From Rule-Based Agents to Agentic AI: A Comprehensive Survey of Mental Health Chatbots

Shravani Jadhav[1], Anupama Tilak[2], Om Jadhav[3]
Rutuja Jadhav[4], Dr. Bharati P. Vasgi[5], Prof. Punam Chavan[6]
[1,2,3,4,5,6]*Marathwada Mitra Mandal's College of Engineering*

***Abstract*—Conversational agents have gained increasing attention in mental health care as scalable, cost-effective, and stigma-free support tools. Evolving from rule-based systems to advanced dialogue models powered by large language technologies, these chatbots are now applied across diverse domains, including general mental health awareness, student support, caregiver assistance, and disorder-specific interventions. Despite their promise, many existing systems lack flexibility, personalization, and the ability to effectively address the needs of heterogeneous user groups particularly caregivers supporting children with early-onset mental health conditions. Recent advances in agentic AI, enabling autonomous reasoning, context-aware interaction, and adaptive decision-making, offer new opportunities to address these limitations. In parallel, retrieval-augmented generation (RAG) is being explored to improve reliability, trustworthiness, and user-centered interaction. This paper surveys current mental health chatbot solutions, categorizing them into general-purpose, student focused, caregiver-oriented, and disorder specific frameworks, and discusses open challenges and future directions for agentic AI–driven chatbot design.**

***Index Terms*—Agentic AI, Human-AI collaboration, mental chatbot, Retrieval Augmented Generation (RAG), Large Language Models (LLMs)**

I. INTRODUCTION

Mental health has become a major public health concern worldwide, with disorders that originate in childhood and adolescence posing particular challenges for families and healthcare systems [4], [8]. Early-onset conditions such as anxiety, depression, and developmental disorders often remain underdiagnosed due to stigma, lack of awareness, or limited access to specialized professionals [8], [5]. In these cases, parents, teachers, and caregivers play a crucial role in identifying symptoms and supporting the affected child [13], [14]. However, resources designed to assist caregivers directly remain scarce [13], [14], and traditional healthcare systems struggle to meet the demand for timely, personalized, and cost-effective support [8], [10].

In parallel, conversational agents commonly referred to as chatbots have emerged as promising digital tools in healthcare [1], [4], [6]. Initially based on rigid, rule driven scripts, these systems have evolved through advances in natural language processing and machine learning into interactive platforms capable of engaging users in meaningful dialogue [6], [7], [9]. Today, mental health chatbots are deployed in multiple contexts:

providing general awareness and psychoeducation [3], supporting students coping with academic and emotional pressures [12], assisting caregivers in their responsibilities [11], and offering condition-specific guidance for disorders such as depression or autism spectrum disorders [8], [15]. Their potential lies in their accessibility, affordability, and ability to deliver stigma free support at scale [4], [5].

Emerging paradigms such as agentic AI and retrieval augmented generation (RAG) offer new opportunities to overcome these barriers [2], [7], [16]. By combining autonomous reasoning, context-aware interaction, and grounding in verified knowledge sources, these approaches have the potential to deliver more reliable, trustworthy, and user-centered conversational agents [16], [17]. This survey paper examines the evolution and current state of mental health chatbots across four domains: general-purpose tools, student-focused applications, caregiver-oriented systems, and disorder specific interventions [6], [11]. The paper highlights how agentic AI and RAG can drive the next generation of chatbot design, supporting early detection, improved caregiver engagement, and more sustainable mental health care delivery [2], [7].

A. Background and Foundations

Childhood and Adolescent Mental Disorders:

Mental disorders emerging in childhood and adolescence constitute a major public health challenge [4], [8]. Conditions such as anxiety, depression, ADHD, and autism spectrum disorders affect emotional development, academic performance, and social integration, making early identification and intervention essential to reduce long-term impacts [5], [8], [13], [15].

Evolution of Conversational Agents in Mental Health: Conversational agents have evolved from rigid rule-based systems to advanced AI-driven platforms capable of understanding natural language and delivering personalized mental health support [1], [6], [7], [16]. Modern chatbots support general awareness, students, caregivers, and disorder-specific guidance, offering scalable, cost-effective, and stigma-free assistance [3], [4], [10], [11].

B. Core Technologies

Natural Language Processing (NLP):

NLP enables chatbots to interpret user input, detect sentiment, and generate coherent responses, ranging from keyword-based methods to transformer-based LLMs [1], [7].

Emotion and Sentiment Recognition:

Emotion detection allows chatbots to adapt responses, identify distress signals, and personalize interactions in mental health contexts [3], [11].

Contextual Reasoning and Retrieval-Augmented Generation (RAG):

RAG grounds chatbot responses in verified knowledge sources, enhancing reliability and reducing hallucinations in sensitive mental health applications [1], [17].

Agentic AI:

Agentic AI supports autonomous reasoning, adaptive decision-making, and long-term context tracking, enabling proactive caregiver support and personalized recommendations [2], [16]. Table 2 summarizes representative systems and their limitations across these domains [11].

II. SURVEY OF EXISTING SYSTEMS

Mental health chatbots developed over the past decade utilize rule-based, machine learning, and LLM-based approaches to support general users, students, caregivers, and disorder-specific needs, providing accessible and stigma-free care [3], [4], [6], [10], [11].

A. Evaluation Metrics for Mental Health Chatbots

Evaluating mental health chatbots extends beyond linguistic accuracy, as traditional NLP metrics such as BLEU or ROUGE fail to assess empathy, therapeutic value, and safety [6], [7], [17]. Effective evaluation therefore includes generic performance, counselling quality, and reliability measures, as summarized in Table 1.

Category	Metric	Implication
General metrics	Helpfulness	Assesses the practical utility of the chatbot’s responses.
	Fluency	Evaluates the naturalness and flow of the chatbot’s language.
	Relevance	Measures how the chatbot’s responses pertain to the context of the dialogue.
	Logic	Determines the logical consistency of the chatbot’s replies.
	Informativeness	Gauges how informative and helpful the chatbot’s responses are.
	Understanding	Assesses the chatbot’s ability to comprehend user queries.
	Consistency	Checks for the chatbot’s ability to provide uniform responses.
	Coherence	Evaluates the chatbot’s ability to maintain topic coherence.
	Empathy	Measures the chatbot’s ability to display understanding and compassion.
	Expertise	Assesses the chatbot’s ability to provide knowledgeable responses.
	Engagement	Evaluates how well the chatbot keeps the user engaged in conversation.

Counseling metrics	Direct guidance	Assesses the chatbot's ability to provide clear therapeutic direction.
	Approval and reassurance	Measures the chatbot's ability to offer affirmation and comfort.
	Restatement	Evaluates the chatbot's skill in paraphrasing to show understanding.
	Reflection	Gauges the chatbot's ability to reflect on the user's statements.
	Listening	Assesses the chatbot's capacity to exhibit active listening cues.
	Interpretation	Measures the chatbot's ability to interpret the user's statements.
	Self-disclosure	Evaluates the chatbot's use of self-disclosure to build rapport.
Reliability metric	Krippendorff's Alpha	Determines the consistency of evaluations across different human evaluators.

Table 1- Technical Evaluation Metrics for Mental Health Chatbots [16].

B. Comparison of Existing Mental Health Chatbots and Digital Support System

Sr. No.	System Name	User Group	Key Features	Core Technology	Limitations
1	Woebot	Adolescents & adults	CBT-based conversational support, mood tracking, psychoeducation	LLM + scripted CBT flows	Primarily designed for patients; limited caregiver support [9]
2	Wysa	Individuals (teens & adults)	Guided exercises, mindfulness, self-help, crisis resources	AI + ML + NLP	No explicit childhood disorder focus; user must interact directly [10]
3	Tess (X2AI)	Patients & caregivers	Automated supportive messaging; integrates into healthcare workflows	Rule-based + ML	General caregiver support, not child-specific [11]
4	Ash (Monash University)	Students / teenagers	School-based wellbeing chatbot; guides, referrals, mental health education	Scripted flows + UX research	Focused on student wellbeing; caregiver module limited [12]
5	Evebot	Students	Campus psychological support; emotion detection	Seq2seq + Bi-LSTM	Pilot study only; adolescent-focused; no caregiver involvement [10]
6	HIGEA	Caregivers	Monitors caregiver burden via conversation; personalized guidance	Dialogue system embedding psychometric tests	Research prototype; child-specific disorders not primary focus [13]
7	BOTANIC	Caregivers	AI-based early detection of caregiver stress; longitudinal tracking	NLP + monitoring	Feasibility study; not child-specific [14]
8	Cognoa / Canvas Dx	Clinicians using caregiver input	Early detection of autism; evaluates videos/ questionnaires from parents	ML on caregiver input; FDA cleared	Limited to autism; not a general-purpose tool [15]
9	Youper	Adolescents & adults	Short guided conversations, symptom monitoring, CBT	AI-driven dialogue + ML	No caregiver module; patient-facing [5]
10	NeuroFlow	Patients, clinicians	Behavioural health engagement; screening, monitoring	Digital platform + analytics	Indirect youth support; no childhood disorder detection [5]

Table 2- Comparative Analysis of Mental Health Chatbots: Features, Technologies, and Limitations

III. OBSERVATIONS AND RESEARCH GAPS

➢ Limited caregiver engagement:
Most chatbots are designed with the assumption that the end-user is the individual experiencing mental health challenges. As a result, parental or guardian involvement particularly crucial for children with early-onset disorders is often minimal or entirely absent, limiting the effectiveness of these tools for child-centered interventions [3], [8], [13].

➢ Underutilization of advanced AI paradigms: Emerging technologies such as agentic AI and retrieval-augmented generation (RAG) are largely unexplored in deployed chatbots. These approaches have the potential to provide context-aware, adaptive, and personalized interactions, which are essential for supporting caregivers and accurately monitoring childhood mental health [2], [7].

➢ Ethical, privacy, and safety concerns: Many commercial chatbots lack comprehensive mechanisms to ensure data privacy, ethical use, and reliability. Issues such as hallucinated or misleading responses, inadequate consent protocols, and cultural insensitivity pose risks, especially when the target population includes vulnerable children and their families [17], [4].

➢ Opportunities for future improvement:
Integrating caregiver-focused design, context sensitive AI, and robust ethical frameworks can enhance the impact, safety, and adoption of mental health chatbots. There is a clear need for research and development that bridges technological innovation with user-centered design and clinical relevance [2].

IV. FUTURE DIRECTIONS AND RECOMMENDATIONS

➢ Caregiver-Centered Design
Future chatbots should prioritize parents, guardians, and other close caregivers as primary users. Unlike conventional patient-focused systems, these tools should provide actionable guidance, structured monitoring, and personalized recommendations based on caregiver observations. Features could include early warning alerts for concerning behaviour, psychoeducational content, and suggested interventions, empowering caregivers to act promptly and effectively [8], [13], [14].

➢ Integration of Agentic AI
Agentic AI can transform chatbot interactions by enabling autonomous reasoning, context-sensitive decision-making, and adaptive conversation flows. In caregiver-mediated systems, agentic AI could analyze patterns of child behaviour, track emotional states over time, and proactively suggest interventions [2], [16].

➢ Retrieval-Augmented Generation (RAG) RAG provides a mechanism for grounding AI generated responses in verified, evidence-based knowledge bases, mitigating risks of hallucination and misinformation. For childhood mental health, RAG can ensure that guidance provided to caregivers is accurate, reliable, and clinically relevant, bridging the gap between AI-driven interaction and professional standards of care [1], [7], [16].

➢ Multimodal Interaction and Early Detection Incorporating multimodal inputs such as text, speech, facial expressions, and behavioral cues can improve context understanding and enhance the accuracy of early disorder detection [2], [16].

V. CONCLUSION

Mental health chatbots provide scalable and accessible support for users and caregivers across general-purpose, student-focused, caregiver-oriented, and disorder-specific domains. Although existing systems enhance awareness and accessibility, most remain patient-centered with limited caregiver engagement and insufficient focus on childhood disorders. Key challenges include personalization, reliability, and ethical data use. Emerging approaches such as agentic AI and retrieval-augmented generation (RAG) offer promising solutions by enabling adaptive, trustworthy, and context-aware interactions, supporting improved early detection, personalized care, and mental health outcomes for children and families [2], [7], [16].

REFERENCES

[1] P. Lewis et al., “Retrieval-augmented generation for knowledge-intensive NLP tasks,” NeurIPS, 2020.
[2] A. Sharma and V. Gupta, “Agentic AI: Autonomous decision-making systems – A comprehensive research review,” IJCET, 2025.
[3] K. S. Inkster, “User perceptions and experiences of an AI-driven conversational agent for mental health support,” mHealth, 2024.
[4] R. Hollis et al., “Effectiveness of AI-driven conversational agents in improving mental health among young people,” JMIR, 2025.
[5] S. Patel and N. Kumar, “AI-powered mental

health chatbot,” IJRASET, 2025.
[6] L. Johnson and R. White, “Charting the evolution of artificial intelligence mental health chatbots from rule-based systems to large language models,” Int. J. Methods in Psychiatric Research, 2024.
[7] M. Rossi et al., “Systematic analysis of retrieval augmented generation-based LLMs for medical chatbot applications,” Machine Learning and Knowledge Extraction, 2024.ssss
[8] P. S. Mehta et al., “Developing mental health support chatbots in India: Challenges and insights,” Annals of Indian Psychiatry, 2025.
[9] A. Fitzgerald and M. Hall, “Effectiveness of a web based and mobile therapy chatbot on depression and anxiety: A randomized controlled trial (Woebot),” JMIR Formative Research, 2024.
[10] S. Banerjee et al., “Wysa: An AI-based mental health chatbot for guided self-help,” Frontiers in Psychiatry, 2025.
[11] A. Williams et al., “Tess (X2AI): Supportive messaging for patients and caregivers,” JMIR Mental Health, 2024.
[12] Monash University, “Ash: A school-based wellbeing chatbot,” Technical Report, 2024.
[13] C. Fernandez et al., “HIGEA: Conversational agent for caregiver burden detection,” Healthcare (MDPI), 2024.
[14] M. O’Connor and J. Singh, “BOTANIC: AI chatbot for caregiver stress monitoring,” Frontiers in Psychiatry, 2025.
[15] U.S. FDA, “Cognoa / Canvas Dx: Digital diagnostic aid for autism,” Approval Report, 2024.
[16] A. Yuan, S. Samtani, and E. G. Colato, "A Review of Large Language Model-based Mental Health Chatbots," ACM Transactions on Management Information Systems, vol. 16, no. 4, pp. 1–24, 2025.
[17] F. Zhao and G. Roberts, “Building trust in mental health chatbots: Safety metrics and LLM-based evaluation tools,” arXiv preprint arXiv:2408.04650, 2024.