Trends
in Psychiatry and Psychotherapy

JOURNAL ARTICLE PRE-PROOF
(as accepted)

Review Article

The Role of Large Language Models in Teaching Psychiatric Semiology: A Systematic Review

Vinícius Vicente Soares, Felipe Francisco de Castro Passos, Isabel Cristina Siqueira da Silva, Flávio Milman Shansis, Juliana Silva Herbert

http://doi.org/10.47626/2237-6089-2025-1237

Original submitted Date: 12-Nov-2025

Accepted Date: 02-Apr-2026

The Role of Large Language Models in Teaching Psychiatric Semiology: A Systematic Review

Short title: LLMs in Psychiatric Semiology Education

Vinícius Vicente Soares[1]*, Felipe Francisco de Castro Passos[2], Isabel Cristina Siqueira da Silva[1], Flávio Milman Shansis[3], Juliana Silva Herbert[1]

[1]Graduate Program in Information Technologies and Healthcare Management, Federal University of Health Sciences of Porto Alegre (UFCSPA), Porto Alegre, RS, Brazil.
[2]Graduate Program in Psychiatry and Behavioral Sciences, Federal University of Rio Grande do Sul (UFRGS), Porto Alegre, RS, Brazil.
[3]Graduate Program in Health Sciences, Federal University of Health Sciences of Porto Alegre (UFCSPA), Porto Alegre, RS, Brazil.

*Corresponding author: viniciusvs@ufcspa.edu.br

Corresponding author: R. Sarmento Leite, 245 - Centro Histórico, Porto Alegre - RS, 90050-170, Brazil; Tel: +55 51 985617546; Fax: N/A; Email: viniciusvs@ufcspa.edu.br.

The Role of Large Language Models in Teaching Psychiatric Semiology: A Systematic Review

Abstract

Introduction: Teaching psychiatric semiology faces challenges such as the limited availability of real patients for educational purposes and a shortage of specialized instructors. This systematic review investigated the applications of Large Language Models (LLMs), exemplified by ChatGPT, to address these gaps. **Methods:** Studies published between November 2022 and September 17, 2025, were identified through searches in PubMed/MEDLINE, Scopus, Web of Science, IEEE Xplore, EBSCO, BVS, and EMBASE databases. Two independent reviewers conducted the systematic review following the PRISMA 2020 checklist, and the methodological quality of the included studies was individually assessed using the Mixed Methods Appraisal Tool (MMAT). **Results:** Of the 1,549 studies initially identified, two met the inclusion criteria, involving medical students and educators in Canada and Switzerland. LLMs demonstrated performance comparable to human experts in creating diagnostic tests and clinical vignettes but exhibited occasional simplifications and algorithmic biases. Participants reported positive perceptions of the tools' efficiency and practicality, but emphasized the need for specialized supervision and attention to potential privacy issues and superficiality in complex cases. **Conclusions:** LLMs show potential as valuable supplementary resources for teaching psychiatric semiology, especially in settings with shortages of teachers and available patients for training. However, their limitations, including cultural and algorithmic biases, potential privacy risks, and superficial content generation, require close monitoring by experienced professionals. Although current evidence is preliminary, prospects are promising, highlighting the need for more robust and long-term studies to evaluate the educational use of LLMs.

Keywords: Artificial Intelligence; ChatGPT; Medical Education; Psychiatric Semiology; Systematic Review.

1. Introduction

Psychiatric disorders represent a significant and increasing global health burden, impacting quality of life and raising the risk of premature mortality [1]. Despite their prevalence, the true burden is often underestimated, highlighting the need for effective preventive and therapeutic strategies to manage these conditions and their associated comorbidities [2].

Training healthcare professionals in psychiatric semiology faces significant challenges. Medical schools often allocate insufficient teaching hours, and traditional methods like presenting real patients can be limiting for both students and patients [3,4]. This lack of innovative, simulation-based methodologies hinders the development of essential clinical skills like critical thinking and empathetic communication [3].

Given these limitations, Large Language Models (LLMs), such as ChatGPT, have emerged as potential tools to address these educational gaps. A recent bibliometric analysis highlights the increasing momentum of chatbots in the digital mental health sector [5]. Powered by artificial intelligence (AI), these models can generate complex texts, including clinical vignettes and simulated dialogues, offer automated feedback, and facilitate personalized learning [6,7]. However, the effectiveness of LLMs remains uncertain, as noted in a recent scoping review [8], and their development is accompanied by potential drawbacks, including algorithmic biases, privacy concerns, and inaccuracies in clinically complex scenarios that require consistent scientific grounding [9,10].

This systematic review investigates the application of LLMs in teaching psychiatric semiology, guided by the following research questions:

- RQ1: Which educational interventions using LLMs have been reported in the literature for teaching psychiatric semiology?
- RQ2: What are students' and educators' perceptions regarding the integration of LLMs in their training?
- RQ3: What barriers and challenges have been identified in adopting these models?
- RQ4: Which ethical aspects require greater attention when incorporating LLMs into the educational context?

2. Methodology

To evaluate the literature about the application of LLMs to teaching psychiatric semiology, a systematic review was chosen and conducted following the PRISMA 2020 checklist guidelines [11].

2.1. Information sources

Relevant databases for medical education and psychiatric training (PubMed/MEDLINE, Scopus, Web of Science, EBSCO, BVS, and EMBASE), as well as for engineering and technology (IEEE Xplore), were utilized, covering publications from November 2022 (the launch date of ChatGPT) to September 17, 2025 (the research date).

2.2. Search strategy

The search strategy was developed through exploratory searches in the aforementioned databases and subsequently reviewed by information technology and medical-psychiatric education experts on the study team. The following descriptors were employed:

- Large Language Models: “Large Language Models”, LLMs, GPT-3, GPT-4, ChatGPT, “Generative AI”;
- Teaching and medical education in psychiatry: “medical education”, “psychiatric education”, “clinical training”, “health education”;
- Psychiatric semiology: “psychiatric semiology”, “psychopathology”, “mental status examination”, “psychiatric diagnosis”, “psychiatry”.

These terms were combined using Boolean operators (AND, OR) to refine the results, following the general formula:

(("Large Language Models" OR LLMs OR GPT-3 OR GPT-4 OR ChatGPT OR "Generative AI")

AND

(teaching OR "medical education" OR learning OR "clinical training" OR "psychiatric education" OR "health education")

AND

("psychiatric semiology" OR psychopathology OR "mental status examination" OR "psychiatric diagnosis" OR psychiatry))

The only filter applied was the publication period from November 2022 to September 2025. Each database required specific adjustments regarding the insertion of descriptors and operators, but the core concepts remained consistent.

2.3. Eligibility criteria

Table 1 outlines the inclusion and exclusion criteria (IC and EC) for study selection:

Table 1. Inclusion and Exclusion Criteria.

Inclusion Criteria (IC)	
IC1	Empirical studies (e.g., quantitative, qualitative, mixed-methods, case studies) relevant to the use of LLMs in psychiatric semiology.
IC2	Participants involved in psychiatric training (medical students, residents, educators).
IC3	Publications in Portuguese, English, or Spanish, from November 2022 to September 2025.
IC4	Posters or abstracts that clearly present methods and results.
Exclusion Criteria (EC)	
EC1	Studies that did not involve LLMs and/or were not related to psychiatric semiology education.
EC2	Educational interventions outside the context of psychiatric semiology.
EC3	Purely theoretical articles, editorials, letters to editors, or opinion papers without empirical data.

2.4. Study selection and data collection

Two independent reviewers conducted study selection, screening titles and abstracts using Rayyan [12]. When titles and abstracts indicated potential inclusion, full-text evaluation followed. In cases of uncertainty or disagreement, a third reviewer was consulted until consensus. After selecting the studies, data were extracted through a standardized form documenting:

- General information (title, author, year, country, journal);
- Methodology (design, type of intervention, instruments);
- Target population (number of participants, participant profile, educational setting);
- Application of LLMs (model used, purpose, parameter settings);
- Results and outcomes, including efficacy, usability, perceptions, barriers, and limitations.

2.5. Data items

The main outcomes defined were:

- Applications of LLMs in creating instructional scenarios and materials (including clinical simulations, question generation, and automated feedback);
- Efficacy and usability of interventions;
- Perceptions of students and educators;
- Educational and technological barriers;
- Ethical considerations and biases inherent in AI tools.

Secondary variables included methodological indicators (bias, study quality, appropriateness to the theme) and evaluation parameters (e.g., comparisons with human experts).

2.6. Study risk of bias assessment

The methodological quality of the included studies was assessed using the Mixed Methods Appraisal Tool (MMAT), version 2018. The MMAT was chosen because it is specifically designed for the appraisal stage of systematic mixed studies reviews, permitting the evaluation of qualitative, quantitative, and mixed methods studies within a single framework [13]. Two independent reviewers applied the specific criteria category corresponding to each study's design, with disagreements resolved by consensus.

2.7. Synthesis methods

Due to the variety of designs and outcomes, descriptive synthesis was selected as the main strategy. Gathering qualitative and quantitative evidence allowed the identification of patterns, trends, and gaps. The heterogeneity of methods, samples, and interventions precluded a formal meta-analysis.

2.8. Reporting bias assessment

To assess reporting bias, the content presented in the articles was compared with any registered protocols, analyzing discrepancies between planned and reported outcomes. For studies without available protocols, internal coherence between methods and results was assessed. Publication bias was also investigated by examining whether there was a tendency to report only positive results, potentially overstating the efficacy of LLMs.

2.9. Confidence assessment

The confidence in the synthesized evidence was evaluated using the GRADE-CERQual (Confidence in the Evidence from Reviews of Qualitative research) approach. This evaluation considered four components: methodological limitations, coherence, adequacy of data, and relevance. Based on these criteria, the overall confidence in the evidence was classified as high, moderate, low, or very low.

2.10. AI- assisted writing disclosure

To improve language clarity and formatting, we used an AI- assisted tool (ChatGPT, OpenAI) limited to copy- editing and style suggestions. No content was generated without author oversight; all authors critically reviewed and approved the final manuscript and accepted full responsibility for its content.

3. Results

3.1. Study selection

The initial database search yielded 1,549 records. After 738 duplicates were removed, 811 unique records were screened by title and abstract. Of these, 795 were excluded, leaving 16 reports for full-text eligibility assessment. Following the full-text review, 14 reports were excluded, resulting in two studies that met the inclusion criteria for this systematic review. The study selection process is detailed in the PRISMA flowchart (Figure 1).

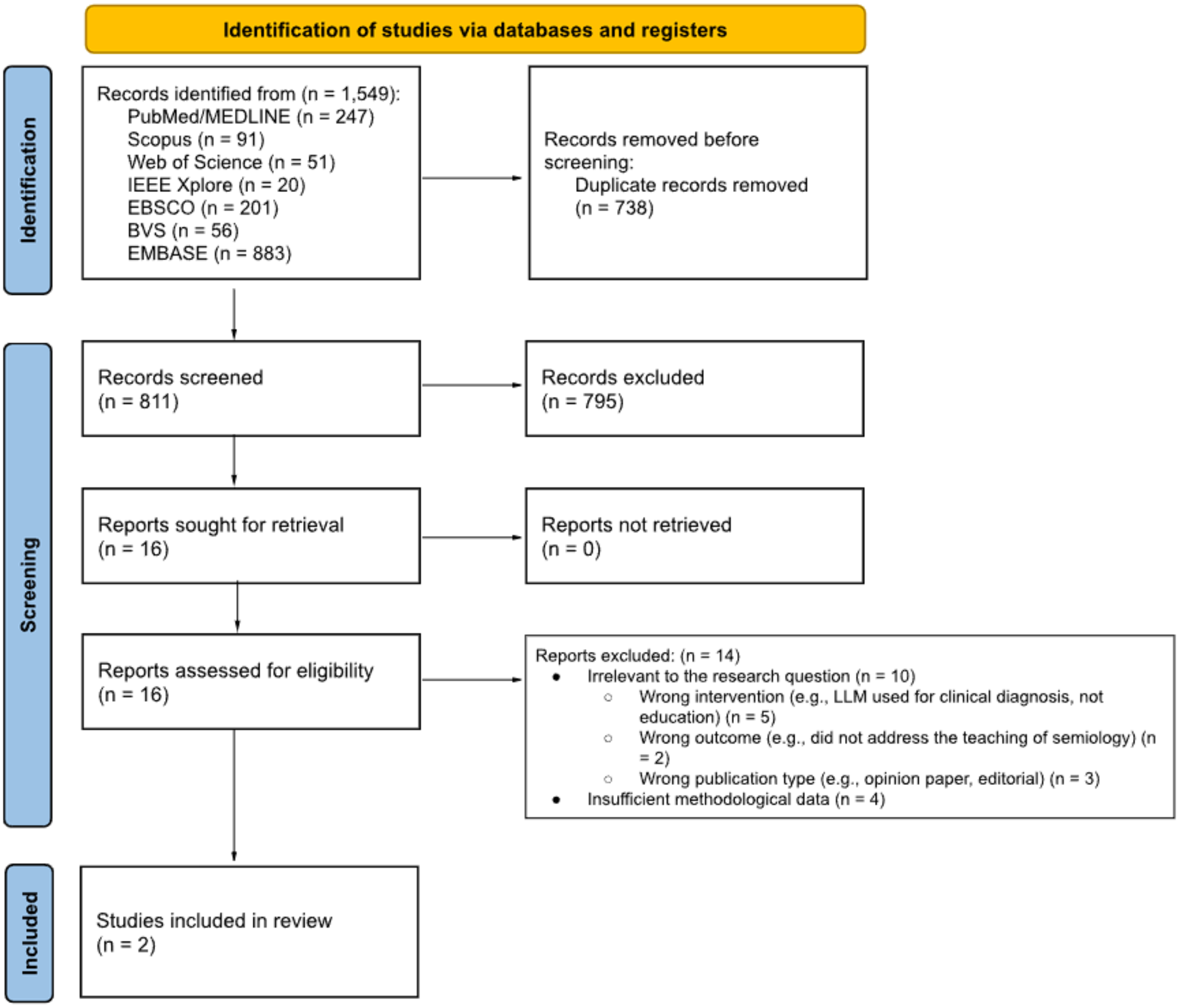

Figure 1. Flowchart illustrating identification, screening, eligibility, and inclusion of studies.

3.2. Study characteristics

The two included studies employed distinct methodologies to explore the role of LLMs in psychiatric education.

A mixed-methods study in Canada compared the quality of Script Concordance Tests (SCTs) generated by ChatGPT with those created by clinical experts [14].

The study recruited 102 psychiatry residents and clinician-educators, who used a web-based survey to evaluate three AI-generated and three expert-created SCTs. The evaluation was guided by a formal conceptual framework, analyzing the clinical scenarios, the quality of the questions, and the expert opinions provided. The qualitative component involved content analysis of open-ended feedback on the perceived strengths and weaknesses of each SCT type.

In Switzerland, a qualitative exploratory study was performed to assess the potential functionalities of ChatGPT (version 3.5) in social psychiatry education [15]. The research team acted as the primary users, prompting the AI to first outline its potential educational roles and then to generate a hypothetical clinical vignette of a migrant patient. The study's data consisted of the direct text output from the AI, which was then qualitatively assessed by the authors for its plausibility, relevance, and utility as a teaching tool for discussing social determinants of health, comparing it against existing educational materials from their institution.

3.3. Methodological quality of included studies

The methodological quality of the included studies was appraised using the MMAT. Hudon et al. was assessed under Category 5 (Mixed Methods) and met all quality criteria, demonstrating an adequate rationale for the design, effective integration of components, and rigorous adherence to the methodological standards of both quantitative and qualitative traditions. Smith et al. was evaluated under Category 1 (Qualitative). While the study presented an appropriate approach and adequate data collection methods for an exploratory study, criteria regarding the substantiation of interpretations and overall analytical coherence were rated as "Can't tell". This was due to the lack of an explicitly described systematic analytical framework, which introduced subjectivity into the interpretation of the AI-generated outputs. Table 2 details the item-by-item appraisal.

Table 2. Methodological Quality Assessment of Included Studies using the Mixed Methods Appraisal Tool (MMAT, version 2018)

Study	Design	MMAT Category	Screening (S1, S2)	C1	C2	C3	C4	C5
Hudon et al. [14]	Mixed-Methods	Category 5	Yes, Yes	Yes	Yes	Yes	Yes	Yes
Smith et al. [15]	Qualitative	Category 1	Yes, Yes	Yes	Yes	Yes	Can't tell*	Can't tell*

Table Legend:

- Screening Questions: S1 (Clear research questions?), S2 (Collected data allow to address the research questions?).
- Category 5 Criteria (Mixed Methods): C1 (Adequate rationale?), C2 (Effective integration?), C3 (Outputs adequately interpreted?), C4 (Divergences addressed?), C5 (Adherence to quality criteria of each tradition?).
- Category 1 Criteria (Qualitative): C1 (Appropriate approach?), C2 (Adequate data collection?), C3 (Findings derived from data?), C4 (Interpretation substantiated by data?), C5 (Coherence between sources, collection, analysis, and interpretation?).
- Notes: (*) Rated as "Can't tell" for Smith et al. due to the absence of a described systematic or rigorous data analysis process, relying primarily on the authors' subjective assessment of the AI output.

3.4. Results of individual studies

In the study by Hudon et al. [14], AI-generated SCTs displayed overall quality comparable to human-produced tests, suggesting potential for optimizing faculty time. However, simplification of cases could compromise the depth of clinical reasoning. Smith et al. [15] found that vignettes created by ChatGPT were perceived as realistic and useful, though concerns were raised about stereotypical representations and methodological inconsistencies hindering scenario reproducibility.

3.5. Synthesis of results

3.5.1. Educational Interventions Using LLMs (RQ1)

In the two analyzed studies, the use of LLMs primarily involved the creation of instructional materials, such as Script Concordance Tests (SCTs) and clinical vignettes. Hudon et al. [14] emphasized the efficacy of ChatGPT-generated SCTs as a resource for training diagnostic reasoning under uncertainty, a crucial element in psychiatric semiology. Similarly, Smith et al. [15] explored thematic vignette creation to facilitate reflections on social determinants in psychiatry, such as the challenges faced by migrant populations. Both studies highlight the potential of LLMs to generate realistic scenarios for educational purposes.

3.5.2. Perceptions of Students and Educators (RQ2)

Regarding acceptance, positive feedback was reported from residents and educators concerning ChatGPT's efficiency in SCT generation, which reduced test preparation time and allowed for more supervision and advanced discussions. Educator satisfaction was also highlighted regarding the variety of AI-generated cases and the opportunity to address often-overlooked psychosocial factors. However, both included studies reported concerns about superficial content and potential biases or stereotypes, emphasizing the need for human supervision to ensure clinical accuracy.

3.5.3. Barriers and Challenges in Adopting LLMs (RQ3)

The review identified technical, pedagogical, and ethical barriers. “Algorithmic hallucinations” wherein models generate imprecise or spurious information, pose risks to content credibility [9,10]. Smith et al. [15] warned about reproducing cultural stereotypes in vignettes, especially involving minority or migrant contexts. Additionally, the absence of specialized supervision might lead to incorrect interpretations, particularly in complex clinical cases. Another challenge was the technological and budgetary infrastructure: institutions with limited resources may lack reliable internet access or servers robust enough to run more sophisticated models. These restrictions hinder sustainable, large-scale LLM implementation and could exacerbate educational inequalities.

3.5.4. Ethical Considerations in Integrating LLMs (RQ4)

Adopting models like ChatGPT in psychiatric semiology education brings ethical implications. Privacy and confidentiality are paramount. Even though simulations do not use real patient data, model training processes might have included sensitive or partial information, resulting in biases. Hudon et al. [14] highlighted the importance of transparency in AI-generated content to avoid confusion about educational material sources. Smith et al. [15] emphasized that a lack of diversity in datasets used for training LLMs results in biased representations, negatively impacting minority groups and perpetuating stereotypes. Institutions are therefore advised to develop ethical usage protocols, incorporate clear student guidelines, and invest in diverse, extensive databases to reduce biases.

3.6. Reporting biases

In this review, publication bias remains significant, as studies reporting difficulties or negative outcomes might not have been captured by the search strategies used. Furthermore, the reviewed articles relied on artificial data, vignettes, or simulations, limiting applicability to real clinical scenarios. LLM training typically occurs from publicly available datasets, which may contain biased or incomplete information, affecting result reliability. Comparisons with human expert standards also do not guarantee neutrality, as these professionals may harbor their own clinical and cultural biases.

3.7. Confidence of evidence

According to the GRADE-CERQual criteria, the overall confidence in the evidence derived from the included studies was low to moderate. Primary factors reducing confidence were methodological limitations in data analysis and the adequacy of data, given the reliance on small or non-participant-based exploratory samples. The absence of longitudinal methodologies limits definitive conclusions regarding the long-term impact of LLMs in psychiatric education.

4. Discussion

The analysis of the two included studies indicates that Large Language Models (LLMs) possess substantial potential to support psychiatric semiology education. Their capacity to generate realistic assessments and clinical vignettes can enrich learning experiences, particularly in settings with restricted exposure to diverse psychopathology. Hudon et al. [14] demonstrated that ChatGPT produced Script Concordance Tests comparable with expert-developed versions, permitting educators to devote more time to mentorship and iterative refinement of AI-generated material. Similarly, Smith et al. [15] reinforced the value of AI-produced vignettes for addressing complex social determinants of health, facilitating deeper debate on migration and vulnerability. One of the most significant findings of this review is the acute scarcity of empirical research specifically focused on the application of LLMs in psychiatric semiology education. The retrieval of only two eligible articles, despite a comprehensive search across seven major databases, underscores the incipient nature of this intersection. Therefore, the scarcity of evidence must itself be considered a central finding, highlighting a critical gap in the literature and signaling an urgent need for methodologically robust studies to build a solid evidence base.

4.1. Contextualizing LLM Applications in General Medical and Health Education

While empirical evidence specifically focused on psychiatric semiology remains scarce, the foundational findings of this review align with the broader integration of LLMs across medical and health education. Beyond the specialized scope of this study, current literature confirms that generative AI is increasingly utilized across medical disciplines to innovate teaching methodologies and generate high-fidelity synthetic patient data [7]. Within the broader landscape of psychiatric training, recent evidence [17] reinforces our observations, demonstrating the viability of LLMs for creating simulated clinical encounters and enhancing self-directed learning.

Furthermore, the positive perceptions of AI efficiency observed among medical residents and educators are consistent with trends in allied mental health disciplines [14, 15]. For instance, research indicates that psychiatric mental health nursing students perceive ChatGPT as a valuable asset for mastering complex theoretical frameworks and completing assignments [16]. This is further supported by bibliometric analyses that map the rapid expansion of chatbots in digital mental health, underscoring the transformative integration of these tools across the educational spectrum [5].

4.2. Limitations and Ethical Imperatives

While these pedagogical advancements are promising, they are accompanied by critical considerations that necessitate a cautious interpretation. As evaluated through the GRADE-CERQual framework, the overall confidence in the current evidence is characterized as low to moderate. This assessment primarily reflects the nascent state of the field, marked by exploratory designs and the inherent absence of longitudinal data typical of rapidly evolving technologies.

The limitations identified in our review — including "algorithmic hallucinations", potential cultural biases, and the risk of reductive analysis in complex diagnostic scenarios [14, 15] — represent structural challenges that require ongoing mitigation. The reliability of AI-generated content remains intrinsically linked to the granularity and representativeness of its training datasets; thus, without specialized oversight and robust institutional protocols, algorithmic bias may inadvertently perpetuate existing inequities [15].

We emphasize that LLMs are designed to augment, rather than replace, clinical judgment and direct patient interaction. They should function as part of a synergistic and diversified pedagogy. Therefore transparency regarding AI deployment, stringent data protection, and explicit disclosure of training set diversity must become standard academic practice. Future research involving larger, multi-center cohorts, comparative architectural analyses, and long-term clinical outcome metrics will be pivotal to scaling these foundational observations. Ultimately, the integration of LLMs into psychiatric training must proceed with ethical stewardship, ensuring that technological innovation remains strictly aligned with the rigorous demands of mental health education.

5. Conclusion

This systematic review demonstrates that LLMs, exemplified by ChatGPT, possess the potential to support psychiatric semiology education in various contexts. Key applications include the generation of simulated clinical scenarios, diagnostic tests, and clinical vignettes, which can provide learners with valuable practice opportunities. Conversely, significant limitations were identified, including the risk of "algorithmic hallucinations", inherent cultural biases, superficial treatment of complex cases, and a scarcity of controlled clinical studies.

To consolidate a secure and effective use of LLMs, rigorous protocols including faculty supervision, ongoing content validation, and data governance guidelines are recommended. Cultural appropriateness and representativeness in training datasets emerge as priorities to minimize biases and ensure the inclusion of diverse populations. Future studies, employing larger samples and robust experimental designs, could elucidate the long-term efficacy of these tools across different educational contexts. A balanced integration of AI with human expertise, combined with ethical and responsible technological development, holds promise for enhancing psychiatric semiology education.

Data availability statement

All data analyzed during this study are included in the published articles cited within the manuscript. No new data were created, and all sources are publicly available through the respective scientific databases and journals referenced.

Acknowledgments

The authors declare that this research was conducted without financial support from public or private institutions or non-profit organizations.

Disclosure

The authors declare no conflicts of interest.

Ethics statement

This article focused on research methods and did not involve human participants or patients. As such, ethics approval and informed consent were not required.

Registration and protocol

This systematic review was previously registered in the International Prospective Register of Systematic Reviews (PROSPERO) under the number CRD42025646591.

Figure legends

Figure 1. PRISMA 2020 flow diagram of study selection. Searches were run in PubMed/MEDLINE (n=247), Scopus (n=91), Web of Science (n=51), IEEE Xplore (n=20), EBSCO (n=201), BVS (n=56), and EMBASE (n=883). Timeframe: November 2022 to 17 September 2025; Duplicates removed: n=738; records screened: n=811; records excluded: n=795; full-text reports assessed: n=16; reports not retrieved: n=0; reports excluded with reasons: n=14 (irrelevant to the research question, n=10: wrong intervention, n=5; wrong outcome, n=2; wrong publication type, n=3; and studies with insufficient methodological data, n=4). Studies included in review: n=2.

Author contributions

Vinícius Vicente Soares: Conceptualization, Methodology, Formal analysis, Writing – original draft, Writing – review & editing. Felipe Francisco de Castro Passos: Conceptualization, Methodology, Formal analysis, Writing – original draft, Writing –

review & editing. Isabel Cristina Siqueira da Silva: Writing – review & editing. Flávio Milman Shansis: Conceptualization, Methodology, Supervision, Writing – review & editing. Juliana Silva Herbert: Conceptualization, Methodology, Supervision, Writing – review & editing.

Handling Editor: Ives Passos

References

1. The Lancet Psychiatry. Global Burden of Disease 2021: mental health messages. Lancet Psychiatry. 2024;11:573.
2. Momen NC, Plana-Ripoll O, Agerbo E, Christensen MK, Iburg KM, Laursen TM, et al. Mortality associated with mental disorders and comorbid general medical conditions. JAMA Psychiatry. 2022;79:444.
3. Cataldo Neto A, Pitta JCDN, Brasil MA, Adad MA, Aguiar RWD, Selbach CR, et al. Panorama nacional do ensino da psiquiatria nas escolas médicas brasileiras. Rev Psiquiatr Rio Gd Sul. 2011;33:8-13.
4. Ferreira AP, Netto VM. O ensino da clínica psicopatológica: o caso da sessão clínica. Rev Latinoam Psicopatol Fundam. 2009;12:481-96.
5. Han Q, Zhao C. Unleashing the potential of chatbots in mental health: bibliometric analysis. Front Psychiatry. 2025;1:1494355.
6. Li Y, Zeng C, Zhong J, Zhang R, Zhang M, Zou L. Leveraging large language model as simulated patients for clinical education. arXiv preprint arXiv:2404.13066. 2024 [cited 2025 Sep 20]. Available from: https://arxiv.org/abs/2404.13066.
7. Zhui L, Yhap N, Liping L, Zhengjie W, Zhonghao X, Xiaoshu Y, et al. Impact of large language models on medical education and teaching adaptations. JMIR Med Inform. 2024;1:e55933.
8. Hua Y, Na H, Li Z, Liu F, Fang X, Clifton D, et al. A scoping review of large language models for generative tasks in mental health care. npj Digit Med. 2025;8:230.
9. Lawrence HR, Schneider RA, Rubin SB, Matarić MJ, McDuff DJ, Jones Bell M. The opportunities and risks of large language models in mental health. JMIR Ment Health. 2024;1:e59479.

10. Liu Z, Bao Y, Zeng S, Qian R, Deng M, Gu A, et al. Large language models in psychiatry: current applications, limitations, and future scope. Big Data Min Anal. 2024;7:1148-68.
11. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ. 2021;372:n71.
12. Ouzzani M, Hammady H, Fedorowicz Z, Elmagarmid A. Rayyan-a web and mobile app for systematic reviews. Syst Rev. 2016;5:210.
13. Hong QN, Fàbregues S, Bartlett G, Boardman F, Cargo M, Dumont P, et al. The Mixed Methods Appraisal Tool (MMAT) version 2018 for information professionals and researchers. Educ Inf. 2018;34(4):285–91. doi: 10.3233/EFI-180221.
14. Hudon A, Kiepura B, Pelletier M, Phan V. Using ChatGPT in psychiatry to design script concordance tests in undergraduate medical education: mixed methods study. JMIR Med Educ. 2024;10:e54067.
15. Smith A, Hachen S, Schleifer R, Bhugra D, Buadze A, Liebrenz M. Old dog, new tricks? Exploring the potential functionalities of ChatGPT in supporting educational methods in social psychiatry. Int J Soc Psychiatry. 2023;69:1882-9.
16. Albikawi ZF, Abuadas MH. Investigating the impact of ChatGPT's utilization on psychiatric mental health nursing learning: a student perspective. Univ J Public Health. 2025;13:659-68.
17. Lee QY, Chen M, Ong CW, Ho CSH. The role of generative artificial intelligence in psychiatric education: a scoping review. BMC Med Educ. 2025;25:438.
18. Omar M, Soffer S, Charney AW, Landi I, Nadkarni GN, Klang E. Applications of large language models in psychiatry: a systematic review. Front Psychiatry. 2024;15:1422807.