Development and Evaluation of a Machine Learning-Based Conversational Agent for Academic Anxiety Screening in Software Engineering Students

Ronald Melgarejo-Solis[1]; Denis Aguilar-Ortega[2]; Iris Hanampa Bellido[3]; Manuel Guzman Chavez[4]; Jeanpiere Burga Montesinos[5]; Albert Huaman Ñahuinlla[6]; Jeanmarco Rosales Trinidad[7]; Ramesh Chander Sharma[8]
[1,3,4,5,6,7]Universidad Nacional Mayor de San Marcos, Perú, *ronald.melgarejo@unmsm.edu.pe, iris.hanampa@unmsm.edu.pe, manuel.guzmanc@unmsm.edu.pe, jeanpiere.burga@unmsm.edu.pe, jesus.raymondes@unmsm.edu.pe, albert.huaman@unmsm.edu.pe, jeanmarco.rosales@unmsm.edu.pe*
[2]Universidad Tecnológica de Honduras, Honduras, *denis.aguilar@uth.hn*
[8]Dr B R Ambedkar University Delhi, India, *rcsharma@aud.ac.in*

***Abstract*—*This paper presents the development and evaluation of a machine learning-based conversational agent for screening academic anxiety among Software Engineering students at the Universidad Nacional Mayor de San Marcos. The system integrates a text-based anxiety-signal classifier, the GAD-7 questionnaire, and a response module designed to provide coping guidance and referral recommendations. Its usability, acceptance, and perceived role as a first point of contact were evaluated through a survey of undergraduate students. The results show high model sensitivity and a positive assessment of the tool, with perceived usefulness emerging as the main factor associated with behavioral intention to use it. The findings suggest that conversational agents are promising as an initial point of contact for anxiety-related help-seeking because they are readily accessible and perceived as less stigmatizing, while serving as a complement rather than a substitute for professional care.*

Keywords*—*academic anxiety, conversational agent, machine learning, screening, usability

I. INTRODUCTION

Concern about university students' mental health continues to increase across higher education. Anxiety is among the most frequent problems they face and is characterized by anticipatory apprehension about potential threats. Unlike stress, which is associated with demanding pressures, and depression, which is linked to persistently low mood, anxiety is a distinct construct that requires specific assessment tools [1], [2]. This distinction is important because it determines how tools for recognizing and addressing anxiety should be designed.

A meta-analysis estimated a pooled anxiety prevalence of 39.65% among undergraduate students, with approximately 20.31% experiencing severe levels [3]. In Peru, a study of university students in Lima reported severe anxiety in 22.6% of participants [4]. Therefore, anxiety is not an isolated phenomenon; it is a widespread condition that affects the well-being and academic performance of university communities [1].

Despite this prevalence, access to psychological support remains limited. Studies identify barriers such as help-seeking stigma, treatment costs, and limited availability of specialists, which delay or prevent early care [5]. These constraints are particularly severe in universities, where student-welfare services are often overwhelmed by growing demand, leaving many students without an initial point of contact that can offer guidance.

In this context, artificial intelligence-based conversational agents, also known as chatbots, have emerged as scalable, accessible, and potentially less stigmatizing options for mental-health support [5]. Meta-analytic evidence also indicates that they can reduce symptoms of anxiety and depression among adolescents and young adults [6], [7]. However, useful support requires accurate assessment. Among validated instruments, the GAD-7 questionnaire is particularly relevant because of its brevity and reliability. In the Peruvian population, it has shown a unidimensional structure and adequate psychometric properties, with a Cronbach's alpha of 0.87 and a McDonald's omega of 0.88 [8], [4]. These findings support its use as a screening tool in the present context and provide a standardized basis for estimating anxiety levels as part of an early-detection approach.

Despite the high prevalence of anxiety, important gaps remain. Nyakhar and Wang [5] reported that the effectiveness of mental-health chatbots among university students has received limited study. Likewise, Hua et al. [9], in a review of 160 studies published between 2020 and 2024, found that only 16% of systems based on large language models had undergone clinical-efficacy testing. Both reviews conclude that the field's rapid growth has not been accompanied by sufficiently rigorous evaluation.

Two opportunities can be identified for this specific population. First, few studies have examined how students accept these tools and the extent to which they view them as an accessible, low-stigma first point of contact [10]. Second, machine learning solutions can detect early indicators of anxiety in user-generated text [11], an approach that remains underexplored in tools designed for university students.

Addressing these opportunities, this study proposes a machine learning-based conversational agent for screening academic anxiety among undergraduate students. The system links three components: a classifier that detects anxiety signals in text, the GAD-7 as a structured screening instrument, and a response module that provides coping strategies and referral to professional care when appropriate. The study evaluates the

tool's usability and acceptance, identifies factors associated with that acceptance, and examines whether it can function as an early, accessible, and low-stigma pathway for help-seeking.

II. OBJECTIVES

A. *General Objective*

To develop and evaluate a machine learning-based conversational agent for academic anxiety screening among Software Engineering students at the Universidad Nacional Mayor de San Marcos, integrating an anxiety-signal classifier, the GAD-7 questionnaire, and a response module, and to assess its usability, acceptance, and perceived role as a first point of contact.

B. *Specific Objectives*

- Train a machine learning model to detect anxiety signals in student text, with emphasis on sensitivity.
- Implement a conversational-agent prototype integrating the model, the GAD-7 questionnaire, and coping and referral modules.
- Evaluate the prototype's usability, acceptance, and perceived role as a first point of contact among undergraduate students.

III. METHODOLOGY

A. *Research Type and Design*

This applied study used a quantitative approach, a descriptive-correlational scope, and a nonexperimental cross-sectional design. It combined two complementary components: development of a software artifact—the conversational agent—and evaluation of its usability, acceptance, and perceived usefulness among students. System development was managed using the Scrum framework, with tasks organized into successive sprints based on a product backlog. This approach enabled incremental iteration of both the classification model and the conversational prototype.

B. *Development of the Conversational Agent*

The agent was built around a machine learning model designed to recognize indications of anxiety in student text. Training used the public Sentiment Analysis for Mental Health dataset [12], available on Kaggle, which contains approximately 53,000 English-language statements labeled by mental-health state. Only the anxiety and normal classes were retained, framing the task as binary classification. This delimitation is consistent with the system's intended use as a screening—not differential-diagnosis—tool and reduces the risk of conflating anxiety with other conditions that may have similar textual expressions.

The training stage followed the pipeline summarized in Fig. 1. After filtering and labeling the statements (anxiety = 1, normal = 0), the data were stratified into training (80%) and test (20%) subsets. Each message was transformed into a vector representation using a pretrained multilingual embedding model [13] with Spanish-language support, and the classifiers operated on these representations.

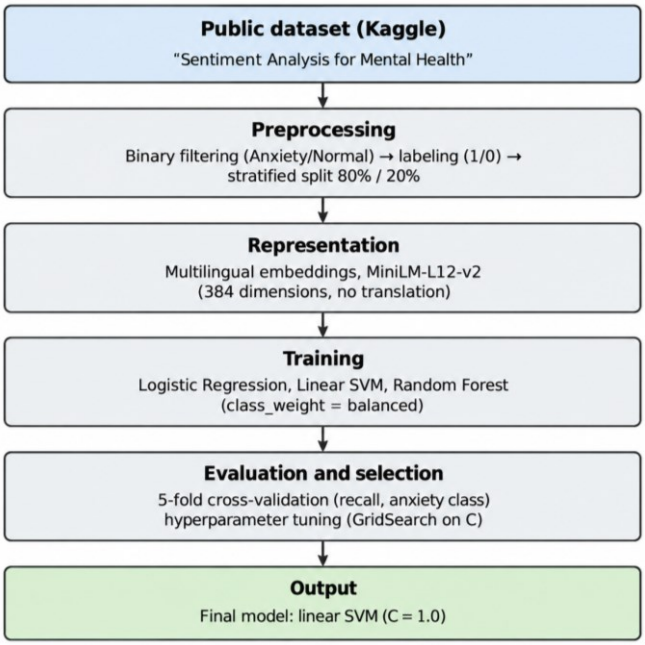


Fig. 1. Training pipeline for the classification model.

Three models were trained on this representation: logistic regression, a linear support vector machine, and Random Forest. Each used balanced class weights to address the natural class imbalance. Model selection and hyperparameter tuning were guided by recall for the anxiety class because it was the less frequent class and the primary class of interest. The selected linear support vector machine evaluates whether a message contains anxiety signals and consequently determines when structured screening should be offered. Fig. 2 shows the Python implementation of the training process and comparison of the three classifiers.

```
Step 6: Training the three models

models = {
    'Logistic Regression': LogisticRegression(class_weight='balanced',
                                              max_iter=1000, random_state=42),
    'SVM (LinearSVC)':     LinearSVC(class_weight='balanced', random_state=42),
    'Random Forest':       RandomForestClassifier(n_estimators=300,
                                                  class_weight='balanced',
                                                  random_state=42, n_jobs=-1),
}

for name, model in models.items():
    model.fit(X_train_emb, y_train)
    print(name, '-> trained')

Step 7: Evaluation and comparison

results = []
for name, model in models.items():
    y_pred = model.predict(X_test_emb)
    recall_anx = recall_score(y_test, y_pred, pos_label=1)
    results.append((name, recall_anx))
    print("\n---", name, "---")
    print(classification_report(y_test, y_pred, target_names=['Normal', 'Anxiety']))

# Summary table
table = pd.DataFrame(results, columns=['Model', 'Recall_Anxiety']) \
          .sort_values('Recall_Anxiety', ascending=False)
print("\nComparison by Anxiety recall:\n", table)

# Confusion matrix of the best model
best_name = table.iloc[0]['Model']
best_model = models[best_name]
cm = confusion_matrix(y_test, best_model.predict(X_test_emb))
ConfusionMatrixDisplay(cm, display_labels=['Normal', 'Anxiety']).plot(cmap='Greys')
plt.title(f'Confusion matrix - {best_name}')
plt.show()
```

Fig. 2. Training and comparison of classifiers in Python (scikit-learn).

The architecture comprises four components: a conversational interface, the text classifier, the GAD-7 screening questionnaire, and a response module that provides anxiety-reduction options or referral to the appropriate support service. The classifier serves only as an activation mechanism; that is, it suggests when the GAD-7 should be administered, whereas

the anxiety level is always determined by the validated questionnaire. This design separates the initial textual signal from the screening result. The prototype was implemented as a web application deployed through the Streamlit service for evaluation. Fig. 3 summarizes the architecture, whereas Fig. 4 shows an actual interaction with the prototype in which a student describes their condition in their own words and the agent offers the GAD-7 response options.

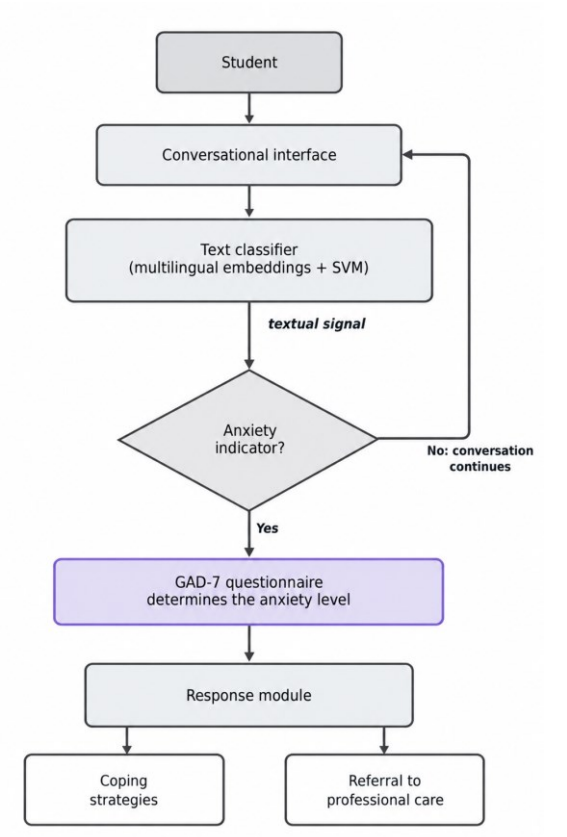


Fig. 3. Conversational-agent architecture.

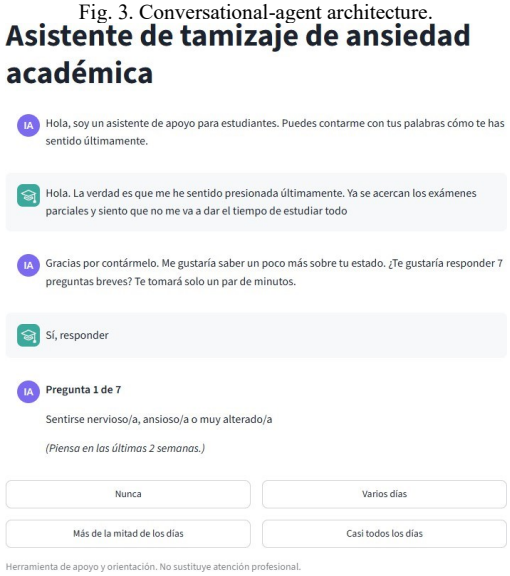


Fig. 4. Interaction with the prototype deployed on Streamlit (interface delivered in Spanish for the target student population).

C. *Population and Sample*

The population consisted of undergraduate Software Engineering students at the Universidad Nacional Mayor de San Marcos. The evaluation used a purposive, nonprobability sample of 40 students enrolled in the program. Students from different academic terms were included to capture perceptions at different stages of their education. Given the sensitive topic, participation was voluntary and anonymous and was preceded by informed consent explaining the study purpose, data safeguards, and the right to withdraw at any time.

D. *Techniques and Instruments*

Evaluation was conducted using a scenario-guided use protocol and a structured questionnaire. To ensure comparable experiences, each participant received a prompt placing the interaction in an academically demanding context, such as the period immediately before final examinations. Participants were asked to converse with the agent as they would in a real situation and complete the interaction flow.

After the interaction, participants completed an online questionnaire organized into three dimensions. Usability was measured using the System Usability Scale (SUS) [14]. Acceptance was evaluated with items based on the Technology Acceptance Model (TAM) [15], which relates perceived usefulness, perceived ease of use, and behavioral intention. Perception of the agent as a first point of contact was assessed through researcher-developed items addressing accessibility and stigma as barriers to help-seeking. Responses used a five-point Likert scale. Table I summarizes the instrument structure.

TABLE I
DIMENSIONS OF THE EVALUATION INSTRUMENT

Dimension	Base instrument	Items	Scale
Usability	SUS	10	Likert 1–5
Acceptance	TAM (usefulness, ease of use, intention)	6	Likert 1–5
Perception as first contact	Researcher-developed	4	Likert 1–5

E. *Variables*

Given the descriptive-correlational scope and nonexperimental design, no variables were manipulated. Instead, study variables were defined at two levels. At the model level, they corresponded to classification-performance metrics, with particular attention to recall for the anxiety class because it was the less frequent class and the primary category of interest. At the prototype-evaluation level, the variables were usability, acceptance, and perception of the agent as a first point of contact. Acceptance was modeled as a construct comprising perceived usefulness, perceived ease of use, and behavioral intention, whose relationships were analyzed.

F. *Statistical Analysis*

The analysis was organized into two components. Model performance was evaluated on the test set using a confusion matrix and classification report, emphasizing recall for the anxiety class. Stratified cross-validation was used to examine

metric stability, with the mean and standard deviation reported for each indicator.

Prototype-evaluation data were analyzed using descriptive and inferential statistics. Frequencies, means, and standard deviations were calculated by item and dimension. The usability score was calculated using the standard SUS procedure (range: 0–100) and interpreted against reference percentiles and the Sauro and Lewis grading scale [16], rather than reported as an isolated value.

Inferential analysis first evaluated instrument reliability using Cronbach's alpha and McDonald's omega for the usability and acceptance scales. Relationships among the acceptance constructs—perceived usefulness, perceived ease of use, and behavioral intention—were examined using Spearman's correlation coefficient, which is appropriate for ordinal data. The Kruskal–Wallis test was also used to compare ratings among participant subgroups based on stage of study.

Model processing was performed in Python using the scikit-learn library [17], and survey data were analyzed using jamovi.

IV. RESULTS

A. *Classification Model Performance*

Comparison of the three trained classifiers showed that the linear support vector machine performed best on the prioritized metric. For recall on the anxiety class, the SVM achieved 0.944, followed by logistic regression at 0.935 and Random Forest at 0.776. Although Random Forest produced the highest precision for the anxiety class (0.96), its lower recall resulted in more undetected anxiety cases, which is undesirable for a screening instrument.

The selected model was evaluated on the test set (n = 4,037). It achieved an overall accuracy of 0.96 and, for the anxiety class, precision of 0.86, recall of 0.94, and an F1-score of 0.90 (Table II). The confusion matrix in Fig. 5 shows that, of 768 actual anxiety cases, the model correctly identified 725 and missed 43. To examine performance stability, five-fold stratified cross-validation was applied to the training set, producing a mean recall of 0.954 for the anxiety class with a standard deviation of 0.007, indicating consistent performance across folds.

TABLE II
SVM MODEL PERFORMANCE ON THE TEST SET

Class	Precision	Recall	F1	n
Anxiety	0.86	0.94	0.90	768
Normal	0.99	0.96	0.98	3269
Overall accuracy	—	0.96	—	4037

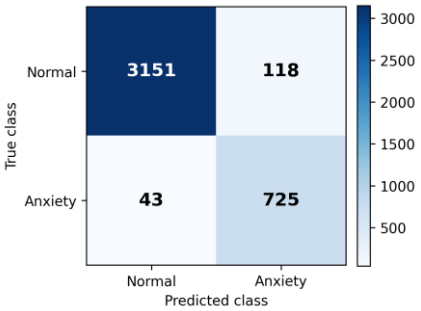


Fig. 5. Confusion matrix for the SVM model on the test set.

B. *Sample Characteristics*

The sample comprised 40 undergraduate Software Engineering students. Most participants were aged 19–21 years (23 participants), followed by those aged 22–24 years (12 participants). The 16–18 and 25-or-older groups included 2 and 3 participants, respectively. Regarding progress through the program, participants were distributed across academic terms, with a concentration in the middle stages: 10 were in early terms (1st–4th), 26 in middle terms (5th–7th), and 4 in advanced terms (8th–10th). This convenience-sample distribution was considered when interpreting subgroup comparisons.

C. *Instrument Reliability*

Before analyzing the evaluation results, internal consistency was examined for the three scales using Cronbach's alpha and McDonald's omega (Table III). The usability scale (SUS) showed acceptable reliability ($\alpha = 0.721$; $\omega = 0.752$). The acceptance and perception scales achieved higher values: the TAM-based acceptance dimension showed excellent internal consistency ($\alpha = 0.904$; $\omega = 0.905$), whereas the researcher-developed first-point-of-contact perception scale showed good reliability ($\alpha = 0.883$; $\omega = 0.889$). All three instruments exceeded the recommended 0.70 threshold, supporting their internal consistency for subsequent analyses.

TABLE III
SCALE RELIABILITY

Dimension	Items	Cronbach's α	McDonald's ω
Usability (SUS)	10	0.721	0.752
Acceptance (TAM)	6	0.904	0.905
Perception	4	0.883	0.889

D. *Perceived Usability (SUS)*

The agent obtained a mean SUS score of 72.4 (SD = 14.5) on the 0–100 scale. According to the Sauro and Lewis curved grading scale [16], this score corresponds to a C+ grade and approximately the 60th percentile, placing it above the industry reference average of 68. The distribution across grade bands was predominantly positive: 22 of the 40 participants placed the agent in bands A or B, 14 in band C, and only 4 in bands D or F. Individual scores ranged from 37.5 to 97.5.

E. *Acceptance and Perception*

Acceptance and perception ratings were positive across all dimensions on a scale from 1 to 5 (Table IV). Within the

acceptance model, perceived ease of use had the highest mean (M = 4.06; SD = 0.86), followed by perceived usefulness (M = 3.83; SD = 0.93) and behavioral intention (M = 3.69; SD = 0.95). Overall acceptance had a mean of 3.86 (SD = 0.82). Perception of the agent as a first point of contact had a mean of 3.84 (SD = 0.88). At the item level, the highest-rated statements in this dimension described the tool as less intimidating than scheduling a professional appointment (M = 4.03) and as potentially encouraging people who do not usually seek help (M = 4.05). The item concerning use of the tool as a first step before seeking psychological services had the lowest mean in the block (M = 3.52).

TABLE IV
DESCRIPTIVE STATISTICS BY DIMENSION

Dimension	Mean	SD
Perceived ease of use	4.06	0.86
Perceived usefulness	3.83	0.93
Behavioral intention	3.69	0.95
Acceptance (overall)	3.86	0.82
Perception as first contact	3.84	0.88

F. *Relationships Among Constructs*

Spearman correlation coefficients were used to analyze relationships among the acceptance constructs (Fig. 6). All correlations were positive and statistically significant. The strongest relationship was between perceived usefulness and behavioral intention ($\rho = 0.80$; $p < .001$), followed by the relationship between perceived usefulness and perception of the agent as a first point of contact ($\rho = 0.74$; $p < .001$). Perceived ease of use was also significantly associated with behavioral intention ($\rho = 0.68$; $p < .001$) and perceived usefulness ($\rho = 0.66$; $p < .001$), although its relationship with perception was the weakest observed ($\rho = 0.42$; $p < .01$). Overall, perceived usefulness had the strongest associations with both behavioral intention and perception of the tool.

Behavioral intention was additionally compared among students at different stages of the program (early, middle, and advanced) using the Kruskal–Wallis test. No statistically significant differences were found among groups ($\chi^2 = 0.469$; $df = 2$; $p = .791$; $\varepsilon^2 = 0.012$), indicating that intention to use the agent was similar regardless of academic term.

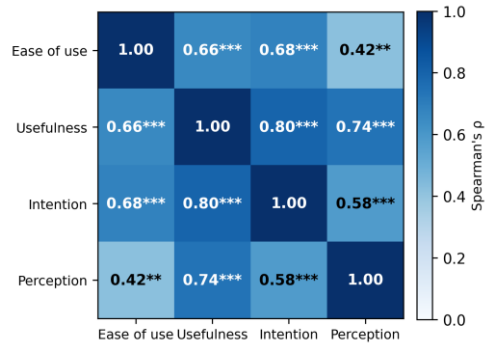


Fig. 6. Spearman correlation matrix for the acceptance constructs (** $p < .01$; *** $p < .001$).

V. DISCUSSION AND CONCLUSIONS

The classification model achieved high recall for the anxiety class on both the test set (0.94) and cross-validation (0.954 ± 0.007), with minimal variation among folds (Table II). This behavior is desirable for a component designed to activate screening because it prioritizes avoiding missed potential anxiety cases, at the cost of more false positives, as shown in the confusion matrix (Fig. 5). The comparison among classifiers further illustrates this trade-off: although Random Forest achieved the highest precision, it had the lowest recall and would therefore have failed to detect a substantial proportion of cases. Selecting the model on the basis of recall rather than overall accuracy is therefore consistent with the system's screening objective [11].

In the user evaluation, the instruments showed adequate internal consistency (Table III), and usability was above the industry reference average, although improvement remains possible, as expected for a prototype. The most relevant finding from the analysis of relationships among constructs was that perceived usefulness had the strongest association with behavioral intention ($\rho = 0.80$; Fig. 6), exceeding the association with ease of use. This finding suggests that students' willingness to use the tool depends primarily on whether they perceive it as useful for their circumstances rather than merely easy to operate. The stronger influence of usefulness than ease of use is consistent with the Technology Acceptance Model [15] and recent studies of behavioral intention and adoption of digital mental-health solutions among university students [18], [19].

Regarding its role as a first point of contact, this dimension received a positive mean rating (Table IV). Participants particularly valued the tool as less intimidating than scheduling professional care and as capable of encouraging people who do not ordinarily seek help, supporting its use as an accessible, low-stigma resource [20]. However, the item concerning use as a step before psychological services received the lowest rating in the block, suggesting that students regard the agent as an orientation complement rather than a replacement for professional care. Finally, the lack of significant differences in behavioral intention among academic stages indicates that acceptance was consistent across the program.

This study has several limitations. First, the sample was small and nonprobabilistic, with unequal groups; consequently, the findings are exploratory and cannot be generalized to the entire population. In addition, individual GAD-7 scores and classifier signals were not recorded, preventing direct assessment of agreement between the two components; this is proposed as future work. Finally, the model was trained using English-language data. An exploratory validation using Spanish phrases showed adequate performance for direct expressions but lower sensitivity to slang and indirect or culturally specific expressions of anxiety, which is a limitation of the cross-lingual approach.

In conclusion, this study developed and evaluated a conversational agent for academic anxiety screening that integrates an anxiety-signal classifier, the GAD-7 questionnaire, and a response module. The proposed model achieved high recall (Table II; Fig. 5), consistent with its screening function, and the tool was perceived as usable, acceptable, and valuable as an accessible, low-stigma first point of contact (Table IV). Perceived usefulness was the principal factor associated with behavioral intention (Fig. 6). The primary contribution is not classifier novelty but empirical evidence concerning how an understudied population—engineering students at a Peruvian public university—responds to this type of tool. Future work should record GAD-7 data to validate the two-stage architecture, train the model with Spanish-language data, expand the sample, and evaluate clinical effectiveness.

REFERENCIAS

[1] A. A. Suárez-Menéndez and J. P. Bitar-Cabezas, "Ansiedad generalizada y estrés académico en estudiantes universitarios," Estudios y Perspectivas Rev. Cient. Acad., vol. 5, no. 1, 2025.

[2] T. Canches, "Ansiedad y estrés académico en estudiantes de una universidad estatal de Lima Metropolitana en el contexto del COVID-19," undergraduate thesis, Faculty of Psychology, Universidad Nacional Mayor de San Marcos, Lima, Peru, 2023.

[3] I. Ahmed, C. M. Hazell, B. Edwards, C. Glazebrook, and E. B. Davies, "A systematic review and meta-analysis of studies exploring prevalence of non-specific anxiety in undergraduate university students," BMC Psychiatry, vol. 23, no. 1, Art. no. 240, 2023.

[4] J. Anicama, F. Ignacio, K. Talla, and B. Bruno, "Propiedades psicométricas de la escala del trastorno de ansiedad generalizada: GAD-7 en universitarios de Lima, Perú," Acta Psicológica Peruana, vol. 6, no. 2, pp. 260–273, 2023.

[5] S. Nyakhar and H. Wang, "Effectiveness of artificial intelligence chatbots on mental health & well-being in college students: a rapid systematic review," Frontiers in Psychiatry, vol. 16, Art. no. 1621768, 2025.

[6] W. Zhong, J. Luo, and H. Zhang, "The therapeutic effectiveness of artificial intelligence-based chatbots in alleviation of depressive and anxiety symptoms in short-course treatments: a systematic review and meta-analysis," Journal of Affective Disorders, vol. 356, pp. 459–469, 2024.

[7] X. Feng, L. Tian, G. W. K. Ho, J. Yorke, and V. Hui, "The effectiveness of AI chatbots in alleviating mental distress and promoting health behaviors among adolescents and young adults: systematic review and meta-analysis," J. Med. Internet Res., 2025.

[8] D. Villarreal-Zegarra, R. Paredes-Angeles, N. Mayo-Puchoc, E. Arenas-Minaya, J. Huarcaya-Victoria, and A. Copez-Lonzoy, "Psychometric properties of the GAD-7 (General Anxiety Disorder-7): a cross-sectional study of the Peruvian general population," BMC Psychology, vol. 12, Art. no. 183, 2024.

[9] Y. Hua et al., "Charting the evolution of artificial intelligence mental health chatbots from rule-based systems to large language models: a systematic review," World Psychiatry, vol. 24, no. 3, pp. 383–394, 2025.

[10] G. N. Rackoff, Z. Z. Zhang, and M. G. Newman, "Chatbot-delivered mental health support: attitudes and utilization in a sample of U.S. college students," Digital Health, vol. 11, 2025.

[11] A. Ahmed, S. Aziz, C. T. Toro, M. Alzubaidi, S. Irshaidat, H. Abu Serhan, A. A. Abd-alrazaq, and M. Househ, "Machine learning models to detect anxiety and depression through social media: a scoping review," Computer Methods and Programs in Biomedicine Update, vol. 2, Art. no. 100066, 2022.

[12] S. Sarkar, "Sentiment Analysis for Mental Health," Kaggle, 2024.

[13] N. Reimers and I. Gurevych, "Sentence-BERT: sentence embeddings using Siamese BERT-networks," in Proc. EMNLP-IJCNLP, Hong Kong, 2019, pp. 3982–3992.

[14] J. Brooke, "SUS: a 'quick and dirty' usability scale," in Usability Evaluation in Industry, P. W. Jordan, B. Thomas, B. A. Weerdmeester, and A. L. McClelland, Eds. London, U.K.: Taylor & Francis, 1996, pp. 189–194.

[15] F. D. Davis, "Perceived usefulness, perceived ease of use, and user acceptance of information technology," MIS Quarterly, vol. 13, no. 3, pp. 319–340, 1989.

[16] J. Sauro and J. R. Lewis, Quantifying the User Experience: Practical Statistics for User Research, 2nd ed. Cambridge, MA, USA: Morgan Kaufmann, 2016.

[17] F. Pedregosa et al., "Scikit-learn: machine learning in Python," Journal of Machine Learning Research, vol. 12, pp. 2825–2830, 2011.

[18] E. F. Gbollie et al., "Intention to use digital mental health solutions: a cross-sectional survey of university students' attitudes and perceptions toward online therapy, mental health apps, and chatbots," Digital Health, vol. 9, Art. no. 20552076231216559, 2023.

[19] L. Li, W. Peng, and M. M. J. Rheu, "Factors predicting intentions of adoption and continued use of artificial intelligence chatbots for mental health: examining the role of UTAUT model, stigma, privacy concerns, and artificial intelligence hesitancy," Telemedicine and e-Health, vol. 30, no. 3, pp. 722–730, 2024.

[20] T. Koulouri, R. D. Macredie, and D. Olakitan, "Chatbots to support young adults' mental health: an exploratory study of acceptability," ACM Trans. Interact. Intell. Syst., vol. 12, no. 2, Art. no. 11, 2022.