Journal of Artificial Intelligence and Computing Applications

ISSN: 3061-8843

MAIKRON

www.maikron.org/jaica January-June 2026 Vol. 4 No. 1

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

Journal of Artificial Intelligence and Computing Applications, Vol. 4, Núm. 1, enero – junio de 2026, 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, eISSN:3061-8843, por el Instituto Nacional del Derecho de Autor (INDAUTOR). Responsable de la última actualización de este número, Mauricio Gabriel Orozco del Castillo, Editor Responsable. Fecha de la última modificación: 22 de junio de 2026, Calle 12 No. 103, Lotes No. 56 y 57, C.P. 97357, Ucú, Yucatán. El contenido de los artículos es responsabilidad de los autores y no refleja el punto de vista de los árbitros, del Editor o de la asociación. Se autoriza la reproducción total o parcial de los textos siempre y cuando se cite la fuente completa y la dirección electrónica de la publicación.

Journal of Artificial Intelligence and Computing Applications
Vol. 4, No. 1, January-June 2026

Contents

Journal of Artificial Intelligence and Computing Applications

MAIKRON

Foreword

Dear Readers, Contributors, and Colleagues,

It is with great pleasure that I introduce Volume 4, Issue 1 of the Journal of Artificial Intelligence and Computing Applications (JAICA). This regular issue marks the beginning of a new volume and continues our mission to publish high-quality research that explores the intersection of artificial intelligence with practical, interdisciplinary, and socially meaningful domains.

A notable feature of this issue is its strong emphasis on the use of artificial intelligence to address complex human and societal challenges. Several contributions focus on mental health and well-being, including the detection of suicidal ideation through machine learning and the assessment of purpose in life among university students using a contextual and data-driven approach. These articles reflect the growing importance of AI-supported tools for understanding psychological risk, contextual opportunity, and student well-being, while also emphasizing that artificial intelligence should complement, rather than replace, human judgment, care, and institutional responsibility.

This issue also broadens the scope of JAICA by including short narrative reviews that examine emerging applications and challenges of artificial intelligence across different fields. One article analyzes the current role of AI in cyberattacks, highlighting both the defensive and offensive implications of intelligent systems in cybersecurity. Another explores the automated generation of multiple-choice questions using large language models, with particular attention to soft skills development and the need for expert validation. A further contribution reviews the application of LSTM models and recurrent neural networks in flood detection using satellite imagery, situating AI within the urgent context of environmental monitoring and hydrometeorological risk management.

Taken together, the articles in this issue illustrate the breadth of contemporary applied AI research. They show how artificial intelligence can support mental health screening, educational innovation, cybersecurity analysis, and environmental risk detection, while also reminding us of the methodological, ethical, and social responsibilities that accompany these technologies.

I extend my sincere thanks to the authors for their contributions, to the reviewers for their thoughtful and meticulous evaluations, and to our sponsor Maikron for their continued support in advancing JAICA's vision.

As we look ahead, JAICA remains committed to providing a platform for researchers, practitioners, and innovators who are working to apply artificial intelligence to the benefit of society.

Warmest regards,

Mauricio G. Orozco-del-Castillo

Editor-in-Chief

Journal of Artificial Intelligence and Computing Applications (JAICA)

Journal of Artificial Intelligence and Computing Applications (2026) 4(1): 1–13

MAIKRON

Journal of Artificial Intelligence and Computing Applications

Research article

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 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 "structured vacation time" ($r = 0.263$, $p = 0.005$). 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

1. Introduction

Mental health plays a crucial role in the development of college students. The World Health Organization defines mental health as *"a state of well-being in which individuals are fully aware of their abilities, can cope with the normal stresses of life, work productively, and contribute actively to their communities"* [1]. Nevertheless, this integral definition of mental health is constantly challenged by the increase of mental disorders in young adults.

In Mexico, the prevalence of mental disorders has increased significantly in recent decades. The *Gaceta Médica de México*, reported in 2021 that 18.1 million people suffered from some form of mental disorder that year, reflecting a 15.4% increase compared to 2019. This phenomenon is especially pronounced among young adults aged 20 to 29, particularly vulnerable to conditions such as anxiety and depression. According to the Mexican Observatory of Mental Health and Addictions [2], in this age group, anxiety affects 23% of women and 19.9% of men.

Regarding depression, nearly 3.6 million Mexican adults reported suffering from it in 2021, with 1% of cases being cataloged as severe enough to require specialized care. The Mexican National Survey of Self-Reported Well-Being [3] reported depressive symptoms in 15.4% of adults, with a higher prevalence among

E-mail address: guillermo.ac@orizaba.tecnm.mx
https://doi.org/10.5281/zenodo.20801725

women (19.5%) than men. The COVID-19 pandemic exacerbated this situation, by 2020, 28% of the adult population exhibited symptoms of depression, showing a strong upward trend compared to previous years. Surveys conducted during the pandemic indicated that 64% of adolescents and 71% of young people showed depressive traits. This deterioration highlighted the structural limitations of psychological and psychiatric care in the country. Especially considering that only one in five individuals with mental disorders receives timely treatment, while 75% of severe cases go untreated [4].

The Mexican National Health and Nutrition Survey in 2021 emphasized that young people aged 18 to 29 are particularly vulnerable to experiencing existential voids, hopelessness, and risk behaviors [5]. Research by Hernández-Torrano et al. [6] reveals that a low sense of life purpose and lack of defined personal goals are associated with higher substance use and school dropout rates. Similar studies by Salazar-Pousa et al. [7] in Colombia and Gutiérrez-García and Londoño [8] in Peru, reinforce the need to address the existential dimension of student well-being as a preventive measure.

The hypothesis of this work is that the structural and contextual characteristics of income and opportunity inequality of Latin American countries might have relevant social effects in the student's sense of life purpose. Therefore, it is important to take a comprehensive approach to mental health issues, where universities play a key role to the identification and prevention of these conditions within their communities.

From a technical perspective, surveys are an affordable option to take on that task. The Purpose in Life Questionnaire (PIL), designed by Crumbaugh and Maholick [9], has been widely used to assess life purpose and its association with emotional well-being. The framework developed by Viktor Frankl's logotherapy, holds that the pursuit of personal life purpose can reduce anxiety and depression, serving as a protective factor against existential emptiness [10]. According to the author, Logotherapy not only seeks to alleviate symptoms but also provides a personal framework that strengthens resilience in the face of life's challenges.

The PIL test is a tool that has also being used in addition to in-depth interviews and other types of tests, reportedly by psychologists and psychiatrists, to confirm signs of depression and other mental disorders that find the logotherapy's framework useful. For this research, a set of selected questions from the in-depth interviews are suggested as an addition to the original questionnaire to integrate set contextual variables into the analysis.

This study adopted a mixed methodological approach combining psychometric evaluation and predictive modeling. First, the PIL was administered to three groups of university students, alongside 18 additional questions designed to capture contextual factors (Contextual Opportunity Index, COI). These questions were organized into four categories: demographics, academic-cultural background, access to services, and lifestyle. Second, the collected data was analyzed using both traditional statistical methods and machine learning techniques. The ML Predictive models were a Naive Bayes classifier and a decision tree; both trained with data from the PIL and COI to assess whether student's *sense of life purpose* could be influenced on their environmental and social conditions.

2. Theoretical and conceptual framework

Life purpose is a central concept in humanistic and existential psychology and represents the clarity with which a person identifies their goals, values, and meaning in life. In clinical and educational contexts, this construct has been considered a protective factor against various mental health issues, including anxiety, depression, and suicidal ideation [11]. From a medical and psychosocial perspective, exploring life purpose can enrich the understanding of an individual's emotional history and guide preventive strategies.

One of the most recognized tools for assessing this construct is the PIL, developed by Crumbaugh and Maholick [9]. The PIL consists of 20 Likert-scale items designed to measure an individual's perception of life purpose, existential direction, clear goals, and personal satisfaction. Its theoretical foundation lies in the principles of logotherapy developed by Viktor Frankl in 1984, which argues that the search for meaning is a primary human motivation. This perspective is particularly relevant in the context of college student's mental health, where a sense of purpose can act as a buffer against emotional suffering.

Hernandez-Torrano et al. [6] conducted a bibliometric mapping study of *"Mental Health and Well-Being of University studies"*. Their study covered more than 5000 journal articles from a Web of Science database for the period 1975-2020. Their key findings showed that studies about mental health of university students have been increasing in the last decades, having an important increase since 2010. Also, that while there are international studies, more than half have been conducted in the United States, recognizing there are multiple research groups participating on psychology, psychiatry and education publications.

2.1 Logotherapy and the existentialist approach in multiple contexts

Logotherapy approaches "life's purpose" as a fundamental force that enables individuals to resist hopelessness and reframe suffering. Frankl [10] argues that even in extreme conditions, it is possible to find meaning, which strengthens coping and resilience mechanisms. In this regard, the PIL provides a way to operationalize the search for meaning through a quantitative evaluation that has been validated across diverse populations.

Existential emptiness, or the lack of meaning, often manifests through demotivation, emotional isolation, and a sense of disorientation. Heisel and Flett [11] have shown that low PIL scores are associated with high levels of depression, hopelessness, and suicidal risk. In the university setting, these manifestations are particularly relevant due to the inherent pressures of transitioning into adult life and how it affects not only their learning capabilities and student performance, but also their social skills and college experience. From a philosophical standpoint, logotherapy shares foundations with the existentialist thinking of authors like Kierkegaard, Nietzsche, and Sartre, who emphasize freedom, the will to meaning, and self-determination as essential human dimensions. These notions have been incorporated into the development and interpretation of the PIL, broadening its clinical and educational usefulness [12].

Schulenberg and Melton [12] found that college students with greater clarity of purpose reported lower stress levels and higher subjective well-being. Studies by Steger et al. [13] and Schnell et al. [14] have shown that a strong sense of life purpose is associated with greater resilience in adversity and lower emotional exhaustion in work environments, respectively. Their findings highlight the importance of strengthening existential life purpose as a priority for the university educative models. Van der Walt [15], described that while the PIL test results also have to be interpret according to the moment of life of the participants, describing that first year students are less likely to be able to have a defined life purpose; but still, their findings on freshman students are relevant to the universities to take actions that guide students in this process.

Despite its psychometric validity, the PIL has limitations when applied in sociocultural contexts different from where it was developed. Vázquez and Castilla [16], noted that in Latin American environments, it was necessary to complement the PIL test with contextual variables; mostly, because of a cultural environment in which students are strongly influenced by family, economic, and community values. The findings of Hernandez-Torrano et al. [6] backs up this perspective as specific traditions of Western countries tend to lead these studies [17] Also considering that our cultural values and traditions do shape how mental health and mental illness are conceptualized [18]; and how individuals manifest communicate and cope with their symptoms as well as the access to different types of treatments.

2.2 AI approaches in mental health

The use of AI tools in data processing has extended broadly in the health field in recent years. Regarding mental health, this data process has different characteristics according to the data type, collection method and processing algorithm that has been chosen to fulfill a specific task. Muetunda et al. [19], explored several examples in which the use of AI algorithms has been used in mental health. Their first classification is regarding the type of AI learning process and algorithm, dividing them in two categories: a) Machine Learning (ML), in which a set of algorithms enable computers to learn from and make predictions or decisions based on selected data sets; and Deep Learning (DL), considering the use of neural networks to model complex patterns in large amounts of data, including unstructured data as audio, images and multiformat text.

Muetunda's second classification, compared various automatic mental diagnosis solutions: a) Using ML methods for automatic diagnosis with self-reported reported questions; b) Using DL methos to study analysis behavior by studying video files of computer vision: c) Using ML algorithms to detect variables, studying stress in working employes to highlight factors of general mental stress; d) Using ML random forests to identify hidden patterns between variables to predict anxiety disorders; and e) Sensor based apps and gadgets as warning systems using biometrics. This examples also have different objectives as diagnostic, predictive or monitoring tools. Finally, as a third remark regarding the type of question in the tools, considered the difference between the type of questions as factoid (closed and specific) and non-factoid (open and more extensive) are also important considerations.

According to the data recollection and the main objective of this study, the possibility to have access to self-reported tests from college students was the ideal option to explore the general state of mind of this population. As the results might add new information to the university's crew of experts and give them a set of topics for their internal prevention campaigns, the general approaches to increase the community's well-being must remain as the main objective of this study rather than an automatic individual diagnostic tool. Nevertheless, for health professional within the universities, the PIL and the COI questionaries, might be useful together as a complement to depth interviews, personal student work sessions and other complementary techniques.

3. Materials and methods

This exploratory and longitudinal study was designed to analyze the relationship between life purpose definition and various contextual privilege factors among Mexican university students. The main instrument was the PIL, developed by Crumbaugh and Maholick [9], which consists of 20 items distributed across dimensions such as existential direction, goal clarity, life satisfaction, and motivation. Interpretation followed the ranges established in the literature, grouping participants by their scores into three categories: defined life purpose, undefined perception, and existential vacuum [20].

To quantify contextual opportunities among university students, a composite index named COI was constructed by 18 complementary questions. The items were grouped into four conceptual categories: a) demographics (sex, age, marital status, economic dependents); b) academic and cultural background (semester, university, field of study, indigenous background, Multilingualism); c) access to public-private services (streaming, cellphone credit, healthcare, transportation, Passport status), and (4) lifestyle (family situation, structured free time, structured vacation time, interpersonal relationships). Each response was converted into an integer value and then multiplied by the number of items in its respective category. These category scores were then normalized to a 7-point scale, aligned with the PIL test format.

A group of students from three universities was selected through non-probability quota sampling to answer the questionnaire. This method was chosen due to its logistical feasibility and its ability to represent groups with heterogeneous academic and socioeconomic characteristics. Data collection was carried out over a 12-month period using Google Forms, ensuring an efficient systematization of the process.

In the first interpretation stage, the data was analyzed using conventional statistical tools and the MATLAB software. The later to perform descriptive and inferential analyses, including correlations between PIL scores and the contextual variables. Subsequently, in the second interpretation stage, two predictive models using machine learning in RapidMiner using Naive Bayes classifier and decision tree. The goal of this stage was to explore the potential of data analysis to automatically predict the level of defined life purpose based on the 18 contextual variables or its four dimensions. Then, a supervised learning approach was used to compare the life purpose categories with the contextual ones, using final score and interpretation of the PIL test as the target variable.

4. Results

4.1 PIL results and individual interpretation

The first part of the analysis focused on the results of the PIL test and its original items. Of the 114 valid responses, 57% of the students demonstrated a defined life purpose, while 26.3% showed existential uncertainty, and 16.7% reported levels related to an existential void. A basic report was created to reflect these results using heatmap tables and circular graphs as shown in Figure 1. A second collection of graphs shows how three subcategories that group some of the items were able to influence the final scores and its interpretation: 1) Cognitive component (1,2,4,6,8,12,16,18); Motivational component (3,5,9,10,11,13,14,17,20); and Affective component (7,15,19), taken from [21].

The original scale of each one of the items remained calculated as in the original test, with a Likert 1 to 7 value response. The final score was interpreted in this study, the overall score on the PIL was interpreted based on a widely used three-level classification system applied in previous research.

Considering the maximum score of 140 points (20 items rated on a 7-point Likert scale); individuals scoring below 90 points are understood to experience a state of existential void, while scores between 90 and 105 reflect uncertainty regarding life purpose. In contrast, scores above 105 indicate a clearly defined sense of meaning and life purpose. This interpretative framework has been adopted in several studies [16, 20].

After the initial results of the test were calculated, an additional correlation process started using the MATLAB software. The 18 additional questions were tested for relevant correlations with the final three value scale of the PIL's interpretation of the scores. In this case, the variable "nuclear family situation" showed a moderate positive correlation with the PIL score ($r = 0.431788$, $p < 0.000001$), indicating that students with stable family relationships tend to report a higher sense of life purpose. Likewise, the variable "structured vacation time" showed a low but significant positive correlation ($r = 0.262523$, $p = 0.004776$), indicating that recreation also influences, though to a lesser extent, existential perception.

Figure 2 presents two boxplots that explore key associations between life purpose and contextual variables. Subfigure 2a shows the relationship between life purpose definition groups and family situation. A clear upward trend is observed: as family ratings increase, so do the PIL scores. Groups 1 and 2 exhibit medians above 106, indicative of a well-defined life purpose, while lower-rated groups cluster in the range associated with existential uncertainty. Subfigure 2b illustrates the correlation between structured vacation time and life purpose. Groups 1 and 2 display higher medians, suggesting a stronger sense of life purpose among individuals with more organized vacation habits. In contrast, Group 1 shows greater variability, indicating a broader range of purpose-related experiences.

4.2 COI results and Individual Interpretation

The COI was designed with the aid of two medical specialists to integrate a multidimensional assessment of structural and material conditions affecting the mental health and life purpose of university students. This index took a critical approach to inequality and considers how each variable reflects disparities in the student's college experience opportunities. The COI scoring is grounded in intersectional and cumulative inequality frameworks, where multiple disadvantage dimensions compound mental health outcomes [22, 23, 24]. Also, some empirical precedents include the Multidimensional Poverty Index; the Social Depravation Index and the School Climate Index [25, 26, 27].

To calculate the COI, a four-step method was conducted. First, the complementary items were divided into four dimensions of structural disadvantage: demographics, academic-cultural background, access to public/private services, and lifestyle. The items were presented in a questionnaire with two types of answers: binary and ordinal. The item descriptions and score ranges are shown in Table 1.

For instance, "multilingualism" and "passport status" exemplify markers of privilege tightly linked to economic stability and social capital in a highly unequal country like Mexico. Multilingual ability points to access to quality education, immersion programs, or private language schools, which are predominately available to students from higher-income families. Similarly, holding a passport not only reflects financial sol-

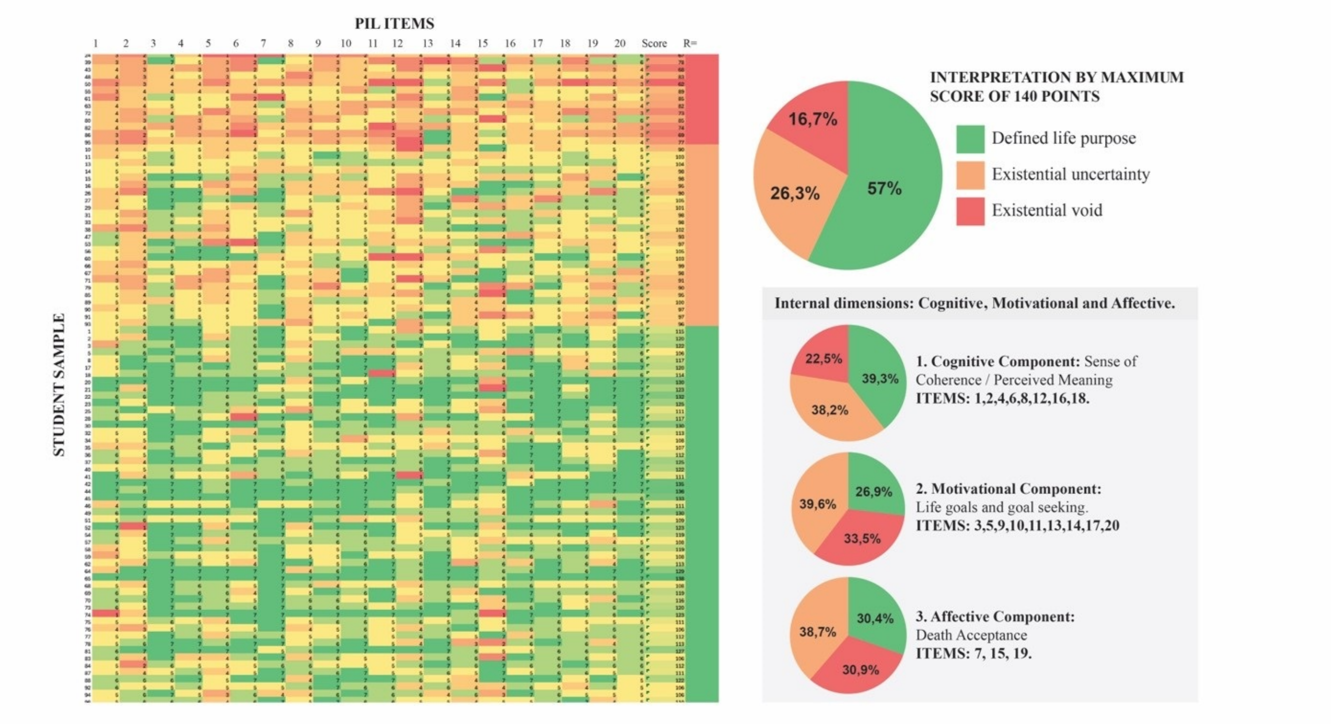
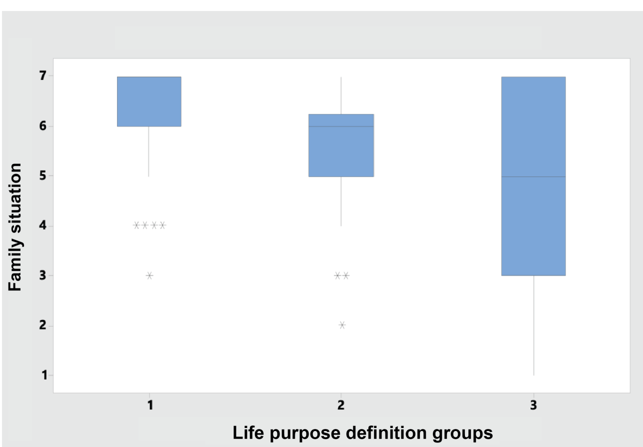
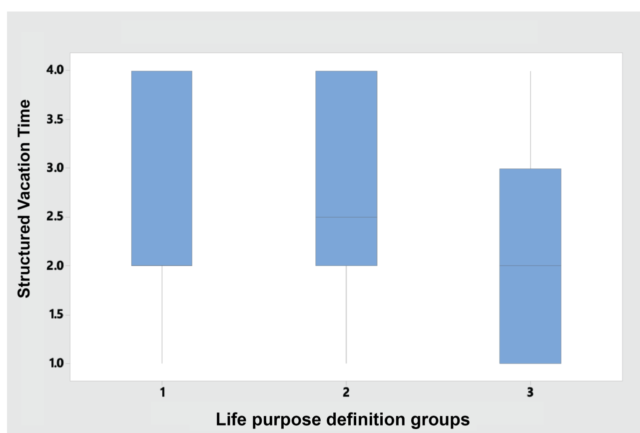


Figure 1. PIL interpretation analysis by final scores and internal dimensions. Heatmap and sector charts created in Microsoft Excel.



(a) Life purpose definition groups vs. Family situation.



(b) Life purpose definition groups vs. Structured vacation time.

Figure 2. Boxplots comparing life purpose definition groups with contextual variables.

Table 1. COI dimensions and item descriptions (2025)

Dimension	Item description
Demographics	Sex; age; marital status; economic dependents
Academic-Cultural Context	University type; field of study; semester; multilingualism; indigenous identity; passport status
Access to Services	Digital services; cellphone coverage/data; health services; transport services
Lifestyle and Routines	Family situation; interpersonal relationships; rest; structured vacation time

vency but also indicates greater economic mobility and exposure to international experiences, elements that foster broader perspectives and opportunities for self-actualization.

The items including marital status, economic dependents, and age group were selected considering that students who bear financial or caregiving responsibilities may face greater challenges in dedicating time and resources to their personal development. Similarly, to reflect academic and cultural background variables, their field of study and indigenous identity, might illuminate inequalities within educational pathways and cultural capital. Students from underrepresented or marginalized groups, such as those from indigenous backgrounds, often encounter systemic obstacles that limit access to resources, mentorship, or academic opportunities.

Secondly, all these dimensions were chosen to collectively capture different aspects of contextual opportunity. Each dimension was first scored separately and then normalized to a 1–7 scale. Together, these dimensions reveal how the intersection of privilege and access creates stratified experiences of meaning-making among university populations.

Thirdly, to quantify contextual differences in opportunity, each response was converted into an integer value and aggregated within its corresponding dimension. To align the dimensions with the PIL test's 7-point scale, each raw dimension score was normalized as shown in Equation 1:

$$\mathrm{Norm}_x = 1 + 6 \times \frac{R_x - R_{\min}}{R_{\max} - R_{\min}} \tag{1}$$

where R_x is the raw score for category x, and $R_{\min}$ and $R_{\max}$ are the theoretical minimum and maximum possible values, respectively.

This transformation ensures that each dimension contributes equally to the composite score and reflects the cumulative impact of contextual opportunity on well-being [28].

Finally, the four resulting scores were added to produce a total index score ranging from 4 to 28, to facilitate analysis and interpretation. The normalized scores for the four dimensions were summed as shown in Equation 2:

$$\mathrm{COI}_{\mathrm{Total}} = N_D + N_C + N_S + N_L \tag{2}$$

where N_D, N_C, N_S, N_L are the normalized scores for the Demographics, Cultural/Academic, Services, and Lifestyle dimensions, respectively.

$$\mathrm{COI}_{\mathrm{Total}} \in [4, 28] \tag{3}$$

The continuous index was then scaled to a [0–1] range and grouped into *five bands of contextual opportunity* according to the results in: Low (0.00–0.15), Moderate (0.16–0.41), Reasonable (0.42–0.58), High (0.59–0.74), and Very high (0.75–1.00) contextual opportunity. Table 2 shows the general results by category and percentage.

The results on Table 2 show that *Low and Moderate opportunities* are more common in the student sample and while it might not necessarily affect their life purpose directly, the next section allowed us to find hidden relations between both PIL and COI results.

After collecting and interpreting the data from all the students, an internal validation of the COI was also conducted. To improve the internal consistency of the COI, a psychometric refinement using a correlation analysis revealed that five items—specifically items 4, 6, 7, 9, and 16—showed weak or even negative correlations with the overall index score. When these items were removed from the scale, it resulted in an increase of the Cronbach's alpha from 0.58 to 0.60. This value represents the minimum threshold commonly accepted for exploratory research and supports the internal coherence of the remaining items. While item 4 and 6 are still relevant for general statistic purposes according to the mental health experts, they might not be part of an updated COI; and items 9 and 16 should be substituted. This refinement serves as a preliminary step toward transitioning from an exploratory framework to a more robust descriptive or explanatory study design in future research, larger groups of students and using probabilistic sampling methods.

4.3 Machine learning algorithms

To explore the predictive relationship between contextual variables and the presence of life purpose among university students, this study employed two supervised

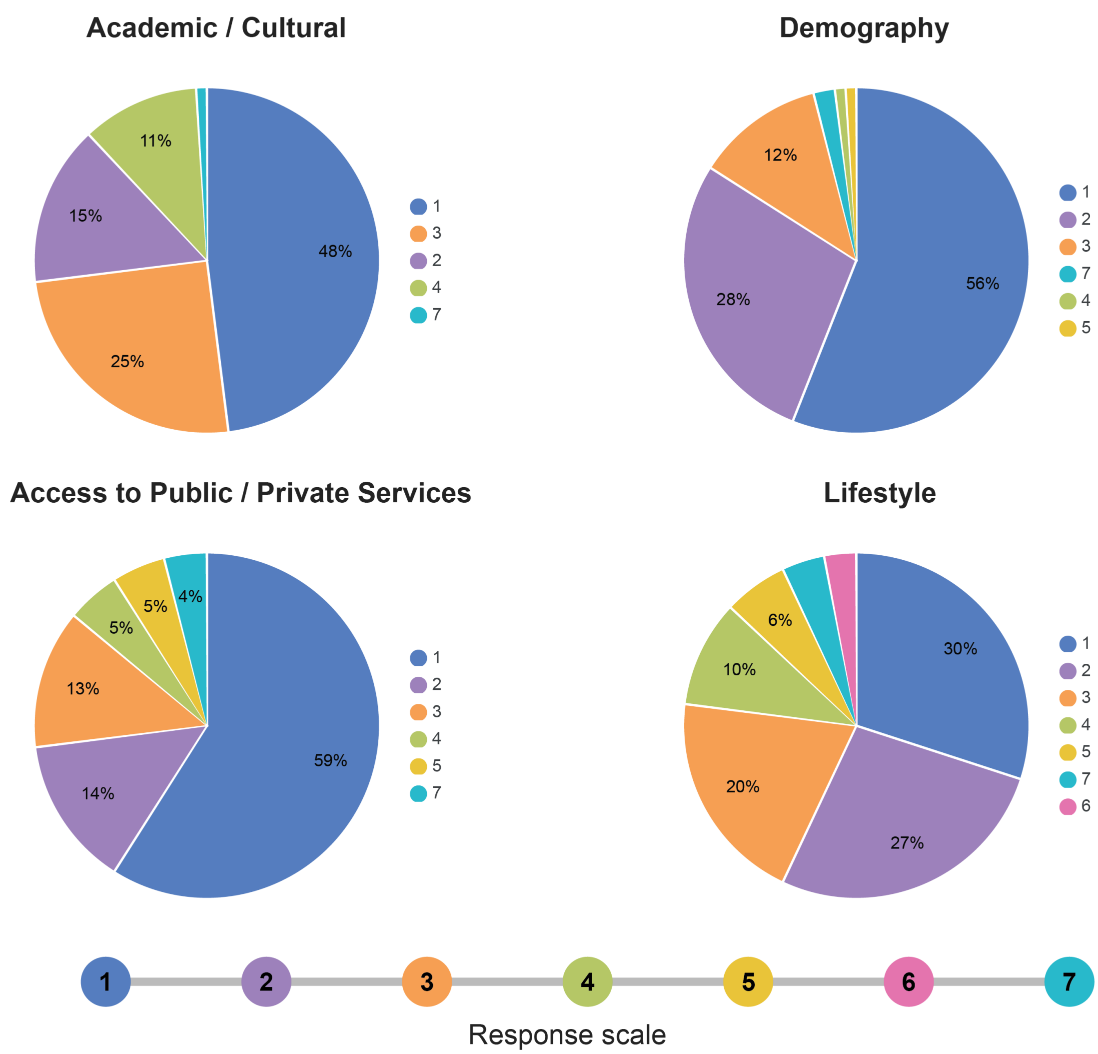


Figure 3. COI scores by category in a 7-point scale created in Microsoft Excel.

Table 2. Opportunity bands by percentage results (2025).

Opportunity Band	Percentage
Low Opportunity	47%
Moderate Opportunity	47%
Reasonable Opportunity	3%
High Opportunity	3%
Very High Opportunity	0%

machine learning models: Naive Bayes and decision tree classifiers. These models were selected for their interpretability and suitability for categorical and ordinal data, which matched the structure of the survey responses.

The Naive Bayes model, grounded in probabilistic reasoning, was particularly useful for handling multidimensional categorical data efficiently and offered rapid results even with a relatively small sample size. In parallel, the decision tree modeling allowed for the identification of nonlinear patterns and interaction effects between COI dimensions as variables, offering interpretable decision rules that aligned with the exploratory goals of the research.

4.3.1 Naive Bayes model in Rapid Miner

In this exploratory sample, the Naive Bayes model demonstrated high classification performance across both iterations. In a first implementation, it achieved an accuracy of 94.05%, with near-perfect classification for the "Defined Life Purpose" and "Existential Vacuum" categories, and minor confusion between "Defined" and "Undefined" cases. The second, improved model reached an accuracy of 96.67%, correctly classifying all cases of "Existential Vacuum" and "Undefined Purpose," with only one misclassification in the "Defined Life Purpose" group. The full process of the implementation is shown in Figure 4. From left to right, 1) an Excel file was imported as a database with four variables, the three categories of the PIL test and the total score of the sum of the COI dimensions, 2) the data was distributed to train the model (70%) and to test it (30%) using two performance tests.

The report visualization tool of RapidMiner showed interesting results, especially in the combination of the PIL results classification (color dots), the COI dimension score and the death acceptance affective component. The results are shown in Figure 5.

4.3.2 Decision tree model in RapidMiner

The decision tree model further reinforced these insights by identifying Services, Demography, Lifestyle, and Cultural engagement, as the most relevant predictors in that order. Participants with greater access to public and private services and with the higher demographic indicators were more likely to exhibit a defined sense of life purpose. In contrast, limited access to services and less structured lifestyle patterns, tended to associate with existential ambiguity. While the visual representation of Figure 6 might be difficult to read, the path from the service access and the demography dimensions at the left side, shows a shorter route.

The Access to private / public services variable serves as a strong initial discriminator, while Academic/Cultural and Lifestyle help refine the classification deeper in the tree. According to the results and considering the 7-scale normalization of the data, a Defined Life Purpose is strongly associated with High levels of Services (> 3.5) and Demography (> 1.5); while with moderate to high levels of Lifestyle and Academic/Cultural scores, especially when Service dimensions is between 2.5 and 3.5 points. It is relevant to see that even when services access might be low, a high Lifestyle score can still predict a defined purpose, potentially linked with the first part of our analysis.

5. Conclusions

The incorporation of the COI alongside the PIL deepened the understanding of university students' sense of purpose by systematically integrating contextual variables, including demographics, academic-cultural background, access to services, and lifestyle. This novel methodological expansion provided an enriched, culturally relevant perspective, allowing the study to link life purpose to broader social and material dimensions. By quantifying the contextual opportunities available to students, the COI highlighted structural inequalities that shape their psychological well-being and existential outlook.

The findings of the study strongly align with Viktor Frankl's logotherapy principles and the existentialist perspective, emphasizing the central role of meaning-making as a protective factor against mental health challenges. Frankl's concept of the will meaning posits that a clear sense of purpose enables individuals to resist despair and find resilience in the face of adversity. This theoretical foundation is evident in the results, which demonstrate how family relationships, structured vacation time, and access to private and public services related to life quality, correlate positively with higher scores of life purpose.

The significant influence of contextual variables, such as structured vacation patterns and nuclear fam-

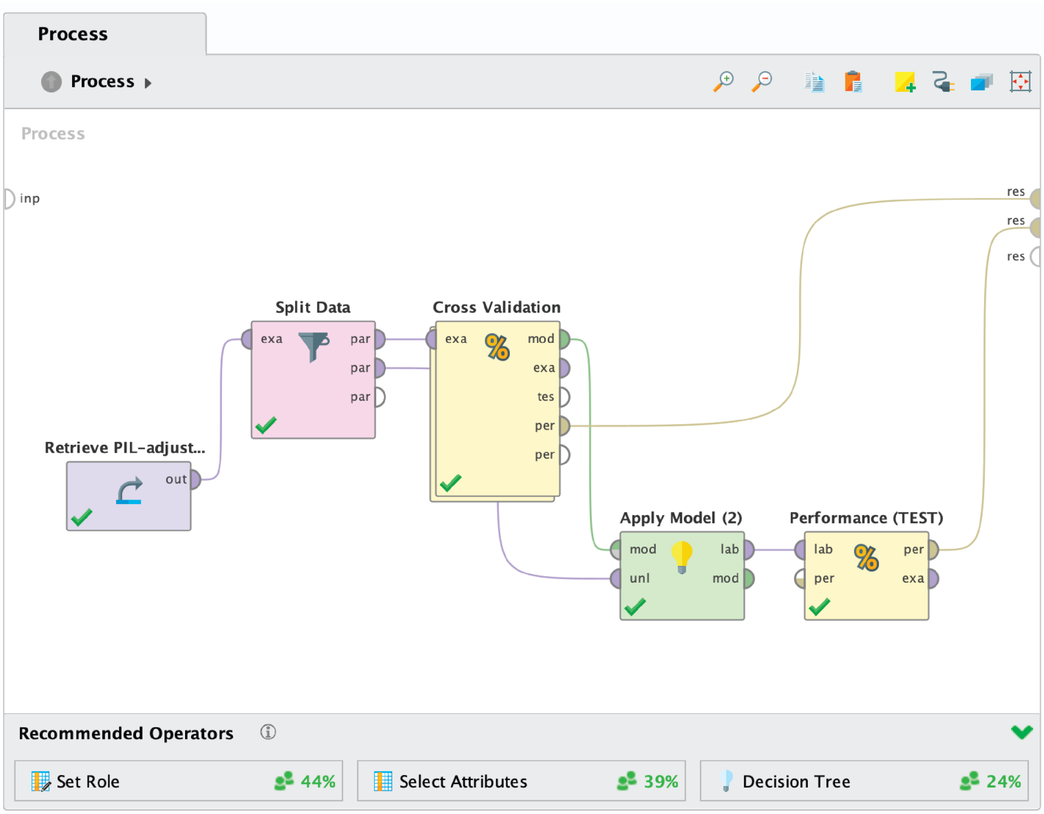


Figure 4. RapidMiner Process Interface with Naive Bayes model.

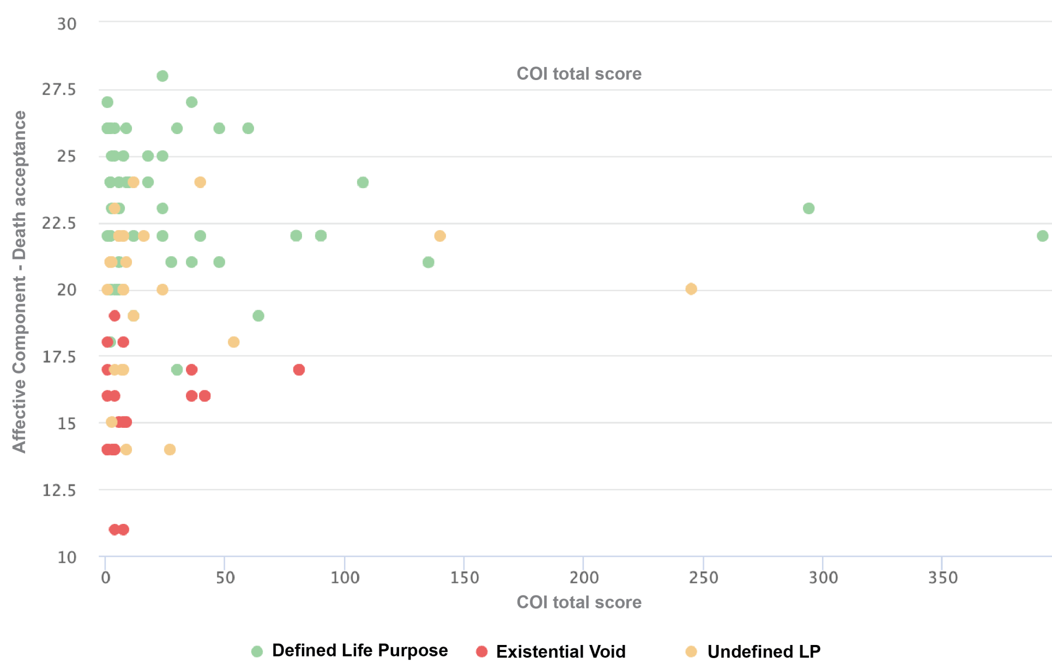


Figure 5. Scatter bubble plot graph generated by RapidMiner comparing COI total score with PIL Score in three colors.

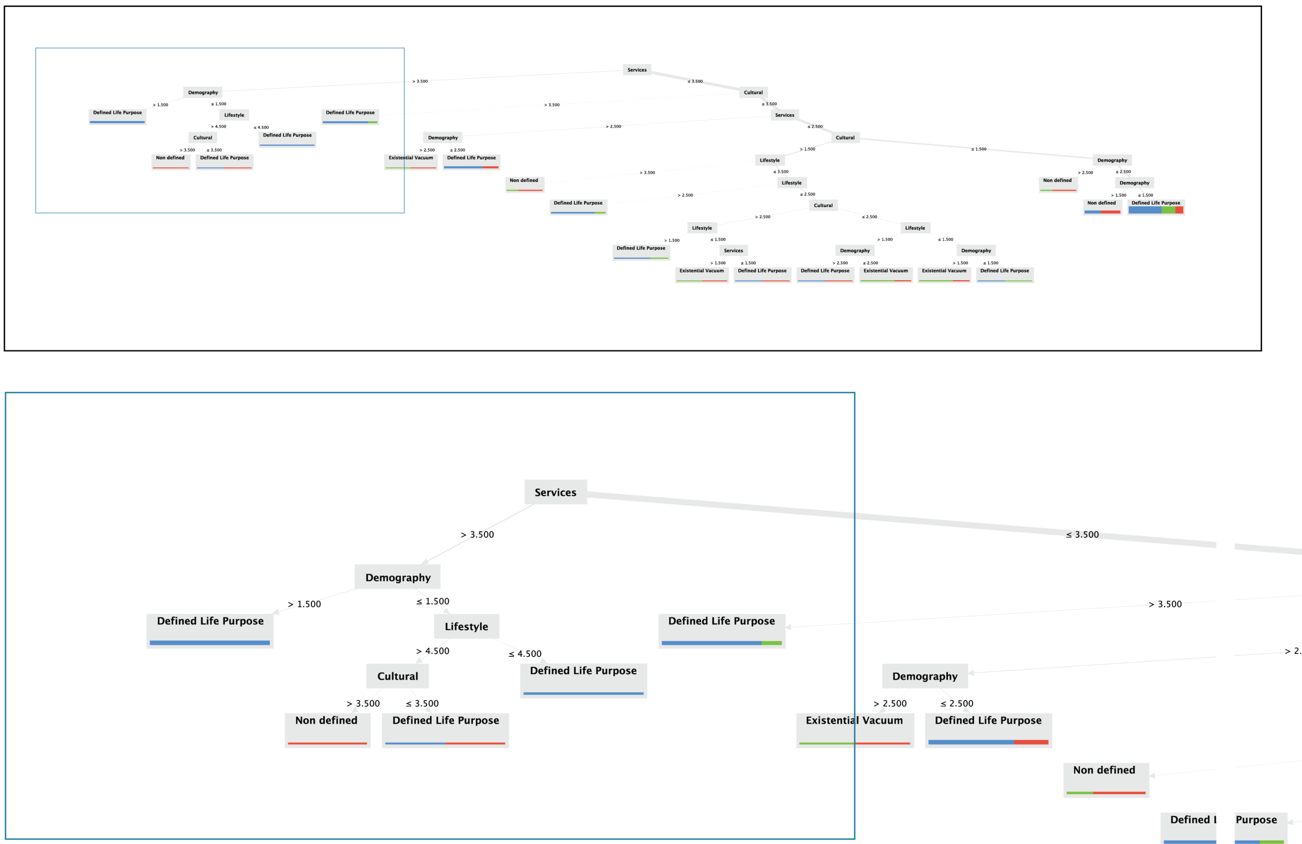


Figure 6. RapidMiner Decision tree results focusing on Services > Demography > Defined Life Purpose.

ily situations, underscores that existential well-being does not arise solely from intrinsic personal effort but is deeply shaped by external social and material conditions. By incorporating the COI, the study not only validates the core tenets of logotherapy but also expands its applicability to diverse cultural and socioeconomic settings, illustrating how structural factors can either hinder or facilitate the pursuit of life purpose. This integration reaffirms the need for culturally informed approaches in applying logotherapy to address mental health issues, particularly within university settings where uncertainty is often magnified.

Regarding key findings for decision-making, the results suggest that access to private and public services, cultural engagement, and structured lifestyle elements may support collaborations with local businesses to offer low-cost transportation, campus food services, and campus life activities. A student life department could directly address these possibilities. Other important elements are the role of the "nuclear family situation" showed a moderate positive correlation with PIL scores ($r = 0.431$), while structured vacation patterns demonstrated a smaller but significant impact ($r = 0.263$). These results underscore the interplay between affective, economic, and organizational factors in cultivating a sense of purpose, emphasizing the need to address these dimensions in university mental health strategies. In this case, the promotion of family activities as community members and of weekend and leisure activities could be added to the list. A list of potential projects for universities is condensed in Table 3.

The study's application of advanced machine learning techniques, including Naive Bayes and decision tree classifiers, validated the relevance of contextual variables in predicting life purpose outcomes. In this exploratory sample, the Naive Bayes model achieved an accuracy of 96.67%, correctly classifying most cases across "Defined Life Purpose," "Existential Vacuum," and "Undefined Purpose" categories. Similarly, decision tree analysis revealed actionable patterns, with "Services," "Demographics," and "Lifestyle" emerging as the most predictive dimensions. These findings suggest the potential of machine learning to generate actionable insights for early identification of at-risk groups and the development of targeted interventions.

As an exploratory study, the 16.7% of participants that experienced significant levels of existential void is an important group to look up for. With broader contextual issues such as lack of support networks, emotional isolation, and limited structured rest contributing to this state, the psychology departments might also consider implementing similar studies to find student with specific need. Overall, these insights call for institutional policies prioritizing connection, belonging, and personal growth within the student body. The results also indicate that high COI scores linked to better access to resources and structured routines can buffer existential uncertainty, reinforcing the importance of addressing structural inequities that impact well-being.

All universities have the potential to replicate this, and other mixed-methods approaches, combining diagnostic tools with machine learning and predictive modeling, to guide innovative and data-driven mental health strategies. These methodologies provide a replicable framework, offering institutions the ability to scale diagnostics from individual interventions to larger populations using digital methods to keep nourishing AI algorithms.

Mental health is not as an individualized challenge but as a collective institutional responsibility encompassing education, family involvement, workload optimization, and normalized self-care practices. The need for cross-cutting institutional strategies that bridge academic, family, and social dimensions is evident to foster meaningful life purpose and emotional resilience among students. As scholars, these conclusions are an invitation to replicate and expand this methodology to additional contextual dimensions, refine intervention models, and promote holistic well-being as a foundational aspect of higher education. By leveraging tools like the COI and predictive analytics, institutions can contribute to addressing the broader systemic challenges impacting mental health and existential purpose in the academic environment.

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 Instituto Tecnológico de Orizaba, TecNM.

CRediT authorship contribution statement

Guillermo Alfredo Arrioja-Carrera: Methodology, Writing – original draft; **Eduardo Roldán-Reyes:** Conceptualization, Supervision; **Pedro Humberto Velázquez-Morales:** Project administration, Writing – review & editing; **Víctor Méndez-García:** Software, Validation.

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

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

Table 3. Potential strategies for university projects, a preventive approach from the results (2025).

Key Finding	Associated Variable	Implication	Institutional Recommendation
57% of students show a defined life purpose, but 16.7% report an existential void	PIL total score	A significant group of students is at risk of existential crisis	Implement periodic PIL screenings to identify at-risk groups and refer them to counseling services
Moderate correlation between family situation and life purpose ($r = 0.43$)	Nuclear family situation	Family environment directly influences life purpose perception	Develop family engagement programs and invite family members to wellness and mental health events
Structured vacation habits positively correlate with life purpose ($r = 0.26$)	Structured vacation time	Rest routines strengthen existential well-being	Establish official academic disconnection weeks and promote healthy leisure planning workshops
94% of the sample showed low or moderate contextual opportunity	COI total score	Students face considerable inequality in material and cultural conditions	Design social mobility programs: transportation, food, scholarships, and access to extracurricular activities
Access to services strongly predicts life purpose according to the decision tree	Access to services (COI)	Basic economic infrastructure enhances sense of life purpose	Provide support for healthcare, internet access, food, and transportation for vulnerable students
Students with high PIL and COI scores show greater well-being	PIL and COI integrated	Psychological well-being and life purpose are deeply shaped by structural factors	Integrate the COI as an institutional well-being indicator to monitor student living conditions
High predictive accuracy of Naive Bayes and decision tree models	AI application	AI can efficiently identify risk profiles	Implement early warning dashboards using AI algorithms to enable personalized prevention

References

[1] World Health Organization, "Promoting mental health: Concepts, emerging evidence, practice," 2004, definition of mental health: "a state of well-being in which individuals are fully aware of their abilities, can cope with the normal stresses of life, work productively, and contribute actively to their communities.". [Online]. Available: https://www.who.int/publications/i/item/9241562943

[2] Observatorio Mexicano de Salud Mental y Adicciones, "Informe anual de salud mental y consumo de sustancias en población joven," Gobierno de México, 2024. [Online]. Available: https://www.gob.mx/salud/saludmental

[3] INEGI, "Encuesta nacional de bienestar autorreportado (enbiare)," 2021. [Online]. Available: https://www.inegi.org.mx/programas/enbiare

[4] M. E. Medina-Mora, G. Borges, C. Benjet, and C. Lara, "La salud mental en méxico: avances y retos," *Salud Pública México*, vol. 65, no. 1, pp. 13–20, 2023. doi: 10.21149/13884 .

[5] M. E. Medina-Mora, S. M. Chainé, J. G. Reyes, E. A. Santillán, R. Sánchez-Huesca, and A. R. Velasco, "Mental health and the covid-19 pandemic: the burden of the disease in the young population and the response of mental health services in mexico," *Salud Pública de México*, vol. 63, no. 6, pp. 613–624, 2021. doi: 10.21149/12738 . [Online]. Available: https://www.scielo.org.mx/scielo.php?script=sci_arttext&pid=S0036-36342021000600813&lng=en&nrm=iso

[6] D. Hernández-Torrano, L. Ibrayeva, J. Sparks, N. Lim, A. Clementi, A. Almukhambetova, Y. Nurtayev, and A. Muratkyzy, "Mental health and well-being of university students: A bibliometric mapping of the literature," *Frontiers in Psychology*, vol. 11, p. 1226, 2020. doi: 10.3389/fpsyg.2020.01226 .

[7] A. Salazar-Pousa, F. Salas, and L. Pineda, "Propósito vital y salud mental en jóvenes universitarios en colombia," *Revista Colombiana de Psicología*, vol. 28, no. 1, pp. 51–67, 2019.

[8] A. G. Gutiérrez-García and C. Londoño, "Sentido de vida y bienestar: Una aproximación desde la logoterapia en estudiantes universitarios," *Revista de Psicología de la Universidad San Marcos*, vol. 23, no. 2, pp. 77–91, 2018.

[9] J. C. Crumbaugh and L. T. Maholick, “An experimental study in existentialism: The psychometric approach to frankl’s concept of noogenic neurosis,” *Journal of Clinical Psychology*, vol. 20, no. 2, pp. 200–207, 1964. doi: 10.1002/1097-4679(196404)20:2¡200::AID-JCLP2270200210¿3.0.CO;2-U .

[10] V. E. Frankl, *Man’s Search for Meaning*. Washington Square Press, 1984.

[11] M. J. Heisel and G. L. Flett, “Purpose in life, satisfaction with life, and suicide ideation in a clinical sample,” *Journal of Psychopathology and Behavioral Assessment*, vol. 26, no. 2, pp. 127–135, 2004. doi: 10.1023/B:JOBA.0000013660.22413.e0 .

[12] S. E. Schulenberg and A. M. A. Melton, “A confirmatory factor-analytic evaluation of the purpose in life test: Preliminary psychometric support for a short form,” *Journal of Happiness Studies*, vol. 11, no. 3, pp. 389–405, 2010. doi: 10.1007/s10902-009-9147-4 .

[13] M. F. Steger, T. B. Kashdan, B. A. Sullivan, and D. Lorentz, “Understanding the search for meaning in life: Personality, cognitive style, and the dynamic between seeking and experiencing meaning,” *Journal of Personality*, vol. 76, no. 2, pp. 199–228, 2008. doi: 10.1111/j.1467-6494.2007.00484.x .

[14] T. Schnell and P. Becker, “Personality and meaning in life,” *Personality and Individual Differences*, vol. 41, no. 1, pp. 117–129, 2006. doi: 10.1016/j.paid.2005.11.030 .

[15] C. V. der Walt, “The relationships between first-year students’ sense of purpose and meaning in life, mental health and academic performance,” *Journal of Student Affairs in Africa*, vol. 7, no. 2, pp. 109–121, 2019. doi: 10.24085/jsaa.v7i2.3828 .

[16] C. Vázquez and D. Castilla, “Adaptación transcultural de instrumentos psicológicos: más allá de la traducción,” *International Journal of Clinical and Health Psychology*, vol. 7, no. 2, pp. 477–492, 2007.

[17] N. Gopalkrishnan, “Cultural diversity and mental health: Considerations for policy and practice,” *Frontiers in Public Health*, vol. 6, p. 179, 2018. doi: 10.3389/fpubh.2018.00179 .

[18] G. E. Vaillant, “Positive mental health: is there a cross-cultural definition?” *World Psychiatry*, vol. 11, no. 2, pp. 93–99, 2012. doi: 10.1016/j.wpsyc.2012.05.006 .

[19] F. Muetunda, S. Sabry, M. L. Jamil, S. Pais, G. Dias, and J. Cordeiro, “Ai-assisted diagnosing, monitoring and treatment of mental disorders: A survey,” *ACM Transactions on Computing for Healthcare*, vol. 5, no. 4, pp. 1–24, 2025. doi: 10.1145/3681794 .

[20] J. García-Alandete, E. R. Martínez, and P. S. Nohales, “Estructura factorial y consistencia interna de una versión española del purpose-in-life test,” *Universitas Psychologica*, vol. 12, no. 2, pp. 517–530, 2013. doi: 10.11144/Javeriana.UPSY12-2.efci .

[21] G. T. Reker and E. J. Peacock, “The life attitude profile (lap): A multidimensional instrument for assessing attitudes toward life,” *Canadian Journal of Behavioural Science*, vol. 13, no. 3, pp. 264–273, 1981. doi: 10.1037/h0081183 .

[22] K. Crenshaw, “Mapping the margins: Intersectionality, identity politics, and violence against women of color,” *Stanford Law Review*, vol. 43, no. 6, pp. 1241–1299, 1991. doi: 10.2307/1229039 .

[23] F. Beroíza-Valenzuela, “The challenges of mental health in chilean university students,” *Frontiers in Public Health*, vol. 12, p. 1297402, 2024. doi: 10.3389/fpubh.2024.1297402 .

[24] G. C. Gee, “Structural racism and health inequities,” *Du Bois Review: Social Science Research on Race*, vol. 10, no. 1, pp. 115–132, 2013. doi: 10.1017/S1742058X1300013X .

[25] S. Alkire and J. Foster, “Counting and multidimensional poverty measurement,” *Journal of Public Economics*, vol. 95, no. 7-8, pp. 476–487, 2011. doi: 10.1016/j.jpubeco.2010.11.006 .

[26] D. C. Butler, S. Petterson, R. L. Phillips, and A. W. Bazemore, “Measures of social deprivation that predict health care access and need within a rational area of primary care service delivery,” *Health Services Research*, vol. 48, no. 2pt1, pp. 539–559, 2013. doi: 10.1111/j.1475-6773.2012.01449.x .

[27] K. J. Zullig, T. M. Koopman, J. M. Patton, and V. A. Ubbes, “School climate: Historical review, instrument development, and school assessment,” *Journal of Psychoeducational Assessment*, vol. 28, no. 2, pp. 139–152, 2010. doi: 10.1177/0734282909344205 .

[28] T. A. DiPrete and G. M. Eirich, “Cumulative advantage as a mechanism for inequality: A review of theoretical and empirical developments,” *Annual Review of Sociology*, vol. 32, no. 1, pp. 271–297, 2006. doi: 10.1146/annurev.soc.32.061604.123127 .

Journal of Artificial Intelligence and Computing Applications (2026) 4(1): 14–22

Journal of Artificial Intelligence and Computing Applications

Research article

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

Gandhi Samuel Hernández-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 Yucatán population, employing machine learning (ML) classification models. Data were 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 79%, achieving 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 suicidal ideation detection.

Keywords: sociodemographic, mental health, suicidal ideation

1. Introduction

The application of machine learning (ML) algorithms for the early detection and prediction of mental disorders has become an increasingly important area of research. This surge in interest is fueled by the growing availability of large datasets, advancements in computational power, and the pressing need to address public mental health challenges.

Several studies have demonstrated the potential of ML in this domain. For instance, [1] explored five ML techniques to predict risk factors for depression and anxiety in Palestinian students, achieving high accuracy with Support Vector Machine (SVM) and Random Forest (RF) models. Other research has shown the effectiveness of neural networks in predicting depressive episodes and the utility of ensemble learning methods in capturing emotional variability. Furthermore, studies have utilized longitudinal data to predict suicide risk and explored the use of biomarkers in conjunction with psychological tests.

While much of the existing research has focused on identifying individuals with mental health disorders or predicting the severity of their symptoms using machine learning, the role of sociodemographic variables in predicting suicidal ideation is less understood. This study aims to address this gap by evaluating the predictive power of sociodemographic variables in conjunction with psychological assessments for the detection of suicidal ideation.

This paper begins by presenting a list of recent work or research related to the use of machine learning-based classification models for the detection of different mental health factors or variables, such as depression, anxiety, and stress. The greatest risk for an individual suffering from these conditions is self-harm or suicide. Therefore, in this paper, ten classification algorithms were trained and tested, focused on detecting individu-

E-mail address: ghernandez@centrogeo.edu.mx
https://doi.org/10.5281/zenodo.20802399
 ISSN: 3061-8843

als without risk (Class 0) and at risk (Class 1) of self-harm based on the C-SSRS Columbia psychological test, which measures suicidal ideation at four levels.

The following section explains the methodology, which consists of the phases of data collection using the APPSI platform, extraction of labeled data from the database, data cleaning and preparation, class balancing, and finally, model training and testing.

The results section below shows the performance of the models after being trained with and without sociodemographic variables. The conclusion section emphasizes the best performing model and explains the importance of the results.

2. Background

In recent years, the use of machine learning (ML) algorithms for the early detection and prediction of mental disorders has gained increasing relevance in scientific research, driven by the availability of large volumes of data, advances in computing, and the urgent need to improve public mental health.

One of the pioneering studies in this field was conducted by [1], who applied five ML techniques: Random Forest (RF), neural networks, decision trees, Support Vector Machine (SVM), and Naive Bayes to predict risk factors associated with symptoms of depression and anxiety in Palestinian students aged 10 to 15. Using a longitudinal database of school behavior, they found that the SVM and RF models achieved remarkable accuracy (SVM: 92.5% for depression and 92.4% for anxiety). Among the most influential factors, they identified school and domestic violence, academic performance, and family income, demonstrating the usefulness of ML for the early prediction of psychological symptoms in school contexts.

That same year, [2] explored the ability of different ML algorithms to predict depressive episodes, finding that neural networks outperformed other models with an accuracy of 97.2%. This study highlights the potential of deep algorithms for classifying complex clinical conditions such as depression, while suggesting the need to incorporate more contextual variables in future models.

Simultaneously, [3] analyzed a dataset based on the results of the DASS42 psychometric test along with demographic variables, comparing multiple ML and ensemble learning algorithms such as AdaBoost, CatBoost, and XGBoost. SVM models achieved the best F1 scores for depression (94%), anxiety (95%), and stress (91%), demonstrating the effectiveness of ensemble and optimal margin models in capturing emotional variability in post-pandemic population contexts.

In another longitudinal study from Japan, [4] developed an ML model based on data from two time points (2016 and 2017) from the National Stress and Health Survey. They trained a suicide risk prediction model using 65 trauma and stress-related variables, achieving an AUC of 0.830. The most relevant risk factors were scores on scales such as the Suicidal Ideation Attributes Scale, the Sheehan Disability Scale, and the PHQ-9. This study not only validated classic risk factors but also proposed new measures related to traumatic experiences as relevant for predicting suicide risk.

Subsequently, [5] used handgrip strength (HGS) readings in combination with the DASS test to detect stress levels. They applied RF and SVM models, concluding that RF performed best with an accuracy of 93.75%. This innovative approach of using physical biomarkers alongside subjective questionnaires offers a path toward more accurate multimodal assessments.

In the work of [6], classical algorithms such as K Nearest Neighbors (KNN), logistic regression, decision tree, Naive Bayes, and SVM were applied to classify anxiety, stress, and depression severity levels in university students. The KNN algorithm proved to be the most effective, highlighting that even traditional methods can yield good results with well-structured data.

In the context of unstructured data analysis, [7] developed deep models to analyze Reddit posts from the "Suicide Watch" and "Depression" communities. With these data, they built models that achieved 95% accuracy in predicting depression. This paper highlights the importance of social media as a valuable source for monitoring psychological risks and proposes an anonymous predictive analytics framework for alerting about suicidal tendencies online.

Karak et al. [8] applied multiple regression and logistic analysis to assess the relationships between suicidality, depression, anxiety, daily stress, and substance use. The analysis revealed a strong positive association between suicidal ideation and depression, followed by anxiety and stress, with a predictive accuracy of 84%. While traditional statistical techniques were used, their work offers key variables to inform more complex ML models in future research.

In a more recent proposal, [9] introduced a longitudinal dataset labeled with DASS21 scores, using acoustic speech samples to predict the severity of depression, anxiety, and stress. They used a one-dimensional convolutional neural network trained on features extracted from the VGG19 model. This approach achieved accurate predictions and suggested that vocal behavior can be a reliable biomarker of psychological disturbances, with potential applications in automated clinical settings.

To improve diagnostic personalization and accuracy, [10] proposed a model based on a multimodal dataset from the NHANES, including demographic, dietary, socioeconomic, and medical data. Five supervised algorithms were implemented: linear regression, SVM, XGBoost, RF, and LASSO. The Random Forest model excelled with an R^2 of 0.93, identifying general and personalized risk factors. This work demonstrates the power of ML in real-life clinical settings and its potential to guide more tailored medical interventions for each patient.

Finally, [11] applied ML and deep learning techniques to detect anxiety in students from EEG signals. They used the BiLSTM algorithm and achieved accuracies of up to 98% on their own database. The finding of increased gamma activity in individuals with high stress highlights new possibilities in computational neuroscience for the study of emotional states.

Taken together, this research demonstrates that machine learning algorithms offer promising solutions for the detection and prediction of mental disorders. The most widely used models include Support Vector Machines, Random Forest, deep neural networks, and ensemble techniques, all of which have been successfully applied in diverse contexts such as school populations, digital platforms, and clinical surveys. The methodological evolution toward the use of longitudinal data, biometric signals, and multimodal analysis reflects a trend toward more accurate, personalized, and clinically applicable models.

3. Methodology

To generate the models presented in this paper, the methodology was divided into several phases.

The first phase consists of data extraction through the APPSI platform. This web-based platform has been adapted to the needs of different institutions but retains the C-SSRS Columbia psychological tests for measuring suicidal ideation, the DASS21 for measuring depression, anxiety, and stress levels, and, finally, the EAYIE (Yucatan Self-Reported Scale of Emotional Intelligence) for detecting protective factors. In addition, the platform includes questionnaires that collect general data such as age, gender, municipality, marital status, number of children, eating habits, sleep, and physical activity, among others.

This information is instantly stored in a relational database, and the responses to the psychological tests are reviewed and scored in real time, making it possible to provide users with their results immediately. Likewise, a data visualization platform was created for each institution, which functions as a mental health observatory for its staff.

The data flow of the APPSI platform can be seen in Figure 1.

Having this information made it possible to select the sociodemographic variables that were used as a basis for defining the characteristics used during the experiments.

The second phase consisted of selecting the sociodemographic variables of greatest interest and extracting information from the database in the tables containing these variables. It is important to note that, as mentioned above, the platform has been used by several institutions, so it was necessary to extract the data separately into CSV files.

Subsequently, using Google Colab, the CSV files were concatenated into a single data frame.

The third phase consisted of cleaning the data to eliminate null data and duplicate records. Because each individual APPSI user is assigned an identification number, this criterion was used to eliminate duplicate records. Furthermore, categorical variables were converted to numeric variables, for example, for the gender and municipality fields.

Because the number of municipalities represented in the sample was large, participants were grouped into two categories: urban and rural. This classification made it possible to compare average levels of depression, anxiety, and stress between both groups. As shown in Figure 2, Figure 3, and Figure 4, participants living in urban areas tended to present higher average scores for depression, anxiety, and stress than those living in rural areas.

These figures report the mean values obtained from the Depression, Anxiety, and Stress Scale–21 (DASS-21). In all three dimensions, scores were represented on a scale from 0 to 4, where 0 indicates the absence of symptoms, 1 indicates mild symptoms, 2 indicates moderate symptoms, 3 indicates severe symptoms, and 4 indicates extremely severe symptoms. Therefore, Figure 2, Figure 3, and Figure 4 illustrate that, within the studied population, urban residence was associated with higher average levels of depression, anxiety, and stress compared with rural residence.

Similarly, for the gender field, only the majority classes, i.e., male and female, were retained because the other classes had limited data and could cause bias.

Furthermore, considering that the Columbia C-SSRS test generates results at four levels of suicidal ideation (0 = No, 1 = Low, 2 = Moderate, 3 = High) and that psychological intervention is necessary at the lowest level, the data were also divided into two classes. Class 0 represents those without risk, and Class 1 represents those at risk (i.e., those with low, moderate, or high levels of ideation).

Finally, given that Class 0 had many more records than Class 1, the final step in this phase consisted of balancing the classes, leaving $n = 919$ records for each class.

The final phase consisted of conducting the experiments. To do this, the selected characteristics were divided into:

- Sociodemographic variables (age, municipality, gender)
- Mental health variables obtained from DASS21 (depression, anxiety, stress)
- Mental health variables obtained from EAYIE (f1, f2, f3)

Due to the differences in the scales, the data were normalized. Ten classification algorithms were tested on these data. The results are shown in the following section.

Using the prepared data, ten classification models were trained using their standard configuration. The results are presented below.

4. Results

This section presents the performance metrics of the ten machine learning classification models evaluated for

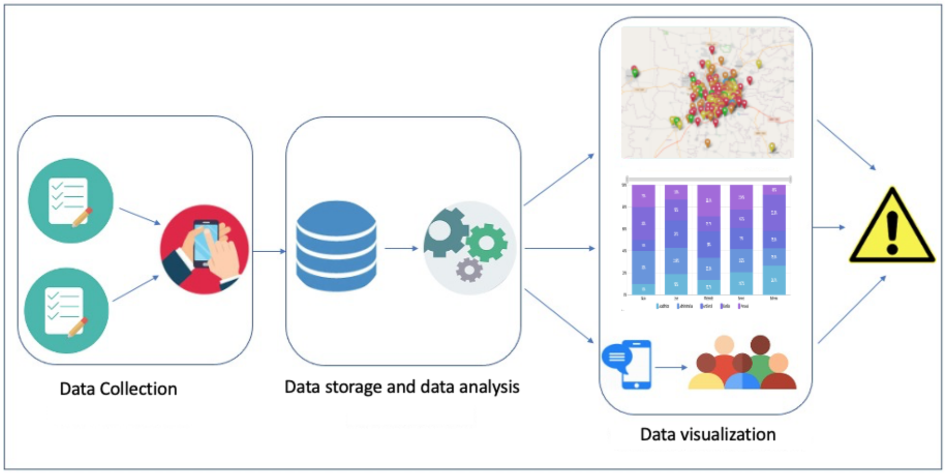


Figure 1. APPSI Data Flow Diagram.

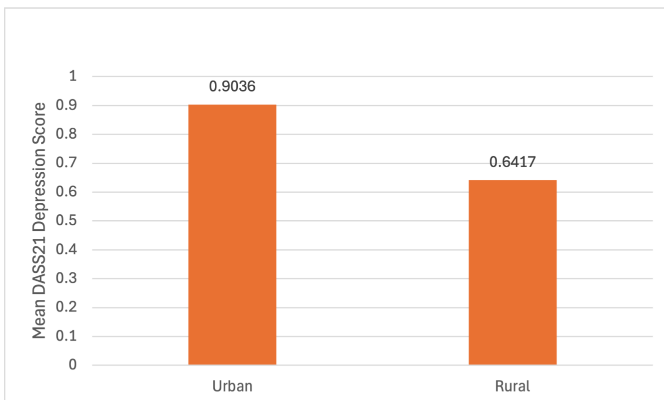


Figure 2. Depression levels difference between rural and urban population.

suicidal ideation detection. The results are divided into two primary experiments: one excluding and one including key sociodemographic variables (age, gender, and municipality). The classification task was binary: Class 0 (without risk) and Class 1 (at risk).

4.1 Model Performance without Sociodemographic Variables

Tables 1 and 2 summarize the performance of the models when trained exclusively on psychological assessment data (DASS21 and EAYIE scores).

In medicine, security, and justice, even small errors can be critical. Sometimes, accuracy is sacrificed to increase recall or precision in the most important class. In Table 1 we can see that the highest recall value is obtained with the Logistic Regression and Naive Bayes models (recall = 0.83), however, the highest accuracy value is 0.78 with the AdaBoost model. Nevertheless, this model generates a recall value of 0.80.

On the other hand, the highest precision and F1-score values were obtained with the SVM and MLP models (precision = 0.78) and Logistic Regression and AdaBoost (F1-score = 0.79).

When evaluating the critical "at risk" (Class 1) category, the highest recall value of 0.77 was achieved by the SVM, MLP, and Random Forest models. The AdaBoost model, however, demonstrated the highest overall accuracy at 0.78.

Table 2 shows the values of the metrics for the "at risk" class after training the models with no sociodemographic variables.

4.2 Model Performance Including Sociodemographic Variables

After training and testing the models without taking into account sociodemographic variables (age, gender, and the urban/rural municipality), these variables were added and the models were retrained. Tables 3 and 4 show the performance metrics after this inclusion.

Table 4 shows the values of the metrics for the "at

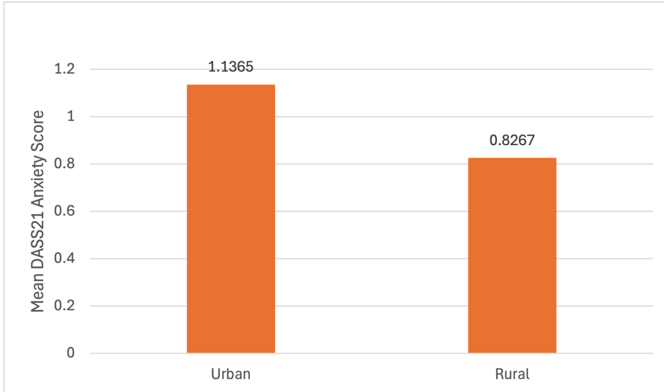


Figure 3. Anxiety levels difference between rural and urban population.

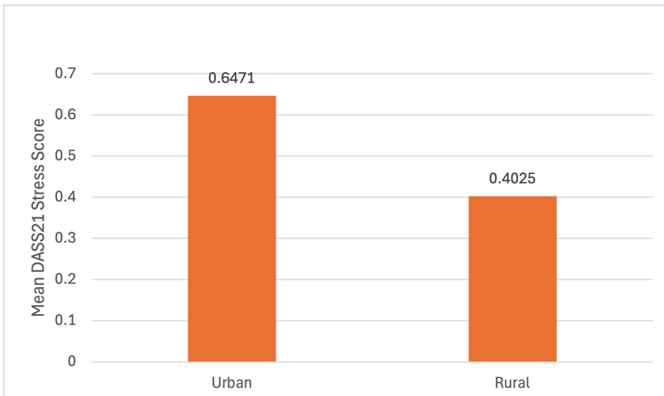


Figure 4. Stress levels difference between rural and urban population.

risk" class after training the models including the sociodemographic variables.

The inclusion of sociodemographic variables produced modest changes in model performance and improved some class-specific metrics, particularly recall for the without-risk class. Random Forest achieved the highest recall for the "at risk" class at 0.79, while Logistic Regression yielded the highest overall accuracy at 0.79. Most notably, the Logistic Regression model's recall for the "without risk" class (Class 0) increased to 0.84.

5. Conclusion

Based on data obtained through the APPSI platform that included psychological variables (depression, anxiety, stress, emotional intelligence) and key sociodemographic variables (age, gender, and urban/rural municipality), ten classification models were trained to classify suicidal ideation in Yucatan population. The classes were balanced to ensure a robust and fair evaluation of model performance. The Logistic Regression model provided the best overall performance across the metrics. Its final metrics, incorporating sociodemographic variables, are presented below in Table 5.

The results underscore the value of the recall metric in this application, as a false negative (failing to identify an individual at risk) carries severe clinical consequences. When models were trained without sociodemographic variables, the maximum recall for the "at risk" class was 0.77. However, the inclusion of sociodemographic variables improved some class-specific metrics. This improvement highlights the critical role of sociodemographic variables in enhancing the predictive power of machine learning models for suicidal ideation detection, placing the model within clinically acceptable performance ranges.

When developing machine learning models to detect individuals at risk of suicide using psychometric scales such as the DASS-21 (Depression, Anxiety, and Stress Scales) and the C-SSRS (Columbia-Suicide Severity Rating Scale), it is essential to evaluate their performance using appropriate metrics. These metrics not only quantify the model's effectiveness but also ensure the clinical safety of its application, given the high cost of a false negative in this field. The recall metric

Table 1. Results for the *without risk* class excluding sociodemographic variables

Model	Precision	Recall	F1-score	Accuracy
Logistic Regression	0.75	0.83	0.79	0.77
Naive Bayes	0.72	0.83	0.77	0.75
KNN	0.70	0.71	0.70	0.69
SVM	0.78	0.77	0.77	0.77
MLP	0.78	0.77	0.77	0.77
CNN	0.76	0.77	0.76	0.76
Decision Tree	0.72	0.68	0.70	0.70
Random Forest	0.77	0.73	0.75	0.75
AdaBoost	0.78	0.80	0.79	0.78
XGBoost	0.70	0.70	0.70	0.69

Table 2. Results for the *at risk* class excluding sociodemographic variables

Model	Precision	Recall	F1-score	Accuracy
Logistic Regression	0.80	0.71	0.75	0.77
Naive Bayes	0.79	0.66	0.72	0.75
KNN	0.69	0.67	0.68	0.69
SVM	0.76	0.77	0.76	0.77
MLP	0.76	0.77	0.76	0.77
CNN	0.75	0.74	0.75	0.76
Decision Tree	0.68	0.72	0.70	0.70
Random Forest	0.73	0.77	0.75	0.75
AdaBoost	0.78	0.76	0.77	0.78
XGBoost	0.69	0.68	0.68	0.69

(Sensitivity or True Positive Rate) indicates how many of the truly positive cases were correctly detected. In mental health, and particularly in suicidology, this is the most critical metric, as failure to detect a case of suicidal risk can have fatal consequences. Clinical literature suggests prioritizing models with high recall, even at the cost of increasing false positives [12].

The model showed useful preliminary screening performance, with recall values of 84% for Class 0 and 74% for Class 1. It is important to note that while artificial intelligence offers powerful support for technical tasks like screening and initial risk assessment, it is not a substitute for the essential human element of psychological practice, which is grounded in empathy, consciousness, and the unique therapeutic relationship. In later stages, the implementation of probabilistic calibration and stratified cross-validation techniques will be explored to ensure its robustness in different cultural contexts.

From a humanistic perspective, artificial intelligence (AI) cannot replace the core essence of psychological work, which is grounded in authentic human relationships characterized by empathy, unconditional acceptance, and genuine presence. While AI can support technical tasks such as data collection, initial screening, and clinical decision-making, it lacks subjectivity, consciousness, and the ability to understand the emotional, symbolic, and ethical depth of the therapeutic encounter. Psychological practice also involves complex moral decisions and a profound recognition of each individual's uniqueness, which AI cannot replicate. Unconscious processes, symbolic communication, and emotional resonance require sensitivity and self-awareness that current AI systems do not possess. Therefore, the most ethical and human-centered approach is not to replace therapists with AI, but to integrate it as a supportive ally—expanding access, reducing technical burdens, and reinforcing the relational core of care. AI should be viewed as a complementary tool, not a substitute for the human therapeutic relationship.

Table 3. Results for the *without risk* class including sociodemographic variables

Model	Precision	Recall	F1-score	Accuracy
Logistic Regression	0.77	0.84	0.80	0.79
Naive Bayes	0.74	0.84	0.79	0.77
KNN	0.74	0.79	0.77	0.75
SVM	0.76	0.79	0.78	0.76
MLP	0.76	0.79	0.78	0.76
CNN	0.77	0.79	0.78	0.77
Decision Tree	0.65	0.54	0.59	0.62
Random Forest	0.78	0.70	0.74	0.74
AdaBoost	0.78	0.80	0.79	0.78
XGBoost	0.78	0.75	0.77	0.76

Table 4. Results for the *at risk* class including sociodemographic variables

Model	Precision	Recall	F1-score	Accuracy
Logistic Regression	0.81	0.74	0.77	0.79
Naive Bayes	0.80	0.69	0.74	0.77
KNN	0.77	0.71	0.74	0.75
SVM	0.77	0.73	0.75	0.76
MLP	0.77	0.73	0.75	0.76
CNN	0.77	0.77	0.77	0.77
Decision Tree	0.60	0.71	0.65	0.62
Random Forest	0.72	0.79	0.76	0.75
AdaBoost	0.78	0.77	0.77	0.78
XGBoost	0.75	0.78	0.76	0.76

Ethics Statement

Given the sensitive nature of the data collected, all procedures conducted in this study strictly adhered to ethical standards to ensure the privacy, confidentiality, and security of participants' information. Participation in the survey was entirely voluntary, and no personally identifiable data were disclosed at any stage. Prior to beginning the study, each participant was presented with an informed consent form detailing the objectives, scope, and purpose of the research, as well as the measures implemented to safeguard their data. Only participants who provided explicit consent proceeded with the study.

CRediT authorship contribution statement

Gandhi Samuel Hernández-Chan: Conceptualization, Methodology, Software, Formal Analysis, Investigation, Resources, Data Curation, Writing—Original Draft Preparation, Writing—Review & Editing. **Matilde Jiménez-Coello:** Conceptualization, Methodology, Validation, Formal Analysis, Investigation, Resources, Writing—Original Draft Preparation, Supervision. **Manuel Sosa-Correa:** Conceptualization, Methodology, Validation, Formal Analysis, Investigation, Resources, Writing—Original Draft Preparation, Supervision. **Sally Vanega-Romero:** Conceptualization, Methodology, Validation, Formal Analysis, Investigation, Resources, Writing—Original Draft Preparation, Supervision. All authors have read and agreed to the published version of the manuscript.

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

No generative artificial intelligence or AI-assisted technologies were used in the drafting, editing, or composition of this manuscript. However, AI-assisted tools were employed for literature search and information retrieval

Table 5. Best model results

Metric	***Without risk*** **class**	***At risk*** **class**
Precision	0.77	0.81
Recall	0.84	0.74
F1-score	0.80	0.77
Accuracy	0.79	0.79

purposes to support the development of the Background section and to review the wording of some paragraphs. All interpretations, analyses, and conclusions presented herein were produced solely by the authors.

Declaration of competing interest

The authors declare no competing interests.

Acknowledgements

The team is extremely grateful to all the more than two thousand individuals who voluntarily agreed to participate in this study, as well as to all the companies and institutions that have joined in the efforts to protect the mental health of their collaborators. We would also like to extend a special thanks to the Sofia Association, which has greatly promoted this project with its strong social commitment. We also thank the Secretariat of Science, Humanities, Technology, and Innovation (SECIHTI) for supporting this project, allowing us to once again put science and technology at the service of society.

References

[1] R. Qasrawi, S. P. V. Polo, D. A. Al-Halawa, S. Hallaq, Z. Abdeen *et al.*, "Assessment and prediction of depression and anxiety risk factors in schoolchildren: machine learning techniques performance analysis," *JMIR formative research*, vol. 6, no. 8, p. e32736, 2022. doi: 10.2196/32736 .

[2] N. K. Trivedi, R. G. Tiwari, D. Witarsyah, V. Gautam, A. Misra, and R. A. Nugraha, "Machine learning based evaluations of stress, depression, and anxiety," in *2022 International Conference Advancement in Data Science, E-learning and Information Systems (ICADEIS)*. IEEE, 2022, pp. 1–5. doi: 10.1109/ICADEIS56544.2022.10037336 .

[3] S. Singh, H. Gupta, P. Singh, and A. P. Agrawal, "Comparative analysis of machine learning models to predict depression, anxiety and stress," in *2022 11th International Conference on System Modeling & Advancement in Research Trends (SMART)*. IEEE, 2022, pp. 1199–1203. doi: 10.1109/SMART55829.2022.10047752 .

[4] P.-H. Chou, S.-C. Wang, C.-S. Wu, M. Horikoshi, and M. Ito, "A machine-learning model to predict suicide risk in japan based on national survey data," *Frontiers in psychiatry*, vol. 13, p. 918667, 2022. doi: 10.3389/fpsyt.2022.918667 .

[5] S. Usman, F. Rusli, S. Z. A. Jalil, and N. A. Bani, "Depression anxiety stress scale and handgrip using machine learning analysis," in *2022 4th International Conference on Smart Sensors and Application (ICSSA)*. IEEE, 2022, pp. 76–80. doi: 10.1109/ICSSA54161.2022.9870948 .

[6] S. S. Malik and A. Khan, "Anxiety, depression and stress prediction among college students using machine learning algorithms," in *2023 Second International Conference on Electrical, Electronics, Information and Communication Technologies (ICEEICT)*. IEEE, 2023, pp. 1–5. doi: 10.1109/ICEEICT56924.2023.10157693 .

[7] Y. Li, "Depression and suicide risk prediction based on machine learning models," *Journal of education, humanities and social sciences*, vol. 15, pp. 302–307, 2023. doi: 10.54097/ehss.v15i.9312 .

[8] G. Karak, K. Ghosh, N. Ansari, S. Das, and A. Chaudhuri, "Prediction of suicidal tendencies using machine learning," *International Journal of Engineering Technology and Management Sciences*, vol. 7, no. 2, pp. 38–46, 2023. doi: 10.46647/ijetms.2023.v07i02.006 .

[9] M. Tasnim, R. D. Ramos, E. Stroulia, and L. A. Trejo, "A machine-learning model for detecting depression, anxiety, and stress from speech," in *ICASSP 2024-2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2024, pp. 7085–7089. doi: 10.1109/ICASSP48485.2024.10446567 .

[10] M. H. Amirhosseini, A. L. Ayodele, and A. Karami, "Prediction of depression severity and personalised risk factors using machine learning on multimodal data," in *2024 IEEE 12th International Conference on Intelligent Systems (IS)*. IEEE, 2024, pp. 1–7. doi: 10.1109/IS61756.2024.10705185 .

[11] V. M. Joshi, D. D. Kulkarni, and N. J. Uke, “Stress and anxiety detection: deep learning and higher order statistic approach,” *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 33, no. 3, p. 1567, 2024. doi: 10.11591/ijeecs.v33.i3.pp1567-1575 .

[12] J. C. Franklin, J. D. Ribeiro, K. R. Fox, K. H. Bentley, E. M. Kleiman, X. Huang, K. M. Musacchio, A. C. Jaroszewski, B. P. Chang, and M. K. Nock, “Risk factors for suicidal thoughts and behaviors: A meta-analysis of 50 years of research.” *Psychological bulletin*, vol. 143, no. 2, p. 187, 2017. doi: 10.1037/bul0000084 .

Journal of Artificial Intelligence and Computing Applications (2026) 4(1): 23–30

MAIKRON

Journal of Artificial Intelligence and Computing Applications

Review article

The current role of artificial intelligence in cyberattacks: a short narrative review

Adrian Yama-Uitz[1], **Miguel Cutz-Pech**[1], **Jose Munoz-May**[1], **Leonardo Romero-Pech**[1], **Jarol Lizama-Chan**[1], **and Zyon Maximo Rodriguez-Gonzalez**[1]

[1]Tecnológico Nacional de México/IT de Mérida, Yucatán, Mexico

ABSTRACT

Artificial Intelligence (AI) exerts a growing impact on the field of cybersecurity, which is characterized by increasing complexity resulting from system interconnectivity, the expansion of the Internet of Things (IoT), and the rise in digital threats. While AI has been widely used to improve defense systems, its application in cyberattacks represents an emerging challenge. Malicious actors are leveraging this technology to automate offensives, adapt to defenses in real time, and exploit vulnerabilities with unprecedented efficiency, raising questions about how it is changing the dynamics of current attacks. This review examines five recent empirical studies published between 2021 and 2025 that address the offensive use of AI in cybersecurity, identifying key techniques such as automated code generation, evasion of detection mechanisms, execution of targeted attacks, and the reduced technical skill requirements for attackers. The results reveal that AI is transforming cybersecurity strategies, enabling more sophisticated and adaptive attacks that pose significant challenges to current detection tools. These threats are particularly critical in sectors such as healthcare, critical infrastructure, and public administration, where disruptions can have systemic consequences. The results point to the need to advance the development of more adaptable and scalable defense systems, as well as to identify key research gaps in areas such as explainability, regulation, and real-time detection of AI-driven threats.

Keywords: artificial intelligence, cyber attacks, cybersecurity vulnerabilities

1. Introduction

Artificial Intelligence (AI) has been recognized for its potential to optimize critical processes in areas as diverse as medicine, engineering, and finance [1, 2]. This transformation is underpinned by its ability to process large volumes of data, identify complex patterns, and optimize decision-making [2]. However, the same potential that drives these advances poses new challenges when applied to the field of cybersecurity, where its capabilities can be leveraged for different purposes [3].

AI is not just a complementary tool in cybersecurity, but a fundamental pillar to address the speed and sophistication of current and future threats [4]. Despite efforts to address this emerging trend, the scale and pace of its implementation, the speed of its advancement, and the specific methods used are not yet fully understood [5, 6]. This lack of understanding highlights the need for an analysis of how AI is redefining offensive capabilities in the field of cybersecurity.

This review seeks to examine the current use of AI in cyberattacks, with particular attention to how malicious actors employ it to automate attacks, evade detection, and exploit vulnerabilities in increasingly complex and interconnected systems. This paper does not propose defense mechanisms but rather synthesizes the current

E-mail address: le25080033@merida.tecnm.mx
https://doi.org/10.5281/zenodo.20802486

literature on the offensive use of AI. By analyzing recent studies, it seeks to shed light on how these applications expand the reach and efficiency of cyberthreats, challenging traditional security paradigms and showing how AI improves the operational execution of modern cyberattacks with minimal human intervention.

The remainder of this article is structured as follows: Section 2 details the approach used for literature selection and analysis, describing the databases consulted, the inclusion and exclusion criteria, and the search strategies employed. Furthermore, the use of the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) model for literature selection is presented, along with the criteria applied. Subsequently, Section 3 presents the main emerging trends and techniques in the use of AI in cyberattacks, including automation methods, detection evasion, and vulnerability exploitation. Section 4 delves into the impact of these findings, assessing their implications for cybersecurity and considering the challenges posed by the use of AI in malicious activities. Finally, Section 5 summarizes the key points of the analysis, addresses potential mitigation strategies, and suggests future directions for research in this field. This framework allows for a comprehensive approach to the role of AI in cyberattacks, combining a critical review of the literature with an analysis of its practical and theoretical implications.

2. Methodology

To ensure an organized, transparent, and reproducible approach to identifying relevant literature, this review adopts the fundamental principles of the PRISMA methodology [7]. While not a systematic review in the strict sense, this framework serves as a methodological guide to structure a clear and replicable selection process, aiming to minimize potential bias in the synthesis of information. The primary objective is to identify and analyze recent relevant studies that significantly contribute to the understanding of the role of AI in the execution of cyberattacks. Through a rigorous selection process, this review ensures that the included papers are relevant, methodologically sound, and reflect the latest advances in this emerging field.

The bibliographic search was conducted using Google Scholar, selected for its broad coverage and ease of access to locate academic articles in different areas of knowledge. A specific search query was designed to ensure the relevance of the retrieved studies: ("artificial intelligence") AND ("cyberattacks") AND ("automation OR detection evasion") AND ("cybersecurity vulnerabilities OR malicious actors") AND ("emerging techniques OR cybersecurity trends"). This query aimed to collect studies from 2021 onwards related to AI-driven cyberattacks, with a particular focus on emerging threats. This search yielded a total of 135 records, which formed the initial dataset for subsequent analysis.

During the first selection phase, a thorough review of the titles and abstracts of the retrieved studies was conducted to eliminate those that did not meet the specific objectives of the review. Three main exclusion criteria were established: first, studies that, although related to technological topics, did not explicitly address the intersection between AI and cyberattacks were eliminated; second, works not published in scientific journals were discarded, ensuring the reliability and quality of the sources consulted; and third, review articles were excluded, given that the purpose of this research was to focus on primary research contributions that offered empirical data and original results. This filtering process ensured that only research of direct relevance to the study of AI-motivated cyberattacks was selected, resulting in the exclusion of 117 studies and the inclusion of 18 articles, which were evaluated for further analysis.

During the eligibility phase, the 18 shortlisted articles were thoroughly reviewed to ensure their methodological rigor, thematic relevance, and potential impact. In this second stage, 13 studies were excluded for the following reasons: one article was excluded due to language restrictions, eleven were not indexed in recognized indexes such as Journal Citation Reports (JCR), and one did not fit the thematic focus of this review. As a result, the final set consisted of five articles that fully met the established criteria. Figure 1 presents a flowchart adapted from the PRISMA model that summarizes the entire process of study identification, selection, eligibility assessment, and inclusion.

The five selected articles represent recent, high-impact contributions at the intersection of AI and cyberattacks. Table 1 presents a detailed summary of these studies, including the article title, journal of publication, and corresponding year. While this review was developed following a structured approach based on the principles of this established methodology, certain methodological limitations were identified. The exclusive use of Google Scholar as a search source could limit the scope and comprehensiveness of the review. Likewise, the focus on research published in recent years could have excluded relevant previous work. Similarly, the exclusion of non-indexed journals could have overlooked relevant contributions from emerging researchers or from specialized fields. Despite these limitations, the selection process, following a systematic approach, ensures that the final set of studies includes high-quality, relevant, and methodologically sound research.

3. Thematic Overview

This section critically organizes, analyzes, and synthesizes key findings from selected literature on the use of AI in cyberattacks, with particular attention to emerging techniques and research gaps. Rather than simply summarizing individual studies, this section identifies key patterns, trends, and areas of debate that emerged during the review. The thematic analysis focuses on four major domains: the automation and escalation of cyberattacks enabled by AI-driven orchestration; the automated generation of malicious code through large language models (LLMs); the use of adaptive AI tech-

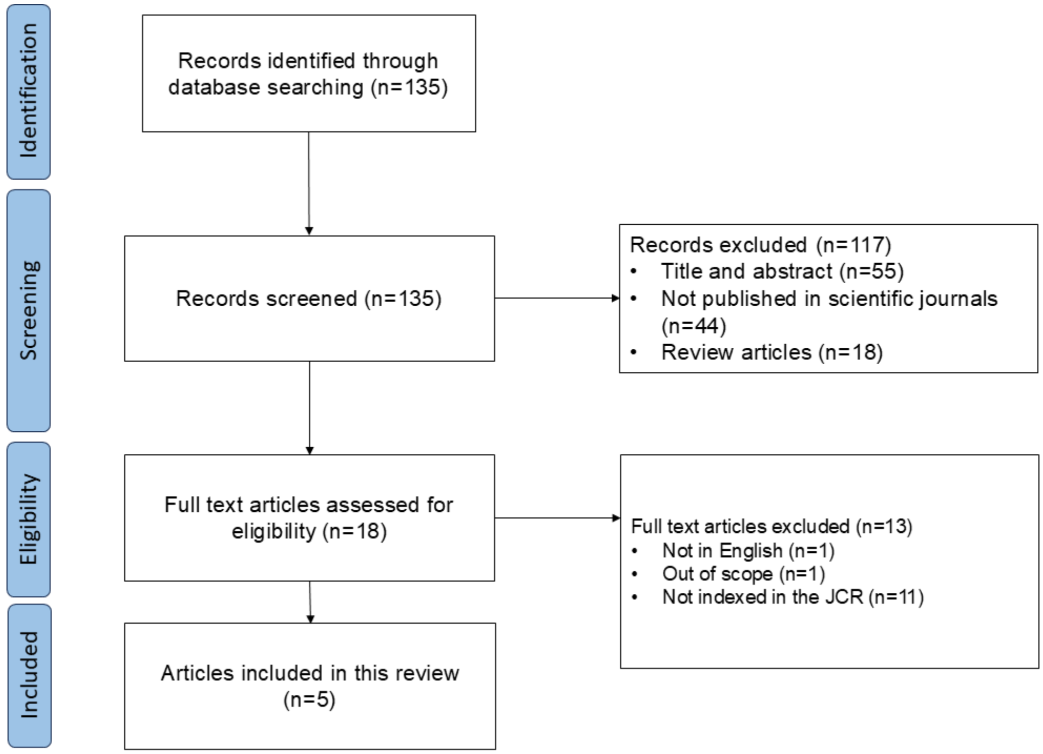


Figure 1. Flowchart of the literature identification, screening, eligibility assessment, and inclusion process.

Table 1. The final list of articles used in this review, including information for title, journal, year of publication and citation.

Title	Journal	Year	Citation
Unleashing offensive artificial intelligence: Automated attack technique code generation	Computers & Security	2024	[1]
Designing and Evaluating a Flexible and Scalable HTTP Honeypot Platform: Architecture, Implementation, and Applications	Electronics	2023	[5]
ChatGPT or Bard: Who is a better Certified Ethical Hacker?	Computers & Security	2024	[8]
Leveraging Explainable Artificial Intelligence in Real-Time Cyberattack Identification: Intrusion Detection System Approach	Applied Sciences-Basel	2023	[9]
A CNN-based automatic vulnerability detection	EURASIP Journal on Wireless Communications and Networking	2023	[10]

niques to evade intrusion detection and monitoring systems; and the challenges and opportunities for AI-Based cyberattack detection and attribution. Each of these themes is examined in depth in the following subsections, highlighting influential findings, unresolved questions, and areas where further research is necessary. A summary of these thematic domains is presented in Table 2.

3.1 Automation and escalation of AI-driven attacks

Advances in AI have made it possible to automate multiple phases of the cyberattack lifecycle, including reconnaissance, exploitation, and evasion of defense systems [8]. Generative models and adaptive architectures allow attackers to dynamically modify their attack vectors, making them difficult to detect by traditional systems [5]. Automation not only accelerates the deployment of attacks but also enables their mass replication without human intervention [1, 8, 5]. This poses a significant threat to environments that are not prepared for immediate responses to thousands of simultaneous events [5].

Current defensive capabilities do not adequately respond to this level of adaptability. While attackers integrate self-tuning and personalizing algorithms, defenses remain reactive, slow, and structurally static, as many intrusion detection systems lack simultaneous adaptability and rely on outdated datasets that do not represent evolving threat patterns [9]. Furthermore, most architectures still rely on fixed rules or non-adaptive models, making them poor at responding to dynamic, behavioral attacks generated by AI [10]. This asymmetry creates an operational gap in cybersecurity that not only compromises current defense strategies [1], but also exposes a broader research challenge: while offensive capabilities are increasingly driven by adaptive automation, existing defensive models have not yet reached a comparable level of responsiveness in concurrent environments [9, 5].

3.2 Automated generation of cyberattack techniques through large language models

AI has reached a point where it can be used not only to automate defensive tasks in cybersecurity, but also to facilitate the execution of cyberattacks [8]. The study [1] shows how generative models like ChatGPT can be prompted to produce malicious code aligned with MITRE ATT&CK techniques. These models allow the creation of specific code fragments to execute advanced attack techniques such as defense evasion, privilege escalation, or initial access, with just a textual instruction [1]. This automation considerably reduces the effort required by attackers and eliminates the need for specialized technical knowledge [1, 9], which broadens the range of users capable of carrying out sophisticated attacks.

In controlled tests, some LLMs have been shown to autonomously generate attack code with a 16% success rate, a figure that rises to 50% when guided by users with cybersecurity expertise [1]. This result highlights the tangible offensive potential of LLMs and the urgent need to develop early mitigation mechanisms against their malicious use [8].

The ability of these models to generate executable code, adapted to different operating systems and offensive tactics, also allows them to be integrated with automated breach and attack simulation platforms [9, 8]. This favors the customization of attacks according to the target environment, making their detection and response more difficult. Furthermore, the dynamic generation of variants of the same attack increases the complexity of threats, forcing defensive systems to evolve rapidly. In this context, LLMs represent a powerful tool that is transforming the offensive cybersecurity landscape, both due to their speed and operational versatility [8].

3.3 Evasion of monitoring and detection systems in cyberattacks using AI

Evading monitoring and detection systems is one of the most complex areas where attackers can leverage AI to optimize their attacks [5]. Traditionally, intrusion detection systems (IDS) and other monitoring tools have been designed to identify anomalous behavior in network traffic or unusual patterns in systems [5, 8]. However, with the use of AI, attackers can now generate adaptive attacks that instantly modify their behavior to evade defense strategies [1, 8].

One of the most advanced techniques is the use of AI algorithms to anticipate defensive responses and adapt attack strategies accordingly, enabling real-time evasion of monitoring mechanisms [1, 8]. This type of dynamic adaptation poses a critical problem for defenses: how to anticipate and detect constantly changing threats. Traditional detection systems are not designed to react so quickly or to predict changes in attack patterns [5, 9]. This creates a security gap, where attackers continue to advance while traditional defenses fall behind.

3.4 Challenges and Opportunities for AI-Based Cyberattack Detection and Attribution

The increasing sophistication of cyberattacks, driven by the integration of AI, poses unprecedented challenges in the detection and attribution of these threats. AI's ability to automate entire phases of the attack lifecycle, from reconnaissance to evasion of defenses, requires a deeper understanding of traditional security strategies [8]. Unlike conventional attacks, the dynamic adaptability inherent in AI-based offensive systems allows for modification of tactics and vectors even during execution, making them considerably more difficult to identify using systems based on static signatures or predefined rules [5]. Studies demonstrate that these adaptive strategies weaken detection models by allowing attacks to modify behavior in real time, circumventing traditional rule-based mechanisms [5].

In this scenario, where attackers are leveraging the automatic personalization and scalability offered by AI, it is crucial to explore both the challenges inherent in detecting and tracking these new forms of evasive attacks and the opportunities that AI itself can offer to develop more advanced, predictive, and effective forensic techniques [9]. A proactive response to these emerging threats lies in the implementation of machine learning models such as convolutional neural networks (CNNs)[10], which have proven effective in automatically detecting vulnerabilities in source code. This technique represents a crucial tool for preventing exploits before they materialize [10], thus helping to reduce the gap in the AI-driven offensive-defensive cycle. However, despite recent advances, existing detection systems still have limitations in real-time attribution, especially when faced with rapidly evolving and adaptive attack patterns [1].

As previously discussed, AI has brought about a substantial shift in cybersecurity strategies, equally affecting offensive and defensive approaches [10]. While AI has improved the detection of and response to cyberthreats, it has also increased the sophistication of attacks, increasing the need for more adaptable and scalable defenses [5]. As AI systems continue to develop, addressing the implications of their use, especially in terms of privacy, accountability, and autonomy, is critical to ensuring responsible and transparent use in cybersecurity [8].

4. Discussion

This section seeks to critically interpret the previously presented thematic findings, analyzing their strategic implications for contemporary cybersecurity [10, 5]. These findings pose significant challenges for current detection tools and, at the same time, suggest the need to review the approach to defense strategies in a rapidly evolving technological environment.

AI-powered cyberattacks not only increase their destructive capacity, but also potentially change the strategic balance between attackers and defenders [10]. The reduction in human intervention required, thanks to intelligent automation, expands the scope and frequency of attacks, opening the door to actors with less technical expertise [1]. Taken together, these advancements represent a paradigm shift in cybersecurity, indicating that traditional static defense mechanisms are increasingly ineffective against highly dynamic AI-driven threats [5].

The practical implications of these findings are especially critical for sectors such as healthcare, critical infrastructure, and public administration, where a disruption can have systemic consequences [9]. The use of AI to increase the adaptability and potential personalization of cyberattacks poses a direct threat to the operational and reputational integrity of organizations, particularly given the potential misuse of sensitive data [5, 1, 8].

This review is consistent with previous studies that have documented the use of AI to increase the complexity and effectiveness of cyberattacks, as well as the limitations of traditional defense systems [8]. However, differences were identified in the methodological approaches and timeframes used. While much of the literature has focused on theoretical analyses or hypothetical scenarios, this work focuses on observable applications and techniques already implemented in actual contexts [5, 10]. The sectors analyzed also contributed to notable differences in the findings, emphasizing the importance of specific contexts in cybersecurity research.

Among the most significant methodological limitations are the exclusive use of Google Scholar as a search source, the exclusion of literature not indexed in JCR, and the focus on recent studies. These restrictions may have reduced the diversity of perspectives or excluded relevant previous references. However, although the findings presented are representative of the current state of the field, they should be interpreted within the framework of these limitations.

There is a need to expand the study of AI-driven cyberthreats to under-researched sectors such as education and logistics, and to specifically focus on the development of explainable defensive AI systems [9, 8]. Furthermore, conducting larger collaborative reviews and longitudinal empirical studies on real AI-driven cyberattacks would significantly improve understanding. Likewise, emphasis should be placed on developing and refining ethical and regulatory frameworks to guide the responsible deployment of AI technologies in cybersecurity contexts [8].

AI-powered cyberthreats have pushed the traditional boundaries of digital defense, introducing more complex, rapid, and adaptive attack scenarios [9, 10, 5]. Bridging this gap requires not only technical innovation, but also international cooperation, proactive cybersecurity policies, and the urgent development of ethical frameworks to guide the responsible use of AI in increasingly interdependent digital environments [6, 8].

5. Conclusion

AI is radically transforming the cyberattack landscape. Its ability to automate attacks, generate malicious code using LLMs, adapt in real time, and reduce reliance on human intervention has increased both the effectiveness and scope of threats. These characteristics represent a critical challenge for traditional detection systems, which fail to keep up with the pace and variability of AI-powered attacks [5].

The findings pose significant implications for cybersecurity, both practical and academic [1, 8]. AI-based offensive technologies can mimic legitimate behavior, dynamically adapt, and execute large-scale personalized attacks, with potential impact not only technical but also psychological, in terms of more effective automated persuasion, emotionally personalized attacks, and reduced user ability to identify compelling deceptions. These advances demand a rethinking of current methodologies for threat detection and assessment.

The offensive use of AI in cyberattacks represents

Table 2. Key thematic domains and insights in AI-enhanced cyberattack research as identified in the literature review, including automation and scaling of attacks, malicious content generation via LLMs, evasion of detection systems, and challenges in attribution and defense.

Thematic Research Domain	**Summary of Key Findings and Strategic Implications**
Automation and Escalation of AI-Driven Attacks	AI enables full-cycle automation of cyberattacks (e.g., reconnaissance, exploitation, evasion), allowing massive, scalable, and adaptive threats. Defensive systems remain mostly static and reactive, resulting in critical operational gaps and higher exposure to AI-augmented offensive strategies.
Automated Generation of Cyberattack Techniques via LLMs	LLMs like ChatGPT can be exploited to generate malicious code and phishing content, lowering the barrier to entry for less-skilled actors. These tools are increasingly integrated into offensive platforms like Breach and Attack Simulation (BAS) systems, improving attack customization and stealth.
Evasion of Monitoring and Detection Systems	AI-driven adversarial techniques allow real-time behavioral adaptation to evade traditional intrusion detection systems (IDS) and firewalls. These strategies significantly reduce the effectiveness of signature-based and rule-based defense models.
Challenges and Opportunities in Detection and Attribution	While AI complicates traceability and undermines conventional forensic methods, it also offers new avenues for advanced threat attribution, behavioral analytics, and proactive cyber defense strategies using predictive models.

a paradigm shift in how digital threats are configured. Understanding its scope and consequences is essential to assessing the risk posed by a technology that, while designed to drive progress, can also be used to amplify damage. Contemporary cybersecurity must recognize this inherent duality of AI: its ability to protect, but also to exploit.

Ethics Statement

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

CRediT authorship contribution statement

Adrian Yama-Uitz: Conceptualization, Methodology, Formal analysis, Investigation, Data curation, Writing – original draft, Writing – review & editing, Visualization. **Miguel Cutz-Pech:** Conceptualization, Methodology, Formal analysis, Investigation, Data curation, Writing – original draft, Writing – review & editing, Visualization. **Jose Munoz-May:** Conceptualization, Methodology, Formal analysis, Investigation, Data curation, Writing – original draft, Writing – review & editing, Visualization. **Leonardo Romero-Pech:** Conceptualization, Methodology, Formal analysis, Investigation, Data curation, Writing – original draft, Writing – review & editing, Visualization. **Jarol Lizama-Chan:** Conceptualization, Methodology, Formal analysis, Investigation, Data curation, Writing – original draft, Writing – review & editing, Visualization. **Zyon Maximo Rodriguez-Gonzalez:** Conceptualization, Methodology, Formal analysis, Investigation, Data curation, Writing – original draft, Writing – review & editing, Visualization.

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

During the preparation of this work, the authors used ChatGPT to assist with language clarity and grammar correction. All final edits and intellectual content were determined 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] E. Iturbe, O. Llorente-Vazquez, A. Rego, E. Rios, and N. Toledo, "Unleashing offensive artificial intelligence: Automated attack technique code generation," *Computers & Security*, vol. 147, p. 104077, dec 2024. doi: 10.1016/j.cose.2024.104077. [Online]. Available: https://doi.org/10.1016/j.cose.2024.104077

[2] C. Colther and J. P. Doussoulin, "Artificial intelligence: Driving force in the evolution of human knowledge," *Journal of Innovation and Knowledge*, vol. 9, no. 4, p. 100625, 2024. doi: 10.1016/j.jik.2024.100625. [Online]. Available: https://doi.org/10.1016/j.jik.2024.100625

[3] A. A. Siam, M. Alazab, A. Awajan, and N. Faruqui, "A Comprehensive Review of AI's Current Impact and Future Prospects in Cybersecurity." *IEEE Access*, vol. 13, no. December 2024, pp. 14 029–14 050, 2025. doi: 10.1109/ACCESS.2025.3528114.

[4] B. Guembe, A. Azeta, S. Misra, V. C. Osamor, L. Fernandez-Sanz, and V. Pospelova, *The Emerging Threat of Ai-driven Cyber Attacks: A Review.* Taylor Francis, 2022, vol. 36, no. 1. doi: 10.1080/08839514.2022.2037254. [Online]. Available: https://doi.org/10.1080/08839514.2022.2037254

[5] M. Rabzelj, L. Š. Južnič, M. Volk, A. Kos, M. Kren, and U. Sedlar, "Designing and Evaluating a Flexible and Scalable HTTP Honeypot Platform: Architecture, Implementation, and Applications," *Electronics*, vol. 12, no. 16, p. 3480, aug 2023. doi: 10.3390/electronics12163480. [Online]. Available: https://doi.org/10.3390/electronics12163480

[6] I. H. Sarker, Y. B. Abushark, F. Alsolami, and A. I. Khan, "IntruDTree: A machine learning based cyber security intrusion detection model," *Symmetry*, vol. 12, no. 5, pp. 1–15, 2020. doi: 10.3390/SYM12050754. [Online]. Available: https://doi.org/10.3390/sym12050754

[7] M. J. Page, J. E. McKenzie, P. M. Bossuyt, I. Boutron, T. C. Hoffmann, C. D. Mulrow, L. Shamseer, J. M. Tetzlaff, E. A. Akl, S. E. Brennan, R. Chou, J. Glanville, J. M. Grimshaw, A. Hróbjartsson, M. M. Lalu, T. Li, E. W. Loder, E. Mayo-Wilson, S. McDonald, L. A. McGuinness, L. A. Stewart, J. Thomas, A. C. Tricco, V. A. Welch, P. Whiting, and D. Moher, "The PRISMA 2020 statement: An updated guideline for reporting systematic reviews," *The BMJ*, vol. 372, 2021. doi: 10.1136/bmj.n71. [Online]. Available: https://doi.org/10.1136/bmj.n71

[8] R. Raman, P. Calyam, and K. Achuthan, "ChatGPT or Bard: Who is a better Certified Ethical Hacker?" *Computers & Security*, vol. 140, p. 103804, may 2024. doi: 10.1016/j.cose.2024.103804. [Online]. Available: https://doi.org/10.1016/j.cose.2024.103804

[9] X. Larriva-Novo, C. Sánchez-Zas, V. A. Villagrá, A. Marín-Lopez, and J. Berrocal, "Leveraging Explainable Artificial Intelligence in Real-Time Cyberattack Identification: Intrusion Detection System Approach," *Applied Sciences*, vol. 13, no. 15, p. 8587, jul 2023. doi: 10.3390/app13158587. [Online]. Available: https://doi.org/10.3390/app13158587

[10] J. H. An, Z. Wang, and I. Joe, “A CNN-based automatic vulnerability detection,” *EURASIP Journal on Wireless Communications and Networking*, vol. 2023, no. 1, p. 41, may 2023. doi: 10.1186/s13638-023-02255-2. [Online]. Available: https://doi.org/10.1186/s13638-023-02255-2

Journal of Artificial Intelligence and Computing Applications (2026) 4(1): 31–39

MAIKRON

Journal of Artificial Intelligence and Computing Applications

Review article

Automated multiple-choice question generation for soft skills development using LLMs: a short narrative review

Luis Cardena-Ley[1]

[1]Tecnológico Nacional de México/IT de Mérida, Yucatán, México

ABSTRACT

The development of soft skills is crucial in higher education and professional training, yet traditional assessment methods often struggle to provide personalized and scalable solutions. This narrative review explores the potential of Large Language Models (LLMs), specifically GPT-based models, in generating multiple-choice questions (MCQs) aimed at soft skills development. A structured analysis of recent adjacent literature suggests that LLMs may offer considerable promise for automating the creation of diverse, contextually relevant, and construct-specific MCQs across various domains, including psychological assessments, medical training, and educational settings. Key techniques such as prompt engineering, model fine-tuning, and retrieval-augmented generation are identified as effective methods for enhancing the accuracy and relevance of AI-generated questions. However, challenges such as biases, factual inaccuracies, and the need for human oversight remain significant concerns. The reviewed studies consistently highlight the importance of integrating human expertise to ensure the validity, fairness, and quality of the generated content. Despite these challenges, the use of LLMs could streamline the creation of personalized learning tools, offering scalable solutions for soft skills assessment. Future research should focus on refining AI techniques, developing robust evaluation frameworks, and exploring the impact of AI-generated MCQs on learners' skill development. This review emphasizes the need for a hybrid approach, combining AI's efficiency with expert validation, to fully realize the potential of LLMs in soft skills education.

Keywords: large language models, multiple-choice questions, soft skills

1. Introduction

The development of soft skills, such as communication, critical thinking, teamwork, and problem-solving, has become increasingly crucial in today's fast-paced and dynamic workforce [1]. Unlike hard skills, which are specific, teachable abilities or knowledge, soft skills are more intangible and harder to quantify [2]. However, they are essential for effective collaboration, leadership, and adaptability in various professional settings [2]. As organizations recognize the value of these competencies, there is a growing demand for innovative methods to nurture and assess them [3]. One promising approach is the use of multiple-choice questions (MCQs), which can be strategically designed to promote reflection and learning [4]. With advancements in artificial intelligence (AI), particularly in Large Language Models (LLMs), there is potential to generate these MCQs using automatic item generation (AIG), creating personalized and effective tools for soft skills development [5].

Despite the recognized importance of soft skills, traditional methods of developing and assessing these competencies often fall short in both engagement and effectiveness [6]. Existing literature highlights several ap-

E-mail address: le15080882@merida.tecnm.mx
https://doi.org/10.5281/zenodo.20802609

proaches to soft skills development, including experiential learning, role-playing, and self-assessment [4], yet there remains a persistent challenge in creating scalable and personalized learning experiences [5]. Foundational work has emphasized the necessity of integrating emotional intelligence and interpersonal skills into professional training programs [3]. However, these approaches often rely on human-mediated interventions, which can be resource-intensive and may lack consistency [2]. More recent research has explored the potential of using technology-based solutions, such as e-learning platforms and virtual simulations [4], but these too have limitations in adaptability and real-time feedback [7]. Notably, there is a limited body of research examining the automatic MCQ generation specifically aimed at soft skills development, particularly leveraging advanced AI models like LLMs [5]. This gap presents an opportunity to explore innovative approaches that can bridge the existing limitations and provide scalable, personalized solutions for soft skills acquisition [4].

The objective of this Short Narrative Review (SNR) is to explore the potential of leveraging artificial intelligence, specifically LLMs, to automatically generate MCQs that facilitate the development of soft skills, hereafter referred to as soft skills MCQs. By addressing the identified gaps in the existing literature, this review aims to provide insights into the capabilities and limitations of LLMs in creating personalized, engaging, and effective soft skills MCQs. The guiding research question for this review is: "Can LLMs be utilized to automatically generate MCQs that help individuals develop specific soft skills?" Through a comprehensive analysis of current AI methodologies, practical applications, and case studies, this review seeks to uncover innovative approaches that can bridge the existing limitations in soft skills development and offer scalable, technology-driven solutions.

Given the vast landscape of research on soft skills, AI, and assessment, this review adopts a focused approach, prioritizing the analysis of the most critical and high-quality contributions to the field. To ensure a rigorous and impactful synthesis, the selection process involved a targeted search using Google Scholar, followed by a meticulous screening of results. Only original research articles indexed in the Journal Citation Reports (JCR) and belonging to the top two quartiles (Q1 and Q2) were included. This deliberate strategy allows for a concentrated examination of influential and peer-validated studies, enabling a deeper and more insightful appraisal of the current state of knowledge regarding the automatic generation of soft skills assessment tools using LLMs.

This review systematically explores the potential of leveraging LLMs for the automated generation of soft skills MCQs. Section 2, *Methodology*, outlines the literature review process, detailing the databases searched, keywords utilized, and selection criteria applied to ensure a transparent, replicable, and credible approach. Section 3, *Thematic Overview*, synthesizes the main findings, identifying emergent themes, trends, and patterns in the use of LLMs for MCQ generation, while critically analyzing their applicability to soft skills development and highlighting gaps in the existing literature. Section 4, *Discussion*, interprets these findings, examining their implications for the use of AI in educational contexts, particularly in designing soft skills MCQs, and addresses limitations through comparative analysis. Finally, Section 5, *Conclusion*, consolidates the key insights, underscores their significance for AI-driven educational tools, and proposes future research directions to advance the development of LLM-generated soft skills MCQs. Through this structured approach, the review aims to contribute to the growing discourse on AI's role in enhancing educational methodologies for skills development.

2. Methodology

The methodological framework for this review follows a structured and transparent approach to synthesize recent and relevant research in the field. While this is not a systematic review, the study selection process aligns with key principles of the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) methodology to ensure clarity, rigor, and replicability [8]. PRISMA guidelines provide a structured approach to literature identification, screening, and inclusion, which enhances the reliability of narrative reviews by minimizing selection bias [8]. Given the SNR format, the focus is not on comprehensive coverage but on curating a targeted subset of impactful studies that provide valuable insights into the research topic [9]. To achieve this, a systematic selection process was implemented, emphasizing recent publications from high-quality, peer-reviewed sources while maintaining transparency in inclusion and exclusion criteria. This methodological rigor ensures that the findings are based on a well-defined and reproducible literature selection process, strengthening the credibility of the review [9].

The identification phase involved a structured search strategy to retrieve relevant studies on the research topic. The search was conducted using Google Scholar, a widely used academic search engine that provides broad access to peer-reviewed journal articles across various disciplines [10]. This platform was selected for its comprehensive coverage of scholarly literature and its ability to index recent and high-impact research in artificial intelligence and computing [11]. A predefined search query, "intitle:("soft skills" OR "social skills" OR "21st century skills" OR "interpersonal skills" OR "skills" OR "skill") AND intitle:(generation) AND intitle:("multiple choice questions" OR "Multiple-choice questions" OR "automated item generation" OR "MCQs" OR "AIG") AND intitle:(LLM OR GPT OR LLMs)", was applied to ensure consistency in retrieval. To maintain relevance, the search was restricted to studies published from 2020 onward, thereby focusing on recent advancements in the field. This process yielded a total of 184 records, which were subsequently assessed through a multi-stage selection process to ensure that only high-quality and thematically relevant studies were included in the review.

The screening phase aimed to refine the dataset by eliminating studies that did not meet the relevance and quality criteria [8]. The initial filtering process was conducted through title and abstract screening, where studies that were unrelated to the research topic were excluded. Additionally, works not published in peer-reviewed scientific journals were removed to ensure that only rigorously vetted research was considered. Review articles were also excluded to maintain a focus on primary research studies that present original findings rather than secondary analyses. As a result of this screening process, 153 studies were excluded, leaving 31 studies for further evaluation in the eligibility phase. This step ensured that the remaining literature was directly relevant, high-quality, and methodologically sound, aligning with the objectives of this Short Narrative Review [9].

The eligibility phase involved a comprehensive full-text assessment of the remaining studies to ensure their alignment with the review's objectives [12]. This stage applied stricter exclusion criteria to further refine the selection [13]. Studies were excluded if they were not written in English, as language consistency is essential for accurate interpretation and synthesis [14]. Additionally, articles that were outside the defined scope of the review were removed to maintain thematic relevance [15]. To ensure scientific rigor, only studies published in journals indexed in the JCR within the first two quartiles were retained [16]. Furthermore, studies with fewer than nine citations per year were excluded to prioritize research with established impact. As a result of this assessment, 23 studies were excluded, leaving a final selection of 8 studies for inclusion. These additional filters ensured that the final dataset consisted of highly relevant, peer-reviewed, and impactful research, strengthening the validity of the review's findings [17].

Following the multi-stage selection process, a final set of 8 studies was identified for inclusion in this review. These studies represent recent, high-impact, and thematically relevant contributions to the field, ensuring that the review reflects the latest advancements. An overview of the selected articles, including their titles, journal names, and publication years, is provided in Table 1. To enhance transparency, the study selection process is visually summarized in Figure 1 (PRISMA diagram), detailing the number of records screened, excluded, and retained at each stage. While this methodology ensures rigor and relevance, certain limitations must be acknowledged. The exclusive use of Google Scholar may have restricted the comprehensiveness of the search, as other databases could yield additional relevant studies [10]. Furthermore, the focus on recent publications may have excluded historically significant research, and the exclusion of lower-quartile or non-indexed journals might have omitted valuable contributions from emerging scholars or specialized domains [18]. Additionally, the citation-based filtering may introduce a bias favoring well-established studies while potentially overlooking influential newer research [19]. Despite these limitations, the structured selection process ensures a transparent, replicable, and methodologically rigorous approach to identifying the most impactful literature for this review [12].

3. Thematic Overview

This section organizes and synthesizes the key findings from the selected literature, grouping them into themes that reflect the major topics, trends, and areas of consensus or debate within the field. The focus is on the use of LLMs such as different versions of GPT to automatically generate soft skills MCQs. The selected articles demonstrate the potential of LLMs to transform AIG across various domains, including psychological assessments, education, and medical training. By structuring the discussion around these themes, we aim to provide a cohesive narrative that highlights significant contributions, contradictions, and gaps in the literature.

The use of LLMs for automatic MCQ generation has been explored in multiple contexts. For instance, [20] demonstrates how GPT-3 can generate personality test items with psychometric properties comparable to human-authored items. Similarly, [21] introduces the Psychometric Item Generator (PIG), which uses GPT-2 to create diverse and high-quality psychological scale items. These studies highlight the scalability and efficiency of LLMs in generating diverse and contextually relevant questions. The implementation of prompt techniques on large language models can facilitate the generation of soft skills MCQs. This is supported by findings from [23], which demonstrate that prompt engineering significantly enhances the quality of AI-generated questions.

The psychometric properties of AI-generated MCQs have been rigorously evaluated in several studies. For example, [22] demonstrates that fine-tuned GPT-2 models can generate construct-specific personality test items with high psychometric quality. Similarly, [24] evaluates the difficulty and discrimination levels of ChatGPT-generated medical questions, showing that they meet acceptable psychometric standards. These findings underscore the importance of construct-specificity in generating MCQs that align with specific soft skills, such as communication, teamwork, and problem-solving. The role of fine-tuning models and prompt engineering is critical for ensuring that generated questions are both relevant and effective for soft skills development.

The practical implications of using LLMs to generate soft skills MCQs are significant. For instance, [25] provides actionable guidance for educators to use ChatGPT in generating high-quality medical questions, reducing their workload and enabling personalized learning. Similarly, [26] demonstrates how LLMs can automate the creation of reading comprehension assessments, saving time and resources. However, challenges such as potential biases, factual inaccuracies, and the need for human oversight are consistently highlighted across studies. For example, [27] emphasizes the impor-

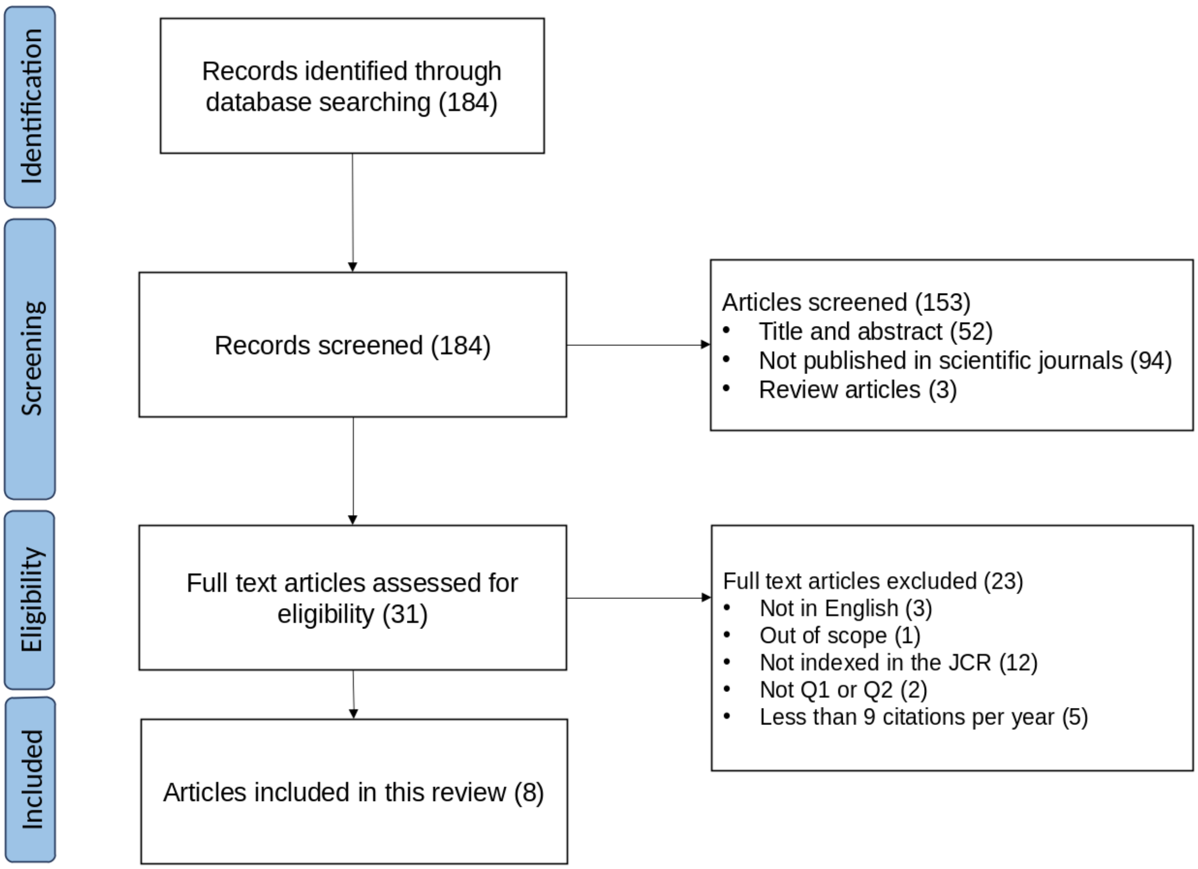


Figure 1. Flowchart of the literature identification, screening, eligibility assessment, and inclusion process.

Table 1. The final list of articles used in this review, including information for title, journal, year of publication and citation.

Title	Journal	Year	Citation
A Paradigm Shift from "Human Writing" to "Machine Generation" in Personality Test Development: an Application of State-of-the-Art Natural Language Processing	Journal of Business and Psychology	2023	[20]
Let the algorithm speak: How to use neural networks for automatic item generation in psychological scale development	Psychological Methods	2023	[21]
Transformer-based deep neural language modeling for construct-specific automatic item generation	Psychometrika	2022	[22]
Few-shot is enough: exploring ChatGPT prompt engineering method for automatic question generation in english education	Education and Information Technologies	2023	[23]
ChatGPT for generating multiple-choice questions: evidence on the use of artificial intelligence in automatic item generation for a rational pharmacotherapy exam	European Journal of Clinical Pharmacology	2024	[24]
Twelve tips to leverage AI for efficient and effective medical question generation: a guide for educators using Chat GPT	Medical Teacher	2024	[25]
The interactive reading task: Transformer-based automatic item generation	Frontiers in Artificial Intelligence	2022	[26]
Biomedical knowledge graph-optimized prompt generation for large language models	Bioinformatics	2024	[27]

tance of integrating knowledge graphs with LLMs to improve accuracy and reduce biases in generated content.

The importance of AI-human collaboration is a recurring theme in the literature. The need for educators to guide LLMs in generating high-quality questions is highlighted in [25], ensuring scientific validity and contextual relevance. Similarly, [26] emphasizes the role of human review in ensuring fairness and bias-free content in AI-generated reading comprehension tasks. Ethical considerations, such as ensuring fairness, avoiding biases, and maintaining transparency in AI-generated content, are also critical. For instance, [27] discusses the ethical implications of relying on LLMs for generating biomedical text and the need for domain-specific knowledge to enhance accuracy.

The themes discussed in this section collectively demonstrate the potential of LLMs to generate high-quality soft skills MCQs. Areas of consensus include the scalability and efficiency of LLMs, the importance of construct-specificity, and the value of AI-human collaboration. However, ongoing debates about biases, factual inaccuracies, and the need for human oversight highlight the challenges that must be addressed to fully realize the potential of AI in this domain. These findings set the stage for the next sections of the review, which will explore the implications of these insights for soft skills development and future research directions. By leveraging the strengths of LLMs while addressing their limitations, we can create innovative and effective tools for soft skills MCQ generation.

4. Discussion

The reviewed literature indicates a significant potential for LLMs, including various GPT versions and ChatGPT, to enhance the efficiency of AIG across diverse assessment domains such as personality testing, psychological scales, medical examinations, and language education [20, 21, 22, 23, 24, 26, 25]. Key AI techniques identified include prompt engineering, ranging from few-shot learning to structured protocols and practical tips [20, 23, 25], and the fine-tuning of pre-trained models for specific constructs [22]. Furthermore, approaches like Knowledge Graph-enhanced Retrieval-Augmented Generation (KG-RAG) are emerging to specifically address the critical challenge of factual accuracy in LLM outputs [27]. Evaluating the effectiveness of these AI-generated items involves various methods, notably quantitative psychometric analyses focusing on properties like factor structure, reliability, difficulty, and discrimination [20, 22, 24], as well as qualitative expert and user reviews assessing content validity, relevance, fairness, and overall quality [23, 26, 25]. Despite the promise, a consistent theme across the studies is the necessity of expert review to address challenges such as potential inaccuracies [24, 27, 25], biases inherited from training data [20, 21, 26], and ensuring the overall quality and appropriateness of the generated items [22, 24, 26, 25]. The purpose of this discussion is therefore to interpret these findings in the specific context of generating effective soft skills MCQs, consider the broader implications, compare the observed methodologies, acknowledge limitations, and propose pertinent directions for future research.

Interpreting these findings for the generation of soft skills MCQs reveals both opportunities and significant challenges. The identified AI techniques—prompt engineering, fine-tuning, and potentially RAG—demonstrate clear pathways for generating assessment content using models like GPT-2, GPT-3, and ChatGPT [20, 22, 23, 24, 25]. Methods emphasizing construct-specificity through fine-tuning [22] or requiring precise, detailed instructions via prompt engineering [23, 25] appear particularly relevant for aligning generated MCQs with specific, often nuanced, soft skills competencies. Defining "effectiveness" for soft skills MCQs, based on the reviewed literature, likely necessitates a hybrid approach; quantitative psychometric properties like item difficulty and discrimination [24] are important for assessment utility, but qualitative expert judgment focusing on construct validity, contextual relevance, fairness, and potential bias [23, 26, 25] is arguably indispensable for ensuring the questions meaningfully address soft skills. The primary challenge lies in translating success from domains focused on cognitive abilities or stable traits [20, 22, 21, 23, 24, 26] to the dynamic, context-dependent, and often behavioral nature of soft skills like communication, teamwork, or problem-solving. It remains an open question how effectively the current methods can generate soft skills MCQs beyond surface-level knowledge. Techniques enhancing factual accuracy, such as KG-RAG [27], might prove valuable for creating reliable scenario-based questions common in soft skills training. Ultimately, the consistent finding across diverse studies [22, 24, 26, 25] that manual validation is essential underscores the interpretation that current AI technologies serve as powerful assistants rather than autonomous creators, particularly for complex and sensitive tasks like generating materials for soft skills development.

The implications of leveraging AI for soft skills MCQ generation are potentially far-reaching. Practically, the most significant implication, echoed across nearly all reviewed studies, is the potential for substantial efficiency gains, saving educators, trainers, and instructional designers considerable time and resources in developing practice materials and assessments [25, 26, 23, 22, 20]. This efficiency could facilitate the creation of larger, more diverse item banks [22], enabling personalized learning pathways or adaptive testing systems tailored to individual soft skills development needs. For the field of educational technology, these findings can guide the development of more sophisticated authoring tools, perhaps integrating user-friendly interfaces similar to the PIG tool concept [21] with robust back-ends for prompting, generation, and validation workflow support. Furthermore, this line of research has implications for AI development itself, highlighting the need for models with improved capabilities in handling nuance, context, ethical considerations, and bias mitigation – all critical aspects when dealing with soft skills. Regarding readiness for application, while studies show promise, such as achieving acceptable psychometric properties for medical MCQs [24], the strong emphasis on mandatory expert review [26, 25, 24, 22] suggests that AI-generated soft skills MCQs should perhaps be initially deployed in lower-stakes formative assessment or practice contexts, rather than high-stakes summative evaluations, pending further validation. The potential beneficiaries include learners gaining more practice opportunities, educators saving valuable time, and organizations obtaining more scalable tools for training and development.

Comparing the findings of the eight reviewed articles with the broader AIG landscape described by recent reviews [28, 29] highlights both continuity and rapid evolution. While [29] demonstrated that earlier AIG methods could produce high-quality MCQs comparable to traditional ones, even for assessing higher-order application of knowledge, the studies in this review predominantly showcase the capabilities of newer LLMs [20, 21, 22, 23, 24, 26, 25]. This focus aligns with the trend towards advanced Natural Language Processing (NLP) techniques noted [28]. Although these LLM-based approaches demonstrate significant potential for generating diverse items efficiently across various domains, the consistent emphasis across multiple articles [22, 24, 26, 25] on the necessity of rigorous expert validation due to concerns about accuracy [24, 27, 25] and bias [20, 21, 26] suggests that achieving the quality parity shown by [29] requires careful human-AI collaboration. The diverse evaluation methods employed in the reviewed articles—spanning psychometrics [20, 22, 24] and expert judgment [23, 26, 25]—reflect the varied evaluation landscape described by [28], reinforcing the importance of expert review also highlighted by [29] and underscoring the need for continued refinement of these powerful but still maturing AI techniques for operational use.

It is crucial to acknowledge the limitations present both within the reviewed studies and in this narrative review itself. A frequently reported limitation across the primary studies is the indispensable need for human expert review to ensure the quality, validity, fairness, and accuracy of the AI-generated items [22, 24, 26, 25]. Concerns regarding factual inaccuracies or "hallucinations" produced by LLMs were also noted [24, 27, 25], alongside the potential for inherent biases present in the models' training data influencing the generated content [20, 21, 26]. Specific studies reported issues like semantic redundancy or poor item quality [22], faced constraints due to small sample sizes or limited scope [24], or highlighted dependency on particular AI models or knowledge graphs [27]. Furthermore, many studies focused on domains other than soft skills or generated item formats other than MCQs (e.g., [20] focused on Likert-scale personality items). This review is also limited by its narrative nature and the small sample size of eight articles, which, despite the PRISMA methodology, may not capture the full breadth of research. Key questions pertinent to the research objective remain largely unanswered by this specific set of literature, including how to reliably generate MCQs that assess higher-order, context-dependent soft skills; how best to evaluate the effectiveness of these soft skills MCQs; and how to systematically mitigate biases in AI-generated content tailored for soft skills.

Based on the synthesis of findings and the identified limitations, several directions for future research emerge. There is a clear need for research directly comparing the efficacy of different AI techniques—prompt engineering, fine-tuning, RAG, or hybrid approaches—specifically tailored for the complexities of generating soft skills MCQs, potentially building upon the methods explored by [23, 22, 27]. Concurrently, robust evaluation frameworks must be developed for this specific context, integrating appropriate psychometric analyses with rigorous expert reviews that explicitly assess construct validity, contextual relevance, fairness, and potential biases, going beyond the general quality checks used in studies like [24, 26, 23]. Crucially, future empirical studies should move beyond evaluating item quality to investigating the actual impact of interacting with AI-generated soft skills MCQs on learners' skill development, ideally through longitudinal research designs. Addressing the persistent challenges of bias and inaccuracy is paramount; research should focus on developing and testing methods for proactively mitigating bias in soft skills content generation and enhancing the factual consistency of scenario-based questions, perhaps by refining KG-RAG techniques [27]. Finally, future work could explore the generation of diverse item formats suitable for soft skills, such as situational judgment test items (as suggested by [20]), investigate the integration of these tools into adaptive learning systems, and develop specific ethical guidelines for the responsible use of AI in soft skills assessment and development.

Overall, the reviewed studies suggests that LLMs offer considerable potential for streamlining MCQ generation, a potential that could, in principle, be extended to the domain of soft skills development. Techniques such as sophisticated prompt engineering, model fine-tuning, and retrieval-augmented generation represent key methodological avenues explored in related fields [23, 22, 27, 25]. However, the reviewed literature consistently underscores that realizing this potential effectively and responsibly hinges upon careful implementation and, critically, the integration of human expertise. Achieving desired levels of quality, accuracy, validity, and fairness currently necessitates a hybrid approach, with AI serving as a powerful assistant requiring human guidance, refinement, and validation [22, 24, 26, 25]. Significant challenges remain, particularly concerning the reliable generation and evaluation of content for nuanced soft skills and the mitigation of potential inaccuracies and biases. Addressing these challenges through focused future research on tailored techniques, robust evaluation methodologies, impact studies, and ethical frameworks will be essential to fully harness the capabilities of AI for soft skills MCQ generation. This review contributes by synthesizing the current state, highlighting both the promise and the necessary precautions based on recent studies investigating AI-driven AIG.

5. Conclusion

This narrative review examined the use of LLMs, particularly GPT-based architectures, for the automated generation of soft skills MCQs. The synthesis of eight recent studies identified core methodological trends, including prompt engineering, model fine-tuning, and RAG, primarily applied in domains such as cognitive assessment, medical testing, and language education. While these approaches demonstrate clear potential, their direct application to soft skills remains underexplored and methodologically challenging due to the inherently contextual and behavioral nature of such competencies.

Evidence from the reviewed literature suggests that LLMs can significantly increase the efficiency of AIG, particularly when integrated with structured prompting strategies or domain-specific fine-tuning. However, ensuring the quality, relevance, and fairness of generated items necessitates rigorous human oversight. Evaluation frameworks that combine psychometric analysis with expert review are essential, particularly when assessing content aimed at complex constructs like communication, teamwork, or ethical reasoning.

The primary implications concern both the scalability of AI-assisted assessment development and the limitations of current models. LLMs offer opportunities to expand item banks and support adaptive learning environments, but concerns about factual inaccuracies, semantic redundancy, and embedded biases highlight the importance of human-in-the-loop design. Current systems should therefore be deployed cautiously, preferably in formative or low-stakes contexts pending further validation.

This review is constrained by its limited sample size and the narrative synthesis methodology, which may not capture the full scope of relevant research. Additionally, most studies reviewed focus on domains other than soft skills or employ formats other than MCQs, limiting direct generalizability. Key gaps remain regarding the reliable generation of context-sensitive items, the empirical evaluation of learning outcomes, and the mitigation of AI-induced biases.

Future research should prioritize the development of specialized generation and evaluation frameworks for soft skills assessments. Comparative analyses of AI techniques, longitudinal impact studies, and bias mitigation strategies are particularly needed. As LLMs continue to evolve, their effective integration into soft skills development will depend on sustained interdisciplinary collaboration, combining advances in AI with domain-specific pedagogical and ethical expertise.

Ethics Statement

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

CRediT authorship contribution statement

Luis Cardena-Ley: Conceptualization, Methodology, Formal analysis, Investigation, Data curation, Writing – original draft, Writing – review & editing, Visualization.

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

During the preparation of this work, the author used ChatGPT to assist with language clarity and grammar correction. All final edits and intellectual content were determined and verified by the author.

Declaration of competing interest

The author declares that he has no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] G. W. Mitchell, L. B. Skinner, and B. J. White, "Essential soft skills for success in the twenty-first century workforce as perceived by business educators." *Delta Pi Epsilon Journal*, vol. 52, no. 1, 2010.

[2] A. F. Hendarman and U. Cantner, "Soft skills, hard skills, and individual innovativeness," *Eurasian Business Review*, vol. 8, pp. 139–169, 2018. doi: 10.1007/s40821-017-0076-6 .

[3] B. Schulz, "The importance of soft skills: Education beyond academic knowledge." *Journal of Language and Communication*, 2008.

[4] L. G. Malicay, "The integration of soft skills in professional education: Exploring the importance of communication, teamwork, and interpersonal skills in professional training and the methods used to incorporate them into educational programs," *International Journal of Advanced Research in Science, Communication and Technology*, vol. 3, no. 1, pp. 836–843, 2023. doi: 10.48175/IJARSCT-11967 .

[5] S. Maity and A. Deroy, "The future of learning in the age of generative ai: Automated question generation and assessment with large language models," *arXiv preprint arXiv:2410.09576*, 2024. doi: 10.48550/arXiv.2410.09576 .

[6] F. S. Mohammed and F. Ozdamli, "A systematic literature review of soft skills in information technology education," *Behavioral Sciences*, vol. 14, no. 10, p. 894, 2024. doi: 10.3390/bs14100894 .

[7] G. Fontaine, S. Cossette, M. A. Maheu-Cadotte, T. Mailhot, M. F. Deschênes, and G. Mathieu-Dupuis, "Effectiveness of adaptive e-learning environments on knowledge, competence, and behavior in health professionals and students: Protocol for a systematic review and meta-analysis," *JMIR Research Protocols*, vol. 6, no. 7, pp. 1–10, 2017. doi: 10.2196/resprot.8085 .

[8] M. J. Page, J. E. McKenzie, P. M. Bossuyt, I. Boutron, T. C. Hoffmann, C. D. Mulrow, L. Shamseer, J. M. Tetzlaff, E. A. Akl, S. E. Brennan *et al.*, "The prisma 2020 statement: an updated guideline for reporting systematic reviews," *bmj*, vol. 372, 2021. doi: 10.1136/bmj.n71 .

[9] R. Ferrari, "Writing narrative style literature reviews," *Medical writing*, vol. 24, no. 4, pp. 230–235, 2015. doi: 10.1179/2047480615Z.000000000329 .

[10] D. Torres-Salinas, R. Ruiz-Pérez, and E. Delgado-López-Cózar, "Google scholar como herramienta para la evaluación científica," *Profesional de la Información*, vol. 18, no. 5, pp. 501–510, 2009. doi: 10.3145/EPI.2009.SEP.03 .

[11] G. Halevi, H. Moed, and J. Bar-Ilan, "Suitability of google scholar as a source of scientific information and as a source of data for scientific evaluation—review of the literature," *Journal of informetrics*, vol. 11, no. 3, pp. 823–834, 2017. doi: 10.1016/j.joi.2017.06.005 .

[12] D. Moher, A. Liberati, J. Tetzlaff, and D. G. Altman, "Preferred reporting items for systematic reviews and meta-analyses: the prisma statement," *PLoS medicine*, vol. 6, no. 7, p. e1000097, 2009. doi: 10.1371/journal.pmed.1000097 .

[13] J. P. Higgins and S. Green, *Cochrane handbook for systematic reviews of interventions.* John Wiley & Sons, 2011. doi: 10.1002/9780470712184 .

[14] A. Dobrescu, B. Nussbaumer-Streit, I. Klerings, G. Wagner, E. Persad, I. Sommer, H. Herkner, and G. Gartlehner, "Restricting evidence syntheses of interventions to english-language publications is a viable methodological shortcut for most medical topics: a systematic review," *Journal of Clinical Epidemiology*, vol. 137, pp. 209–217, 9 2021. doi: 10.1016/j.jclinepi.2021.04.012 . [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0895435621001347

[15] D. Gough, S. Oliver, and J. Thomas, *An introduction to systematic reviews.* SAGE, 2017. doi: 10.4135/9781036234942 .

[16] M. Callaham, R. L. Wears, and E. Weber, "Journal prestige, publication bias, and other characteristics associated with citation of published studies in peer-reviewed journals," *JAMA*, vol. 287, no. 21, pp. 2847–2850, 2002. doi: 10.1001/jama.287.21.2847.

[17] E. C. Martinez, P. E. G. Hasbun, V. P. S. Vargas, O. Y. García-González, M. D. F. Madera, D. E. R. Capistrán, T. C. Carmona, C. S. Cruz, and C. T. Hooper, "A comprehensive guide to conduct a systematic review and meta-analysis in medical research," *Medicine*, vol. 104, p. e41868, 8 2025. doi: 10.1097/MD.0000000000041868. [Online]. Available: https://journals.lww.com/10.1097/MD.0000000000041868

[18] N. R. Haddaway, P. Woodcock, B. Macura, and A. Collins, "Making literature reviews more reliable through application of lessons from systematic reviews," *Conservation Biology*, vol. 29, no. 6, pp. 1596–1605, 2015. doi: 10.1111/cobi.12541.

[19] A. V. Kulkarni, B. Aziz, I. Shams, and J. W. Busse, "Comparisons of citations in web of science, scopus, and google scholar for articles published in general medical journals," *JAMA*, vol. 302, no. 10, pp. 1092–1096, 2009. doi: 10.1001/jama.2009.1307.

[20] P. Lee, S. Fyffe, M. Son, Z. Jia, and Z. Yao, "A paradigm shift from "human writing" to "machine generation" in personality test development: an application of state-of-the-art natural language processing," *Journal of Business and Psychology*, vol. 38, pp. 163–190, 2023. doi: 10.1007/s10869-022-09864-6. [Online]. Available: https://doi.org/10.1007/s10869-022-09864-6

[21] F. M. Götz, R. Maertens, S. Loomba, and S. van der Linden, "Let the algorithm speak: How to use neural networks for automatic item generation in psychological scale development," *Psychological Methods*, vol. 29, pp. 494–518, 2024. doi: 10.1037/met0000540.

[22] B. E. Hommel, F. J. M. Wollang, V. Kotova, H. Zacher, and S. C. Schmukle, "Transformer-based deep neural language modeling for construct-specific automatic item generation," *Psychometrika*, vol. 87, pp. 749–772, 2022. doi: 10.1007/s11336-021-09823-9. [Online]. Available: https://doi.org/10.1007/s11336-021-09823-9

[23] U. Lee, H. Jung, Y. Jeon, Y. Sohn, W. Hwang, J. Moon, and H. Kim, "Few-shot is enough: exploring chatgpt prompt engineering method for automatic question generation in english education," *Education and Information Technologies*, vol. 29, pp. 11 483–11 515, 2024. doi: 10.1007/s10639-023-12249-8. [Online]. Available: https://doi.org/10.1007/s10639-023-12249-8

[24] Y. S. Kıyak, Özlem Coşkun, I. İrem Budakoğlu, and C. Uluoğlu, "Chatgpt for generating multiple-choice questions: Evidence on the use of artificial intelligence in automatic item generation for a rational pharmacotherapy exam," *European Journal of Clinical Pharmacology*, vol. 80, pp. 729–735, 2024. doi: 10.1007/s00228-024-03649-x.

[25] I. R. Indran, P. Paranthaman, N. Gupta, and N. Mustafa, "Twelve tips to leverage ai for efficient and effective medical question generation: A guide for educators using chat gpt," *Medical Teacher*, vol. 46, pp. 1021–1026, 2024. doi: 10.1080/0142159X.2023.2294703.

[26] Y. Attali, A. Runge, G. T. Laflair, K. Yancey, S. Goodwin, Y. Park, and A. A. V. Davier, "The interactive reading task: Transformer-based automatic item generation," *Frontiers in Artificial Intelligence*, vol. 5, 7 2022. doi: 10.3389/frai.2022.903077.

[27] K. Soman, P. W. Rose, J. H. Morris, R. E. Akbas, B. Smith, B. Peetoom, C. Villouta-Reyes, G. Cerono, Y. Shi, A. Rizk-Jackson, S. Israni, C. A. Nelson, S. Huang, and S. E. Baranzini, "Biomedical knowledge graph-optimized prompt generation for large language models," *Bioinformatics*, vol. 40, 11 2024. doi: 10.1093/bioinformatics/btae560. [Online]. Available: http://arxiv.org/abs/2311.17330

[28] R. Circi, J. Hicks, and E. Sikali, "Automatic item generation: foundations and machine learning-based approaches for assessments," 2023. doi: 10.3389/feduc.2023.858273.

[29] D. Pugh, A. D. Champlain, M. Gierl, H. Lai, and C. Touchie, "Can automated item generation be used to develop high quality mcqs that assess application of knowledge?" *Research and Practice in Technology Enhanced Learning*, vol. 15, 2020. doi: 10.1186/s41039-020-00134-8.

Journal of Artificial Intelligence and Computing Applications (2026) 4(1): 40–47

MAIKRON

Journal of Artificial Intelligence and Computing Applications

Review article

Application of LSTM models and recurrent neural networks in flood detection using satellite images: a short narrative review

Christopher Abraham Arevalo-Erosa[1], Jose Emilio Ek-Chay[1], Rodrigo Kalel Rodriguez-Chavez[1], Máximo Arturo Guzmán-Ascencio[1], Raul Humberto Rodriguez-Mendez[1], Romell Gerardo Mejia-Ramón[1], and Misael Garcia[1]

[1]Tecnológico Nacional De México / IT de Mérida, Yucatán, México

ABSTRACT

The detection of floods using satellite imagery is essential for strengthening monitoring systems and early-warning mechanisms in vulnerable regions. This Short Narrative Review examines the effectiveness of Long Short-Term Memory (LSTM) models and other recurrent neural networks (RNNs) in comparison with traditional methods and deep learning architectures based on convolutional neural networks (CNNs). The reviewed studies indicate that CNNs remain the most widely used and report strong performance in the spatial segmentation of flooded areas, particularly with architectures such as U-Net and DeepLab. Moreover, some articles highlight that LSTMs provide advantages by integrating temporal information and hydrological variations that CNNs do not capture, which may support the interpretation of events with temporal dynamics. The most common limitations include environmental variability, sensor dependence, and lack of methodological consistency across studies. Although the evidence is still limited, the reviewed literature suggests that LSTMs may serve as a complementary approach for flood detection based on satellite imagery.

Keywords: flood detection, recurrent neural networks, long short-term memory

1. Introduction

In the field of applied artificial intelligence, Recurrent Neural Networks (RNNs) have become established as one of the most effective architectures for processing sequential data [1], enabling the modeling of complex temporal dynamics in natural phenomena and physical systems. Their ability to retain contextual information and learn long-term dependencies has driven their adoption across multiple areas of scientific computing [2], ranging from weather forecasting [3] to environmental monitoring [4]. Within this landscape, flood detection using satellite imagery has emerged as a high-impact field, where the automated analysis of time series is essential for anticipating extreme events and reducing human and material losses. In a global context marked by the intensification of hydrometeorological disasters, the integration of deep learning algorithms with Earth-observation technologies in optical and satellite radar sensors represents a promising pathway for strengthening early-warning systems and resilience to climate change.

Despite advances in deep learning applied to Earth observation, the automated detection of floods continues to face significant limitations in terms of accuracy, generalization, and scalability. Recent literature reveals that many traditional approaches, based on spectral

E-mail address: le25081007@merida.tecnm.mx
https://doi.org/10.5281/zenodo.20802668
 ISSN: 3061-8843

thresholds or supervised classification algorithms, exhibit inconsistent performance when confronted with the spatial and temporal variability of satellite data. Although RNNs and, in particular, Long Short-Term Memory (LSTM) models have demonstrated notable potential for capturing spatiotemporal relationships in image sequences, their application in flood detection remains fragmented and methodologically heterogeneous. This dispersion of approaches hinders the comparison of results and the consolidation of evaluation standards, creating a knowledge gap regarding the actual effectiveness of these architectures across diverse scenarios and under changing environmental conditions. In this context, it is necessary to conduct a critical and up-to-date review that synthesizes the most relevant findings and contributes to clarifying the role of RNNs and LSTM models in improving flood detection systems using satellite imagery.

In response to the identified limitations, this review aims to analyze and synthesize recent scientific evidence on the use of deep learning architectures for flood detection using satellite imagery, considering primarily convolutional neural networks (CNNs) based approaches, which are the most representative in the reviewed literature, and also incorporating studies that employ LSTM models and other RNNs when available. The analysis seeks to examine the predominant methodological approaches, the types of satellite data used, and the performance metrics reported, with the purpose of identifying patterns, strengths, and persistent gaps in the field. Under a Short Narrative Review framework, the study prioritizes both the analysis and the critical synthesis of high-quality research published over the past five years, guiding the discussion toward understanding how different architectures contribute to improving the accuracy and predictive capacity of early-warning systems. In this manner, the study is situated at the intersection of artificial intelligence, remote sensing, and climate-risk management, proposing an integrative perspective that fosters future applied research directions.

The review attains scientific relevance by demonstrating the lack of stable and comparable criteria in the methods used for flood detection through satellite imagery, particularly given the diversity of sensors, architectures, and validation strategies employed in recent literature. More than reiterating the potential of recurrent models, the central contribution of this work lies in clarifying the current state of this methodological heterogeneity and in assessing its impact on reproducibility, the generalization of results, and the operational capacity of hydrometeorological monitoring systems. By organizing and contrasting dispersed approaches, the review helps delineate actual advances, remaining challenges, and opportunities for standardizing analytical processes in future studies. In this way, it provides a conceptual foundation that guides the development of more consistent and useful research for integrating artificial intelligence techniques into flood-risk management.

To contextualize the scope of the analysis developed in the following sections, this article is structured in a manner consistent with the objectives of the review. Section 2, Methodology, details the procedure used for the identification, selection, and critical evaluation of the literature, specifying the criteria that allow for examining model performance, operational applicability, comparison among architectures, and geographic transferability. In Section 3, Thematic Overview, the main findings are presented, organized according to methodological trends, model types, and geospatial applications. Section 4, Discussion, provides an interpretative analysis that contrasts the results, identifies significant advances, and discusses persistent limitations in the field. Finally, Section 5, Conclusion, synthesizes the central contributions of the review and outlines research directions aimed at strengthening intelligent systems for flood detection.

2. Methodology

The methodology employed in this review aims to ensure a fair, transparent, and replicable process for the identification, selection, and analysis of the most relevant literature related to the topic of study. Google Scholar was used as the search engine due to its extensive repository of scientific articles and its accessible interface, which facilitates the retrieval of up-to-date academic sources. This platform is particularly useful for locating works from diverse disciplines, including the field of neural networks and machine learning, which constitute the main focus of our research.

The search was conducted using specific keywords: ("long short-term memory" OR "recurrent neural networks") AND satellite AND "flood detection", with the aim of encompassing the available literature related to the application of LSTM and RNN models in flood detection using satellite imagery. The search strategy was designed with clearly defined inclusion and exclusion criteria to ensure the quality and relevance of the selected articles.

Articles that met the following parameters were considered for inclusion:

Published in peer-reviewed scientific journals indexed in the Journal Citation Reports (JCR), specifically in Q1 or Q2 journals written in English. Published within the last five years, in order to include both foundational research and the most recent advancements. With a minimum of 10 citations per year, reflecting their impact and recognition within the scientific community.

In the identification phase, 636 records were obtained. After reviewing titles and abstracts, 592 records were excluded during the screening phase because they did not meet the initial relevance, language, or publication-type criteria. The remaining 44 articles were then evaluated in the eligibility phase according to the predefined quality criteria. At this stage, 33 articles were excluded: 12 were not classified as Q1 or Q2, and 21 had fewer than 10 citations per year. Ultimately, 11 articles met all the established parameters and were selected for inclusion. Any discrepancies were reviewed by a team of seven evaluators, with disagreements resolved through consensus.

Articles were excluded based on the following criteria: Not written in English ($n = 0$). Not directly related to the topic after reviewing the title and abstract ($n = 0$). Not indexed in the JCR or not meeting the required quality level ($n = 0$). Not classified as Q1 or Q2 ($n = 12$). Having fewer than 10 citations per year ($n = 21$).

The selection process illustrated in Figure 1 describes the systematic filtering from the initial search to the final inclusion of the studies. The selected articles were qualitatively analyzed to identify approaches, methodologies, and relevant findings on the use of LSTM and RNN networks in flood detection using satellite imagery.

To ensure the rigor of the identification process, the search was carried out in Google Scholar using the complete strategy: ("long short-term memory" OR "recurrent neural networks") AND satellite AND "flood detection", applying publication-year filters for the period 2020–2025. The query was conducted on January 15, 2025, and the results were cleaned through the manual removal of duplicates before proceeding to the initial screening. Subsequently, it was verified that each record met the methodological criteria previously described, thus forming the final set of selected studies. The articles that passed this process are presented in Table 1, where their main characteristics are synthesized to facilitate the subsequent comparative analysis.

3. Thematic Overview

The collection of analyzed studies focuses on the application of artificial intelligence and deep learning for advanced water-risk management, encompassing two essential lines of research: post-disaster event detection and mapping, and predictive monitoring or susceptibility mapping. In remote sensing-based flood detection, the predominant methodological approach is the implementation of CNNs for image classification and segmentation [6]. U-Net and its variants have become established as key tools for refining segmentation and improving mapping accuracy [11, 14]. Regarding data acquisition platforms and sensor types, the reviewed studies show a clear diversity of sources: some employ unmanned aerial vehicles (UAVs) equipped with RGB cameras to obtain high spatial resolution and rapid response in targeted areas [6], while others rely on SAR imagery, Sentinel-1 data, multisensor optical imagery, Very High Resolution (VHR) images, or Sentinel-2 multispectral data to ensure broader coverage and greater robustness under different environmental conditions [11, 14]. Complementarily, the predictive axis includes the integration of Artificial Intelligence of Things (AIoT) with in situ sensors for real-time overflow prevention [8], as well as machine learning models such as Random Forest (RF) for generating susceptibility maps based on topographic, hydrological, and environmental variables derived from satellite or GIS-based data [8, 15]. Based on the primary acquisition source reported in each study, SAR or Sentinel-1 imagery represents 36.4% of the reviewed articles, followed by satellite or GIS-derived environmental variables with 27.3%, optical VHR or Sentinel-2 multispectral imagery with 18.2%, UAV or RGB imagery with 9.1%, and AIoT or in situ sensing with 9.1%. This distribution, summarized in Figure 2, shows that the analyzed literature is dominated by satellite-based and geospatial data sources, while UAV and in situ sensing approaches are less frequent but operationally relevant.

In addition to the general trends observed, the comparative analysis reveals substantial methodological differences that affect both model performance and the comparability of results among studies. One of the most relevant divergences concerns the balance between spatial resolution and coverage extent. Some studies prioritize the high spatial resolution obtained through UAVs equipped with RGB cameras, which facilitates detailed flood characterization and damage assessment in localized areas [6]. In contrast, other studies rely on satellite sensors such as Sentinel-2, VHR, and SAR to ensure broader and more continuous regional observation, although generally with a lower level of spatial detail [11, 14]. At the algorithmic level, contrasts are also observed: Random Forest exhibits better generalization capacity in scenarios with significant geomorphological and climatic variability, particularly when environmental and topographic factors derived from satellite data are combined, whereas U-Net and related architectures are more strongly oriented toward fine spatial segmentation [15]. Similarly, the reviewed studies differ in their operational objectives, since most focus on post-disaster flood identification, while others adopt preventive approaches through AIoT-based systems and in situ sensors for real-time hydrological monitoring [8]. This methodological heterogeneity is especially relevant for assessing LSTM and RNN architectures, because their effectiveness depends on the availability of coherent multitemporal series with stable acquisition frequencies. Therefore, variability in spatial resolution, coverage, data availability, and evaluation protocols limits the possibility of directly comparing recurrent models with traditional spatial models or convolutional approaches.

The integrated analysis reveals that one of the most significant gaps in the literature is the lack of well-documented reference datasets, which hinders a rigorous evaluation of model performance and stability [10]. This deficiency limits the ability to compare approaches under homogeneous conditions and generates biases derived from differences in resolution, spatial extent, and atmospheric conditions, particularly affecting the transferability of the proposed methods [15]. At the methodological level, difficulties also persist in scenarios where the hydrological signal is subtle, such as spectrally homogeneous landscapes, or where events occur at very small scales [10], which restricts the generalization capacity of several architectures based on spatial segmentation. From a critical perspective, it is observed that

Table 1. List of analyzed articles on flood detection.

Title	Journal	Year	Reference
Deep Learning-Based Improved WCM Technique for Soil Moisture Retrieval with Satellite Images	Remote Sensing	2023	[5]
Uavs in disaster management: Application of integrated aerial imagery and convolutional neural network for flood detection	Sustainability	2021	[6]
Sentinel-1 SAR Images and Deep Learning for Water Body Mapping	Remote Sensing	2023	[7]
An Integrated Artificial Intelligence of Things Environment for River Flood Prevention	Sensors	2022	[8]
A Near-Real-Time Flood Detection Method Based on Deep Learning and SAR Images	Remote Sensing	2023	[9]
Flood Detection in Dual-Polarization SAR Images Based on Multi-Scale Deeplab Model	Remote Sensing	2022	[10]
Flood Detection Using Multiple Chinese Satellite Datasets during 2020 China Summer Floods	Remote Sensing	2021	[11]
Flood Susceptibility Assessment Using Artificial Neural Networks in Indonesia	Artificial Intelligence in Geosciences	2021	[12]
Flood Susceptibility Mapping Using Multi-Temporal Sentinel-1 Data and eXtreme Deep Learning Model	Geoscience Frontiers	2024	[13]
Mapping Pluvial Flood-Induced Damages with Multi-Sensor Optical Remote Sensing: A Transferable Approach	Remote Sensing	2023	[14]
Integrating Satellite Images and Machine Learning for Flood Prediction and Susceptibility Mapping for the Case of Amibara, Awash Basin, Ethiopia	Remote Sensing	2024	[15]

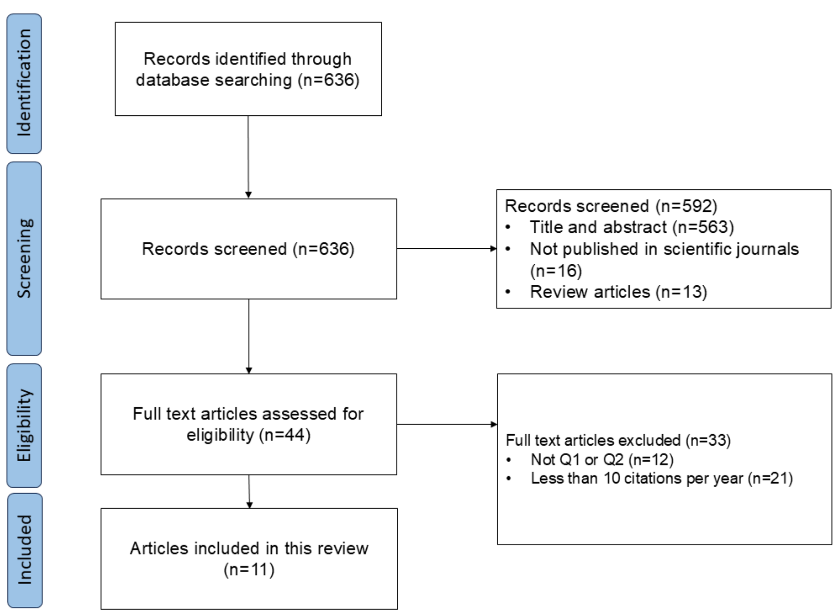


Figure 1. Flow diagram of the literature identification, selection, eligibility assessment, and inclusion process.

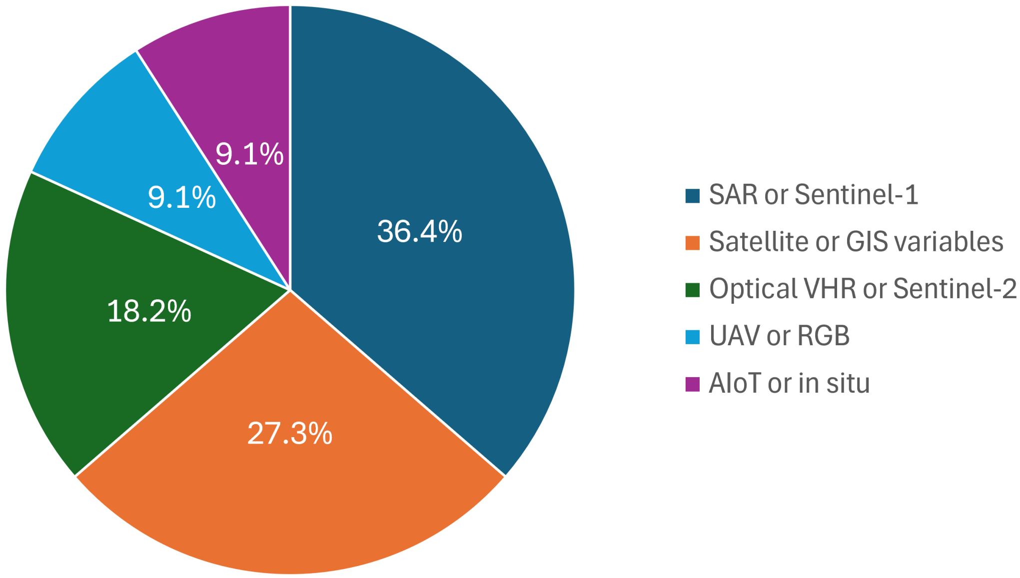


Figure 2. Distribution of the reviewed studies according to their primary acquisition platform or sensor type. The categories summarize the dominant data source reported in each article.

most of the reviewed studies emphasize techniques oriented toward spatial analysis, such as CNN, RF, or UAV sensors, whereas temporal approaches based on LSTM and RNN are constrained by the limited availability of consistent multitemporal series, making it difficult to properly contrast them with purely spatial models. This methodological imbalance suggests the need to promote more balanced experimental designs, as well as comparative frameworks that integrate both spatial and temporal information. In this regard, the integration of complementary sensors (radar, shortwave infrared) and the expansion of AIoT systems for continuous monitoring represent promising directions for generating more robust multitemporal datasets, a necessary condition for more accurately evaluating the real potential of recurrent architectures in flood detection.

4. Discussion

The evidence gathered shows a significant advance in the use of deep learning techniques for flood detection through satellite imagery; however, the methodological domain observed is strongly concentrated in convolutional architectures oriented toward spatial analysis. This predominance introduces a clear bias in the literature, as it limits the ability to fairly evaluate alternative approaches such as recurrent neural networks, including LSTMs, whose performance cannot be directly inferred due to the absence of rigorous comparative studies within the analyzed corpus. Thus, although recurrent models are theoretically suitable for representing the temporal evolution of hydrological variables, the review indicates that their practical application remains marginal and methodologically underexplored, which prevents the formulation of conclusive claims regarding their superiority or operational advantage. This lack of empirical evaluation is complemented by recurrent methodological limitations, data heterogeneity, scarcity of consistent multitemporal series, and the absence of comparable metrics that constrain critical analysis and hinder the generalization of the findings. In totality, these observations allow for a reinterpretation of the study's objective: rather than validating the performance of LSTM/RNN architectures, the results reveal a systematic gap in their implementation and evaluation, underscoring the need for research that jointly integrates temporal and spatial analysis to strengthen future monitoring and early warning systems.

The comparison of the study's findings with the analyzed articles reveals a clear methodological convergence, insofar as most investigations highlight the effectiveness of convolutional models and multisensor fusion strategies to improve accuracy in flood detection. This trend aligns with the high levels of accuracy reported in UAV-based applications [6], in SAR imagery from Sentinel-1 [9], in Copernicus program products, and in deep architectures such as U-Net or DeepLab [10], which demonstrate a notable capacity to process complex spatial information. However, a relevant divergence is observed regarding temporal approaches: although several studies employ historical data, multitemporal series, or continuous monitoring [13], the explicit application of RNNs or LSTMs is scarce, indicating a gap between the potential identified in the present study and current practices in the literature. Even so, the presence of spatiotemporal dynamics in the reviewed procedures suggests that the field is in a transitional phase toward more integrated approaches, which reinforces the theoretical relevance of recurrent models and their possible contribution to future research.

The results of the study show that CNNs continue to be the main theoretical reference in the literature due to their high performance in extracting spatial patterns, achieving reported accuracy values between 91% and 95% in flood detection tasks using well-established architectures such as U-Net or DeepLab [10]. (It should be noted that these values correspond to general performance metrics reported in the studies primarily accuracy and IoU which are not directly comparable across publications due to differences in datasets and experimental protocols.) Nevertheless, the increase in the availability of multitemporal series highlights the need to incorporate models capable of capturing temporal dependencies, such as LSTM and RNN architectures, whose theoretical benefits for modeling hydrological variability are not yet reflected in the reviewed literature. This absence reveals a methodological gap: although the studies work with sequential or multitemporal data, few employ recurrent approaches, which limits the exploitation of the dynamic component of flood processes. At the practical level, the use of UAVs stands out due to their ability to access hard to reach areas, reduce operational costs, and provide rapid and accurate detections [6]. Likewise, the integration of sensors such as Sentinel-1 and satellites from the Copernicus program expands geographic coverage [11], although with slight losses in spatial precision compared to very-high-resolution sensors. These approaches allow affected areas to be identified, soil moisture to be estimated, and water levels to be detected [5], thereby strengthening early warning systems. However, limitations persist, such as high computational requirements, sensitivity to atmospheric factors, and dependence on extensive and continuously updated databases. In summary, the methodological advances enrich the understanding of hydrometeorological processes and provide tools with a direct impact on risk management.

The limitations identified in the literature significantly affect the interpretation and scope of the results, as numerous studies depend on favorable environmental conditions and extensive datasets to ensure adequate performance. Factors such as variations in illumination, the presence of clouds, shadows, or watermarks introduce uncertainty that compromises the precise segmentation of flooded areas [7], weakening the internal validity of the models. Added to this is the reduced transferability to regions with different hydrological, topographic, or climatic characteristics, which restricts the generalization of the findings [14]. There also remains a lack of homogeneous methodological criteria both in the choice of sensors and in model design, which makes it difficult to rigorously compare approaches based on CNNs, LSTMs, or other deep learning architectures. Taken together, these limitations outline a scenario in which the effectiveness of the models depends closely on the quality of the available data, the geographic context, and the methodological consistency applied.

In totality, the evidence discussed demonstrates that the field of flood detection is moving toward more integrated approaches, where the combination of deep models, heterogeneous sensors, and multitemporal monitoring strategies constitutes the current axis of innovation. Although convolutional architectures continue to dominate due to their performance in spatial tasks, the gaps identified regarding the treatment of the temporal dimension reveal a clear opportunity for the systematic incorporation of recurrent models such as LSTM and RNN. At the same time, methodological limitations, particularly the lack of standardization, variability in data quality, and reduced geographic transferability introduce constraints that must be considered in any attempt at operational implementation. In this regard, the review offers a decisive contribution by clarifying the current state of the field, identifying its main gaps, and establishing the elements necessary to empirically strengthen future developments in monitoring and early warning systems.

5. Conclusion

The review shows that, although recurrent models and LSTM architectures possess theoretical potential for capturing relevant temporal dynamics in hydrological processes, CNNs and variants such as U-Net remain the predominant techniques due to their strong performance in spatial segmentation tasks. Likewise, the analyzed studies highlight the usefulness of UAVs for acquiring high-resolution imagery, the robustness of Sentinel-1 in complex atmospheric scenarios, and the improvements derived from multisensor integration. Cumulatively, these advances confirm the central role of deep learning in the rapid and accurate detection of floods; however, the limited adoption of recurrent models reveals a methodological gap that opens a clear avenue for future research aimed at systematically evaluating the contribution of RNNs and LSTMs in real operational contexts and with more demanding multitemporal datasets.

In light of the central objective of this work to examine the current and potential role of LSTM and RNN models in flood detection using satellite imagery the findings indicate that the contribution of these architectures remains incipient within the high-impact literature analyzed and still underrepresented in top-tier journals. Their limited adoption does not reflect a lack of theoretical relevance, but rather the absence of rigorous comparative studies that evaluate their usefulness against established models such as U-Net. In this regard, the review provides a conceptual clarification of where the temporal capabilities of recurrent models are actually positioned within the current methodological landscape, offering a more precise foundation to guide future research toward analytical frameworks that explicitly integrate the temporal and spatial dimensions of the flood phenomenon.

The findings of the review show that the integration of deep learning techniques with satellite sensors and UAV platforms constitutes an essential component for strengthening hydrometeorological risk management. The relevance of this approach responds to the increase in flood events and the need for more precise and operational tools. Nevertheless, limitations persist related to the dependence on large volumes of data, sensitivity to environmental conditions, and the limited generalization across regions with heterogeneous characteristics.

Taken together, the review makes it possible to more precisely delineate the incipient yet conceptually relevant role of LSTM models and other recurrent architectures within the current landscape of flood detection. Rather than offering definitive conclusions, the findings indicate the need for systematic comparative evaluations, reproducible experimental frameworks, and standardized multitemporal datasets that allow for rigorous measurement of the actual contribution of these models relative to well-established architectures such as U-Net. Based on this, several lines of future research are identified: (1) developing controlled studies that explicitly compare CNNs and LSTMs in operational scenarios; (2) expanding open repositories with high-quality temporal annotations; (3) designing hybrid approaches that integrate spatiotemporal dynamics within a single model; and (4) advancing toward unified methodological protocols that facilitate transferability across regions. These directions constitute necessary steps for consolidating the application of deep models in monitoring and early warning systems, contributing to a more robust and verifiable technological development.

Ethics Statement

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

CRediT authorship contribution statement

Christopher Abraham Arevalo-Erosa: Conceptualization, Methodology, Formal analysis, Investigation, Data curation, Writing – original draft, Writing – review & editing, Visualization. **Jose Emilio Ek-Chay:** Conceptualization, Project administration, Supervision. **Rodrigo Kalel Rodriguez-Chavez:** Conceptualization, Project administration, Supervision. **Máximo Arturo Guzmán-Ascencio:** Conceptualization, Project administration, Supervision. **Raul Humberto Rodriguez-Mendez:** Conceptualization, Project administration, Supervision. **Romell Gerardo Mejia-Ramón:** Conceptualization, Project administration, Supervision. **Misael Garcia:** Conceptualization, Project administration, Supervision.

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

During the preparation of this work, the authors used ChatGPT to assist with language clarity and grammar correction. All final edits and intellectual content were determined 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] P. A. B. Andrade, P. E. C. Zamora, A. E. C. Lara, and J. P. P. Piedra, "A comprehensive evaluation of ai techniques for air quality index prediction: Rnns and transformers," *Ingenius*, 2025. doi: 10.17163/ings.n33.2025.06.

[2] E. Heidari, V. Samadi, and A. A. Khan, "Leveraging recurrent neural networks for flood prediction and assessment," *Hydrology*, vol. 12, no. 4, 2025. doi: 10.3390/hydrology12040090.

[3] A. M. Rugină, "The use of recurrent neural networks (s-rnn, lstm, gru) for flood forecasting based on data extracted from classical hydraulic modeling," *Modelling in Civil Environmental Engineering*, vol. 18, no. 3, pp. 1–18, 2023. doi: 10.2478/mmce-2023-0011.

[4] C. Chen, D. Luan, S. Zhao, Z. Liao, Y. Zhou, J. Jiang, and Q. Pei, "Flood discharge prediction based on remote-sensed spatiotemporal features fusion and graph attention," *Remote Sensing*, vol. 13, no. 24, 2021. doi: 10.3390/rs13245023.

[5] G. S. Nijaguna, D. R. Manjunath, M. Abouhawwash, S. S. Askar, D. K. Basha, and J. Sengupta, "Deep learning-based improved wcm technique for soil moisture retrieval with satellite images," *Remote Sensing*, vol. 15, 2023. doi: 10.3390/rs15082005.

[6] H. S. Munawar, F. Ullah, S. Qayyum, S. I. Khan, and M. Mojtahedi, "Uavs in disaster management: Application of integrated aerial imagery and convolutional neural network for flood detection," *Sustainability (Switzerland)*, vol. 13, 2021. doi: 10.3390/su13147547.

[7] F. Pech-May, R. Aquino-Santos, and J. Delgadillo-Partida, "Sentinel-1 sar images and deep learning for water body mapping," *Remote Sensing*, vol. 15, pp. 1–22, 2023. doi: 10.3390/rs15123009.

[8] Z. Boulouard, M. Ouaissa, M. Ouaissa, F. Siddiqui, M. Almutiq, and M. Krichen, "An integrated artificial intelligence of things environment for river flood prevention," *Sensors*, vol. 22, pp. 1–20, 2022. doi: 10.3390/s22239485.

[9] X. Wu, Z. Zhang, S. Xiong, W. Zhang, J. Tang, Z. Li, B. An, and R. Li, "A near-real-time flood detection method based on deep learning and sar images," *Remote Sensing*, vol. 15, 2023. doi: 10.3390/rs15082046.

[10] H. Wu, H. Song, J. Huang, H. Zhong, R. Zhan, X. Teng, Z. Qiu, M. He, and J. Cao, "Flood detection in dual-polarization sar images based on multi-scale deeplab model," *Remote Sensing*, vol. 14, pp. 1–22, 2022. doi: 10.3390/rs14205181 .

[11] L. Zhang and J. Xia, "Flood detection using multiple chinese satellite datasets during 2020 china summer floods," *Remote Sensing*, vol. 14, p. 51, 12 2021. doi: 10.3390/rs14010051 .

[12] S. Priscillia, C. Schillaci, and A. Lipani, "Flood susceptibility assessment using artificial neural networks in indonesia," *Artificial Intelligence in Geosciences*, vol. 2, pp. 215–222, 2021. doi: 10.1016/j.aiig.2022.03.002 .

[13] R. Al-Ruzouq, A. Shanableh, R. Jena, M. B. A. Gibril, N. A. Hammouri, and F. Lamghari, "Flood susceptibility mapping using a novel integration of multi-temporal sentinel-1 data and extreme deep learning model," *Geoscience Frontiers*, vol. 15, 5 2024. doi: 10.1016/j.gsf.2024.101780 .

[14] A. Cerbelaud, G. Blanchet, L. Roupioz, P. Breil, and X. Briottet, "Mapping pluvial flood-induced damages with multi-sensor optical remote sensing: A transferable approach," *Remote Sensing*, vol. 15, pp. 1–27, 4 2023. doi: 10.3390/rs15092361 .

[15] G. K. Wedajo, T. D. Lemma, T. Fufa, and P. Gamba, "Integrating satellite images and machine learning for flood prediction and susceptibility mapping for the case of amibara, awash basin, ethiopia," *Remote Sensing*, vol. 16, p. 2163, 6 2024. doi: 10.3390/rs16122163 .

Journal of Artificial Intelligence and Computing Applications

MAIKRON

Volume 4 · Issue 1 · January–June 2026 · Open Access (CC BY 4.0) · eISSN 3061-8843 ·
DOI 10.5281/zenodo.20804305 · Published by Maikron · maikron.org/jaica