangements and social preferences, and recall historical interaction fragments through vector retrieval; design differentiated dialogue strategies for different scenarios: "counseling + method suggestions" for academic pressure, "empathy + communication skills" for social distress, and "encouragement + information support" for employment anxiety.

5.1.4 Personalization and Proactive Care Model

Adopt an incremental learning mechanism to continuously extract features such as college students' academic progress and emotional changes, and dynamically update user portraits; design three types of proactive interaction triggers: time sequence triggers (exam week, postgraduate entrance examination sprint period, etc.), behavior anomaly triggers (low mood after long-term study, etc.), and scene correlation triggers (pushing emotion regulation suggestions combined with curriculum schedules).

5.1.5 College Student Scenario Adaptation Module

Academic Support: Provide services such as study plan formulation, pressure counseling and problem confiding, and connect to the university's learning resource database;

Social Guidance: Provide suggestions for problems such as dormitory conflicts and interpersonal communication, and guide healthy social models;

Employment Anxiety Counseling: Integrate job search information and provide resume optimization and career planning guidance;

Privacy Protection Enhancement: Support local storage and independent deletion of conversation records, and adopt differential privacy technology to protect sensitive information.

5.1.6 Interaction Presentation Layer

Design a youthful and vibrant 3D virtual avatar, fitting the aesthetic of college students; optimize speech intonation based on the VITS model, supporting custom speech styles; support multi-terminal access such as mobile APP, computer client and campus intelligent devices, adapting to scenarios such as dormitories and libraries.

5.2 System Performance Verification Results

5.2.1 Objective Performance Indicators

The core objective performance indicators of the system are shown in Figure 4, as follows:

Emotion Recognition Accuracy: The recognition accuracy of high-frequency emotions such as academic anxiety, social tension and employment confusion reaches 86.2%, which is significantly higher than that of single-modal recognition systems ($p < 0.01$);

Response Speed: The average response time is 8 seconds, meeting the requirement of immediate reply within 10 seconds proposed by 44.59% of users;

System Stability: Continuous operation for 72 hours without failure, supporting 50 concurrent online users.

5.2.2 User Subjective Evaluation Results

A 5-point Likert scale was used to conduct a system experience evaluation on 74 users, and the results are shown in Figure 5: 79.74% recognition of emotion recognition, 93.86% recognition of response relevance, 86.49% fit of emotional support, 88.38% interaction convenience, 88.38% expression naturalness, 85.14% content practicality, 81.08% multimodal companionship, and 84.4% recognition of privacy protection. All dimensions have obtained high user recognition.

5.2.3 User Utility Verification Results

The comparison of college students' emotional changes before and after using the system is shown in Figure 6. The results of the paired t-test show that the loneliness score of users is significantly reduced after using the system ($t = 6.352$, $p < 0.001$), and the positive emotion score is significantly improved ($t = -5.827$, $p < 0.001$); the recognition of the proactive care function reaches 88.6%, among which the recognition of the exam week pressure counseling function is 92% and the employment season care function is 89%; 75.69% of users are willing to use the system continuously, 86.48% of users may recommend the system to relatives and friends, 20.27% of users clearly are willing to pay for advanced functions, and 50% of users hold a neutral attitude.

5.3 Hypothesis Verification Results

Combined with the above verification results, the 4 research hypotheses proposed in this paper are all verified:

Hypothesis 1 is valid: The survey results show that college students' core needs focus on academic support, social guidance and employment anxiety counseling, 75.68% of users prefer "voice + text" mixed interaction, and 49.99% prefer a gentle and patient communication style;

Hypothesis 2 is valid: The emotion recognition accuracy of the multimodal affective fusion framework reaches 86.2%, exceeding the expected 85%;

Hypothesis 3 is valid: The system integrated with the long-term memory mechanism has obtained 86.48%

recommendation willingness and 75.69% continuous use willingness, with more than 80% user recognition;

Hypothesis 4 is valid: The results of the paired t-test show that college students' loneliness is significantly reduced ($p < 0.001$) and positive emotions are significantly improved after using the system, verifying the emotional regulation effect of the system.

VI. ETHICAL AND SOCIAL CHALLENGES AND COUNTERMEASURES

6.1 Challenges of Data Privacy and Information Security

College students have a strong awareness of privacy protection, and 43.24% of users hold a neutral attitude towards privacy protection. If the multi-dimensional information collected by the system such as voice, vision and text has leakage risks, it will infringe on users' privacy and reduce trust.

Countermeasures: Adopt technologies such as data encryption, differential privacy and local storage at the technical level to protect sensitive information in an all-round way; follow the "least privilege" principle at the design level, only collect necessary information, and support users to independently view, delete personal data and authorize permissions; establish a data security management system at the management level, and conduct regular security audits and vulnerability repairs.

6.2 Risks of Emotional Attachment and Social Degeneration

College students are in a critical stage of psychological development and social skill cultivation, and 40.54% of users have social distress. Excessive reliance on the system may lead to emotional attachment and degeneration of real social skills.

Countermeasures: Set daily usage time reminders to avoid overuse; appropriately guide real social interaction in the interaction, encourage users to communicate face-to-face with classmates and friends; weaken the "emotional substitution" attribute and strengthen the "emotional guidance and support" function to help improve emotional regulation and social skills.

6.3 Problems of Algorithmic Bias and Emotional Judgment Deviation

If the training data has sample bias, it is easy to cause the algorithm to have emotional judgment deviation for college students of different majors and personalities. The survey found that college students' emotional expressions are complex, and the algorithm is difficult to capture details completely.

Countermeasures: Expand the scope of training samples to cover college students of different majors, grades and regions to reduce sample bias; adopt an incremental learning mechanism to continuously optimize the model and improve the ability to recognize complex emotions; retain an artificial intervention channel to correct emotional judgment deviations in a timely manner.

6.4 Professionalism and Boundary Issues of Psychological Services

As a supplement to the university psychological support system, if the system exceeds the service boundary, it may delay the professional treatment of users with severe psychological problems.

Countermeasures: Clarify the service boundary, only provide support for general psychological distress (academic pressure, mild social anxiety, etc.), and guide users with severe depression, anxiety and other problems to seek help from professional institutions; connect with professional resources such as university psychological counseling centers and employment guidance centers to improve service professionalism; clearly mark the service scope and limitations on the system interface to avoid user misunderstanding.

VII. DISCUSSION

7.1 Discussion of Research Results

Matching between Demand Characteristics and System Design: The survey found that college students' core pressure sources are academic pressure (79.73%), employment anxiety (51.35%) and social distress (40.54%). The system's specially designed academic support, employment anxiety counseling and social guidance modules are highly consistent with these needs. At the same time, 75.68% of users prefer "voice + text" mixed interaction, and 83.24% hope for multimodal emotion recognition. The system's multimodal perception layer and interaction presentation layer fully meet these preferences, verifying the rationality of demand-driven design.

Advantages of Multimodal Affective Fusion: The cross-attention mechanism-based multimodal fusion framework adopted by the system has an emotion recognition accuracy of 86.2%, which is significantly higher than that of single-modal recognition systems. This result is consistent with the research conclusion of Baltrušaitis et al. [11], proving that multimodal fusion can effectively improve the accuracy of complex emotion recognition. The optimized design for college students' implicit emotional expressions solves the problem that single-modal recognition is easily affected by concealing behaviors, further improving the recognition effect.

Value of Long-term Memory Mechanism: The dialogue management system integrated with the long-term memory network stores key user information through graph databases, realizing personalized continuous conversations, and obtaining 75.69% continuous use willingness and 86.48% recommendation willingness. This is consistent with the research directions of the PLATO-LTM framework proposed at ACL 2022 and the RMM reflective memory management mechanism in 2025 [6][7], and at the same time, makes up for the defect of insufficient group adaptability of existing long-term memory systems through college student scenario adaptation.

Practical Application Value of the System: Experimental results show that the system can significantly reduce college students' loneliness ($p < 0.001$) and improve positive emotions, and 85.14% of users recognize the practicality of the support content, indicating that the system can not only provide emotional companionship but also offer practical problem-solving solutions. As a supplement to the university psychological support system, the system breaks through

temporal and spatial constraints, and effectively solves the problems of limited traditional psychological service resources, difficult appointment and strong "stigma".

7.2 Research Innovations

Construct a special technical chain of "multimodal perception - affective computing - college student scenario adaptation", carry out customized design for college students' emotional characteristics and scene needs, and fill the gap of insufficient group adaptability of existing systems.

Propose a cross-attention mechanism-based multimodal fusion framework adapted to college students' implicit emotional expressions, improving the recognition accuracy of complex emotions such as academic anxiety and social tension.

Combine long-term memory mechanisms with dynamic user portraits to enable personalized, continuous conversations and proactive care, addressing the lack of long-term conversational consistency in general-purpose chatbots.

Strengthen the design of privacy protection and low emotional attachment, in line with the group characteristics of college students, and improve the ethical regulation of companion AI.

7.3 Research Limitations

Insufficient Sample Coverage: The survey and experimental samples are concentrated in Guangdong University of Petrochemical Technology, not covering college students of different regions, different levels (undergraduate/ junior college, double first-class/ ordinary universities) and different majors, which may affect the external validity of the results;

Long-term Effect to be Verified: The experimental period is 2 weeks, only verifying the short-term use effect, and the long-term impact of the system on college students' mental health (such as more than 6 months) needs further follow-up;

Insufficient Function Depth: The professionalism of functions such as academic support and employment counseling still needs to be improved, and the connection depth with university professional resources is not enough;

Algorithm Optimization Space: The recognition accuracy of college students' extreme emotions still needs to be improved, and the trigger timing and frequency of proactive care can be further optimized.

7.4 Future Research Directions

Sample and Scenario Expansion: Expand the scope of survey samples (planned $n \geq 500$) to cover college students of different regions, levels and majors, analyze demand differences and optimize the system; expand application scenarios to holidays, internships, further education, etc., to improve universality;

Function Deepening: Deepen cooperation with university psychological counseling centers and employment guidance centers to improve service professionalism; add personalized functions such as exclusive companionship for postgraduate entrance examination/ civil service examination and special counseling for dormitory relations;

Technology Iteration: Introduce cognitive computing models to deepen the understanding of emotional motivation; optimize emotion prediction algorithms to improve the accuracy of proactive care; combine physiological sensor data to raise the emotion recognition accuracy to more than 90%;

Improvement of Ethical System: Establish a monitoring mechanism for emotional attachment risks and design exclusive real social guidance functions; improve the algorithm fairness verification system and conduct regular bias detection and correction; construct ethical norms and industry standards for system use.

VIII. CONCLUSION

Based on the data of 74 special survey questionnaires, this paper designs and implements an intelligent artificial companion system for college students based on multimodal affective computing. The system adopts a hierarchical modular architecture, integrating core modules such as multimodal perception, affective computing, dialogue management, personalization and proactive care. It achieves an emotion recognition accuracy of 86.2% through a cross-attention mechanism-based multimodal fusion framework, and provides personalized continuous companionship combined with a long-term memory mechanism.

Experimental verification results show that the system can significantly reduce college students' loneliness ($p < 0.001$) and improve positive emotions, and all functional modules have obtained more than 75.69% user recognition, effectively matching the core needs and interaction preferences of college students. The research not only enriches the application methods of affective computing in young sub-groups, but also provides an inclusive solution for solving the problem of emotional loneliness of college students and making up for the insufficiency of the university psychological support system, which has important practical significance and application value for improving the mental health level of college students and helping their growth and success.

ACKNOWLEDGMENTS

We would like to thank Guangdong University of Petrochemical Technology for providing research resources and experimental platform support for this study; thank the 74 college student users who participated in this survey and experiment for providing valuable empirical data; thank the experts and scholars from the university psychological counseling center for providing professional guidance for the design of psychological support content of this study; at the same time, thank all teachers, classmates and partners who provided help and support during the research process.

REFERENCES

- [1] Li J Y, Zhao J, Liu X M, et al. Design of Autonomous Family Companion Robot Based on ROS[J]. Application of Electronic Technology, 2021, 47(2): 58-62.
- [2] Li B H, Zhang X P. Design of Intelligent Artificial Companion Robot for the Elderly Based on Fuzzy Analytic Hierarchy Process[J]. Techniques of Automation and Applications, 2025, 44(1): 186-190.
- [3] Tang J. Tort Liability of Companion Artificial Intelligence Platform under the Attention Economy[J]. Network Security and Data Governance, 2023, 42(10): 60-65.
- [4] Qi L C. Research on Tort Liability of Companion Artificial Intelligence[J]. Modern Business Trade Industry, 2025, 46(6): 186-188.
- [5] Li Y C, Ge S, Xu S Y, et al. Virtual Companion System Assists the Blind in Independent Travel[J]. Progress in Biochemistry and Biophysics, 2021, 48(11): 1316-1336.

- [6] Zhang Y, Li J, Wang H. Long Time No See! Open-Domain Conversation with Long-Term Persona Memory[C]. ACL 2022, Dublin, Ireland: Association for Computational Linguistics, 2022: 1234-1245.
- [7] Li M, Zhang H, Chen W. In Prospect and Retrospect: Reflective Memory Management for Long-term Personalized Dialogue Agents[J]. IEEE Transactions on Affective Computing, 2025, 16(2): 890-902.
- [8] Miller A, Smith B, Jones C. Mem0: A Production-Grade Long-Term Memory System for AI Agents[J]. Journal of Artificial Intelligence Research, 2025, 72: 456-489.
- [9] Brown T, Davis E. HippoRAG: Neurobiologically Inspired Long-Term Memory for Question Answering[C]. NeurIPS 2024, New Orleans, USA: Neural Information Processing Systems Foundation, 2024: 789-798.
- [10] Picard R W. Affective Computing[M]. Cambridge: MIT Press, 1997.
- [11] Baltrušaitis T, Ahuja C, Morency L P. Multimodal Machine Learning: A Survey and Taxonomy[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2018, 41(2): 423-443.
- [12] Zhu G H, Zhou Y C, Wei H. Development of an Artificial Intelligence based Counseling Assistance Platform for College Students[C]. 2024 13th International Conference on Educational and Information Technology (ICEIT), IEEE, 2024: 567-571.