ISSN: 3061-8843

Journal of Artificial Intelligence and Computing Applications

ICAIMH2025

Mérida, Yucatán, México

July 2-3, 2025

MAIKRON

Special Issue – September 2025 Vol. 3 No. 2

www.maikron.org/jaica

Journal of **Artificial Intelligence** and **Computing Applications**

Journal of Artificial Intelligence and Computing Applications, Vol. 3, Núm. 2, Edición Especial, es una publicación semestral editada por la Asociación para el Avance de las Aplicaciones Inteligentes y Tecnologías con Impacto Social, Calle 12 No. 103, Lotes No. 56 y 57, Ucú, Yucatán, C.P. 97357, Tel. (55) 28 55 70 20. URL: https://www.maikron.org/jaica, correo electrónico: contact@maikron.org. Editor responsable: Mauricio Gabriel Orozco del Castillo. Certificado de Reserva de Derechos al uso Exclusivo del Título: 04-2025-073112454300-102 por el Instituto Nacional del Derecho de Autor (INDAUTOR), eISSN:3061-8843. Responsable de la última actualización de este número, Mauricio Gabriel Orozco del Castillo, Editor Responsable. Fecha de la última modificación: 30 de septiembre de 2025, Calle 12 No. 103, Lotes No. 56 y 57, C.P. 97357, Ucú, Yucatán.

Journal of Artificial Intelligence and Computing Applications
Vol. 3, No. 2, Special Issue

Contents

Foreword	i

Conference Reports

Conference Report on the 2025 International Conference on Artificial Intelligence for Mental Health (ICAIMH 2025)	1

Conference Abstracts

A methodology for the development of serious games for the cognitive stimulation of elderly people with mild cognitive impairment	3
Appraising cognitive status in dementia via touch-based reaction time: a preliminary machine learning study	4
Mapping the scientific landscape of artificial intelligence in mental health	5
Analysis of the sleep quality of college students from different knowledge areas using a data mining approach	6
Relationship between academic engagement and burnout syndrome in Mexican students: a PLS-SEM analysis	7
Del algoritmo a la interpretación clínica de la ansiedad en el parto: análisis y explicabilidad de modelos predictivos obstétricos basados en indicadores psicológicos	8
Multimodal emotion recognition for empathic virtual agents in mental health interventions	9
Analyzing municipal patterns of suicide and depression in Mexico: a multilayer network approach	10
The role of main sociodemographic variables in suicide ideation detection using machine learning	11

Purpose in life assessment with artificial intelligence: a contextual approach to mental health in university students	12

Expanded Abstracts

Level of attachment to artificial intelligence in university students	13
Detection of people with social phobia using pupillary position classification algorithms	18
Automated detection of academic anxiety levels: proposal of a machine learning-based expert system for educational and clinical settings	25
Automatic feedback through natural language processing using a chatbot-based simulated patient (PEPE) for the training of mental health professionals	33
On-device conversational agent for psycho-oncology based on acceptance and commitment therapy manuals	38

Journal of Artificial Intelligence and Computing Applications

MAIKRON

Foreword

Dear Readers, Contributors, and Colleagues,

It is with great enthusiasm that I introduce Volume 3, Issue 2 of the Journal of Artificial Intelligence and Computing Applications (JAICA)—a Special Issue dedicated to the 2025 International Conference on Artificial Intelligence for Mental Health (ICAIMH 2025), held at Universidad Anáhuac Mayab in Mérida, Yucatán, México. This issue serves as the official proceedings of the conference and reflects our continued commitment to disseminating high-quality research that applies artificial intelligence to socially significant domains.

The papers and abstracts presented here embody the central mission of ICAIMH: to advance dialogue and innovation at the intersection of artificial intelligence and mental health. The collection includes ten long papers—three selected for consideration in Inteligencia Artificial and seven published or forthcoming in JAICA—as well as five expanded abstracts representing emerging lines of inquiry. Together, they showcase diverse approaches to topics such as diagnostic and detection tools, conversational agents, predictive modeling, and pattern identification in mental health data.

This special issue stands as the result of close collaboration between JAICA and ICAIMH, supported by the Association for the Advancement of Intelligent Applications and Technologies with Social Impact (Maikron) and the International Institute for Intelligent Technologies (IIIT). I would like to express my sincere appreciation to the authors, the conference program committee, and the journal's reviewers for their dedication and thoughtful contributions throughout the review and editorial process.

Beyond documenting the outcomes of a single event, this issue represents the consolidation of a growing research community that recognizes mental health as a field where artificial intelligence can make a genuine and positive impact. The studies presented here illustrate how technical innovation, when guided by ethical principles and interdisciplinary collaboration, can contribute not only to scientific progress but also to human well-being.

The synergy between JAICA and ICAIMH continues to strengthen our shared vision: to create a space where technological innovation meets human-centered care. As this issue demonstrates, interdisciplinary collaboration remains essential to ensuring that artificial intelligence evolves not only as a scientific field but also as a force for ethical and socially meaningful progress.

Warmest regards,

Mauricio G. Orozco-del-Castillo

Editor-in-Chief

Journal of Artificial Intelligence and Computing Applications (JAICA)

Conference Reports

Conference Report on the 2025 International Conference on Artificial Intelligence for Mental Health (ICAIMH 2025)

Mauricio G. Orozco-del-Castillo
Tecnológico Nacional de México / IT de Mérida, México

Juan Recio-Garcia
Universidad Complutense de Madrid, Spain

Pedro Ortiz-Sánchez
Tecnológico Nacional de México / IT de Mérida, México

The International Conference on Artificial Intelligence for Mental Health (ICAIMH) 2025 was held from July 2 to July 3, 2025, at Universidad Anáhuac Mayab in Mérida, Yucatán, México. As the third edition of ICAIMH, the event continued to build upon the momentum and success of the 2023 and 2024 conferences, further consolidating its role as a leading forum in this emerging field. Maintaining its overarching focus on exploring how artificial intelligence can contribute to the understanding, detection, and improvement of mental health, ICAIMH 2025 brought together researchers, practitioners, and students from diverse disciplines to share knowledge, foster dialogue, and advance innovative solutions at the intersection of AI and mental health.

ICAIMH 2025 was organized by the Association for the Advancement of Intelligent Applications and Technologies with Social Impact (Maikron), reaffirming its mission to leverage artificial intelligence for societal benefit. The conference also received important support from institutional partners, particularly the Asociación para la Salud Mental de Yucatán (ASMY) and the International Institute for Intelligent Technologies (IIIT). In addition, active participation from the AAAIMX Mexican Student Chapter and the ACM/ITM Student Chapter served to engage the next generation of researchers. A wide range of institutions also collaborated to make the conference possible, further enriching the diversity and reach of ICAIMH 2025.

The program of ICAIMH 2025 was coordinated by the Program Chairs: Mauricio Gabriel Orozco-del-Castillo, Juan Antonio Recio-García, and Pedro Ortiz-Sánchez, with the support of an extensive program committee. Among its key organizers, Lucía Alejandra Dzul-Sánchez, Carlos Bermejo-Sabbagh, Nora Cuevas-Cuevas, and Ana Martín-Casado played central roles in shaping the academic content and ensuring the conference's interdisciplinary orientation. Unlike the 2024 edition, which included an associated track with ICCBR, this year the conference program was structured around four main sessions: (1) AI-powered diagnostic and detection tools, (2) Conversational AI and virtual environments for mental health, (3) Mental health prediction using AI, and (4) Identification of mental health patterns using AI.

The call for papers of ICAIMH 2025 invited submissions in the form of long papers and expanded abstracts, reflecting the conference's aim to encourage both mature contributions and emerging ideas. The publication outcomes demonstrated the quality and diversity of the work received. Among the long papers, three were selected for consideration in the journal *Inteligencia Artificial* as recipients of the Best Paper Award, while the remaining seven were accepted for publication in regular issues of the *Journal of Artificial Intelligence and Computing Applications* (JAICA), with five already published in Volume 3(1) and two forthcoming. In addition, this special issue of JAICA (Volume 3(2)) serves as the official proceedings of the conference, comprising the present report, the conference abstracts of all ten long papers, and the five expanded abstracts in their complete form.

Official banner of ICAIMH 2025 displayed at Universidad Anáhuac Mayab, Mérida, Yucatán, México.

https://doi.org/10.5281/zenodo.17281575
 ISSN: 3061-8843

The first day of ICAIMH 2025 opened with the invited keynote lecture by Dr. Raúl Alelú Paz (Universidad Francisco de Vitoria, Spain; EVER3; Thera4All), titled "When the Mind Meets the Algorithm: Artificial Intelligence and the Future of Mental Health." A multidisciplinary scientist with dual doctorates in medicine and psychology, Dr. Alelú Paz highlighted the transformative role of artificial intelligence in shaping the future of mental health, drawing from his extensive expertise in neuroscience, psychotherapeutic practice, and computational modeling. Later that day, Dr. Eduardo Barbará Morales (Universidad Anáhuac Mayab, México) presented his invited talk "The Impact of COVID-19 on Brain Morphology: A Quantitative Approach." With a distinguished background in biomedical engineering and neuroimaging, Dr. Barbará Morales analyzed the neurological consequences of the COVID-19 pandemic and discussed how AI-driven quantitative methods can shed light on structural brain changes.

The second day of ICAIMH 2025 continued with two thought-provoking invited lectures. Mtra. Mariana Sologuren López (Universidad Modelo, México) delivered the talk "Can a Machine Hold the Soul? Mental Health and Human Bonds in the Age of Artificial Intelligence." Drawing from her clinical experience in psychotherapy and education, she reflected on the implications of AI for human relationships, emphasizing the need to preserve authentic interpersonal connections in technology-mediated contexts. The conference concluded with the closing invited lecture by Dr. Juan Antonio Recio-García (Universidad Complutense de Madrid, Spain; International Institute for Intelligent Technologies), titled "When Machines Decide: Ethics and Responsibility in the Age of Artificial Intelligence." Dr. Recio-García, an expert in explainable AI and software engineering, examined the ethical challenges posed by autonomous systems and emphasized the importance of accountability and transparency as AI becomes increasingly embedded in mental health practice and beyond.

ICAIMH 2025 brought together a diverse and engaged community of participants. The conference featured 6 invited speakers, 17 authors, and a total of 90 assistants, comprising 20 guests, 38 students, and 32 professionals. This broad participation reflected the interdisciplinary and inclusive spirit of the event. Attendees represented a variety of institutions, including the Universidad Autónoma de Yucatán (UADY), the Universidad Nacional Autónoma de México (UNAM, FES Iztacala and IIMAS), the American University of the Middle East, the Centro de Investigación en Ciencias de Información Geoespacial (CentroGeo), the Universidad Complutense de Madrid, the Instituto Nacional de Astrofísica, Óptica y Electrónica (INAOE), the Universidad de La Rioja, the Universidad Anáhuac Mayab, the Instituto Tecnológico Autónomo de México (ITAM), and the Centro de Investigación Científica y de Educación Superior de Ensenada (CICESE), among others.

Beyond its academic contributions, ICAIMH 2025 fostered an environment of inclusivity and collaboration. The conference welcomed participants from academia, healthcare, industry, and public service, creating opportunities for dialogue across sectors that seldom converge. Networking activities and open forums encouraged attendees to share perspectives and build connections that extend beyond the conference itself. The strong involvement of students, both as authors and as participants, highlighted the role of the next generation in advancing research and practice at the intersection of AI and mental health.

The success of ICAIMH 2025 was made possible through the contributions of many individuals and institutions. We extend our gratitude to the Association for the Advancement of Intelligent Applications and Technologies with Social Impact (Maikron) for leading the organization of the conference, and to our secondary partners, the Asociación para la Salud Mental de Yucatán (ASMY) and the International Institute for Intelligent Technologies (IIIT), for their valuable support. We also acknowledge the participation of the AAAI Student Chapter Mexico (AAAIMX) and the ACM/ITM Student Chapter, whose involvement reflected the commitment of emerging researchers and professionals. Special thanks go to the program committee members, whose expertise and dedication shaped the academic content of the conference, as well as to our invited speakers, who enriched the program with their knowledge and perspectives. Finally, we thank the numerous collaborating institutions that supported ICAIMH 2025 and contributed to fostering an environment of interdisciplinary exchange and collaboration.

ICAIMH 2025 highlighted the transformative potential of artificial intelligence in addressing pressing challenges in mental health. The discussions and contributions throughout the conference reaffirmed that meaningful progress in this area requires not only technological innovation but also sustained collaboration across disciplines. Emphasis was placed on the need for ethical and human-centered approaches, ensuring that advances in AI serve to strengthen care, support practitioners, and benefit individuals and communities. Looking ahead, the next edition, ICAIMH 2026, is expected to take place in July in Mérida, Yucatán, México, with further details to be announced soon. For updates and information about upcoming editions, participants are encouraged to visit the official website at www.icaimh.org.

Journal of Artificial Intelligence and Computing Applications (2025) - **Special Issue** - 3(2): 3

MAIKRON

Journal of Artificial Intelligence and Computing Applications

Conference abstract

A methodology for the development of serious games for the cognitive stimulation of elderly people with mild cognitive impairment

Luisa Andrea Morales-García[1,*], **Luis Enrique Sucar**[1], **Delia Irazú Hernández Farias**[2], **and Alberto L. Moran Y Solares**[2]

[1]Instituto Nacional de Astrofísica, Óptica y Electrónica, Puebla, Mexico.
[2]Universidad Autónoma de Baja California, Ensenada, Mexico.

ABSTRACT

Mild cognitive impairment affects many older adults and can lead to severe dementia. Early detection and intervention are key to slowing its progression. Serious games offer a promising way to stimulate cognitive function. However, there is a lack of clear methodologies for developing effective serious games for cognitive stimulation. In this paper, we introduce a methodology for developing serious games to help people with mild cognitive impairment. We also include a case study of this methodology through the development of a serious game that underwent usability testing with older adults. The obtained results provide evidence that, by following the proposed methodology, it is possible to develop serious games that are well received by the target population.

Keywords: serious games, cognitive impairment, cognitive stimulation

This work corresponds to a paper presented at the International Conference on Artificial Intelligence for Mental Health (ICAIMH) 2025. The complete version has been published in the Journal of Artificial Intelligence and Computing Applications (JAICA) and is available at: https://maikron.org/jaica/index.php/ojs/article/view/63.

E-mail address: luisa.morales@inaoep.mx
https://doi.org/10.5281/zenodo.17180274
 ISSN: 3061-8843

MAIKRON

Journal of Artificial Intelligence and Computing Applications

Conference abstract

Appraising cognitive status in dementia via touch-based reaction time: a preliminary machine learning study

Marco Esquer-Rochin[1,*], **Luis-Felipe Rodriguez**[1], **and J. Octavio Gutierrez-Garcia**[2]

[1]Instituto Tecnologico de Sonora (ITSON)
[2]Instituto Tecnológico Autónomo de México (ITAM)

ABSTRACT

People with dementia (PwD) perform cognitive-based therapeutic activities. Literature reports a variety of studies exploring relationships between the cognitive status of PwD as determined by the Mini-Mental State Examination (MMSE) and their reaction times from a myriad of stimuli incorporated into cognitive activities. Nevertheless, these technology-supported activities usually include distracting elements, complex instructions, and unfamiliar devices for older adults, introducing bias into reaction times. The objective of this work is to appraise the cognitive status of people with dementia using reaction times from touch interaction tasks. For this purpose, a relatively simple cognitive activity (involving the intuitive tap gesture) and a 32-inch wide touchscreen were designed and implemented. Afterward, 21 PwD from a day center located in Sonora, Mexico were recruited. The participants were instructed to carry out a cognitive activity consisting of five consecutive taps and their reaction times were recorded. The collected data was analyzed using (i) a correlation analysis, (ii) a bootstrap evaluation of machine learning classification models, and (iii) a logistic regression analysis. From the empirical results, it can be concluded that there is a negative relationship between the MMSE score of PwD and the reaction times from taps. In addition, the bootstrapped mean accuracy results of the classifiers suggest that it may be feasible to automatically classify PwD.

Keywords: dementia, cognitive tasks, machine learning

This work corresponds to a paper presented at the International Conference on Artificial Intelligence for Mental Health (ICAIMH) 2025. The complete version has been published in the Journal of Artificial Intelligence and Computing Applications (JAICA) and is available at: https://maikron.org/jaica/index.php/ojs/article/view/73.

E-mail address: octavio.gutierrez@itam.mx
https://doi.org/10.5281/zenodo.17196762
 ISSN: 3061-8843

MAIKRON

Journal
of
Artificial Intelligence
and
Computing Applications

Conference abstract

Mapping the scientific landscape of artificial intelligence in mental health

Ivanna Shubina[1,*] **and Adrian Jarema Dzido**[2]

[1]Liberal Arts Department, American University of the Middle East, 54200, Egaila, Kuwait
[2]Computing Science Department, Radboud University, Nijmegen, the Netherlands

ABSTRACT

Artificial Intelligence (AI) has gained increasing popularity in contemporary scientific research; however, its application in mental health still requires a consolidated understanding of existing findings regarding effectiveness. This bibliometric study aims to synthesize current knowledge and explore research trends related to AI's role in mental health. It investigates how advancements in modern technologies are used to predict, prevent, and treat mental disorders, and evaluates their effectiveness. A literature search was conducted using Lens software to retrieve peer-reviewed empirical studies in English from highly ranked databases, covering the period from 2005 to 2025. A total of 97 relevant publications were identified and analyzed for patterns, trends, and associations using the Bibliometrix package in R. Results reveal a sharp increase in publications after 2020. Clinical and applied psychology emerged as dominant fields. *Eating and Weight Disorders* is the leading journal (n=22), followed by the *Journal of Psychopathology and Behavioral Assessment* (n=19) and *Cognitive Therapy and Research* (n=17). The United States is both the most productive (n=149) and most cited country (n=8,896). AI has demonstrated promise in detecting symptoms of depression and suicidal behavior, preventing mental health disorders, and enhancing traditional psychological interventions. Nonetheless, several gaps remain, including the underrepresentation of diverse populations and a limited understanding of factors influencing user acceptance of AI-based tools. This study provides researchers with an overview of publication trends, collaboration networks, keyword analysis, and future research directions. It also supports practitioners in selecting appropriate AI-based interventions to improve mental health outcomes and overall well-being within healthcare systems.

Keywords: artificial intelligence, mental health, depression

This work corresponds to a paper presented at the International Conference on Artificial Intelligence for Mental Health (ICAIMH) 2025. The complete version has been published in the Journal of Artificial Intelligence and Computing Applications (JAICA) and is available at: https://maikron.org/jaica/index.php/ojs/article/view/84.

E-mail address: ivanna.shubina@aum.edu.kw
https://doi.org/10.5281/zenodo.17196884
 ISSN: 3061-8843

MAIKRON

Journal of Artificial Intelligence and Computing Applications

Conference abstract

Analysis of the sleep quality of college students from different knowledge areas using a data mining approach

Andrea Morales-Robles[1], **Víctor Menéndez-Domínguez**[1], **and Héctor Rubio-Zapata**[1]

[1]Universidad Autónoma de Yucatan

ABSTRACT

Sleep is an essential physiological process involved in memory consolidation, metabolic and endocrine homeostasis, and immune system regulation. Therefore, good sleep quality is vital for maintaining physiological homeostasis. Poor sleep quality is prevalent among college students and affects both physical and mental health. Conventional statistical methods, such as logistic regression, are commonly used to generate predictive models of sleep quality and have been extensively applied to Health Sciences students, but their use has been less studied among students from other disciplines, such as Engineering and Exact Sciences. Data mining can help overcome certain limitations of these conventional methods, such as multicollinearity, by uncovering associations that might otherwise have gone unnoticed. In this study, we separately analyzed two samples of students from Health Sciences and Engineering and Exact Sciences. We found significant correlations between sleep quality and attributes such as perceived sleep quality, sleep latency, sleep duration, drug use, and the use of medication for depression and anxiety. Decision trees identified different predictive attributes between the two samples. These findings offer a novel insight into sleep quality among college students and may support informed decision-making and targeted interventions.

Keywords: sleep quality, college students, data mining

This work corresponds to a paper presented at the International Conference on Artificial Intelligence for Mental Health (ICAIMH) 2025. The complete version has been published in the Journal of Artificial Intelligence and Computing Applications (JAICA) and is available at: https://maikron.org/jaica/index.php/ojs/article/view/57.

E-mail address: andreamoralesrobles@outlook.com
https://doi.org/10.5281/zenodo.17156435
 ISSN: 3061-8843

MAIKRON

Journal of Artificial Intelligence and Computing Applications

Conference abstract

Relationship between academic engagement and burnout syndrome in Mexican students: a PLS-SEM analysis

Irving A. Ramírez-Muñoz[1], Vanesa Avalos-Gaytán[1,*], Igor Barahona[2], Valeria Soto-Mendoza[1], and Gabriela Linares Acuña[3]

[1]Universidad Autónoma de Coahuila, Centro de Investigación en Matemáticas Aplicadas, Unidad Saltillo, Mexico.
[2]King Fahd University of Petroleum and Minerals, Department of Information Systems and Operations Management, Dhahran, Saudi Arabia.
[3]Universidad Autónoma de Coahuila, Facultad de Psicología, Unidad Saltillo, Mexico.

ABSTRACT

Academic burnout is a growing concern in higher education, characterized by emotional exhaustion, cynicism, and a reduced sense of accomplishment. In contrast, academic engagement – defined as a positive, energetic, and committed state toward learning – has been identified as a protective factor and even an antidote to burnout. While most studies in this area have focused either on theoretical model development or on validating measurement instruments, few address both simultaneously. Moreover, research using Structural Equation Modeling (SEM) has predominantly been conducted in Europe and the United States, leaving Latin American contexts underexplored. A literature review revealed only nine studies on academic burnout in Mexico, underscoring the need for further investigation in the region. This study aims to bridge that gap by validating adapted versions of the Maslach Burnout Inventory-Student Survey (MBI-SS) and the School Engagement Measure, and by developing a SEM to examine how academic engagement influences burnout levels among students at a public university in northern Mexico. The findings are expected to contribute to the understanding of student well-being in Latin America and to offer validated tools for measuring and addressing academic burnout.

Keywords: burnout syndrome, academic engagement, structural equation model

This work corresponds to a paper presented at the International Conference on Artificial Intelligence for Mental Health (ICAIMH) 2025. The complete version has been published in the Journal of Artificial Intelligence and Computing Applications (JAICA) and is available at: https://maikron.org/jaica/index.php/ojs/article/view/61.

E-mail address: vanesaavalos@uadec.edu.mx
https://doi.org/10.5281/zenodo.17156510
 ISSN: 3061-8843

Journal of Artificial Intelligence and Computing Applications (2025) - **Special Issue** - 3(2): 8

Journal of Artificial Intelligence and Computing Applications

Conference abstract

Del algoritmo a la interpretación clínica de la ansiedad en el parto: análisis y explicabilidad de modelos predictivos obstétricos basados en indicadores psicológicos

Ana Martin-Casado[1] **and Juan A. Recio-Garcia**[2,*]

[1]Universidad Internacional de la Rioja, UNIR
[2]Department of Software Engineering and Artificial Intelligence, Instituto de Tecnologias del Conocimiento. Universidad Complutense de Madrid

ABSTRACT

La ansiedad durante el embarazo constituye un factor relevante que puede influir significativamente en el desarrollo del parto. Este estudio presenta un enfoque novedoso basado en inteligencia artificial para predecir tanto el tipo como la duración del parto utilizando indicadores psicológicos de ansiedad previos al alumbramiento. Empleando datos de 235 gestantes a término de dos hospitales españoles, desarrollamos un modelo de perceptrón multicapa para clasificar partos eutócicos y distócicos, alcanzando una capacidad para identificar el 88% de los partos distócicos. Adicionalmente, implementamos un modelo de regresión que predice el tiempo de parto con un error medio de 2 horas, prediciendo correctamente el 86% de los casos con un margen de error inferior a 3 horas. La aplicación de técnicas de inteligencia artificial explicable permite comprender la influencia específica de cada factor de ansiedad en el desarrollo del parto. Estos resultados demuestran el potencial de los modelos de IA para mejorar la atención obstétrica y optimizar la asignación de recursos sanitarios.

Keywords: anxiety prediction, explainable artificial intelligence, birth

This work corresponds to a paper presented at the International Conference on Artificial Intelligence for Mental Health (ICAIMH) 2025, where it was selected as one of the recipients of the **Best Paper Award**. The complete version is expected to be published soon after the conference.

E-mail address: jareciog@ucm.es
https://doi.org/10.5281/zenodo.17229146
 ISSN: 3061-8843

Journal of Artificial Intelligence and Computing Applications (2025) - **Special Issue** - 3(2): 9

MAIKRON

Journal of Artificial Intelligence and Computing Applications

Conference abstract

Multimodal emotion recognition for empathic virtual agents in mental health interventions

Marcelo Alejandro Huerta-Espinoza [1,*], **Ansel Yoan Rodríguez González** [1], **and Juan Martinez-Miranda** [1]

[1]CICESE Unidad Académica Tepic

ABSTRACT

Depression and anxiety disorders affect millions of individuals globally and are commonly addressed through psychological interventions. A growing technological approach to support such treatments involves the use of embodied conversational agents that employ motivational interviewing, a method that promotes behavioral change through empathic engagement. Despite its critical role in therapeutic efficacy, empathy remains a significant challenge for virtual agents to emulate. Emotion Recognition (ER) technologies offer a potential solution by enabling agents to perceive and respond appropriately to users' emotional states. Given the inherently multimodal nature of human emotion, unimodal ER approaches often fall short in accurately interpreting affective cues. In this work, we propose a multimodal emotion recognition model that integrates verbal and non-verbal signals (text and video) using a Cross-Modal Attention fusion strategy. Trained and evaluated on the IEMOCAP dataset, our approach leverages Ekman's taxonomy of basic emotions and demonstrates superior performance over unimodal baselines across key metrics such as accuracy and F1-score. By prioritizing text as the main modality and dynamically incorporating complementary visual cues, the model proves effective in complex emotion classification tasks. The proposed model is designed for integration into an existing conversational agent aimed at supporting individuals experiencing emotional and psychological distress. Future work will involve embedding the model in the conversational agent platform for emotionally distressed users, aiming to assess its real-world impact on engagement, user experience, and perceived empathy.

Keywords: emotion recognition in conversation, deep learning, multimodal classification

This work corresponds to a paper presented at the International Conference on Artificial Intelligence for Mental Health (ICAIMH) 2025, where it was selected as one of the recipients of the **Best Paper Award**. The complete version is expected to be published soon after the conference.

E-mail address: marcelo.huerta@cicese.edu.mx
https://doi.org/10.5281/zenodo.17229360
 ISSN: 3061-8843

MAIKRON

Journal of Artificial Intelligence and Computing Applications

Conference abstract

Analyzing municipal patterns of suicide and depression in Mexico: a multilayer network approach

Jorge Manuel Pool Cen[1,*], **Hugo Carlos Martínez**[1], **Gandhi Samuel Hernández Chan**[1], **Alfredo Alejandro Montero Arciniega**[2], **Pedro Alberto Mendoza Pablo**[3], **and Martha Cordero Oropeza**[4]

[1]Centro de Investigación en Ciencias de Información Geoespacial (Centro Geo); Laboratorio Nacional de Geointeligencia
[2]EPO 339; Colegio de Bachilleres
[3]Escuela Militar de Graduados de Sanidad
[4]Instituto Nacional de Psiquiatría "Ramón de la Fuente Muñiz"

ABSTRACT

This study analyzes the territorial patterns of suicide and depression in Mexican municipalities from 2015 to 2020 using a multilayer network approach. A panel dataset was constructed with variables grouped into three categories: mental health cases (suicide and depression rates), substance use (alcoholism, tobacco use, and drug use), and infrastructure availability (mental health services and resources). Cosine similarity was used to generate multilayer graphs, and the Infomap algorithm was applied to identify clusters of municipalities with similar structural characteristics. The results show clear differences across clusters in both levels and temporal dynamics of the indicators. Notably, Clusters 1 and 2 consistently exhibited higher rates of substance use and mental health indicators. These findings highlight the value of network-based approaches to understanding the territorial and temporal dimensions of mental health.

Keywords: clustering, multilayer graph, mental health

This work corresponds to a paper presented at the International Conference on Artificial Intelligence for Mental Health (ICAIMH) 2025, where it was selected as one of the recipients of the **Best Paper Award**. The complete version is expected to be published soon after the conference.

E-mail address: jpool@centrogeo.edu.mx
https://doi.org/10.5281/zenodo.17229663
 ISSN: 3061-8843

MAIKRON

Journal of Artificial Intelligence and Computing Applications

Conference abstract

The role of main sociodemographic variables in suicide ideation detection using machine learning

Gandhi Samuel Hernandez-Chan[1,*], **Matilde Jiménez-Coello**[2], **Manuel Sosa-Correa**[2], and **Sally Vanega-Romero**[2]

[1]Centro de Investigación en Ciencias de Información Geoespacial
[2]Universidad Autónoma de Yucatán

ABSTRACT

This study investigates the role of key sociodemographic variables in the detection of suicidal ideation within a Yucatan population, employing machine learning (ML) classification models. Data was obtained from the APPSI platform, encompassing psychological assessments (C-SSRS, DASS21, EAYIE) and sociodemographic information (age, gender, municipality, etc.). Ten classification algorithms were trained and evaluated to predict suicidal ideation. The Logistic Regression model demonstrated the strongest performance, achieving an F1 score of 0.77 and an accuracy of 81%, with a recall of 74% for the at-risk group. Results highlight the potential of ML to support mental health decision-making and the importance of sociodemographic factors in suicide ideation detection.

Keywords: sociodemographic, mental health, suicide ideation

This work corresponds to a paper presented at the International Conference on Artificial Intelligence for Mental Health (ICAIMH) 2025. The complete version **is expected to be published** in the *Journal of Artificial Intelligence and Computing Applications (JAICA)* and will be available at: https://www.maikron.org/jaica/.

E-mail address: ghernandez@centrogeo.edu.mx
https://doi.org/10.5281/zenodo.17202717
 ISSN: 3061-8843

MAIKRON

Journal of Artificial Intelligence and Computing Applications

Conference abstract

Purpose in life assessment with artificial intelligence: a contextual approach to mental health in university students

Guillermo Alfredo Arrioja-Carrera[1,*], Eduardo Roldán-Reyes[1], Pedro Humberto Velázquez-Morales[2], and Víctor Méndez-García[1]

[1]Tecnológico Nacional de México - Tecnológico de Orizaba
[2]Universidad Veracruzana

ABSTRACT

The Purpose in Life Questionnaire (PIL) assesses life purpose and its connection to depression, anxiety and other mental disorders. This exploratory, longitudinal study investigates the sense of life purpose among 114 university students from three institutions in Mexico, using the PIL and 18 additional context-based questions. These extra items were grouped into a Contextual Opportunity Index (COI) consisting of four dimensions: demographics, academic-cultural background, access to public-private services, and lifestyle. Results show that 57% of participants reported a defined life purpose, 26.3% experienced existential uncertainty, and 16.7% showed signs of an existential void. The lifestyle dimension showed the strongest correlation with life purpose—particularly "nuclear family situation" ($r = 0.431$, $p < 0.000001$) and "travel and leisure" ($r = 0.263$, $p = 0.005$). Predictive modeling using Naive Bayes and Decision Tree classifiers revealed that students' attitudes toward death acceptance, heavily influenced final PIL scores. Additionally, the dimensions Access to private services and Demography variables were linked to a higher PIL score. These findings suggest that both internal and contextual factors contribute to students' psychological well-being and life meaning.

Keywords: purpose in life, predictive modeling, mental health

This work corresponds to a paper presented at the International Conference on Artificial Intelligence for Mental Health (ICAIMH) 2025. The complete version **is expected to be published** in the *Journal of Artificial Intelligence and Computing Applications (JAICA)* and will be available at: https://www.maikron.org/jaica/.

E-mail address: guillermo.ac@orizaba.tecnm.mx
https://doi.org/10.5281/zenodo.17203619
 ISSN: 3061-8843

Journal of Artificial Intelligence and Computing Applications (2025) - **Special Issue** - 3(2): 13–17

MAIKRON

Journal of Artificial Intelligence and Computing Applications

Expanded abstract

Level of attachment to artificial intelligence in university students

José J. Karam[1,*] **and Gabriel Urzaiz**[1]

[1]Anahuac Mayab University

ABSTRACT

Artificial Intelligence (AI) is currently a widely used tool, but confusion between humans and computers could lead to erroneous behaviors and attitudes. Specifically in the field of education, AI can be a powerful tool, but its application carries a risk when it invades human functions. This article proposes the concepts of "Level of Attachment to AI," "Attributions of Human Qualities," and "Catastrophic Ideas about AI." It presents the results of a series of studies conducted over the past three years with the aim of thoroughly understanding these concepts, as well as the implications that over-involvement and excessive dependence on this new technology could have. The results show significant levels in each of the categories, which encourages further study, as well as the development and implementation of measures aimed at the healthy adoption of AI techniques and tools.

Keywords: level of attachment, artificial intelligence, university students

1. Introduction

In recent years, especially after the pandemic, the use and abuse of technology, social media, and recently, artificial intelligence have generated two major poles among professionals in various fields [1], especially mental health. A large group supports its use and utilization in moderation to enhance education, entertainment, and communication from a very early age, while another important sector views its incorporation with caution, wary of its potential harm to communication, creativity, and even people's mental health.

Artificial Intelligence (AI) is now a widely used tool, and the development of this technology has reached the point where it can now be used without a deep understanding of the technology. It's now possible to use AI on a computer or mobile phone without having a clear understanding of what it is or how it works, similar to driving a car without any mechanical knowledge.

Indeed, to take advantage of a tool, it's not necessary to know how to build or repair it, and in many cases, knowing how to operate it may be enough. But in the case of AI, simply knowing how to operate it isn't enough, because this technology is capable of emulating some unique human behaviors. For this reason, it's necessary to thoroughly understand what AI is and what it's used for.

A good way to begin to understand the true essence of AI is to remember that this name was coined in 1956 when a group of scientists led by John McCarthy, Minsky, and others, met [2] at Dartmouth College in New Hampshire, USA, to try to establish the domain of knowledge in which they worked and coined the term AI to refer simply to the possibility of machines exhibiting intelligent behavior. The so-called "artificial intelligence" is not intelligence, but rather intelligent behavior, the fruit of natural human intelligence.

Human intelligence is truly intelligence; it is the ability to reason and to read things deeply. How-

E-mail address: jose.karam@anahuac.mx

https://doi.org/10.5281/zenodo.17228049

 This article is part of the Special Issue on ICAIMH 2025. ISSN: 3061-8843

ever, when we talk about AI, its essence is a machine—circuits, software, or other things—but ultimately, it is not intelligence. Its essence is electronics and mechanical parts. Although it may appear to be an animal, like a kangaroo or an android, although it may have a physical appearance when it speaks (or rather, when it simulates speaking), it is nothing more than machines and in no way animals or artificial people, because their essence simply cannot be changed.

Confusion between humans and computers could lead to computers and robots being given powers that are exclusively human. This leads to erroneous behaviors and attitudes.

American psychologist Jonathan Hidt [3] has strongly promoted the idea that excessive use of screens is producing a significant increase in depression and perceived anxiety disorders among the new generation, and that it is urgent to reverse this phenomenon. The proliferation of terms such as robophilia, robophobia, and syndromes such as Hikikomori seem to support this cautious view. Research studies are looking at phenomena such as the romantic and sexual relationships of a recent number of adolescents and young adults with AI, through specific applications [4, 5].

At our institution, a collaborative effort has been proposed between the schools of Engineering, Humanities, and Psychology to thoroughly understand the level of attachment, erroneous beliefs, and attributions of human qualities, as well as the implications that the over-involvement of younger generations and their excessive dependence on this new generative tool could have.

Specifically in the field of education, AI can be a powerful tool to support the teaching-learning process, but the problem begins when AI invades human functions.

A first risk lies in the fact that this technology is capable of performing activities that students necessarily need to perform and experience as a fundamental part of their learning. For example, having the AI solve exercises assigned to the student to practice, or having the AI conduct bibliographic research and write an essay instead of having the student practice searching for information and structuring their ideas, etc. The quality of the work produced with AI may be excellent, but it loses sight of the main objective: for the student to learn and grow through practice activities. Proceeding in this manner prevents the student from practicing and developing their skills.

A second risk exists in the way students interact with the various people involved in the learning process. The use of technology can be of great help in supervising, advising, and supporting students, especially in cases where physical presence is difficult or impossible (due to distance) or personal attention (due to time constraints or the number of students). However, the support of a machine, no matter how useful, timely, pertinent, and warm, will never be comparable to the relationship between people. Except in the case of self-study, the success of the educational process depends largely on the way students interact with their teachers and peers.

2. Methodology

The first step was to define the categories of interest, and then move on to developing the questionnaire and finally proceeded to conduct studies to estimate the category levels.

2.1 Definition of interest categories

The following three interest categories were defined to assess their respective levels during the studies:

- Level of Attachment to AI
- Attributions of Human Qualities
- Catastrophic Ideas about AI

The Level of Attachment to AI category [6, 7, 8, 9, 10] refers to the signs and symptoms of AI dependence or addiction. Most people show a mix of attachment and dependence towards technology, especially in caring for devices and the need for constant connection. However, not everyone feels emotionally attached or dependent on technology, and many can use it without experiencing too much perceived anxiety or impairment.

It is important to mention that from the very first meetings, the team debated whether the first category, or the development of an excessively close bond with AI, should be called "attachment" or "dependence." Although the term "attachment" is associated with the attachment between two human beings (according to Ainsworth and Bowlby's initial attachment theory), finding articles in the literature detailing frames of reference and even degrees of "attachment to technology" and Artificial Intelligence contributed to defining the concept in this study as "AI Attachment."

The category of Attributions of Human Qualities [11] is an indicator of the confusion between humans and machines. This category is associated with the perception of Artificial Intelligence as a technology dependent on and controlled by humans, lacking the autonomous capacity to develop intentions, emotions, or judgments. It allows for an analysis of the existence or absence of a practical and controlled vision of AI, viewing it as a tool under human control, ruling out the possibility of the technology evolving to the point of posing human risks.

The category of Catastrophic Ideas about AI is associated with fears of a potential negative, even catastrophic, impact of AI on humanity. This relates to issues such as concerns about job replacement and the challenge of upskilling, concerns about data privacy and security, distrust of AI's creativity and emotional intelligence, and opinions about AI autonomy and human control.

2.2 The questionnaire

The 25 questionnaire items were originally written by the research team, ensuring their appropriateness for

measuring each category. The questionnaire was reviewed by five experts, professors of research methodology and instrument validation. A pilot test was then conducted with 30 students, whose opinions were taken into consideration, and any necessary adjustments were made before administering the survey.

The following are the items associated with the Level of Attachment to AI category:

- When I interact with an AI, I feel I can explore new ideas or questions without fear of judgment.
- I feel that AI is a reliable source of information when I need help.
- I frequently use AI to understand myself and grow as a person.
- In difficult times, I often seek the company of an AI to feel better.
- I believe that an AI can effectively provide emotional support.
- I feel I cannot complete a school assignment without first consulting an AI for advice.
- It causes me stress when I want to use an AI and cannot for some external reason.
- The thought that AIs might stop working or be banned causes me anxiety.
- I believe that if I tell an AI about my problems, it can respond empathetically.
- I believe that AI chats can foster bonds that users perceive as friendships.

The following are the items associated with the Attributions of Human Qualities category:

- When I face emotional difficulties, I prefer to share my problems with an AI rather than talking to friends or family.
- I enjoy spending time with an AI and actively seek out opportunities to interact with it.
- I would like to have an AI near me in social or stressful situations.
- I find comfort in having constant access to an AI, even when I don't need it at the moment.
- I believe that AI has the ability to learn and evolve autonomously, similar to humans.
- I believe that artificial intelligence can develop its own intentions without the need for human intervention or programming.
- I believe that AI expresses value judgments (opinions) based on the information it processes.
- In the future, machines will be like people, with their own ways of being and a place in society.
- As technology advances, machines should be granted certain labor rights.
- I tend to be kind to artificial intelligences and use words like "please" and "thank you" when interacting with them.

The following are the items associated with the category of Catastrophic Ideas about AI:

- I believe that society should limit the development of artificial intelligence to avoid the risk of it dominating humans in the future.
- I believe that in the future, machines will outnumber human workers in businesses and healthcare services.
- I think some AI applications could pose a security risk to humanity.
- I believe that artificial intelligence will allow humans to enjoy greater well-being and more free time by reducing their workload.
- I believe that AI will drive human evolution.

2.3 Description of the studies carried out

All the studies mentioned here were conducted at Anahuac Mayab University, in Merida, Yucatan, Mexico.

Starting in August 2022, a preliminary study was conducted with a group of students, asking them five questions that asked whether they "felt capable of establishing emotional relationships with a robot," whether they were afraid that robots would one day take over the world or take our jobs, among other questions. The results of this initial study showed clear indications that encouraged further research.

Later, and more formally, a research team made up of teachers and students began meetings with the aim of developing projects that would help gain a deeper understanding of these issues starting in August 2023. That year, the team designed a 10-item questionnaire and set out to evaluate categories such as "robophilia and robophobia index" and "level of erroneous beliefs about AI" among 100 students surveyed.

During this semester, from January to May 2025, the team refined and piloted the instrument and defined a meaningful sample. Considering the University's population of 8,500 students, the sample was established at 380 students surveyed. With the support of the Department of Educational Innovation, faculty members were selected to reserve the first 10 minutes of their classes for team members to administer the instrument, and the 380 surveys were completed for convenience. The team met to capture and obtain results, as well as present them to a group of faculty members and the vice-rector's office, with the goal of disseminating the findings presented here.

Table 1. Attachment dynamics

Level	Percentage
Low	49%
Medium	43%
High	8%

Table 2. Attribution of human qualities

Level	Percentage
Low	37%
Medium	54%
High	9%

3. Results and Discussion

After the first questionnaire was administered in 2023, it was found that 43% of the students surveyed had a severe level and 42% a moderate level of erroneous beliefs about AI.

In subsequent years, 2024 and 2025, the teams, with new students, focused these concepts more on "Catastrophic Beliefs" and "Attribution of Human Qualities to AI." And the instrument was refined to focus a greater number of items on the "Level of Attachment to AI" category. It was also found that 75% of the students surveyed did not have an extreme level, but 20.7% had extreme robophilia rates, and 4.3% had extreme robophobia rates in the sample studied.

This year, in May 2025, after reformulating the questionnaire, a total of 25 items were included, using a 5-point Likert scale, from "strongly disagree" to "strongly agree", 10 for the category "Level of attachment to AI", another 10 items for the category "Attributions of human qualities" and 5 items for a final category, "Catastrophic ideas about AI".

The overall results confirm that 8% of the total number of students surveyed have attachment levels considered extremely high, as can be seen in Table 1.

Likewise, in relation to the attribution of human qualities, 9% of the students surveyed have high levels, as can be seen in Table 2.

And 31% of participants showed a high rate of catastrophic beliefs about AI, as shown in Table 3.

4. Conclusion

In conclusion, AI is a very practical and widely used tool, but it's not enough to simply know how to use it; it's also necessary to clearly understand what it is and what it's used for. So-called AI is not intelligence itself; rather, the term refers to the construction of machines and programs that exhibit intelligent behavior. AI is a powerful tool that should always be used to contribute to a person's growth and fulfillment, never to limit, stunt, diminish, or enslave them. Specifically in the field of education, AI should not be used to prevent students from exercising their skills, nor as a substitute for the teacher's personal work as a fundamental part of supervision, counseling, and mentoring activities.

This article presents the results of several studies conducted over the past three years aimed at understanding levels of attachment to AI, as well as attribution of human qualities and the existence of catastrophic ideas about AI. The results of the studies show significant levels in each of these categories.

Future work includes conducting a more in-depth study of the student population, as well as developing and implementing measures aimed at promoting the healthy adoption of AI techniques and tools.

Ethics Statement

The authors acknowledge that ethical approval was not obtained for this study. The authors take full responsibility for any ethical considerations and are willing to cooperate with post-publication review if necessary.

CRediT authorship contribution statement

José J. Karam: Conceptualization, Methodology, Validation, Formal analysis, Investigation, Resources, Data curation, Writing – original draft, Writing – review & editing, Visualization, Supervision, and Project administration. **Gabriel Urzaiz:** Conceptualization, Writing – original draft, Writing – review & editing, and Supervision.

Declaration of Generative AI and AI-assisted technologies in the writing process

This manuscript was written without the assistance of generative AI tools. All content, including figures and text, was produced by the authors. The authors confirm that no generative AI or AI-assisted tools were used during the preparation of this article.

Declaration of competing interest

The authors declare no competing interests.

Acknowledgements

We express our deep gratitude to Mara Patricia Acuña Molina, Jeanet Castillo Mancera, Edgar Arturo Villorín Galaviz, José Ángel Chan Estrella and Emiliano Jesús Vázquez Castillo, all of them students of the Anahuac Mayab University. We recognize and appreciate their invaluable help with the planning, piloting and implementation of the instrument, as well as their support in capturing responses to obtain statistics on the results

Table 3. Catastrophic ideation

Level	Percentage
Without ideation	69%
With ideation	31%

References

[1] F. Kaya, F. Aydin, A. Schepman, P. Rodway, O. Yetişensoy, and M. Demir Kaya, "The roles of personality traits, AI anxiety, and demographic factors in attitudes toward artificial intelligence," *International Journal of Human–Computer Interaction*, vol. 40, no. 2, pp. 497–514, 2024. doi: 10.1080/10447318.2022.2151730 .

[2] J. McCarthy, M. Minsky, N. Rochester, and C. Shannon. (1955) A proposal for the dartmouth summer research project on artificial intelligence. [Online]. Available: http://jmc.stanford.edu/articles/dartmouth/dartmouth.pdf

[3] J. Haidt and V. Puertollano, *La generación ansiosa: por qué las redes sociales están causando una epidemia de enfermedades mentales entre nuestros jóvenes*, primera edición impresa en méxico ed. Ciudad de México: Paidós, 2024.

[4] Hopelab and Common Sense Media and NORC at the University of Chicago. (2024) A double-edged sword: How diverse communities of young people think about the multifaceted relationship between social media and mental health. [Online]. Available: https://www.commonsensemedia.org/research/double-edged-sword-how-diverse-communities-of-young-people-think-about-social-media-and-mental-health

[5] Hopelab and Common Sense Media and Harvard Graduate School of Education. (2024) Teen and young adult perspectives on generative AI: Patterns of use, excitements, and concerns. [Online]. Available: https://www.commonsensemedia.org/research/teen-and-young-adult-perspectives-on-generative-ai-patterns-of-use-excitements-and-concerns

[6] N. Rabb, T. Law, M. Chita-Tegmark *et al.*, "An attachment framework for human-robot interaction," *International Journal of Social Robotics*, vol. 14, pp. 539–559, 2022. doi: 10.1007/s12369-021-00802-9 .

[7] T. Xie and I. Pentina, "Attachment theory as a framework to understand relationships with social chatbots: A case study of replika," in *Proceedings of the 55th Hawaii International Conference on System Sciences*, 2022. doi: 10.24251/hicss.2022.258 .

[8] M. Dziergwa, M. Kaczmarek, P. Kaczmarek *et al.*, "Long-term cohabitation with a social robot: A case study of the influence of human attachment patterns," *International Journal of Social Robotics*, vol. 10, pp. 163–176, 2018. doi: 10.1007/s12369-017-0439-2 .

[9] S. Deng, J. Zhang, Z. Lin, and X. Li, "Service staff makes me nervous: Exploring the impact of insecure attachment on AI service preference," *Technological Forecasting and Social Change*, vol. 198, p. 122946, 2024. doi: 10.1016/j.techfore.2023.122946 .

[10] J. Guerreiro and S. M. Correia Loureiro, "I am attracted to my cool smart assistant! analyzing attachment-aversion in AI-human relationships," *Journal of Business Research*, vol. 161, p. 113863, 2023. doi: 10.1016/j.jbusres.2023.113863 .

[11] J. Kim, S. Kang, and J. Bae, "Human likeness and attachment effect on the perceived interactivity of AI speakers," *Journal of Business Research*, vol. 144, pp. 797–804, 2022. doi: 10.1016/j.jbusres.2022.02.047 .

Journal of Artificial Intelligence and Computing Applications (2025) - **Special Issue** - 3(2): 18–24

MAIKRON

Journal of Artificial Intelligence and Computing Applications

Expanded abstract

Detection of people with social phobia using pupillary position classification algorithms

Víctor Rangel-Fajardo[1], **David Cruz-Villavicencio**[1], **Erick Mpangi-Musungu**[1], **and Jesús Martínez-Castro**[1,*]

[1]Centro de Investigación en Computación, Instituto Politécnico Nacional

ABSTRACT

This study analyzes the visual attention of people with and without social phobia toward different images (photos of faces) depicting emotions such as normality, sadness, anger, and happiness. The analysis is divided into four main parts: The first part consists of identifying participants, both those with and without social phobia, using a psychometric instrument known as the BFNE-II. The second part focuses on the presentation of images composed of facial expressions, as well as the detection of the subjects' pupillary positions when observing these images. In the third part, the participants' pupillary positions are classified into characteristic points called fixations. Finally, in the fourth part of the study, these fixations are used as a measure to quantify the visual attention of people with and without social phobia, with the aim of identifying potential biases in both groups. Based on the analysis of visual attention to the different images, it was concluded that people with social phobia showed significantly greater visual attention to facial expressions of anger, compared to those without social phobia. On the other hand, non-phobic people tended to focus their attention primarily on images that depicted happy facial expressions.

Keywords: eye tracking, attention, social phobia

1. Introduction

Social phobia (SP) is defined as an unpleasant emotional state accompanied by somatic and psychological changes. It can present as an adaptive reaction or as a symptom associated with various medical and psychiatric conditions [1].

Social phobia is the second most common mental illness in Mexico, with a ratio of 3.6 women to every man. It is worth noting that only between 10% and 30% of all people affected by some type of phobic disorder seek help. Both social phobia and specific phobias manifest primarily during youth.

Various studies suggest that social phobia (social anxiety disorder) has a high comorbidity with other psychiatric disorders, especially depression, dysthymia, generalized anxiety disorder, and panic disorder. For clinical psychology professionals, it is essential to have valid and reliable instruments to detect this disorder. However, there is a lack of validated instruments in the clinical setting, due to the difficulty and financial resources involved in designing, evaluating, and validating them for different study populations.

For his part, Ressler [2] [3], in the Diagnostic and Statistical Manual of Mental Disorders (DSM-5) of the American Psychiatric Association (2013), established new standards for diagnosis and research on anxiety and related disorders. The DSM-5 classifies anxiety disor-

E-mail address: jamartinezc@ipn.mx
https://doi.org/10.5281/zenodo.17228364
 This article is part of the Special Issue on ICAIMH 2025. ISSN: 3061-8843

ders into eleven categories: 1) separation anxiety disorder, 2) selective mutism, 3) specific phobia, 4) social anxiety disorder (social phobia), 5) panic disorder, 6) agoraphobia, 7) generalized anxiety disorder, 8) substance- or medication-induced anxiety disorder, 9) anxiety disorders due to another medical condition, 10) other specified anxiety disorder, and 11) unspecified anxiety disorder.

Individuals with social anxiety show sustained attention biases towards socially threatening stimuli [4, 5, 6], so it is proposed that, by measuring the visual attention of individuals on different visual stimuli in the form of images, people with social phobia can be differentiated from those without social phobia.

This paper analyzes the visual attention of Mexican adults with and without social phobia by detecting their gaze when visual stimuli (in the form of an image) are presented to them on a monitor.

To identify individuals with social phobia, a psychometric instrument, the BFNE-II test, was used. This instrument consists of 12 questions in a 5-point Likert-type response format and measures a person's fear of being negatively evaluated by others.

2. Eye movements

This paper describes visual attention through two types of movements: voluntary movements of moving from one point to another (saccades) and gaze-holding movements (fixation). The movements that comprise a fixation are: tremor, drift, and microsaccade [7].

3. Pupillary positions, fixations and classification algorithms

The classification of pupillary positions into fixations is the grouping into n sets of the pupillary positions in such a way that they statistically describe the physiological behaviors of eye movements [8]. This process significantly reduces the size of the pupillary positions to representative tuples (fixations or characteristic points), where each tuple is composed of the arithmetic mean of the ocular positions in the group, the sum of the duration of each element, and the number of elements that make up the group.

Classification is a fundamental procedure in any analysis of eye movements, as poor classification can result in tuples that do not represent gaze behavior [9, 10].

The algorithms for classifying pupillary positions into fixations must evaluate statistical characteristics and physical magnitudes of the pupillary position samples, such as point-to-point velocity and/or acceleration of the samples, as well as data dispersion, among others. On the other hand, the physiological characteristics of eye movement behavior must be considered, such as the minimum duration time of a fixation, the minimum duration time of a saccade, and the maximum and minimum amplitude of a repositioning movement, among others. To achieve this, identification and classification algorithms for pupillary positions in fixations are used [11]. Some of the algorithms used are: dispersion-threshold identification algorithm (I-DT), velocity-threshold identification algorithm (I-VT), hidden Markov model identification algorithm (I-HMM), and the adaptive algorithm.

4. Materials and methods

There are different eye-tracking methods, which can be grouped into two major categories: invasive and non-invasive. The first category obtains the pupillary position directly by incorporating special devices into the eyeball, which implies physical interference with the subject and is therefore considered invasive. The second category estimates the pupil position indirectly, that is, it uses different techniques to estimate the position (one of them is through image analysis and known reference points), and thus does not physically interact with the subject (non-invasive). Invasive methods offer high precision and accuracy in measurement and are capable of detecting very small movements, such as tremors. However, they interfere with the natural characteristics of the subject's movements (for example, they reduce movement speed), in addition to causing discomfort. In contrast, indirect methods present lower precision and accuracy in measurement, but the movements obtained can be considered natural. Some non-invasive eye-tracking methods include the electrooculogram, the search coil, and infrared reflection. For this study, the infrared reflection method was chosen.

In [12], the twelve leading companies in the commercialization of eye-tracking products are presented, ordered by the number of publications from January 2016 to January 2017. For this work, the sensor *The Eye Tribe* was selected. For data capture, a communication module was developed between the sensor and the computer, capable of transmitting information at 30 and 60 Hz.

To calibrate the sensor, the subject is asked to look at nine points on the screen (see Figure 1), while the position of the pupil and the corneal reflection are determined at a given moment in time [13]. Current literature proposes various error metrics [14]. For a monitor with a resolution of 1280×1024 pixels, we use a value of ε between 30 and 60 pixels.

5. Procedure

The test begins with the *"Calibration"* operation, in which the subject is instructed to observe the reference points. The subject is asked to avoid moving their face during the test to prevent loss of sensitivity and high variance in the sample. Afterwards, they are informed that 10 images containing 4 faces each will be presented, and they should observe the face that draws their attention the most *"Visual Stimuli Presentation"*.

The ten *test images* are presented automatically, so the subject does not need to perform any additional action beyond viewing the stimuli. At the end of the presentation of the *test images*, a message will appear

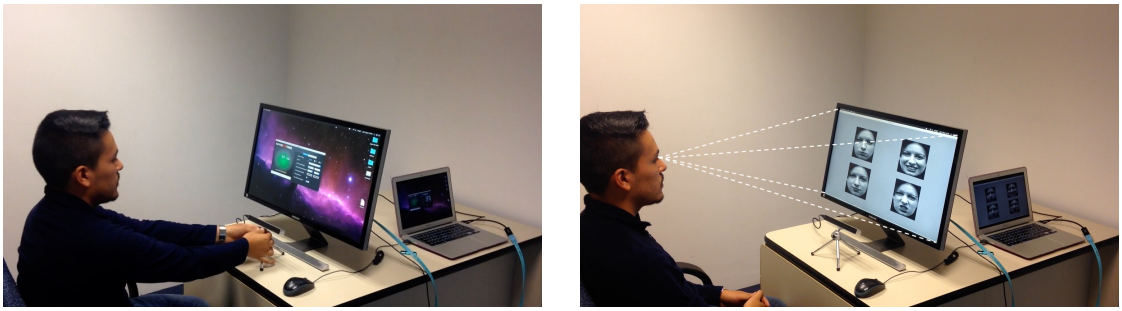

Figure 1. Experimental setup during its calibration stage. A point is randomly selected and viewed for one second; after this time, another point is selected. The process is repeated until all nine calibration points are displayed.

on the screen indicating the end of the test. At that moment, the subject may leave the testing room. The duration of the test ranges between four and ten minutes, depending on the calibration process.

The test images were taken from a sample of 145 undergraduate students, with an average age of 21.3 years. Of them, 62.7% were women. The image capture was carried out under controlled lighting and sound conditions.

The selected visual stimuli are based on work similar to the studies conducted by Paul Ekman on basic emotions [15, 16].

On the other hand, n_i images were presented, each composed of k different stimuli. The images were shown to individuals with social phobia (SP) and without social phobia (No-SP). During a time period t_i, the eye movements of the individuals with SP and No-SP were recorded as the n_i images were presented to them.

As part of the proposed hypothesis, it is suggested that individuals with social phobia exhibit sustained attention biases toward stimuli perceived as threatening. The stimuli that have been shown to be effective in tests with individuals with social phobia are faces with different facial expressions, as described in [4].

In this work, the image set of facial expressions "AU-Coded Facial Expression Database," developed by Cohn-Kanade [17], was used. This repository contains 2000 images of 200 individuals aged between 18 and 30 years. The image set is composed of different facial expressions (neutral, joy (happiness), sadness, anger, surprise, fear, and disgust) categorized according to the "FACS" standard [18]. From the Cohn-Kanade image set, ten faces (five women and five men) were used. An oval crop was applied to each face in order to suppress fixations centered on protruding elements of the stimulus (hair, earrings, ears, etc.). The goal was to ensure that the patients' attention was focused on the facial expression and not on distracting elements.

With the selected stimuli, ten images were composed and presented to the subjects. Each test image was composed of four stimuli (neutral, joy, sadness, and anger) arranged in a 2×2 grid (see Figure 2), where each cell is randomly assigned a stimulus.

6. Results and discussion

The expression of anger was the most focused on by individuals with SP, while the expression of joy was the most focused on by individuals without SP. The remaining expressions (neutral and sadness) are included to obtain better discrimination in attention levels between individuals with SP and No-SP.

Cutoff points (total score on the Likert-type test from which a subject is considered to present symptoms of social phobia) on fear scales vary depending on the population and social environment. In Mexico, in the state of Michoacán de Ocampo, Marcelo A. et al. *et al.* [19] validated the instrument (BFNE-II test) for the adult population of Michoacán, finding a cutoff point of 42 with a sensitivity of 94% for having social phobia. During this research, a version of the original BFNE-II test was applied and evaluated based on the cutoff point published by Marcelo A. et al. The results are described in [12].

The conventional analysis of eye movements was carried out through the study of fixations (pauses in regions of interest) or saccadic movements (rapid movements between fixations). To this end, the four algorithms described in Section 3 were evaluated with the aim of determining the optimal one for our objective. The four algorithms consider velocity as the main discriminative feature, so the calculation of velocities between samples must be unified for each one. Likewise, it is necessary to calculate noise-free velocities; for this purpose, the equations proposed by *Engbert and Kliegl* [20] are used, where the calculation is performed on the x and y coordinates for the left and right eyes independently.

One parameter of the pupillary position algorithms is the velocity threshold, from which it is determined whether a position $(x_i,\ y_i)$ with velocity $\dot{v}_i$ belongs to a fixation group or not. This threshold varies depending on the type of test and the sampling rate. The specialized literature proposes a threshold of 1.5 px/ms [8], but does not specify the type of experiment. In this work, a velocity threshold is calculated for the actual experimental conditions at a sampling frequency of 60 Hz. For this, the mean velocities per test image were ana-

Figure 2. The images represent four stimuli (neutral, joy, sadness, and anger), among which the subject had to primarily select one.

lyzed for each person. That is, each person contributes ten mean velocity values to the analysis, where each mean velocity describes the subject's general behavior during the presentation of a test image.

In Figure 3, the behavior of mean velocities for 145 subjects is presented. The data show a positively skewed distribution with a heavy tail, suggesting a high concentration of velocities close to zero. According to theory, low velocities are due to fixation movements (microsaccades), while high velocities correspond to repositioning movements (saccades). Once the velocity threshold parameter ε is determined, the classification algorithms for fixations are tested and evaluated.

The decision criteria for determining the optimal algorithm among I-VT, I-DT, I-HMM, and the Adaptive algorithm are described in [12].

6.1 Attention by stimulus

Using the attention tuples T_{At}, the attention time for each stimulus was measured, distinguishing between individuals with SP and without SP. The attention times per stimulus of the subjects with SP and without SP are compared using quartiles to generate the boxplots shown in Figure 4.

While individuals with SP focus their attention on the threatening stimulus, individuals with No-SP direct their attention to the happiness (joy) stimulus.

Boxplots are an efficient comparison method that allows subtle trends in the data to be observed; however, they do not allow a trend to be validated conclusively. For this purpose, a normality analysis was conducted (see [12]).

Finally, quantile-quantile (Q-Q) plots are presented with the aim of comparing the actual distribution with a theoretical normal distribution. In each plot, two distributions are compared, where each corresponds to the attention time of a group of individuals (SP and No-SP) for the same stimulus k_i.

From Figure 5, it is possible to state that a significant difference does exist: in the case of the Joy (happiness) stimulus, individuals with No-SP exhibit longer attention times compared to individuals with SP (see [12]), which suggests that the hypothesis of the study was confirmed for the study population.

7. Conclusions

This research presents results that could be useful for the eventual confirmation of the hypothesis proposed at the beginning of this work for the individuals examined.

It was verified through statistical methods that the level of attention in individuals with SP is higher for stimuli considered threatening (Anger), where the mean attention time was 1.6s, compared to individuals without SP, whose mean attention time for the threatening stimulus was 1.05s. A similar trend was observed when comparing the opposite stimulus (Happiness), where individuals without SP recorded a mean of 1.88s, while individuals with SP recorded a mean attention time of 1.26s.

The Neutral and Sadness stimuli did not show significant changes; the same level of attention was maintained for both types of individuals analyzed. A possible cause of this is the similarity that exists in the facial expressions of these stimuli.

It is important to mention an issue regarding the experimental design. Initially, the test was designed to display 30 test images for 10 seconds each. After the first trials, we received feedback indicating how exhausting this procedure was, since the observer had to avoid moving their face or body during the entire test (5 minutes). As a result of this overexposure to the images, the quality of the fixations detected in the final test images was poor. The main reason was that most individuals changed their posture after a certain number of images had been shown. This change in posture affected sensor detection and, consequently, the measurement. The final experiment was designed to use 10 test images, which equates to a stimulus exposure of 1.66 minutes. With this number of stimuli, uniform detection was achieved throughout the entire test.

In this research, visual attention was quantified as the time a subject devotes to an area of interest. An-

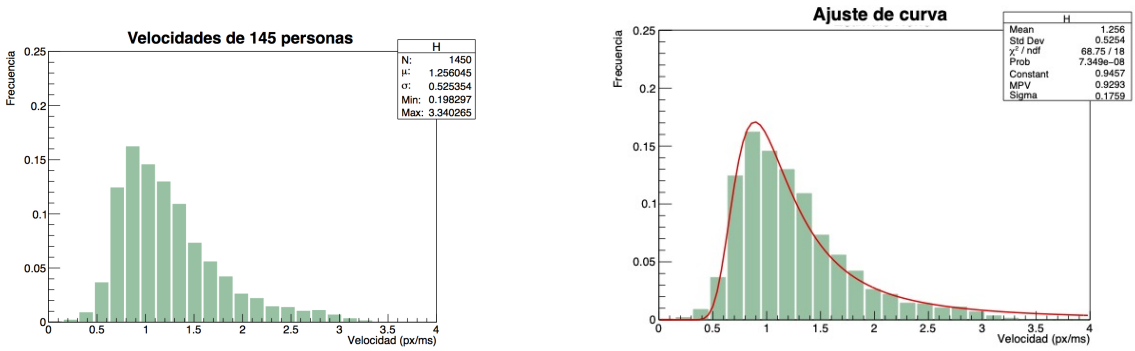

Figure 3. Velocity distribution and its fitted curve.

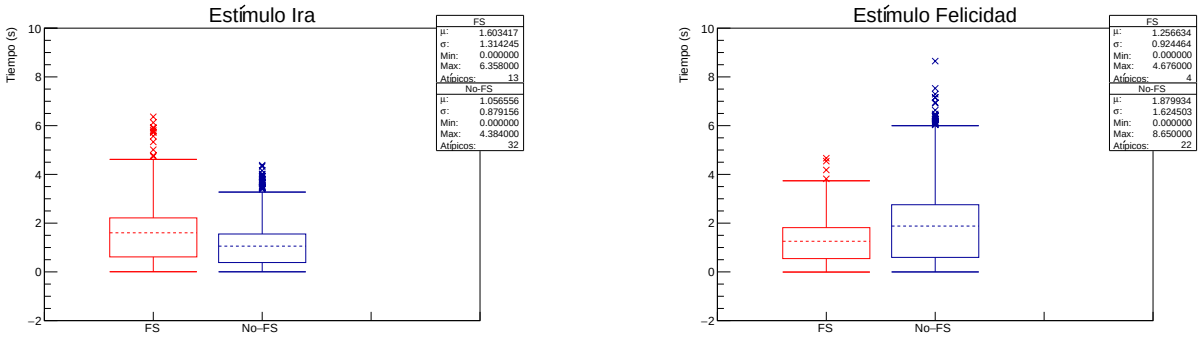

Figure 4. Boxplot for the anger and happiness stimuli.

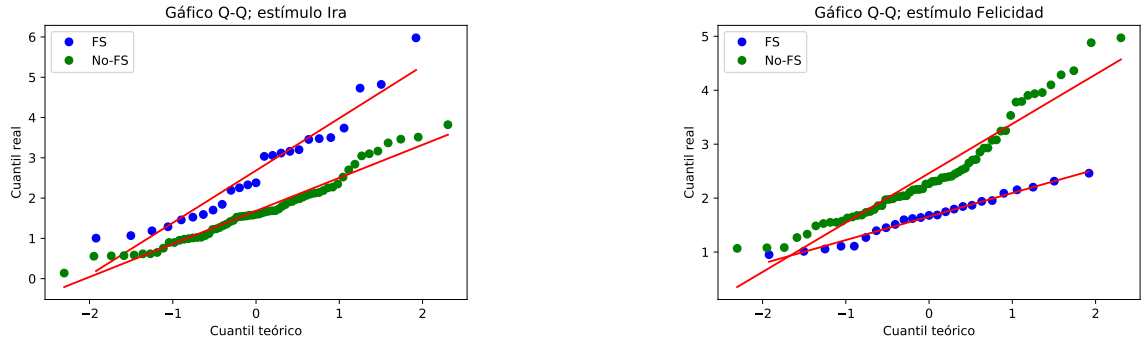

Figure 5. Q-Q plot for the anger and happiness stimuli.

other contribution of this work was the evaluation of the quality of the fixations detected by each evaluated algorithm, in the absence of a reference standard. The main proposed measure was the calculation of the nearest neighbor index, which assesses the dispersion of the data within a fixation.

Ethics Statement

This study used anonymized data. No personally identifiable information was collected. Ethical approval was deemed unnecessary according to the guidelines of Instituto Politénico Nacional.

CRediT authorship contribution statement

Víctor Rangel-Fajardo: Conceptualization, Methodology, Software, Investigation, Validation, Resources, Formal analysis, Project administration, Data curation. **David Cruz-Villavicencio:** Software, Writing - review & Validation. **Erick Mpangi-Musungu:** Writing - review & editing. **Jesús Martínez-Castro:** Conceptualization, Methodology, Formal analysis, Investigation, Writing original draft, Writing - review & editing, Visualization, Supervision, Project administration.

Declaration of Generative AI and AI-assisted technologies in the writing process

The authors utilized Grammarly and ChatGPT to refine sentence structure and enhance readability. No content was generated by AI; all scientific insights and original ideas are the authors' own.

Declaration of competing interest

The authors declare no competing interests.

Acknowledgements

The authors thank the IPN and Secihti for their support in completing this work.

References

[1] American Psychiatric Association, *Diagnostic and Statistical Manual of Mental Disorders: DSM-5-TR*. Washington, DC: American Psychiatric Association Publishing, 2022. doi: 10.1176/appi.books.9780890425787. [Online]. Available: https://www.psychiatry.org/psychiatrists/practice/dsm

[2] K. Ressler, *Anxiety Disorders*. Oxford: Oxford University Press, Incorporated, 2015. [Online]. Available: http://ebookcentral.proquest.com/lib/anahuac-ebooks/detail.action?docID=2012685

[3] A. Moukheiber, G. Rautureau, F. Perez-Diaz, R. Soussignan, S. Dubal, R. Jouvent, and A. Pelissolo, "Gaze avoidance in social phobia: Objective measure and correlates," *Behaviour Research and Therapy*, vol. 48, no. 2, pp. 147–151, 2010. doi: 10.1016/j.brat.2009.09.012. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/19863948/

[4] T. Bantin, S. Stevens, A. L. Gerlach *et al.*, "What does the facial dot-probe task tell us about attentional processes in social anxiety? a systematic review," *Journal of Behavior Therapy and Experimental Psychiatry*, vol. 50, pp. 40–51, 2016. doi: 10.1016/j.jbtep.2015.04.009.

[5] C. A. Schofield, A. L. Johnson, A. W. Inhoff, and M. E. Coles, "Social anxiety and difficulty disengaging threat: Evidence from eye-tracking," *Cognition and Emotion*, vol. 26, no. 2, pp. 300–311, 2012. doi: 10.1080/02699931.2011.602050. [Online]. Available: https://www.tandfonline.com/doi/abs/10.1080/02699931.2011.602050

[6] N. K. Sahu, N. S. Harshit, R. Uikey, and H. R. Lone, "Beyond questionnaires: Video analysis for social anxiety detection," 2024. [Online]. Available: https://arxiv.org/abs/2501.05461

[7] S. Martinez-Conde, S. L. Macknik, and D. H. Hubel, "The role of fixational eye movements in visual perception," *Nature Reviews Neuroscience*, vol. 5, no. 3, pp. 229–240, 2004. doi: 10.1038/nrn1348.

[8] D. Salvucci and J. Goldberg, "Identifying fixations and saccades in eye-tracking protocols," in *Proceedings of the Eye Tracking Research and Applications Symposium*, Jan 2000, pp. 71–78. doi: 10.1145/355017.355028.

[9] A. Lazarov, D. Basel, S. Dolan, D. G. Dillon, D. A. Pizzagalli, and F. R. Schneier, "Increased attention allocation to socially threatening faces in social anxiety disorder: A replication study," *Journal of Affective Disorders*, vol. 290, pp. 169–177, 2021. doi: 10.1016/j.jad.2021.04.063. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0165032721003955

[10] L. R. Lidle and J. Schmitz, "Assessing visual avoidance of faces during real-life social stress in children with social anxiety disorder: A mobile eye-tracking study," *Child Psychiatry & Human Development*, vol. 55, no. 1, pp. 24–35, Feb 2024. doi: 10.1007/s10578-022-01383-y. [Online]. Available: https://doi.org/10.1007/s10578-022-01383-y

[11] R. Andersson, L. Larsson, K. Holmqvist *et al.*, "One algorithm to rule them all? an evaluation and discussion of ten eye movement event-detection algorithms," *Behavior Research Methods*, vol. 49, no. 2, pp. 616–637, Apr 2017. doi: 10.3758/s13428-016-0738-9. [Online]. Available: https://doi.org/10.3758/s13428-016-0738-9

[12] V. Rangel-Fajardo, "Análisis de la fobia social mediante la clasificación de las posiciones pupilares ante estímulos visuales," Tesis de Maestría, Centro de Investigación en Computación, Ciudad de México, México, Diciembre 2018.

[13] A. T. Duchowski, *Eye Tracking Methodology: Theory and Practice*, 3rd ed. London: Springer, 2007.

[14] S. A. Johansen, J. San Agustin, H. Skovsgaard, J. P. Hansen, and M. Tall, "Low cost vs. high-end eye tracking for usability testing," in *CHI'11 Extended Abstracts on Human Factors in Computing Systems*. ACM, 2011, pp. 1177–1182. doi: 10.1145/1979742.1979744 .

[15] P. Ekman, "An argument for basic emotions," *Cognition & Emotion*, vol. 6, no. 3-4, pp. 169–200, 1992. doi: 10.1080/02699939208411068 .

[16] ——, "Basic emotions," *Handbook of Cognition and Emotion*, pp. 45–60, 1999. doi: 10.1002/0470013494.ch3 .

[17] J. F. Cohn, T. Kanade, and Y. Tian, "Au-coded facial expression database," Digital Repository, 2000, obtenido de http://www.pitt.edu/˜emotion/ck-spread.htm.

[18] P. Ekman, "Facial action coding system (facs)," *A Human Face*, 2002.

[19] M. A. Bravo, F. González Betanzos, Castillo Navarro, and F. Padrós Blázquez, "Escala de miedo a la evaluación negativa versión breve (bfne): propiedades psicométricas en muestras clínicas y de universitarios en méxico," *Acta Colombiana de Psicología*, vol. 18, no. 1, pp. 69–77, 2015. doi: 10.14718/ACP.2015.18.1.4 .

[20] R. Engbert and R. Kliegl, "Microsaccades uncover the orientation of covert attention," *Vision Research*, vol. 43, no. 9, pp. 1035–1045, 2003. doi: 10.1016/S0042-6989(03)00084-1 .

Journal of Artificial Intelligence and Computing Applications (2025) - **Special Issue** - 3(2): 25–32

MAIKRON

Journal of Artificial Intelligence and Computing Applications

Expanded abstract

Automated detection of academic anxiety levels: proposal of a machine learning-based expert system for educational and clinical settings

José Manuel Sánchez Sordo[1]

[1]Universidad Nacional Autónoma de México

ABSTRACT

Public speaking anxiety is a significant barrier to academic performance and social integration among university students. Despite its prevalence, research on targeted technological solutions remains limited. This study presents the development of an expert system designed to automatically classify academic anxiety levels related to oral expression using a supervised machine learning approach. A sample of 159 undergraduate students completed a 30-item psychometric questionnaire assessing cognitive dimensions associated with public speaking anxiety. Participants' scores were standardized into Z-scores and categorized into three anxiety levels: Low, Normal, and High. The Naive Bayes algorithm was employed due to its simplicity, efficiency, and interpretability. Model parameters were estimated using Python 3.11 and the scikit-learn library, with performance assessed via stratified 10-fold cross-validation, achieving an overall accuracy of 95.6% and an AUC above 0.99 across all categories. The system was implemented in an interactive Streamlit application, linked to Airtable for automatic data storage. Results indicate that the expert system reliably detects characteristic response patterns associated with each anxiety level, offering immediate, personalized clinical interpretations to users. This study demonstrates the potential of integrating classical psychometric methods with machine learning to develop accessible, explainable, and ethically responsible digital tools for mental health assessment and early intervention in educational settings.

Keywords: expert systems, social anxiety, machine learning

1. Introduction

Public speaking anxiety is a significant issue that impacts the teaching-learning process. It is among the most common fears in the general population and a prominent manifestation in individuals with social anxiety. According to Blöte et al. [1], public speaking anxiety can be classified as a specific and clinically relevant subtype of social anxiety, which is defined as intense fear or anxiety in situations where one may be subject to social evaluation [2]. Among university students, 20–30% report considerable difficulties speaking in public [3], and over 75% have avoided such situations at some point in their academic lives [4], with oral presentations being among the most stressful academic experiences [5]. Numerous studies have linked this form of anxiety to factors such as poor academic performance [6], explicit social evaluation, and high levels of self-demand [7]. At the cognitive level, individuals with social anxiety often overestimate social demands and underestimate their ability to cope with them [8]. They

E-mail address: jose.sordo@iztacala.unam.mx
https://doi.org/10.5281/zenodo.17228680
 This article is part of the Special Issue on ICAIMH 2025. ISSN: 3061-8843

also tend to excessively focus their attention on their performance and process information in a biased way that confirms their fears [9]. These difficulties can lead to reduced classroom participation [10] and, in some cases, academic dropout [11], negatively affecting students' emotional well-being, academic identity, and social integration [12, 13]. Despite its relevance, research specifically addressing this type of anxiety in university settings remains limited [14].

From an applied perspective, various psychotherapeutic approaches have proven effective in addressing this problem, including graduated exposure, cognitive restructuring, and social skills training [15, 16]. However, the integration of emerging technologies such as artificial intelligence (AI) opens new possibilities for the diagnosis, monitoring, and early intervention in mental health [17]. In light of this need, it becomes pertinent to explore innovative approaches to assess and classify this form of anxiety, including AI-based technological solutions. In particular, machine learning techniques have proven useful for predicting psychological conditions using structured data [18, 17]. Supervised machine learning is a branch of AI in which an algorithm learns to predict a category or value from a previously labeled dataset. This involves training the model on examples with known outcomes (classes or labels), so it can later classify new instances based on the learned patterns [19]. This methodology has shown considerable promise in clinical and educational psychology, particularly for identifying complex patterns within large datasets derived from questionnaires, scales, or records [20].

One of the most widely used supervised algorithms for classification tasks due to its simplicity, efficiency, and performance is Naive Bayes. This model is based on Bayes' theorem and assumes conditional independence among the predictor variables—meaning each variable contributes independently to the probability of a given class [21]. While this assumption is a simplification, Naive Bayes has proven effective in practice even when predictors are moderately correlated. Its probabilistic interpretation supports the development of transparent and interpretable expert systems suitable for educational and clinical applications.

The present study aims to present the development and implementation of an expert system for the automatic classification of academic anxiety related to oral expression, based on a Naive Bayes model trained on response patterns to 30 valid and reliable items assessing cognitive variables associated with public speaking anxiety. Through the recognition of psychological data patterns, the system classifies users into anxiety categories (low, normal, or high) based on their responses to specific items in a digital questionnaire. The system interface was programmed in Python 3, implemented using the Streamlit framework, and connected to an Airtable database for automatic result storage.

The use of machine learning algorithms in psychology goes beyond classical approaches based on normative scores or z-transformations, offering more dynamic, tailored, and explanatory alternatives. Such models not only enhance diagnostic accuracy but also support the design of interactive, user-adapted evaluation and intervention systems [18]. In this context, the present study contributes to the growing field of machine learning applications in mental health data analysis, emphasizing the utility of expert systems as support tools in clinical, educational, and psychological research settings.

2. Methodology

This study is based on a sample of 159 university students from various undergraduate programs of the National Autonomous University of Mexico, obtained through non-probabilistic convenience sampling. Participation was voluntary and anonymous, carried out through online forms. Participants completed a 30-item psychometric questionnaire assessing academic anxiety related to oral expression. The instrument demonstrated strong psychometric properties (CVI = .97, α = .90; [22]). Each item was rated on a Likert scale ranging from 0 (never) to 5 (always), allowing for significant variability in responses.

Subsequently, each participant's raw scores were transformed into Z-scores. Based on these standardized values, normative thresholds were established to classify participants into three levels of anxiety: Low ($z < -1$), Normal (z between -1 and 1), and High ($z > +1$). These categories served as supervised labels for training the machine learning model. This procedure provided a psychometric criterion for defining the classes used by the expert system.

2.1 Applied AI model

For the development of the expert system, the Naive Bayes algorithm was selected—a classical supervised machine learning technique that offers advantages such as fast training, interpretability, robustness with small samples, and performance comparable to that of more complex models [21, 19]. This algorithm is based on Bayes' theorem and assumes conditional independence among predictor variables, thereby simplifying the calculation of class membership probabilities.

The model parameters (mean and standard deviation per class for each item) were estimated using Python 3.11 with the scikit-learn library, employing the Gaussian Naive Bayes function and stratified 10-fold cross-validation. These parameters were subsequently exported and integrated into an interactive application developed in Python 3.11, using Streamlit as the visual interface and Airtable for data and results storage.

The system calculates, for each user, the log-likelihood of membership in each class (High, Normal, Low) and assigns the classification according to the class with the highest logarithmic value. This classification is presented alongside an automated and personalized clinical interpretation, displayed within the user interface.

2.2 Experimental setup and evaluation metrics

A training set of 159 labeled instances was constructed, with Z-score-based norms serving as the target classes. The model's performance was assessed using stratified 10-fold cross-validation, yielding an overall accuracy of 95.6%, a Kappa coefficient of 0.934, and F-measure values exceeding 0.94 for all categories. These results suggest a high level of effectiveness in distinguishing between different levels of academic anxiety.

In addition, individual metrics for each class were reported:

- High Anxiety: Precision = 0.981, Recall = 1.00, F-Measure = 0.991
- Normal Anxiety: Precision = 0.979, Recall = 0.887, F-Measure = 0.931
- Low Anxiety: Precision = 0.912, Recall = 0.981, F-Measure = 0.945

These findings indicate that the expert system is particularly effective in detecting cases of high and low anxiety, while maintaining sufficient precision for normal levels. Furthermore, the practical implementation in Streamlit allowed for testing the application with real users and verifying its functionality as a digital diagnostic tool.

3. Results and discussion

The developed expert system demonstrated a high level of accuracy in the automatic classification of academic anxiety related to oral expression. Through the implementation of the Naive Bayes algorithm, users were successfully classified into three distinct levels—High, Normal, and Low—based on their responses to 30 psychometric items.

The 10-fold cross-validation performed during the modeling phase yielded an overall accuracy of 95.6%. The model exhibited areas under the ROC curve (AUC) close to 0.99 across all categories, indicating outstanding predictive performance.

3.1 Classification rules

The Naive Bayes algorithm used infers the most probable class for each user by calculating the joint probability of their responses in specific items under the estimated normal distributions for each class. Broadly speaking:

- Users with high response averages (between 3 and 5) on items related to fear of public speaking and oral expression tend to be classified as "High."
- Users with moderate response averages (around 2) tend to be classified as "Normal."
- Users with very low response averages (between 0 and 1) are classified as "Low."

Table 1 presents a summary of the central tendency (means) and dispersion (standard deviations) for each class, based on a representative subset of psychometric items.

These systematic differences enable the expert system to reliably detect characteristic patterns associated with each level of anxiety. For the complete list of items and their descriptive statistics, see Table A1.

3.2 Clinical interpretation presented to the user

After classification, the system provides an automated interpretation based on the user's profile:

- High: "Your profile suggests a high level of academic anxiety associated with oral expression. It is likely that you frequently experience fear of being judged by others, even when you are well-prepared, and that you tend to avoid participating or presenting due to fear of ridicule or disapproval."
- Normal: "Your profile indicates a moderate or normal level of academic anxiety. You may experience occasional doubts or insecurities in oral academic contexts, but they do not significantly interfere with your performance."
- Low: "Your profile reflects a low level of academic anxiety in oral expression contexts. You are likely to feel comfortable participating, presenting, or engaging in class discussions."

This feedback was designed to be accessible and useful, enabling immediate understanding by the user without requiring technical knowledge of psychometrics or clinical psychology.

The results obtained in the present study indicate that the supervised machine learning approach is effective in detecting latent patterns within psychological data. Despite its simplicity, the Naive Bayes algorithm can achieve performance comparable to, and in some cases surpassing, that of more complex methods reported in prior research on psychological classification [18, 17]. Among the main advantages observed, the system's high interpretability stands out, as the probabilistic structure of the model facilitates a detailed explanation of each classification decision, which is particularly valuable in clinical and educational settings. Furthermore, the system exhibited scalability, as it can be readily applied to larger datasets while maintaining its accuracy and efficiency. In terms of clinical applicability, the system provides immediate user profiles that can serve as complementary tools for psychological diagnosis or educational intervention.

However, some inherent limitations of the model were also identified. One of the primary limitations is that the assumption of conditional independence among items—an intrinsic feature of the Naive Bayes algorithm—is not always strictly met in real psychological responses, which may introduce certain distortions in

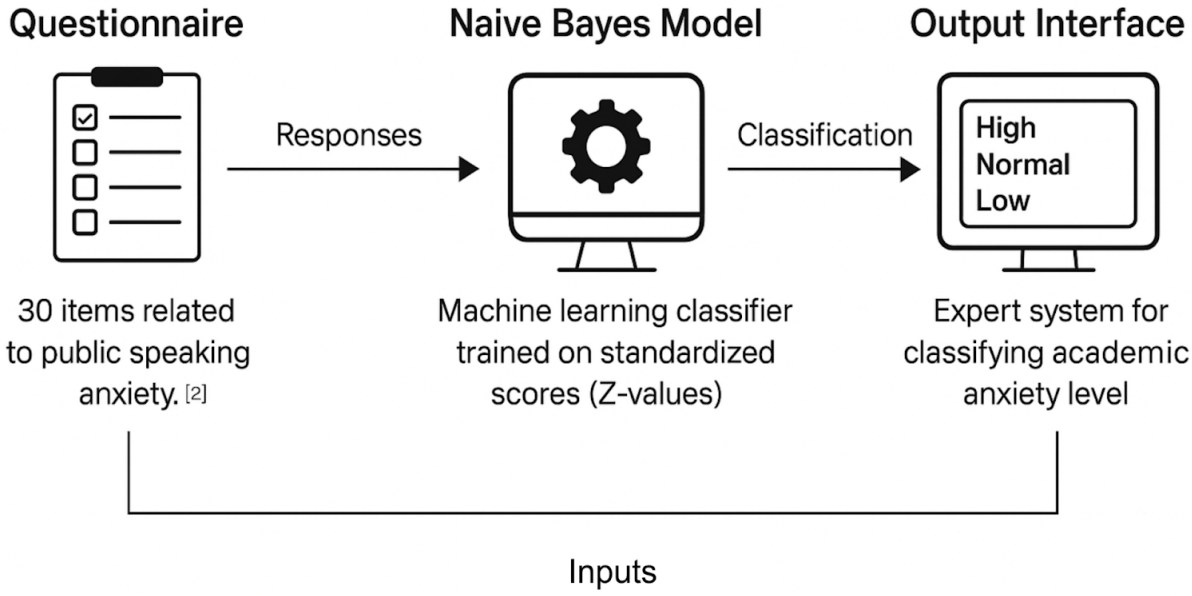

Figure 1. System architecture diagram. The diagram illustrates the operational flow of the system: users complete a digital questionnaire, whose values (inputs) feed into a Naive Bayes classification model previously trained on standardized psychological data. The model calculates the probability of belonging to each category (High, Normal, or Low) and generates an immediate interpretative output, including a diagnostic label and a descriptive profile. The data are automatically stored in an Airtable database for further analysis. This modular design allows for flexible integration in clinical or educational contexts.

Table 1. Representative Psychometric Items by Anxiety Level

Summarized Item	**High** ($\mu \pm \sigma$)	**Normal** ($\mu \pm \sigma$)	**Low** ($\mu \pm \sigma$)
Fear of being ridiculed during a presentation	3.28 ± 0.73	1.39 ± 0.91	0.09 ± 0.35
Fear of answering questions	3.79 ± 0.45	1.52 ± 0.96	0.37 ± 0.65
Thinking that one's question is foolish	3.35 ± 0.70	1.21 ± 1.01	0.11 ± 0.31
Belief in one's inability to present	3.35 ± 0.70	1.43 ± 1.00	0.24 ± 0.43
Arguing within a group	3.58 ± 0.52	1.88 ± 0.92	0.69 ± 1.20

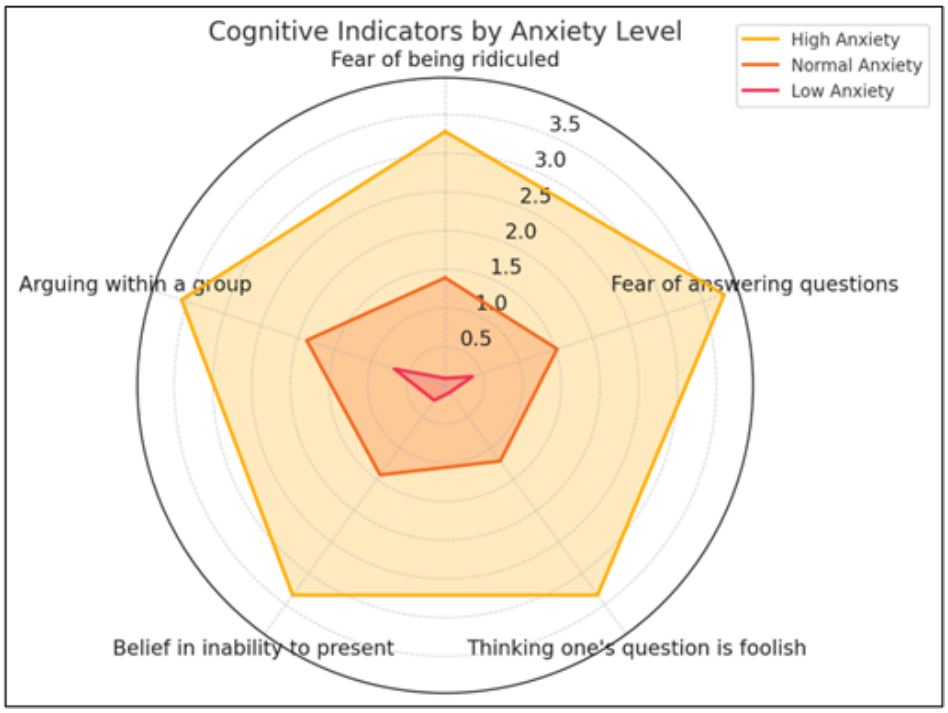

Figure 2. Radar chart displaying the average scores of five key cognitive indicators associated with academic social anxiety, differentiated by anxiety level (High, Normal, and Low). A clear pattern of higher intensity in dysfunctional thoughts is observed in the High Anxiety group.

specific cases. Additionally, although the sample used was suitable for developing a functional prototype, it would be necessary to expand and diversify the sample in future studies to achieve more robust and generalizable validations across different population contexts.

4. Conclusion

The expert system developed in this study represents a successful convergence of classical psychometric principles and contemporary machine learning methodologies, illustrating that data-driven models can be effectively adapted to create accessible, interpretable, and ethically responsible solutions within the domains of mental health and higher education. Leveraging the Naive Bayes algorithm, the system enabled the automatic, transparent, and real-time classification of academic anxiety levels related to oral expression, delivering immediate, evidence-based feedback grounded in validated psychological constructs.

The principal contributions of this work include: the design of a novel hybrid evaluation model that seamlessly integrates classical psychometric frameworks, grounded in Z-score norms, with the predictive capabilities of supervised machine learning techniques, enabling pattern recognition within psychological datasets; and the development of an accessible and open-source platform (utilizing Streamlit and Airtable) that promotes its applicability in both clinical and educational contexts. In addition, the system's ability to generate user-friendly, personalized interpretations enhances its potential as a complementary tool for early diagnosis, psychoeducational intervention, and longitudinal monitoring of academic anxiety.

Nevertheless, several challenges and avenues for future research remain. To enhance the external validity and generalizability of the model, it is imperative to expand the dataset to encompass more diverse and representative university populations. Moreover, future studies could explore methodological alternatives that relax the conditional independence assumption inherent in Naive Bayes, employing more sophisticated algorithms such as Random Forests, neural networks, or deep learning architectures, while maintaining a critical balance between predictive accuracy and model interpretability. Incorporating capabilities for longitudinal tracking of anxiety profiles and adaptive feedback mechanisms would further strengthen the system's clinical and educational utility. Finally, the integration of automated intervention modules, offering personalized recommendations based on users' profiles, represents a promising direction for extending the system's impact.

In sum, this study opens new pathways for the ethical, efficient, and scientifically grounded application of artificial intelligence in psychological assessment, contributing meaningfully to the development of innovative, user-centered tools for promoting mental health and academic well-being in educational environments. Overall, this research highlights the value of explainable

AI in digital mental health tools and lays the groundwork for scalable interventions that support emotional resilience in academic population.

Ethics Statement

This study used anonymized data collected via an online survey. No personally identifiable information was collected. Ethical approval was deemed unnecessary according to the guidelines of UNAM.

CRediT authorship contribution statement

José Manuel Sánchez Sordo: Conceptualization, Methodology, Software, Validation, Formal analysis, Data curation, Visualization, Writing – original draft, Writing – review & editing.

Declaration of Generative AI and AI-assisted technologies in the writing process

The authors utilized ChatGPT to refine sentence structure and enhance readability. No content was generated by AI; all scientific insights and original ideas are the authors' own.

Declaration of competing interest

The authors declare no competing interests.

References

[1] A. W. Blöte, M. J. W. Kint, A. C. Miers, and P. M. Westenberg, "The relation between public speaking anxiety and social anxiety: A review," *Journal of Anxiety Disorders*, vol. 23, no. 3, pp. 305–313, 2009. doi: 10.1016/j.janxdis.2008.11.007 .

[2] American Psychiatric Association, *Diagnostic and Statistical Manual of Mental Disorders*, 5th ed. Arlington, VA: American Psychiatric Publishing, 2013. doi: 10.1176/appi.books.9780890425596 .

[3] A. Bados, "Trastorno de ansiedad social: Conceptualización y tratamiento psicológico," *Análisis y Modificación de Conducta*, vol. 35, no. 154, pp. 325–348, 2009.

[4] S. Orejudo, M. Puyuelo, and T. Fernández-Turrado, "La ansiedad social en el aula y su relación con la comunicación académica," *Revista de Psicodidáctica*, vol. 13, no. 1, pp. 103–114, 2008.

[5] J. Polo, "La comunicación oral en el aula universitaria: Dificultades de los alumnos y estrategias didácticas," *Revista de Educación*, no. 311, pp. 81–99, 1996.

[6] Y. Pacheco Neiva, H. Rabelo Bezerra Ferreira, V. E. Moran, and L. Guimaraes Cardoso de Sa, "Adaptação brasileira do teste de ansiedade social para universitários (tasu)," *Revista iberoamericana de psicología*, vol. 14, no. 1, pp. 47–58, jan 2021.

[7] A. Bados, *Trastornos de ansiedad*. Barcelona: Editorial UOC, 2008.

[8] S. G. Hofmann, *Cognitive Behavioral Therapy for Social Anxiety Disorder: Evidence-Based and Disorder-Specific Treatment Techniques*. New York: Routledge, 2014. doi: 10.4324/9780203067659 .

[9] M. M. Antony, "Anxiety disorders," in *Handbook of Depression*, 2nd ed., I. H. Gotlib and C. L. Hammen, Eds. New York: Guilford Press, 2010, pp. 628–644.

[10] E. I. Pinargote Macías and L. C. Caicedo Guale, "La ansiedad y su relación en el rendimiento académico de los estudiantes de la carrera de psicología de la universidad técnica de manabí," *Espirales Revista Multidisciplinaria de Investigación Científica*, vol. 3, no. 28, 2019.

[11] E. Y. Strahan, "The effects of social anxiety on academic performance," *Journal of Social Psychology*, vol. 143, no. 2, pp. 219–229, 2003. doi: 10.1080/00224540309598442 .

[12] P. Sotomayor Soloaga, C. Burgos Videla, F. Martínez Nehme, and N. Caradeuc, "Desarrollo de la competencia de liderazgo en la educación superior: perspectivas de estudiantes, egresados y docentes en una universidad chilena," *Perspectiva Educacional*, vol. 63, no. 2, pp. 179–201, jul 2024.

[13] A. Christy, Jufri, and Mukhaiyar, "The effect of speaking anxiety on students' performance in speech class," in *Proceedings of the 9th International Conference on Language and Arts (ICLA 2020)*, 2021, pp. 241–245. doi: 10.2991/assehr.k.210325.043 .

[14] M. Tillfors, T. Furmark, I. Marteinsdottir, H. Fischer, A. Pissiota, B. Långström, and M. Fredrikson, "Cerebral blood flow in subjects with social phobia during stressful speaking tasks: A pet study," *The American Journal of Psychiatry*, vol. 158, no. 8, pp. 1220–1226, 2001. doi: 10.1176/appi.ajp.158.8.1220 .

[15] V. E. Caballo, *Manual de técnicas de terapia y modificación de conducta*. Madrid: Siglo XXI, 2007.

[16] E. Fonseca-Pedrero *et al.*, "Tratamientos psicológicos empíricamente apoyados para la infancia y adolescencia: Estado de la cuestión," *Psicothema*, vol. 33, no. 3, pp. 386–398, 2021. doi: 10.7334/psicothema2021.56 .

[17] X. Liu *et al.*, "A comparison of deep learning performance against healthcare professionals in detecting diseases from medical imaging: A systematic review and meta-analysis," *The Lancet Digital Health*, vol. 1, no. 6, pp. e271–e297, 2019. doi: 10.1016/S2589-7500(19)30123-2 .

[18] A. B. R. Shatte, D. M. Hutchinson, and S. J. Teague, "Machine learning in mental health: A scoping review of methods and applications," *Psychological Medicine*, vol. 49, no. 9, pp. 1426–1448, 2019. doi: 10.1017/S0033291719000151 .

[19] C. M. Bishop, *Pattern Recognition and Machine Learning*. New York: Springer, 2006. doi: 10.1007/978-0-387-31073-2 .

[20] J. M. Sánchez-Sordo, "Data mining techniques for the study of online learning from an extended approach," *Multidisciplinary Journal for Education, Social and Technological Sciences*, vol. 6, no. 1, pp. 1–24, 2019. doi: 10.4995/muse.2019.11482 .

[21] H. Zhang, "The optimality of naive bayes," in *Proceedings of the 17th International Florida Artificial Intelligence Research Society Conference*, 2004, pp. 562–567.

[22] J. M. Sánchez-Sordo and A. Pérez, "Tratamiento de ansiedad social con realidad virtual e inteligencia artificial," Poster presented at XLII Coloquio de Investigación, FES Iztacala, UNAM, Estado de México, México, aug 2024.

Table A1. Complete List of Psychometric Items by Anxiety Level

Summarized Item	**High** ($\mu \pm \sigma$)	**Normal** ($\mu \pm \sigma$)	**Low** ($\mu \pm \sigma$)
Fear of being ridiculed during a conference	3.28 ± 0.73	1.39 ± 0.91	0.09 ± 0.35
Fear of answering questions	3.79 ± 0.45	1.52 ± 0.96	0.37 ± 0.65
Thinking that one's question is foolish	3.35 ± 0.70	1.21 ± 1.01	0.11 ± 0.31
Belief in one's inability to present	3.35 ± 0.70	1.43 ± 1.00	0.24 ± 0.43
Arguing within a group	3.58 ± 0.52	1.88 ± 0.92	0.69 ± 1.20
Being ridiculed during social interactions	3.00 ± 1.22	1.09 ± 0.95	0.00 ± 0.17
Fear of ridicule during a presentation	3.18 ± 0.82	1.26 ± 0.89	0.26 ± 0.64
Fear of not being understood in debates	3.37 ± 0.62	1.77 ± 0.76	0.67 ± 0.86
Thinking that others mock when participating	3.08 ± 0.74	1.62 ± 0.91	0.35 ± 0.55
Fear of not understanding during an oral exam	3.17 ± 0.98	2.09 ± 0.83	0.52 ± 0.96
Avoiding giving opinions in a team	3.26 ± 0.89	1.38 ± 1.20	0.07 ± 0.26
Making incorrect comments in class	3.21 ± 0.65	1.69 ± 0.81	0.67 ± 0.63
Lack of knowledge within a team	2.96 ± 1.06	1.18 ± 0.84	0.26 ± 0.44
Not being accepted by a team	2.85 ± 1.20	0.96 ± 1.00	0.00 ± 0.17
Fear that the professor gets angry when participating	2.91 ± 1.05	1.01 ± 0.85	0.17 ± 0.37
Failing to capture attention during a presentation	3.15 ± 1.05	1.20 ± 0.93	0.30 ± 0.66
Fear of doubts arising during a conference	3.33 ± 0.77	1.79 ± 1.01	0.45 ± 0.81
Fear of raising one's hand	3.30 ± 0.79	1.84 ± 1.10	0.52 ± 0.63
Avoiding expressing ideas in a team	2.96 ± 1.00	0.86 ± 1.01	0.11 ± 0.31
Fear of not finding the right words during an exam	3.54 ± 0.76	2.13 ± 0.97	1.01 ± 0.94
Fear of appearing like a charlatan at a symposium	2.88 ± 1.25	0.75 ± 0.82	0.17 ± 0.50
Fear of responding in public	3.01 ± 1.10	1.47 ± 1.00	0.41 ± 0.76
Having uninteresting conversations	3.00 ± 1.24	1.37 ± 0.99	0.54 ± 0.90
Not being taken into account by the team	2.56 ± 1.28	0.96 ± 1.08	0.11 ± 0.31
Lowering one's voice out of nervousness	2.71 ± 1.15	1.66 ± 0.75	0.56 ± 0.87
Fear of participating in class	3.03 ± 0.93	1.88 ± 0.81	0.98 ± 0.78
Not feeling part of the group	3.24 ± 0.94	1.57 ± 1.26	0.66 ± 0.86
Negative evaluation from classmates	2.96 ± 1.08	1.50 ± 0.96	0.67 ± 0.84
Thinking that one does not know the answer during an oral exam	3.22 ± 1.03	1.84 ± 1.07	0.83 ± 0.84
Fear that others will not pay attention when speaking	2.86 ± 1.19	1.39 ± 1.18	0.32 ± 0.60

Journal of Artificial Intelligence and Computing Applications (2025) - **Special Issue** - 3(2): 33–37

MAIKRON

Journal of Artificial Intelligence and Computing Applications

Expanded abstract

Automatic feedback through natural language processing using a chatbot-based simulated patient (PEPE) for the training of mental health professionals

Ricardo Cruz[1,*], **Violeta Felix Romero**[2], **Marcela Rosas**[2], **Diana Patricia Tzek**[2], and **Ivan Meza**[1]

[1]Instituto de Investigaciones en Matemáticas Aplicadas y en Sistemas, UNAM
[2]Facultad de Psicología, UNAM

ABSTRACT

Mental, neurological, and substance use disorders are highly prevalent worldwide; however, it is estimated that between 75% and 95% of affected individuals lack access to treatment. Therefore, it is essential to promote the development of effective strategies for the training and evaluation of mental health professionals. There is evidence supporting the effectiveness of using standardized simulated patients to train healthcare professionals. One of the main challenges is the extent to which these simulations can authentically and validly represent real patients. The use of Artificial Intelligence has recently been explored to enhance the ecological validity of simulations through Natural Language Processing (NLP). NLP enables various applications, including chatbots and virtual assistants that can engage in natural, human-like conversations. In healthcare, chatbots can serve as valuable tools for training and providing feedback to professionals on a wide range of topics. This project aims to evaluate the impact of automated feedback—delivered through a simulated patient chatbot named PEPE—on the training of healthcare professionals at the Addiction Prevention Center, Faculty of Psychology, UNAM, in areas such as depression, anxiety, and substance abuse.

Keywords: LLMs, chatbots, mental health professionals

1. Introduction

The use of simulations in the assessment and training of health professionals implies the strengthening of competencies for the management of mental health risk indicators in the population [1]. Simulations provide participants with immediate and detailed feedback on their performance, allowing practice in a controlled environment, reducing the likelihood of making mistakes with real patients. Training health professionals to address issues such as substance abuse, depression, and anxiety at the primary level is crucial to ensure the implementation of effective interventions. The use of technological resources based on natural language processing ensures that the simulation is as close as possible to the reality faced daily by mental health professionals.

There are examples of chatbots that are used to diagnose potential mental health problems based on the person's profile and symptoms, offering help in areas such as stress, depression, sleep, and self-esteem [2, 3, 4]. However, very few focus on being used as a virtual patient for the training of health professionals, and even

E-mail address: carlosricardocm@gmail.com
https://doi.org/10.5281/zenodo.17227862
 This article is part of the Special Issue on ICAIMH 2025. ISSN: 3061-8843

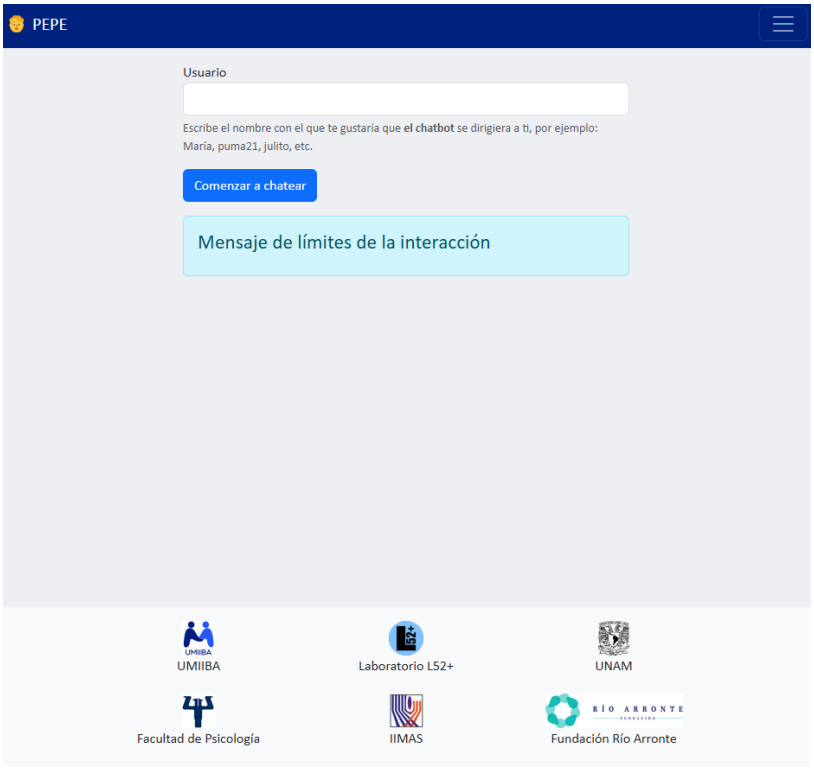

Figure 1. Main Page of the Chatbot Simulated Patient (PEPE)

fewer in the Spanish language [5].

Therefore, one of the main contributions of this work will be the development of a chatbot in Spanish that provides automatic feedback to mental health professionals on depression, anxiety, and substance use. This will allow professionals in training to identify areas for improvement, facilitating continuous learning and the refinement of clinical skills.

2. Methodology

With the aim of training mental health specialists from the Addiction Prevention Center at the Faculty of Psychology of UNAM in providing counseling on depression, anxiety, and substance abuse, a chatbot named PEPE (Spanish acronym for Programa de Entrenamiento con Pacientes Estandarizados) was developed. PEPE integrates various technologies, such as an API for querying ChatGPT-4o, as well as a web system developed at the Institute of Research in Applied Mathematics and Systems at UNAM. This system uses Python as the programming language, along with a socket-based architecture for thread management, database handling, dialogue engine, security, authentication, and graphical interface as showed in Figure 1. Additionally, an administration web page was created to store the data and training sessions of the mental healthcare professionals as shown in Figure 2.

ChatGPT-4o was pre-trained using transcriptions of simulated patient-student session conversations on topics of substance use, depression, and anxiety, creating a script for the fine-tuning of the Large Language Model (LLM). The simulated cases as shown in Fig 3 were developed based on the analysis of videos and files of real cases, obtaining a sample of the representative sociodemographic and clinical characteristics of typical cases who are treated at a primary care training center. Once the case was developed, the indicators and dialogues were validated by expert judges, and finally, PEPE was trained to represent these patients in simulated interactions, providing standardized information and presenting all the stimuli of the case [6].

The functionalities currently available in PEPE are as follows: First use case for evaluating a healthcare professional. The use case includes elements such as signs, symptoms, and strategies to reduce alcohol consumption, as mentioned in the Mental Health Gap Action Programme (mhGAP) Intervention Guide [7]. In this use case, PEPE simulates a 30-year-old male patient who presents with concerns related to excessive alcohol consumption (See Figure 4).

3. Results and Discussion

Currently, PEPE can engage in dialogue with a mental healthcare professional; presenting general information about the case such as personal details, the reason for consultation, and triggers and consequences associated with alcohol use. However, the following objectives are yet to be achieved:

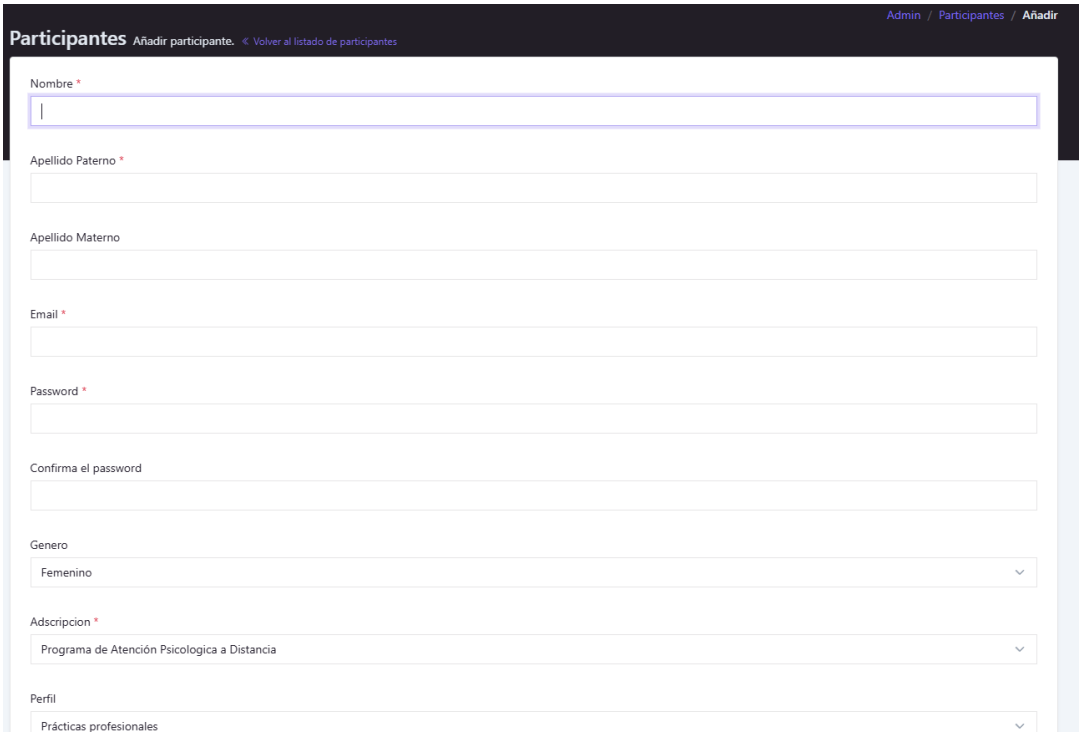

Figure 2. Webpage for registering mental health professionals in PEPE's administrative interface.

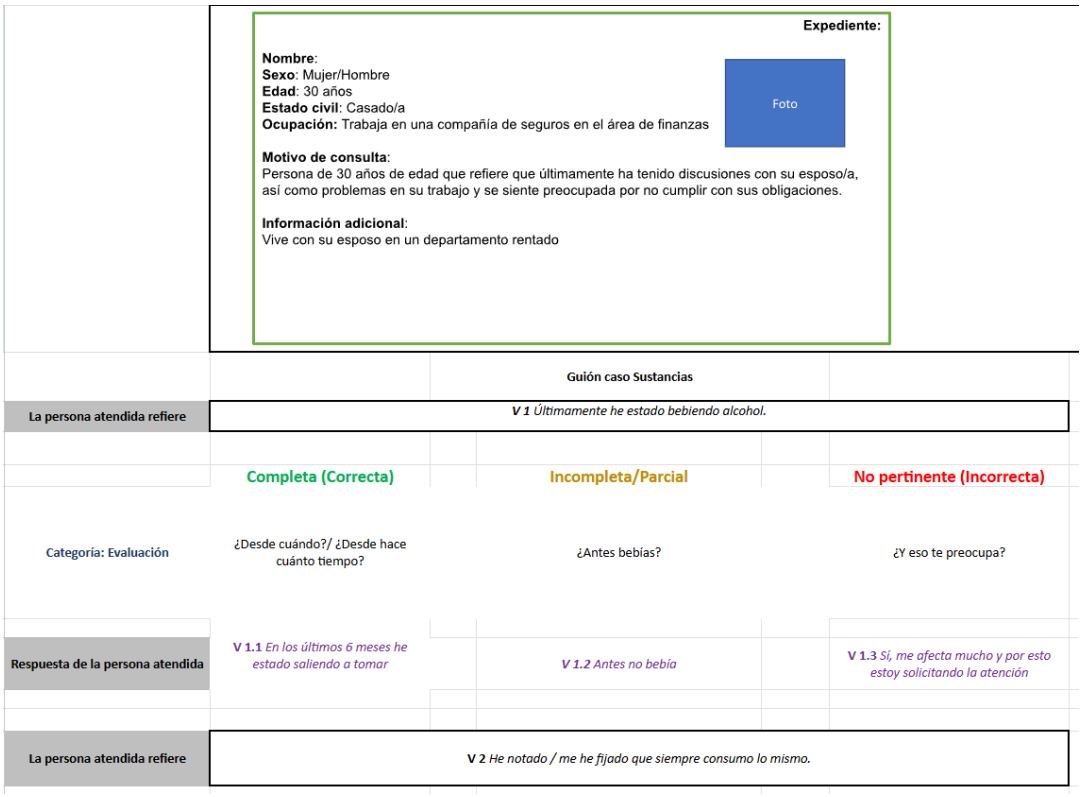

Figure 3. Example of a simulated alcohol abuse case that includes incorrect, incomplete, and non-pertinent answers.

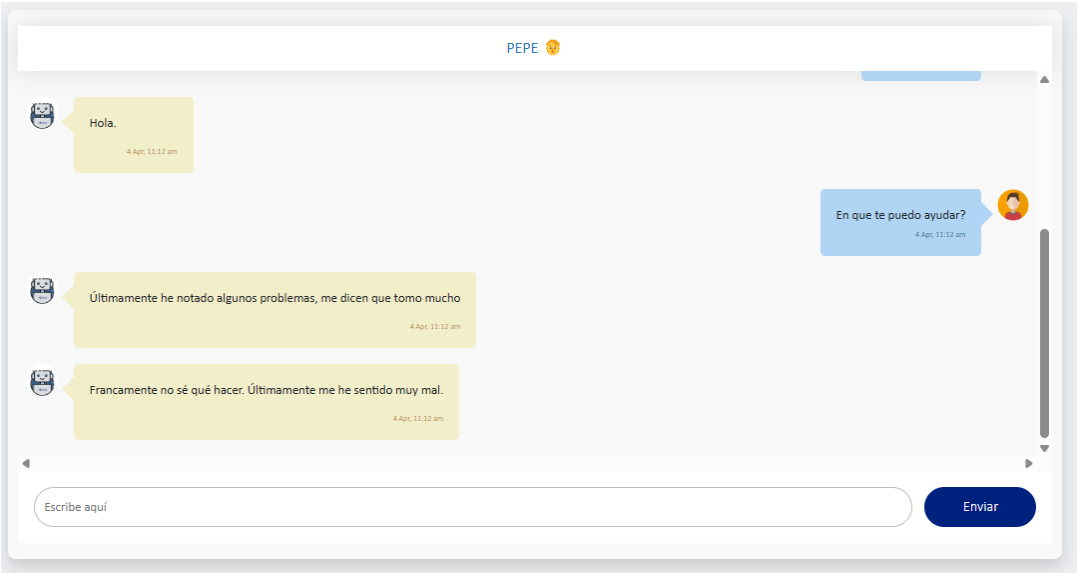

Figure 4. Screenshot showing a dialogue in Spanish between the chatbot PEPE, simulating a 30-year-old male patient expressing concerns about alcohol abuse, and a trainee mental healthcare professional.

- Automatic Feedback for Mental Health Professionals. For each question posed by PEPE, the healthcare professional will provide a response, which will be immediately evaluated by PEPE, receiving instant feedback and an assessment.
- Improvements to the conversation model to make it more adaptive.
- Improvements to the historical storage of conversations to contextualize the responses.
- Evaluation of improvements at the conversation flow level.
- Evaluation of the system with end users.
- Creation of a results log.
- Adding the use cases for depression and anxiety.

4. Conclusion

The research highlights the development and application of a chatbot-based simulated patient, PEPE, to train mental health professionals in addressing issues such as depression, anxiety, and substance abuse. The implementation of automatic feedback via Natural Language Processing allows for real-time assessment and continuous learning. Future work will focus on refining the conversation model to enhance its adaptability, evaluating healthcare professionals' responses, and improving the storage of conversation histories for better context. Furthermore, end-user evaluations will be conducted, and a results log will be created to track progress.

The integration of AI technologies into mental health training, while highly promising, also presents important ethical considerations. A central concern is the protection of personal data generated during simulated interactions. In this project, data privacy is safeguarded by ensuring that all training information is provided anonymously by healthcare professionals from the Addiction Prevention Center at UNAM. This approach protects sensitive information, promotes ethical data handling, and minimizes the risk of compromising individual privacy in educational and research contexts.

Furthermore, the implementation of AI systems like PEPE demands transparency regarding the accuracy and limitations of the feedback they provide. It is essential that users understand that AI-generated evaluations are based on algorithmic models rather than clinical judgment, and should be interpreted as supportive tools to guide learning—not as definitive assessments. Despite these limitations, PEPE demonstrates strong potential to enhance the training and skill development of mental health professionals, ultimately contributing to more effective and accessible mental healthcare delivery. However, future evaluations are necessary to assess the system's effectiveness and its actual impact on educational outcomes.

Ethics Statement

This study did not involve human participants or animals and therefore did not require ethical approval.

CRediT authorship contribution statement

Ricardo Cruz: Conceptualization, Software, Writing – original draft. **Violeta Felix Romero:** Investigation, Writing – review editing. **Marcela Rosas:** Investigation, Writing – review editing. **Diana Patricia Tzek:** Investigation, Writing – review editing. **Iván Meza:** Conceptualization, Software, Writing – review editing.

Declaration of Generative AI and AI-assisted technologies in the writing process

The authors utilized Grammarly and ChatGPT to refine sentence structure and enhance readability. No content was generated by AI; all scientific insights and original ideas are the authors' own.

Declaration of competing interest

The authors declare no competing interests.

Acknowledgements

The authors would like to thank UNAM, especially for the support received from the Support Program for Research and Technological Innovation Projects (PAPIIT) TA300124.

References

[1] C. Osborn and R. Cash, "Effects of interview training with simulated patients on suicide, threat, and abuse assessment," *Athens Journal of Social Sciences*, vol. 8, no. 4, pp. 245–258, oct 2021. doi: 10.30958/ajss.8-4-3 .

[2] G. Cameron, D. Cameron, G. Megaw, R. Bond, M. Mulvenna, S. O'Neill, C. Armour, and M. McTear, "Assessing the usability of a chatbot for mental health care," in *Internet Science*, S. S. Bodrunova, O. Koltsova, A. Følstad, H. Halpin, P. Kolozaridi, L. Yuldashev, A. Smoliarova, and H. Niedermayer, Eds. Cham: Springer International Publishing, 2019, pp. 121–132. doi: 10.1007/978-3-030-17705-8$_1$1 .

[3] A. F. Ur Rahman Khilji, S. R. Laskar, P. Pakray, R. A. Kadir, M. S. Lydia, and S. Bandyopadhyay, "Healfavor: Dataset and a prototype system for healthcare chatbot," in *2020 International Conference on Data Science, Artificial Intelligence, and Business Analytics (DATABIA)*, 2020, pp. 1–4. doi: 10.1109/DATABIA50434.2020.9190281 .

[4] C. Gudmundsen Høiland, A. Følstad, and A. Karahasanovic, "Hi, can i help? exploring how to design a mental health chatbot for youths," *Human Technology*, vol. 16, no. 2, pp. 139–169, Aug 2020. doi: 10.23987/ht.88715 . [Online]. Available: https://ht.csr-pub.eu/index.php/ht/article/view/5

[5] A. Suárez, A. Adanero, V. Díaz-Flores García, Y. Freire, and J. Algar, "Using a virtual patient via an artificial intelligence chatbot to develop dental students' diagnostic skills," *International Journal of Environmental Research and Public Health*, vol. 19, no. 14, p. 8735, 2022. doi: 10.3390/ijerph19148735 . [Online]. Available: https://www.mdpi.com/1660-4601/19/14/8735

[6] V. Félix Romero, D. Ortiz Gómez, S. Morales Chainé, and C. Uriarte Rojo, "Caso simulado estandarizado: Evaluación conductual en profesionales de la salud en adicciones," *Acta de Investigación Psicológica*, vol. 11, no. 3, pp. 87–98, dic 2021. doi: 10.22201/fpsi.20074719e.2021.3.395 . [Online]. Available: https://www.revista-psicologia.unam.mx/revista_aip/index.php/aip/article/view/395

[7] World Health Organization, *mhGAP Intervention Guide for Mental, Neurological and Substance Use Disorders in Non-specialized Health Settings: Version 2.0*, World Health Organization, Geneva, 2016. [Online]. Available: https://www.who.int/publications/i/item/9789241549790

Journal of Artificial Intelligence and Computing Applications (2025) - **Special Issue** - 3(2): 38–43

MAIKRON

Journal of Artificial Intelligence and Computing Applications

Expanded abstract

On-device conversational agent for psycho-oncology based on acceptance and commitment therapy manuals

Samara Acosta-Jiménez[1], **Miguel M. Mendoza-Mendoza**[1], **Gerardo N. Rivera-Rojas**[1], **Jorge I. Galván-Tejada**[1], **José M. Celaya-Padilla**[1], **and Carlos E. Galván-Tejada**[1,*]

[1]Unidad Académica de Ingeniería Eléctrica, Universidad Autónoma de Zacatecas, Jardín Juarez 147, Centro, Zacatecas 98000, Mexico

ABSTRACT

Psychological distress is highly prevalent among individuals diagnosed with cancer, yet access to specialist psycho-oncology services remains limited by workforce shortages, geographical barriers, and concerns about privacy. This study presents a fully *on-device* conversational agent that combines Retrieval-Augmented Generation with a curated corpus of Acceptance and Commitment Therapy (ACT) manuals to deliver evidence-grounded emotional support without reliance on cloud resources. Multiple ACT guides and workbooks are parsed using *PyMuPDF*, segmented into 250-character chunks via a token-aware recursive splitting strategy, and embedded with *all-MiniLM-L6-v2* sentence transformers. The resulting are indexed in a FAISS *IndexFlatIP* store. At inference, LangChain retrieves and ranks the top-k passages, while LangGraph enforces source fidelity before passing the context to a locally hosted `llama3` model served via `ChatOllama`. Preliminary interactions suggest that the agent delivers concise, empathic responses referencing core ACT processes, while attaching inline citations that trace each claim to a specific passage. All responses are generated in real time on CPU-only hardware, preserving user privacy and making the system viable in low-resource clinical environments. Although current evaluation remains qualitative, no hallucinated citations or clinically unsafe statements are observed, indicating robust factual grounding at this early stage. Future work will add psycho-oncology texts, enable optional web search for unseen queries, and run concordance studies with psycho-oncologists to measure accuracy, tone, and usability. This lightweight offline pipeline therefore paves the way for privacy-preserving chatbots that enhance psychosocial care in oncology and other mental-health settings.

Keywords: cancer, chatbot, mental health

1. Introduction

Cancer constitutes a pressing global health challenge: the World Health Organization reports 20 million new diagnoses and 9.7 million deaths in 2022, while five-year prevalence already exceeds 53 million survivors and annual incidence is projected to rise above 35 million by 2050 [1, 2]. Beyond its somatic burden, the disease inflicts a profound psychological toll. Meta-analytic evidence indicates that roughly one third of adults with cancer meet criteria for an anxiety, depressive, or adjustment disorder [3], and narrative syntheses confirm that up to one half of patients with advanced disease experience clinically significant distress [4]. Such symp-

E-mail address: ericgalvan@uaz.edu.mx
https://doi.org/10.5281/zenodo.17237143
 This article is part of the Special Issue on ICAIMH 2025. ISSN: 3061-8843

toms translate into poorer treatment adherence, increased unplanned hospitalizations, diminished quality of life, and an 85% elevation in suicide mortality relative to the general population [5]. Although international guidelines mandate routine psychosocial screening, workforce shortages, geographical barriers, financial constraints, and stigma restrict access; fewer than one-fifth of distressed patients ultimately receive specialized mental-health care [6, 7, 8].

Digital interventions provide a scalable means to narrow this gap. Web- and app-based programs yield modest yet significant reductions in depression and anxiety, but engagement often wanes once human guidance is withdrawn. Fully automated conversational agents offer real-time empathic interaction; however, generic large-language-model chatbots are vulnerable to hallucination and lack domain grounding—limitations that pose unacceptable risks in oncology. Retrieval-Augmented Generation (RAG) mitigates these concerns by coupling neural text generation with dense semantic retrieval so that every output is explicitly supported by external evidence [9]. Early mental-health chatbots such as Woebot demonstrate that cognitive-behavioral strategies can be delivered safely and effectively through text-based conversation [10], yet no system to date targets the complex psychosocial needs of people living with cancer using a rigorously grounded RAG architecture.

Acceptance and Commitment Therapy (ACT)—a third-wave behavioral intervention centered on psychological flexibility—offers a mechanism-based framework ideally suited to this purpose. ACT integrates six interlocking processes (acceptance, cognitive defusion, present-moment awareness, self-as-context, values clarification, and committed action) and has accumulated robust empirical support across chronic-illness populations, including oncology.

This study introduces an innovative on-device chatbot that couples RAG with a multilingual corpus of ACT manuals. By fusing rapid semantic retrieval with contextualized natural language generation, the system delivers empathic, evidence-backed answers without cloud reliance, reducing privacy and hallucination risks. The tool positions itself as a scalable adjunct for psycho-oncology support; its effectiveness and safety will be assessed through rigorous clinical validation in future work.

2. Methodology

The notebook implements a RAG workflow executed entirely on-device. Figure 1 shows the LangGraph control flow: text chunks are embedded, stored in a local `FAISS` index, retrieved on demand, and passed to a locally served `llama3` model via `ChatOllama`.

To embed this evidence-based approach into an accessible digital tool, the present study curates a multilingual ACT corpus extracted from clinician-endorsed guides and manuals, including *Terapia de Aceptación y Compromiso* [11], the therapist and patient workbooks *Duelo: Tratamiento basado en ACT* [12], and Hayes' *Proceso del Cambio Consciente* [13]. To transform the curated ACT corpus into a machine-readable knowledge base, each PDF is parsed with `PyMuPDF`, cleaned, and split into 250-character chunks via a `RecursiveCharacterTextSplitter`. Every chunk is embedded as a 384-dimensional sentence vector using `all-MiniLM-L6-v2`. The resulting matrix is stored in a `FAISS IndexFlatIP` structure—FAISS (Facebook AI Similarity Search) is an open-source C++/Python library from Meta that performs sub-millisecond nearest-neighbour search in dense-vector spaces even on commodity CPUs, making it ideal for strictly on-device deployments[14].

When a question arrives, the system collects the most relevant text snippets and removes any that fail a basic quality check. The remaining snippets are sent to `llama3`, Meta's freely available language model[15], running locally through the lightweight `ChatOllama` program. This setup is chosen because it (i) keeps all data on the device for maximum privacy, (ii) avoids ongoing cloud costs, and (iii) connects directly to the retrieval code through LangChain's built-in ChatOllama link [16].

The methodology is presented as a fully on-device RAG pipeline. First, the ACT manuals undergo cleansing, segmentation into 250-character chunks, embedding with `all-MiniLM-L6-v2`, and storage in a `FAISS` index. Next, a `LangChain/LangGraph` workflow retrieves the most relevant passages, verifies their relevance, and forwards them to a locally hosted `llama3` model served through `ChatOllama`, which generates citation-grounded responses. Finally, the approach outlines the hallucination-mitigation guardrails, an optional web-search extension slated for future work, and the precise steps required to reproduce the experiment offline.

2.1 Imported libraries

The code relies on a deliberately small dependency set, so the prototype can be replicated on any machine with `pip`, Python's built-in package installer that fetches and installs libraries from the Python Package Index (PyPI).

Table 1 lists every third-party package referenced in an `import` statement; any module not shown belongs to the Python standard library. All packages install cleanly on Python 3.10 with no GPU support required.

2.2 Corpus preparation

Seven ACT manuals and workbooks (PDF) form the knowledge base. Each file is opened using `fitz`; headers, footers, and non-text objects are removed, yielding clean UTF-8 text. A `RecursiveCharacterTextSplitter` (`chunk_size=250`, `chunk_overlap=0`) divides the text into fixed-length segments that preserve local coherence while remaining within the context window of the target model. Every segment is embedded as a 384-dimensional vector using `HuggingFaceEmbeddings("all-MiniLM-L6-v2")`

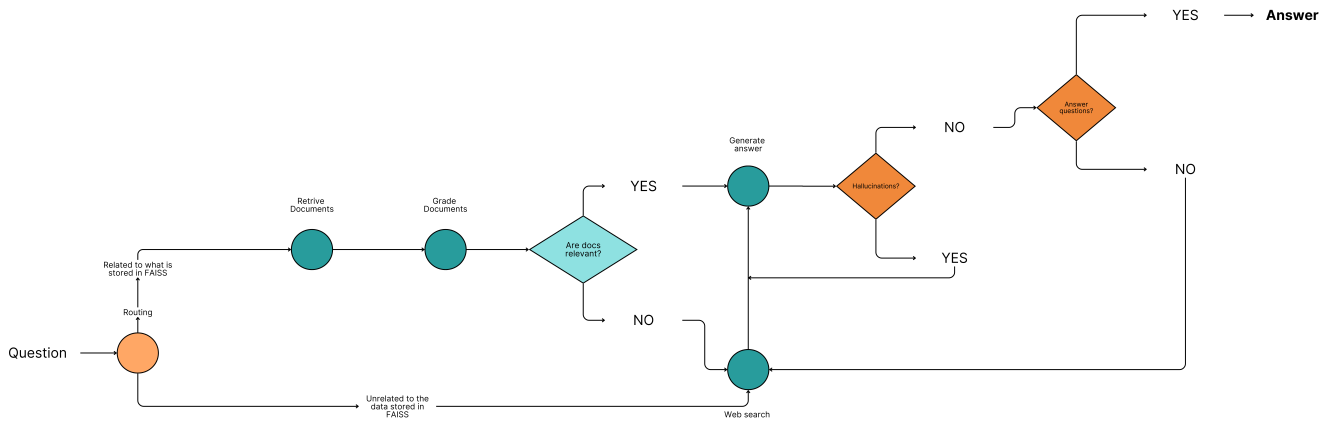

Figure 1. LangGraph control flow used in the notebook. "Retriever" queries FAISS; "LLM" invokes ChatOllama.

Table 1. External Python packages referenced in the notebook

Package	Role
PyMuPDF (*fitz*)	PDF text extraction
python-dotenv	Load `.env` variables
TextBlob	Lightweight text inspection
sentence-transformers	Embeddings (`all-MiniLM-L6-v2`)
faiss-cpu	Vector index (`IndexFlatIP`)
langchain	Core RAG primitives
`langchain_community`	FAISS wrapper, ChatOllama client
`langchain_core`	Prompt/schema abstractions
`langchain_huggingface`	HF embedding adapter
langgraph	Branching and validation nodes
ollama	Serve local `llama3` model
tavily_search	Optional web-search tool

and inserted into a `faiss.IndexFlatIP` index, enabling cosine-similarity retrieval directly on CPU with no cloud dependencies.

2.3 Retrieval and answer generation

1. *Router* – a keyword filter routes oncology and mental-health questions to the ACT corpus; unrelated queries return an *out-of-scope* message.
2. *Retriever* – `vectorstore.as_retriever()` returns the top-k = 4 passages; no similarity threshold is applied.
3. *Document grader* – a structured prompt plus `ChatOllama(model="llama3", temperature=0, format="json")` assigns a relevance score (1–5) to each passage; passages rated < 4 are discarded.
4. *Answer generator* – the filtered passages are passed to the same model to produce a plain-language answer that includes inline citations.

2.4 Hallucination guard

A validation node re-queries the `FAISS` index; if any citation string in the generated answer does not match the retrieved passages, the response is discarded and regenerated using a narrower context window.

2.5 Reproducibility

Running the notebook sequentially performs: PDF parsing → chunking → embedding → FAISS index build → LangGraph agent launch. Once embeddings are stored, the entire pipeline runs fully offline, ensuring both data privacy and reproducibility in low-resource environments.

To enhance reproducibility and clarity, the entire methodology has been structured as a modular pipeline, where each stage performs a clearly defined function. From PDF parsing and chunk-based segmentation to vector embedding, FAISS indexing, LangGraph routing, passage grading, and citation-grounded generation, ev-

ery component is selected to balance performance with transparency. This architecture not only ensures that all operations remain fully offline and privacy, preserving, but also facilitates replication across different clinical or research settings. Figure 1 provides a visual reference to support comprehension of the full process.

3. Results and Discussion

The prototype is exercised through a set of representative pilot interactions embedded in the development notebook. Each user query triggered the complete RAG pipeline: the question is routed to the ACT corpus, relevant passages are retrieved from the `FAISS` index, graded for topical fit, and filtered before being passed to a locally served `llama3` model via `ChatOllama`, which returned a citation-backed answer.

Table 2 presents user prompts and representative excerpts from the replies; inline citations are omitted here for brevity, although the system includes them in actual output.

Manual inspection confirms that replies consistently reference core ACT constructs, including acceptance, values clarification, and psychological flexibility, without veering into prescriptive or unsafe medical advice. Each factual claim is grounded in a source passage, and the validation node re-queries FAISS to ensure all citations match retrieved content. In all pilot interactions, citations resolved correctly, suggesting that the retrieval–grading mechanism effectively constrains hallucinations at this early stage.

From a user experience perspective, responses are generated in real time on CPU-only hardware and use clear, empathetic language. The system successfully addresses open-ended emotional queries (e.g., "Why do I feel overwhelmed?") while maintaining ACT consistency, indicating feasibility for deployment in privacy-sensitive, patient-facing scenarios.

Current evaluation remains qualitative and limited to ACT materials; integration of oncology-specific psychoeducational content and conditional web search—already scaffolded in the code via the `tavily_search` interface, are planned next.

A structured concordance study involving licensed psycho-oncologists is underway to assess answer quality in terms of factual accuracy, therapeutic tone, and clinical adequacy. Quantitative metrics such as retrieval precision, citation overlap, and response time will also be reported once a broader question set is available.

Overall, these preliminary results demonstrate that a locally hosted RAG pipeline can deliver transparent, ACT-grounded support for individuals experiencing cancer-related distress. The modular architecture supports incremental expansion of both the corpus and the evaluation framework, laying the foundation for future clinical validation.

While this work focused on establishing a proof of concept, we recognize that ethical safeguards are essential before real-world deployment. Future iterations will incorporate critical features such as crisis escalation protocols, automated disclaimers, and pathways for optional clinical supervision. These measures are necessary to ensure the responsible use of the system and to support patient safety, especially in emotionally vulnerable scenarios.

Also, future iterations of the system will include testing with a larger and more diverse dataset, incorporating broader user characteristics and use-case contexts to assess generalizability and robustness.

4. Conclusion

This work introduces a privacy-preserving, end-to-end RAG chatbot that delivers ACT guidance tailored to the psycho-oncology context. By embedding seven authoritative ACT manuals into a `FAISS` index and orchestrating retrieval, grading, and generation via `LangChain` and `LangGraph`, the prototype demonstrates that high-quality, citation-backed responses can be generated locally using an open-source `llama3` model. Once embeddings are created, the system operates entirely offline—an essential design choice for mental-health applications where confidentiality, data locality, and cost control are critical.

From a methodological perspective, the contribution is threefold. First, it outlines a lightweight, replicable pipeline, PDF parsing, overlap-free chunking, and MiniLM-based embedding, that runs on CPU-only hardware. Second, it shows how LangGraph validation nodes can enforce strict source fidelity, effectively preventing hallucinated citations in the tested interactions. Third, it offers an extensible scaffold: the codebase already includes a Tavily search module for future integration of dynamic web evidence, and its modular architecture allows seamless expansion to oncology-specific psycho-education or additional psychotherapeutic frameworks.

Qualitative results indicate that the chatbot produces concise, empathetic responses aligned with core ACT principles while avoiding prescriptive or clinically unsafe advice—an encouraging sign of therapeutic relevance. However, this evaluation remains preliminary; larger query sets, quantitative retrieval metrics, and structured expert review by psycho-oncologists are essential before considering patient-facing deployment. Any future rollout must also address ethical safeguards, including disclaimers, crisis-response escalation, and real-time monitoring.

In summary, the project illustrates a viable pathway toward evidence-grounded conversational agents that augment psychosocial care for individuals affected by cancer. By prioritizing on-device execution and reproducibility, it reduces technical and regulatory barriers for institutions seeking to explore AI-assisted mental-health support. Ongoing work, expanding the corpus, activating conditional retrieval, and conducting expert concordance studies, will advance the prototype from proof of concept to a clinically validated, deployable system.

Table 2. Representative pilot exchanges with the prototype

User prompt	Prototype reply (excerpt)
How does a cancer diagnosis affect mental health?	A diagnosis often triggers anxiety, sadness, and uncertainty. ACT invites acknowledging these emotions while taking small, value-guided steps to maintain psychological flexibility.
How can I learn to live with grief after treatment?	Grief is a natural response. ACT recommends accepting difficult feelings without avoidance and committing to actions that honour personal values.
What does acceptance really mean in ACT?	Acceptance refers to making space for unpleasant thoughts and emotions instead of fighting them, while continuing to pursue meaningful goals.
Why do I feel emotionally overwhelmed even after being cured?	Recovery does not always bring immediate emotional closure. ACT normalises lingering discomfort and encourages value-based engagement despite inner struggle.

Ethics Statement

This study did not involve human participants or animals and therefore did not require ethical approval.

CRediT authorship contribution statement

Samara Acosta-Jiménez: Conceptualization, Methodology, Data curation, Formal analysis, Visualization, Writing – original draft, Writing – review & editing. **Miguel M. Mendoza-Mendoza:** Methodology, Formal analysis, Validation, Writing – review & editing. **Gerardo N. Rivera-Rojas:** Methodology, Investigation, Resources, Writing – review & editing. **Jorge I. Galván-Tejada:** Validation, Resources, Writing – review & editing. **José M. Celaya-Padilla:** Supervision, Writing – review & editing. **Carlos E. Galván-Tejada:** Conceptualization, Supervision, Project administration, Writing – review & editing.

Declaration of Generative AI and AI-assisted technologies in the writing process

During the preparation of this work, the authors used Grammarly to assist with grammar correction and language clarity. No generative AI tools were used. All scientific content, data analyses, figures, and tables are the sole work of the authors, and all final edits were performed and verified by the authors.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] World Health Organization, "Global cancer burden growing, amidst mounting need for services," News Release, Geneva, 2024. [Online]. Available: https://www.who.int/news/item/01-02-2024-global-cancer-burden-growing--amidst-mounting-need-for-services

[2] F. Bray, J. Ferlay, I. Soerjomataram, R. L. Siegel, L. A. Torre, and A. Jemal, "Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries," *CA: A Cancer Journal for Clinicians*, vol. 74, no. 3, pp. 229–263, 2024. doi: 10.3322/caac.21834 .

[3] A. J. Mitchell, M. Chan, H. Bhatti, M. Halton, L. Grassi, C. Johansen, and N. Meader, "Prevalence of depression, anxiety, and adjustment disorder in oncological, haematological, and palliative-care settings: a meta-analysis of 94 interview-based studies," *The Lancet Oncology*, vol. 12, no. 2, pp. 160–174, 2011. doi: 10.1016/S1470-2045(11)70002-X .

[4] A. Pitman, S. Suleman, N. Hyde, and A. Hodgkiss, "Depression and anxiety in patients with cancer," *BMJ*, vol. 361, p. k1415, 2018. doi: 10.1136/bmj.k1415 .

[5] M. Heinrich, L. Hofmann, H. Baurecht, P. M. Kreuzer, H. Knüttel, M. F. Leitzmann, and C. Seliger, "Suicide risk and mortality among patients with cancer," *Nature Medicine*, vol. 28, no. 4, pp. 852–859, 2022. doi: 10.1038/s41591-022-01745-y .

[6] J. C. Holland, B. Andersen, W. S. Breitbart, L. O. Buchmann, B. Compas, T. L. DeShields, M. M. Dudley, S. Fleishman, C. D. Fulcher, D. B. Greenberg *et al.*, "Distress management," *Journal of the National Comprehensive Cancer Network*, vol. 11, no. 2, pp. 190–209, 2013. doi: 10.6004/jnccn.2013.0027 .

[7] F. I. Fawzy, "Psychosocial interventions for patients with cancer: what works and what doesn't," *European Journal of Cancer*, vol. 35, no. 11, pp. 1559–1564, 1999. doi: 10.1016/S0959-8049(99)00191-4 .

[8] S. L. Ehlers, K. Davis, S. M. Bluethmann, L. M. Quintiliani, J. Kendall, R. M. Ratwani, M. A. Diefenbach, and K. D. Graves, "Screening for psychosocial distress among patients with cancer: implications for clinical practice, healthcare policy, and dissemination to enhance cancer survivorship," *Translational Behavioral Medicine*, vol. 9, no. 2, pp. 282–291, 2019. doi: 10.1093/tbm/iby119 .

[9] P. Lewis, E. Perez, A. Piktus, F. Petroni, V. Karpukhin, N. Goyal, H. Küttler, M. Lewis, W.-T. Yih, T. Rocktäschel *et al.*, "Retrieval-augmented generation for knowledge-intensive NLP tasks," in *Advances in Neural Information Processing Systems*, vol. 33, 2020, pp. 9459–9474. [Online]. Available: https://proceedings.neurips.cc/paper/2020/hash/6b493230205f780e1bc26945df7481e5-Abstract.html

[10] K. K. Fitzpatrick, A. Darcy, and M. Vierhile, "Delivering cognitive behavior therapy to young adults with symptoms of depression and anxiety using a fully automated conversational agent (woebot): a randomized controlled trial," *JMIR Mental Health*, vol. 4, no. 2, p. e7785, 2017. doi: 10.2196/mental.7785 .

[11] M. Luciano and K. Wilson, *Terapia de aceptación y compromiso: Un tratamiento conductual orientado a los valores*. Madrid: Pirámide, 2002.

[12] J. I. Cruz Gaitán, R. O. Sánchez, and P. L. Gutiérrez, *Duelo: Tratamiento basado en la Terapia de Aceptación y Compromiso*. Ciudad de México, Mexico: Editorial El Manual Moderno, 2017.

[13] S. C. Hayes, K. D. Strosahl, and K. G. Wilson, *Acceptance and Commitment Therapy: An Experiential Approach to Behavior Change*. New York, NY, USA: Guilford Press, 1999.

[14] J. Johnson, M. Douze, and H. Jégou, "Billion-scale similarity search with GPUs," *IEEE Transactions on Big Data*, vol. 7, no. 3, pp. 535–547, 2021. doi: 10.1109/TBDATA.2019.2921572 .

[15] Meta AI, "Meta Llama 3 — technical report," Meta AI, Tech. Rep. arXiv:2404.14219, 2024. [Online]. Available: https://arxiv.org/abs/2404.14219

[16] LangChain Developers. (2025) ChatOllama integration guide. [Online]. Available: https://python.langchain.com/docs/integrations/chat/ollama