Emotionally Intelligent Chatbots in Mental Health: A Review of Psychological, Ethical, and Developmental Impacts

Ruwini Herath
International University of Applied Science
Stockholm, Sweden

ABSTRACT
Use of emotionally intelligent chatbots is increasing in mental health settings to provide support by recognizing and reacting to users' emotions. This review has a closer look at 59 peer-reviewed studies from 2017 to 2024, with a focus on systems like Woebot and Wysa. It maps out how affective computing, psychological frameworks like cognitive behavioral therapy (CBT), and human-computer interaction theories shape these systems. While there is early evidence of benefits like reduced anxiety and better emotional self-awareness, many issues remain unresolved. These include weak long-term evidence, cultural bias in emotion recognition, and potential over-dependence on AI. We also highlight the risks of collecting and using emotional data without sufficient oversight. Based on this, we suggest future research should move toward multicultural, longer-term, and ethically grounded studies. The goal should be to create emotionally intelligent systems that support, not replace, genuine human connection, especially in vulnerable populations.

General Terms
Artificial Intelligence, Affective Computing, Human–Computer Interaction, Sentiment Analysis, Natural Language Processing, Machine Learning, Ethics in AI

Keywords
Empathic AI, Affective computing, Mental-health chatbots, Artificial empathy, Human–computer interaction, Emotion recognition

1. INTRODUCTION
The development of empathic artificial intelligence (AI) is a welcome advancement in how machines interface with humans more emotionally. Emotionally astute chatbots, machines crafted to recognize, understand, and interact with corresponding emotions, are being researched as part of the curriculum for instructional psychology and coaching [1], [2] voice, text, and facial recognition help them identify emotions and respond with empathy. Their integration into mental health applications and chatbots is changing the landscape for support accessibility, particularly in situations where face-to-face therapy is not possible.

These models encounter significant challenges when constructing responses demonstrating social and technological empathy. Many models use fixed emotion types along with pre-recorded texts, thus failing to cope with more intricate or changing emotional expressions [2], [3]. Moreover, there are still issues regarding the quality of data, the "cultural" bias in emotive frameworks, and the sociological ethics of machines designed to demonstrate empathy [4]. Systems that supplement therapeutic practices coexist with those marketed as therapeutic companions or mental health aids. All of these, however, raise significant concerns about the emotional impact of such technologies regarding attachment and human connection.

AI's wider uses in healthcare and education have shown how effectively and personally tailored these sectors can become [5]. AI's use includes but is not limited to diagnostic and rehabilitative processes and administrative tasks in medicine, while education benefits from adaptive teaching and monitoring. Still, as much care and sensitivity to emotion as the administration of mental health support requires, ethical governance, precise design requirements, and transdisciplinary approaches are heightened.

This literature review explores research on emotionally intelligent chatbots, particularly their applications in mental health and their impact on human emotional development. It discusses the psychological advantages and ethical concerns emerging from these technologies and identifies gaps in the available literature regarding their societal and developmental consequences. By integrating findings across disciplines, this review seeks to guide the design of AI systems with empathy in mind, ensuring that they are ethically defensible and psychologically robust.

1.1 Methodology
This literature review takes an integrative and narrative approach to analyzing existing research from psychology, computer science, and AI ethics. Research literature was obtained using online databases such as PubMed, Scopus, IEEE Xplore, arXiv, and Google Scholar. Search strategies used were combinations of "emotionally intelligent chatbots," "affective computing," "artificial empathy," "sentiment analysis," "mental health AI," and "attachment theory in AI." Only peer-reviewed articles, systematic reviews, and relevant preprints published between 2017 and 2024 were included, focusing on publications from 2020 onwards to capture developments in LLMs and emotion AI technologies, as most published post-2020 focused on these technologies. Selectively, some grey literature, which included conference proceedings and ethics papers, was considered if it dealt with underrepresented populations and critical emerging risks. Over 70 sources were reviewed, but only 59 were considered recent, relevant, and sufficiently diverse in discipline. This review seeks to trace patterns and identify conceptual frameworks and gaps in research about emotionally intelligent chatbot design and their influence on mental health and socio-emotional development.

2. THEORETICAL FOUNDATION.

2.1 Understanding Empathic AI
Empathic artificial intelligence (AI) is the computing capability of recognizing, understanding, and appropriately responding to human emotions in each context. It is one of the essential elements for achieving emotionally intelligent responses in

human-AI interactions. Artificial empathy (AE), a branch of affective computing, begins with the identification of emotional signs (e.g., facial expression, prosody, and textual emotion) and ends with relevant response formulation [6]. These tasks usually depend on advanced machine learning methods to handle multimedia data like convolutional neural networks used in facial recognition (VGGFace), emotion-simulating autoencoders, and sentiment-sensitive language models [6]

Even with technological advancements, achieving genuine empathy remains an overwhelming task. These systems still operate with predefined categories of emotions, which are often scripted and do not change with the flow of interaction [7], [3]. Moreover, the absence of well-defined benchmarks for empathy evaluation in AI makes it difficult to compare emotional efficacy across different models [6]. Still, other studies highlight that even basic empathetic actions, such as simulating concern or other emotions, improve user engagement and trust, especially in healthcare and therapeutic settings [7].

Newer studies have also included pain recognition in the physical and emotional sense within the scope of artificial empathy. Some AI has been created to execute plans and goals in a human-like empathic fashion [8]. Other systems focus on computational pain detection to allow for better sympathy in the medical field [7]. These changes highlight the increasing multi-disciplinary focus on artificial empathy's technological and psychological frameworks as more utilities are utilized in healthcare, elder care, and digital mental health services.

2.2 Emotionally Intelligent Chatbots

Right margins should be justified, not ragged. Emotionally intelligent chatbots make supportive and more engaging interactions possible. Emotional recognition features enable providing better assistance to users. These systems combine NLP, Machine Learning, and deep learning to identify affective features in user input and provide them with responses appropriate to the emotions expressed. For providing emotion-specific responses, open-ended generative models make use of enhanced Sequence-to-Sequence (Seq2Seq) frameworks and transformer-based architectures like Dialogpt [9], [10]. For those seeking assistance with mental health issues, these chatbots have proven effective due to their ease of access, lack of judgment, and customized interactions. Some studies suggest that users with extraverted personality traits will likely experience positive mood changes and emotional self-examination if interacting with an emotionally responsive system [10].

Despite the advances that have been made, there are still underlying issues that have not been successfully addressed. While users often report feeling a sense of empathy emanating from these systems, there is little empathy due to the emotional labels and learned verbal patterns. [10], [11]. The debate is centered on issues around data security and emotional exploitation, particularly about mental health services. Critics contend that without proper ethical guidelines, emotionally responsive chatbots might create unnecessary bonds or foster unwarranted reliance on a machine, particularly among the most vulnerable [11].

Despite the inherent challenges, the potential uses in therapy, customer service, and digital companionship highlight the incredible opportunity that emotionally intelligent chatbots present. Further study aims to enhance technology concerning the depth and authenticity of empathy expressed while focusing on user safety, ethical clarity, transparency, and psychological realism.

2.3 Affective Computing and Analysis of Sentiment

One defining characteristic of empathic AI is ‘sentiment analysis’, an AI's capability to recognize the emotions underlying a user's sentiment. As a subfield of NLP, it classifies emotions in text as positive, negative, or neutral using machine learning classifiers, such as Support Vector Machines, Naïve Bayes, or lexicon-based and even hybrid methods [12]. More sophisticated frameworks apply deep learning with neural networks, fuzzy logic, and ontological approaches to semantics and emotional intensity, enabling more precise and context-aware interpretation of user sentiment [12].

Research in this area goes beyond analyzing text to include voice, facial, and physiological signals. For example, facial expression analysis has been incorporated into LLMS to improve empathy within stress monitoring systems [13]. Others have investigated the ethical implications of using deepfake technology to mask emotions within datasets to train AI for healthcare or to compose automated music for users based on their feelings [14]. Sentiment analysis and affective computing create the foundation technology of empathic AI. While these fields enable more personalized and emotionally intelligent interactions from machines by fine-tuning how human emotions are recognized and treated, they also pose additional challenges regarding authenticity, fairness, and the technology's psychosocial impact.

2.4 Theoretical Frameworks in Empathic AI Research

Several interdisciplinary theories have developed around research on emotionally responsive chatbots and empathic AI. These frameworks assist in developing systems that can interact with human emotions on a deeper level and serve as guiding or answering mechanisms for why people use such systems.

Human-Computer Interaction (HCI) is an especially critical area of study, as it deals with interactive computing systems' design, operation, and usability issues. The base of HCI, which emerged in the 1960s, lies in the disciplines of computer science, cognitive psychology, and design, but has also grown to include considerations of system functionality, human actions, thoughts, and social context [15]. Other vital aspects of HCI include technology interface and user interaction through input and output devices, as computing aids are increasingly integrated into more environments and made part of the daily routine [16]. Applying HCI principles in designing systems sensitive to emotions aims to make such systems useful, socially appropriate, and responsive to emotional stimuli.

Media Equation Theory sheds light on how people interact with computers as though they are social entities. Reeves and Nass first proposed this, and the theory states that even when individuals are conscious of the systems as non-human, they extend social scripts such as courtesy or cooperation to them [17]. The computers back this Are Social Actors (CASA) framework, highlighting the instinctive and automatic nature of these “social” reactions. However, the theory does have some drawbacks. For example, some studies indicate that people’s relationships with robots or AI agents are context-dependent; in some cases, such as obedience situations, people exhibit less compassion toward machines than humans [18]. Additionally, prior experience with computers seems to alter how these social actions are evoked [19].

The human attachment theory, first devised to explain the emotional connections between babies and their parents, can now be used to analyze a person's relationship with AI. Attachment styles, such as anxiety and security, have been shown to make specific predictions about the trust placed in AI systems [20]. Individuals experiencing loneliness or emotional distress may form strong emotional connections to chatbots, which can provide comfort but may also interfere with real-life relationships or increase dependence [21]. This extension of attachment theory to human-AI interaction builds on earlier psychological models of love and trust, incorporating insights from Bowlby's foundational work and recent proposals to model attachment dynamics using Bayesian reasoning and cognitive control systems [22].

Cognitive Behavioral Therapy (CBT), a widely used psychological framework for managing depression and anxiety, has informed the design of therapeutic chatbots. Digital CBT systems aim to deliver cognitive restructuring, behavioral activation, and self-monitoring through AI-driven dialogue. Meta-analyses suggest that digital CBT can be as effective as traditional face-to-face therapy when properly implemented [23]. Recent innovations include mental health apps and chatbots that provide real-time CBT-based support, improving users' psychological skills and emotional resilience [24], [25]. However, digital CBT also introduces ethical concerns, such as managing risk without a live therapist, ensuring equitable access, and maintaining user confidentiality [26]. Future research should aim to improve engagement and address the "implementation gap" between technological potential and real-world use.

These frameworks provide essential lenses for understanding empathic AI's psychological, social, and design implications. They inform how such systems are constructed and how users interpret and emotionally relate to them, making them crucial to theoretical and applied research in this field.

3. APPLICATIONS IN MENTAL HEALTH

AI chatbots have garnered significant attention due to their potential as easy-to-use options for mental health care. Separate from traditional therapy, these chatbots provide cost-effective forms of treatment. AI chatbots and autonomous conversational agents powered by natural language processing promise to take over interventions such as mood logging, stress relief, and cognitive reframing. Research indicates that chatbots like Woebot, Wysa, and Youper significantly alleviate depression and anxiety, especially among users seeking anonymous, on-demand support [27].

In addition to delivering CBT, chatbots provide remarkable advantages regarding accessibility, stigma, and personalization. Chatbots can mitigate emotional distress without the judgment, cost, or geographical barriers associated with traditional therapy. Some users even report that deeply personalized systems such as Replika alleviate feelings of isolation [28]. In other healthcare areas, emotionally responsive chatbots are employed for patient education, triage, adherence support, and symptom monitoring. These systems improve clinical communication, decrease the mental workload associated with healthcare tasks, and enhance participation in public health initiatives [29], [30]. Chatbots have been trialed for use in chronic disease management and clinical trial coordination, primarily through smartphone applications and text-based interfaces to maximize accessibility [31].

However, their implementation faces difficulties. Although utilization is high, emotionally aware systems generally do not embody human therapists' emotional nuance, responsibility, and clinical judgment. Important constraining factors are the failure to interpret sarcasm, slang, and culturally specific phrases, the lack of uniform and reliable expression of answers, and privacy issues about users' personal information [32], [33]. To resolve these problems, researchers highlight the need to combine the skills of mental health practitioners, technologists, and ethics specialists. Standards such as ISO/IEC 25010 can appraise chatbot effectiveness in usability, reliability, and emotional impact [34]. In addition, further training on validated psychological and medical datasets will improve their value and versatility [35]. As outlined above, providing access to mental health services through emotionally intelligent chatbots presents significant opportunities. However, these systems require profound validation, ethical scrutiny, and refinement to adapt to the multifaceted user demands before they can be clinically deployed.

Table 1Comparative evaluation of key emotionally intelligent chatbot systems used in mental health applications. While these tools offer promising emotional support features, long-term efficacy, cultural sensitivity, and ethical robustness remain ongoing challenges

Chatbot	Core Technology	Therapeutic Framework	Target Users	Reported Benefits	Known Limitations
Woebot	NLP, rule-based, CBT modules	Cognitive Behavioural Therapy (CBT)	Young adults, general population	Reduced anxiety and depression in short-term use; 24/7 availability	Limited emotional depth; lacks cultural personalization; short evaluation periods
Wysa	AI + human-in-the-loop; emotion tracking	CBT, Dialectical Behavioural Therapy (DBT), mindfulness	Stress and anxiety sufferers: some clinical support	Scalable, customizable, supports emotional journaling and mood tracking	Emotional responses are scripted; concerns over user data privacy
Replika	Transformer-based, generative AI	Human–AI companionship, affective mirroring	Users seeking emotional support or conversation	High user engagement and perceived empathy reduce loneliness	Risk of overdependence; not clinically validated; anthropomorphism issues

4. IMPACTS ON HUMAN EMOTIONAL DEVELOPMENT

The rapid development of emotionally intelligent chatbots has raised questions concerning their impact on the user's emotional awareness, empathy, and social-emotional development. Some studies indicate that such systems may provide therapeutic assistance and actively foster emotional intelligence and affective skills in educational and clinical contexts.

Empath.ai, empathy-based chatbots, have been shown to improve users' emotional understanding and self-regulation using conversational cognitive behavioral therapy [1]. Research shows that chatbot feedback's emotional and motivational impact tends to be on par, in some cases, with that of human feedback within digital learning environments, particularly for emotional self-regulation and goal attainment [36]. [37] It has also been noted that interaction with chatbots capable of responding emotionally alters the user's ability to identify and modulate emotionally empathic responses.

Unlike adult users, emotionally intelligent AI is increasingly used to assist with social-emotional learning (SEL) in children and adolescents. AI technologies like virtual tutors, educational chatbots, and digital gaming platforms have been assimilated into SEL frameworks for personalized emotion training and simulated social interactions [38], [39]. These applications target critical emotional skills such as empathy, self-management, and socio-interactive conflict resolution. For example, social-emotional AI has been designed based on early childhood empathic development, with soft conversational turns and scenario-based interactions designed to emulate empathic interactions [40]. An unusual study that placed preschoolers with robotic and real dogs contrasted the two responses they evoked. Both stimulated empathic learning in distinctly different ways: conversation-based with the robot and emotionally reactive with the live animal [41]. While these approaches are encouraging, they pose concerns regarding their cultural sensitivity and the possible lack of appropriate educator training to use AI tools in emotionally sensitive contexts.

Other researchers have cautioned regarding the social and emotional consequences, even with these advantages. As interaction with artificial intelligence becomes more fluid and emotionally responsive, users will likely form strong emotional attachments to AI friends and chatbots. While these attachments can be comforting in the face of loneliness, they can also increase emotional dependence [42], [21]. The attachment theory research indicates that users with insecure attachment styles, usually characterized by high attachment anxiety, tend to over-trust or under-trust AI systems and emotionally exploit them, which may gradually undermine their emotional resilience [20].

Moreover, deepening emotionally responsive AI into therapy creates new ethical, ontological, and developmental questions. Some researchers contend that as AI systems evolve from being able to analyze to possessing some level of empathetic intelligence, human empathy is impossible to substitute [43], [44]. AI devices can process and respond to emotions through advanced sentiment analysis, facial recognition, and voice tone analysis [45]. However, such technologies lack genuine emotion, contextual understanding, and empathy, which are foundational to humanity. This restriction may jeopardize the authenticity of the provision of emotional support and heighten the possibility of emotional exploitation or manipulation.

The existing body of literature presents what can be considered a balanced controversy. For instance, emotionally intelligent AI can aid in developing emotional literacy, awareness, and support. Simultaneously, emotionally intelligent AI poses psychological and ethical concerns in attachment, autonomy, and the authenticity of emotional experiences. Such AI requires interdisciplinary governance and design alongside humanity, focusing on system interactions and deep-seated human emotion and development aspects.

5. ETHICAL AND SOCIETAL CONSIDERATIONS

Engaging with emotionally intelligent chatbots in social services, education, and even healthcare may enhance performance results. Still, because of their design, automation, and potential future applications, such systems pose significant societal and ethical concerns. Privacy, bias, emotional manipulation, inequality, anthropomorphism, and similar factors make the compassionate AI debate more difficult. The use of emotionally responsive chatbots in sensitive areas like mental health care and education raises immediate ethical concerns. While these technologies could provide much-needed relief and support, they also pose risks of privacy invasion, identity bias, and emotional exploitation.

Acoustic tremors, minute facial expressions, and text sentiment entail the capture of private emotion data, analyzing AI emotions unprecedentedly. Unlike conventional health data, capturing emotion cues can disclose more than intended. This intelligence creates loopholes for exploitation [46]. Examples include the public outcry over Amazon's emotion recognition patents, which were condemned for facilitating workplace emotional surveillance [47]; the sensitive therapy chatbot log disclosures the public was exposed to [48]. These examples highlight how emotional data often exists in an unregulated state, lacking the protective mechanisms found in biometric or medical data, as evidenced by the violence of the exposed logs.

Table 2. Mitigation of ethical risks

Risk	***Example***	***Mitigation***
Covert Data Harvesting	Apps extracting mood data without consent	Adopt "emotional data minimization" (collect only essentials)
Re-identification	Voice recordings linked to identities	Federated learning (process data locally; no central storage)
Third-party Sharing	Mental health apps are selling data to advertisers	Ban commercial use of emotional data (EU AI Act, 2024)

The policy gap is troubling. Data reflecting a person's emotional state is not yet recognized by law as distinct, and therefore, emotional data is not treated differently from other biometric information. As a result, collecting emotional data from individuals is not explicitly protected under HIPAA.

Emotion recognition algorithms and bias issues, specifically those related to facial emotion recognition (FER) systems, pose another concern. Such models are frequently subjected to the biases of their training data. In healthcare and education, where unwarranted discrimination due to overgeneralised prejudice poses a significant risk, such biases are problematic. Studies have shown commercial FER systems have a heightened propensity to misinterpret neutral facial expressions of Black males as "anger" 35% more [49]. In addition, Research

indicates that East Asian cultural norms emphasizing emotional restraint and low assertiveness can lead to misinterpretation and underperformance in Western contexts. [50], leading to therapeutic misdiagnosis and employment discrimination.

Recent studies highlight glaring issues such as cultural biases and commercial ethics within automatic facial emotion recognition (FER) systems. In addition, FER technologies are based on contentious psychological frameworks that require facial expressions to be universal. These models overlook cultural differences and are challenged by variations in emotional expression and perception [51]. Research shows that East Asians and Westerners interpret facial expressions differently; for example, East Asians often perceive 'fear' as 'surprise' [52]. Such cultural differences also apply to how specific facial regions are interpreted, which can affect the accuracy of facial expression recognition (FER) systems [53]. Additionally, dataset demographic biases introduce inequity into the underlying machine learning frameworks of the FER system. Most FER datasets contain statistical biases concerning the demographic groups represented within the dataset [54]. These findings highlight the growing demand for more complete evaluation frameworks alongside diverse datasets and the consideration of culture when constructing and applying FER systems to reduce bias and address ethical concerns. As a solution, policymakers propose the participatory design of overlooked social groups, like the AI for Mental Health Initiative in India, accompanied by mandatory bias audits using IBM's AI Fairness 360 methodology.

Positive empathy exhibits a distinct form of well-being risk due to high susceptibility to manipulation and emotional treachery, particularly when chatbots simulate emotional understanding. Replika and similar systems pretend to offer companionship, which 41% of users reportedly accepted and believed the AI "loved" them [28]. This has the potential to exacerbate harmful overreliance. In Japan, elderly citizens reported replacing human interactions with AI pet companions, which has led to a decrease in social interactions. [21]. These outcomes reflect concerns posed by [55] warning against exploitative artificial empathy bonds. Core protective measures include moderation policies (e.g., "This AI cannot feel emotions" and similar phrases, AI-generated) and usage caps recommending high-risk users to human caregivers as primary responders.

Implementing empathy in AI systems through anthropomorphism raises greater ethical concerns. Adding human characteristics could improve user experience and lead to affective disorientation. For instance, users explained to chatbots how their actions could cause emotional damage and apologised for "hurting their feelings" [11]. Engagement from Woebot's cartoon face enhanced participation, but emotional dependence was also exacerbated [55], [56] suggests that these design decisions interface with the moral and emotional spectrum, overriding users' trust in systems they should not. Guidelines should prevent humanoid clinical design and establish protocols that establish consent processes that remind users of the AI's cons, aside from being an entity.

Applying empathic AI across different global contexts introduces equity gaps. Most tools for sentiment analysis do not go beyond English and Chinese [57]. Many low- and Middle-Income Countries (LMICs) lack the infrastructure to access AI-mediated teletherapy [58]. If left unaddressed, these gaps pose risks of advancing a bifurcated system that primarily benefits the Global North. More equitable approaches include AfroSent, an open-source project focused on Natural Language Processing (NLP), and grassroots initiatives like Sangath NGO in India.

6. DISCUSSION AND FUTURE WORK

While the field of chatbots with emotional recognition skills is evolving, drastic gaps remain regarding psychological and ethical frameworks. Addressing these cultural gaps is essential for developing and deploying empathic AI in a socially responsible and inclusive way.

Absence of longitudinal and ecologically authentic investigations is a critical gap. Most available research focuses on interactions over brief periods within controlled laboratory settings. This provides limited insight into the psychosocial impacts that span longer durations. For example, 89% of studies around therapy chatbots such as Woebot and Wysa have user tracking data for less than eight weeks [27]. This short-duration tracking fails to illuminate potential relapse or dependency risks during the periods of unmonitored time. Along similar lines, there is a notable lack of studies examining the impact of prolonged AI companionship, such as with Replika, on the development of empathy and social skills among adolescents over sustained periods [39]. In developmental psychology, addressing these gaps requires five-year cohort studies comparing users of AI-based therapy with those receiving traditional care, along with the use of ecological momentary assessment (EMA) to evaluate the emotional impact of AI interactions in real-world settings.

Another unresolved issue is socio-cultural and linguistic exclusion. Emotion AI technologies are primarily shaped by WEIRD (Western, Educated, Industrialized, Rich, and Democratic) societies, making them less applicable—and potentially unjust—for more diverse global populations. For instance, dominant sentiment analysis models often misinterpret emotional restraint common in collectivist cultures as disengagement or low emotional involvement [39]. These gaps highlight the scarcity of open-source emotional lexicons for African dialects, such as AfroSent, and underscore the need for collaboration with researchers from the Global South to culturally center the design of AI systems, as exemplified by India's Sangath Model.

Moreover, understudied populations like neurodivergent individuals and elderly users are also rarely the focus of emotion AI research. No studies assess whether contemporary emotion recognition mechanisms consider accommodating autistic users' atypical facial features and expressions, and dementia-friendly design is often absent from elder care trials, despite many seniors with dementia actively using AI companion tools [40]. Addressing this imbalance involves incorporating community-based participatory research (CBPR) strategies to engage these populations and the creation of more flexible user interfaces, such as text-only options designed for dyslexic users.

Gaps focusing on specific populations are available; additionally, the legal frameworks governing emotion AI are still nascent and under-researched. There is currently no legal carve-out for emotional data that distinguishes it from ordinary biometrics, which poses potential risks regarding privacy and data exploitation [46]. Moreover, ISO/IEC 24027 and other existing standards do not account for cultural bias as discrimination in using affective computing technologies [34]. Emotion logs should be treated as sensitive health information protected by laws such as GDPR to rectify these gaps. In contrast, clinical emotion AI applications should undergo bias audits and approval pathways akin to FDA regulations.

Finally, the practical application of emotionally intelligent AI systems integrates only as far as the core concepts of CBT and attachment theory, beyond which there is little to no

integration. Self-determination theory and social learning theory, other valid psychological theories, are largely ignored. For instance, chatbots rarely model prosocial behaviours like conflict de-escalation or provide autonomy-supportive contingent feedback to facilitate emotional growth [36]. Future innovations should apply psychodynamic frameworks, such as examining transference dynamics in AI-user bonds, and explore the integration of positive psychology principles through features like gratitude journaling bots.

6.1 A Proposed Framework for Future Research

Advancing emotionally intelligent chatbots requires an integrated approach that combines technological, psychological, and ethical perspectives. Future research should be structured around these key areas. First, AI empathy needs to be redefined past sentiment analysis to include multicultural contextual perspectives and dynamic emotional adaptation through hybrid models (e.g., techno-sociological blends where transformer-based NLP intersects with psycholinguistic theories). Second, comprehensive longitudinal studies (3-5 years) on emotionally self-regulated, socially skilled users of AI across diverse cultures need to be conducted to understand the socio-emotional developmental implications of AI. Third, active participation of underrepresented design and audit communities, including neurodivergent users and non-WEIRD populations, must ally with algorithmic bias mitigation using model tests of AI fairness, such as IBM's AI Fairness 360, anchored in systemic design principles. Fourth, real-world testing is required for ethical guardrails, including classification of emotional data protection under GDPR and "empathy caps" to contain AI over-reliance. Finally, interdisciplinary collaboration is essential—bringing together clinicians, ethicists, and engineers to create meaningful benchmarks for evaluating genuine AI empathy. These standards should go beyond traditional therapeutic rapport measures to help distinguish genuine empathy from mere simulation, ultimately supporting more authentic therapeutic interactions. Human-first design ensures AI integration serves to augment, not replace, human connections in support of fundamental relational intersections.

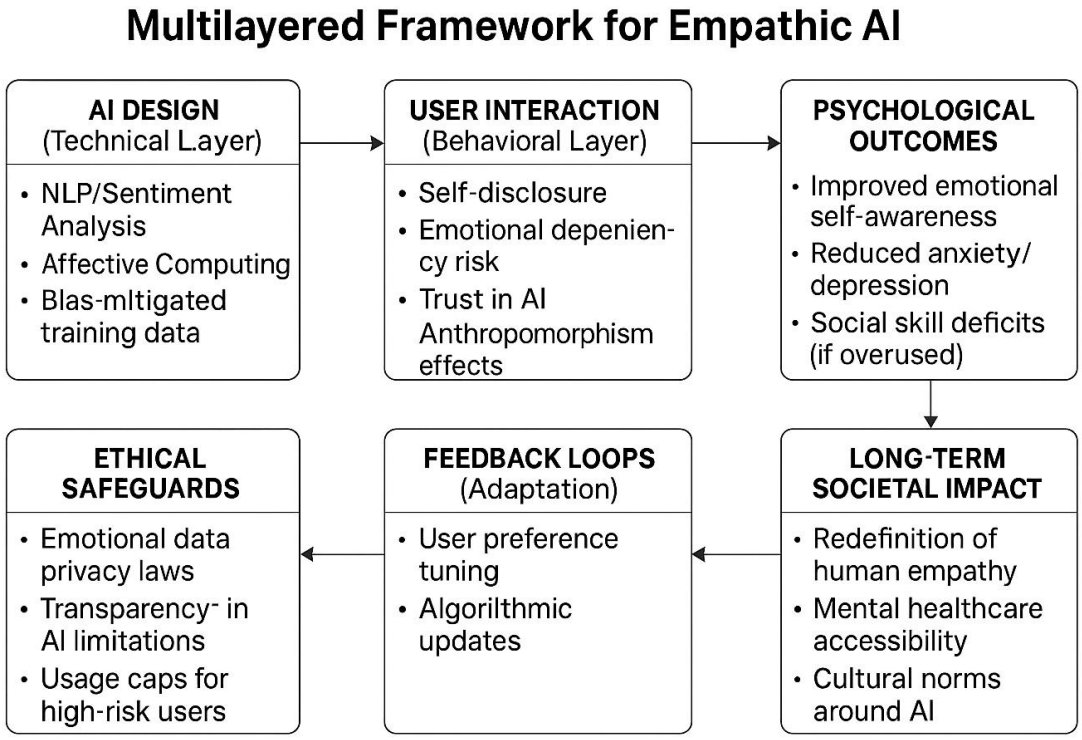

Figure 1: Multilayered Framework for Empathic AI: From design to societal impact. (Source: Author's elaboration based on reviewed literature)

7. CONCLUSION

The emergence of emotionally intelligent chatbots marks a new epoch in human-technology interaction, particularly in mental health and emotional development. These systems demonstrate clear potential in delivering affordable, scalable, and emotionally attuned support by integrating affective computing, sentiment analysis, and psychological frameworks. However, simulated empathy remains fundamentally distinct from genuine emotional depth. Without robust ethical guidelines, inclusive design practices, and longitudinal validation, such technologies risk undermining personal autonomy, emotional regulation, and the authenticity of human relationships.

Future research should prioritize several key directions, such as conducting long-term and ecologically valid studies to assess the sustained psychosocial impact of chatbot usage, developing ethical evaluation frameworks for AI empathy, especially in clinical and educational deployments, designing systems that are inclusive for neurodivergent users and elderly populations, ensuring accessibility and emotional safety and addressing cultural variation in emotional expression to reduce bias and improve the global relevance of emotion AI.

As the field advances, emotionally intelligent AI must be directed with care, ensuring that human values are preserved and augmented, rather than replaced, in the pursuit of empathetic technological progress.

8. ACKNOWLEDGMENT

This literature review was independently conceived and written as a personal academic contribution. I want to acknowledge the open-access researchers and communities whose work made this exploration possible. Their insights formed the foundation for a deeper understanding of emotionally intelligent AI and its psychological implications. I also wish to thank those in my personal and academic circles who offered informal support, thoughtful conversations, and encouragement throughout this process. Lastly, I acknowledge the lived experience of neurodivergence that shaped the motivation and the methodology behind this review. It is offered as an independent academic contribution grounded in both evidence and lived experience.

9. REFERENCES

[1] N. Kallivalappil, K. D'souza, A. Deshmukh, C. Kadam, and N. Sharma, "Empath.ai: A context-aware chatbot for emotional detection and support," in *Proc. 14th Int. Conf. Comput. Commun. Netw. Technol. (ICCCNT)*, 2023, pp. 1–7. doi:10.1109/ICCCNT56998.2023.10306584.

[2] T. Spring, J. Casas, K. Daher, E. Mugellini, and O. A. Khaled, "Empathic response generation in chatbots," *CONVERSATIONS Workshop*, Amsterdam, 2019. [Online]. Available: https://arxiv.org/abs/1911.12315

[3] S. Devaram, "Empathic Chatbot: Emotional Intelligence for Mental Health Well-being," in *IEEE ICAC3*, Bournemouth University, UK, 2020. [Online]. Available: https://arxiv.org/abs/2012.09130

[4] S. B. Velagaleti, "Empathetic algorithms: The role of AI in understanding and enhancing human emotional intelligence," *J. Electr. Syst.*, vol. 20, no. 3s, pp. 2051–2060, 2024. doi:10.52783/jes.1806

[5] S. Zeb, N. FNU, N. Abbasi, and M. Fahad, "AI in Healthcare: Revolutionizing Diagnosis and Therapy," *Int. J. Multidiscip. Sci. Arts*, vol. 3, no. 3, 2024. doi:10.47709/ijmdsa.v3i3.4546

[6] S. Tahir, S. A. Shah, and J. Abu-Khalaf, "Artificial Empathy Classification: A Survey," *arXiv preprint*, arXiv:2310.00010, 2023. [Online]. Available: https://arxiv.org/abs/2310.00010

[7] S. Cao et al., "Pain recognition and pain empathy from a human-centered AI perspective," *iScience*, vol. 27, no. 8, p. 110570, 2024. doi: 10.1016/j.isci.2024.110570

[8] M. Shvo and S. A. McIlraith, "Towards Empathetic Planning and Plan Recognition," in *Proc. AIES '19*, 2019, pp. 525–526. doi:10.1145/3306618.3314307

[9] G. Bilquise, S. Ibrahim, and K. Shaalan, "Emotionally intelligent chatbots: A systematic review," *Hum. Behav. Emerg. Technol.*, pp. 1–23, 2022. doi:10.1155/2022/9601630

[10] A. Ghandeharioun, D. McDuff, M. Czerwinski, and K. Rowan, "Towards understanding emotional intelligence for behavior change chatbots," *arXiv preprint*, arXiv:1907.10664, 2019. doi:10.48550/arXiv.1907.10664

[11] M. Rostami and S. Navabinejad, "Artificial empathy: User experiences with emotionally intelligent chatbots," *AI & Tech. Behav. Soc. Sci.*, vol. 1, no. 3, pp. 19–27, 2023. doi:10.61838/kman.aitech.1.3.4

[12] P. Borele and D. A. Borikar, "An approach to sentiment analysis using artificial neural networks," *IOSR J. Comput. Eng.*, vol. 18, no. 2, pp. 64–69, 2016. doi:10.9790/0661-1802056469

[13] P. Chakriswaran et al., "Emotion AI-driven sentiment analysis," *Appl. Sci.*, vol. 9, no. 24, p. 5462, 2019. doi:10.3390/app9245462

[14] H. S. Yang et al., "AI chatbots in clinical laboratory medicine," *Clin. Chem.*, vol. 69, no. 11, pp. 1238–1246, 2023. doi:10.1093/clinchem/hvad106

[15] A. R. Mathew, A. Al Hajj, and A. Al Abri, "Human-computer interaction (HCI): An overview," in *IEEE Int. Conf. Comput. Sci. Autom. Eng.*, 2011, pp. 99–100. doi:10.1109/CSAE.2011.5953178

[16] B. Myers et al., "Strategic directions in human-computer interaction," *ACM Comput. Surv.*, vol. 28, no. 4, pp. 794–809, 1996. doi:10.1145/242223.246855

[17] R. J. Lee-Won, Y. K. Joo, and S. G. Park, "Media Equation," *Int. Encycl. Media Psychol.*, 2020. doi:10.1002/9781119011071.iemp0158

[18] C. Bartneck, C. Rosalia, R. Menges, and I. Deckers, "Robot abuse – A limitation of the media equation," Eindhoven Univ. Technol., n.d. [Online]. Available: http://www.bartneck.de

[19] D. Johnson and J. Gardner, "The media equation and team formation," *Int. J. Hum.-Comput. Stud.*, vol. 65, no. 2, pp. 111–124, 2007. doi:10.1016/j.ijhcs.2006.08.007

[20] O. Gillath et al., "Attachment and trust in AI," *Comput. Hum. Behav.*, vol. 115, p. 106607, 2021. doi: 10.1016/j.chb.2020.106607

[21] T. Xie and I. Pentina, "Attachment theory for chatbot relationships: A case study of Replika," in *Proc. HICSS*, 2022. doi:10.24251/HICSS.2022.258

[22] D. Petters and E. Waters, "AI, attachment theory, and secure base simulation," *AISB 2010 Convention*, 2010.

[23] L. Kambeitz-Ilankovic et al., "Review of digital and face-to-face CBT for depression," *npj Digit. Med.*, vol. 5, p. 144, 2022. doi:10.1038/s41746-022-00677-8

[24] H. M. Jackson et al., "Skill enactment in digital CBT," *J. Med. Internet Res.*, vol. 25, p. e44673, 2023. doi:10.2196/44673

[25] G. R. Thew, A. Rozental, and H. D. Hadjistavropoulos, "Advances in digital CBT," *Cogn. Behav. Ther.*, vol. 15, p. e44, 2022. doi:10.1017/S1754470X22000423

[26] L. Lawlor-Savage and J. L. Prentice, "Digital CBT in Canada: Ethical considerations," *Can. Psychol.*, vol. 55, no. 4, pp. 231–239, 2014. doi:10.1037/a0037861

[27] M. Farzan et al., "AI-powered CBT chatbots: A review," *Iran. J. Psychiatry*, 2024. doi:10.18502/ijps.v20i1.17395

[28] B. Maples et al., "GPT3-enabled chatbots and suicide prevention," *npj Ment. Health Res.*, vol. 3, p. 4, 2024. doi: 10.1038/s44184-023-00047-6

[29] E. Gabarron, D. Larbi, K. Denecke, and E. Årsand, "Chatbots in public health," *Stud. Health Technol. Inform.*, IOS Press, 2020.

[30] V. K. Voola et al., "AI chatbots in clinical trials," *Int. J. Res. Publ. Seminar*, vol. 13, no. 5, pp. 323–337, 2022. doi:10.36676/jrps.v13.i5.1505

[31] L. T. Car et al., "Conversational agents in health care: Scoping review and conceptual analysis," *J. Med. Internet Res.*, vol. 22, no. 8, p. e17158, 2020. doi:10.2196/17158

[32] M. Laymouna et al., "Roles, users, benefits, and limitations of chatbots in health care: Rapid review (preprint)," 2024. doi: 10.2196/preprints.56930

[33] D. S. Parikh and H. Raval, "Limitations of existing chatbots: An analytical survey," *Int. J. Innov. Res. Sci. Eng. Technol.*, vol. 7, no. 2, 2020.

[34] V. S. Barletta et al., "Clinical-chatbot AHP evaluation based on 'quality in use' of ISO/IEC 25010," *Int. J. Med. Inform.*, vol. 170, p. 104951, 2023. doi:10.1016/j.ijmedinf.2022.104951

[35] H. S. Yang et al., "AI chatbots in clinical laboratory medicine: Foundations and trends," *Clin. Chem.*, vol. 69, no. 11, pp. 1238–1246, 2023. doi:10.1093/clinchem/hvad106

[36] E. Ortega-Ochoa et al., "The effectiveness of empathic chatbot feedback in online higher education," *Internet Things*, vol. 25, p. 101101, 2024. doi:10.1016/j.iot.2024.101101

[37] M. Rostami and S. Navabinejad, "Artificial empathy: User experiences with emotionally intelligent chatbots," *AI Tech Behav. Soc. Sci.*, vol. 1, no. 3, pp. 19–27, 2023. doi:10.61838/kman.aitech.1.3.4

[38] R. Indellicato, "Artificial intelligence and social-emotional learning: What relationship?" *J. Mod. Sci.*, vol. 60, no. 6, pp. 460–470, 2024. doi:10.13166/jms/196765

[39] S. S. Sethi and K. Jain, "AI technologies for social-emotional learning," *J. Res. Innov. Teach. Learn.*, vol. 17, no. 2, pp. 213–225, 2024. doi:10.1108/JRIT-03-2024-0073

[40] M. I. Gómez-León, "Development of empathy through socioemotional AI," *Papeles del Psicólogo*, vol. 43, no. 3, p. 218, 2022. doi:10.23923/pap.psicol.2996

[41] K. Heljakka, P. Ihamäki, and A. I. Lamminen, "Empathic responses to robot dogs vs. real dogs in learning," in *Proc. CHI PLAY '20*, pp. 262–266, 2020. doi:10.1145/3383668.3419900

[42] C. Akbulut et al., "All too human? Mapping and mitigating risks from anthropomorphic AI," *AIES Conf.*, vol. 7, pp. 13–26, 2024. doi:10.1609/aies.v7i1.31613

[43] C. Montemayor, J. Halpern, and A. Fairweather, "In principle, there are obstacles for empathic AI in healthcare," *AI & Society*, vol. 37, no. 4, pp. 1353–1359, 2022. doi:10.1007/s00146-021-01230-z

[44] M. Rubin, H. Arnon, J. D. Huppert, and A. Perry, "Considering human empathy in AI-driven therapy (preprint)," 2024. doi:10.2196/preprints.56529

[45] R. Agrawal and N. Pandey, "Developing rapport with emotionally intelligent AI assistants," *Int. J. Res. Appl. Sci. Eng. Technol.*, vol. 12, no. 3, pp. 1473–1480, 2024. doi: 10.22214/ijraset.2024.59015

[46] A. McStay, "Emotional AI and privacy," *Big Data & Society*, vol. 7, no. 1, p. 205395172090438, 2020. doi:10.1177/2053951720904386

[47] K. Roemmich, F. Schaub, and N. Andalibi, "Emotion AI at work," in *Proc. CHI Conf. Hum. Factors Comput. Syst.*, pp. 1–20, 2023. doi:10.1145/3544548.3580950

[48] E. Sedenberg and J. Chuang, "Smile for the camera: Privacy implications of emotion AI," *UC Berkeley School of Information*, n.d. [Online]. Available: https://www.ischool.berkeley.edu/research/publications/smile-camera-privacy-and-policy-implications-emotion-ai

[49] L. Rhue, "Racial influence on automated perceptions of emotions," *SSRN Electron. J.*, 2018. doi:10.2139/ssrn.3281765

[50] M. Yoshie and D. A. Sauter, "Cultural norms in nonverbal emotion expression," *Emotion*, vol. 20, no. 3, pp. 513–517, 2020. doi:10.1037/emo0000580

[51] M. Mattioli and F. Cabitza, "Ethics in automatic face emotion recognition," *Mach. Learn. Knowl. Extr.*, vol. 6, pp. 2201–2231, 2024. doi: 10.3390/make6040109

[52] M. Nagata and K. Okajima, "Observer culture and facial expression recognition," *PLoS ONE*, vol. 19, no. 10, p. e0313029, 2024. doi:10.1371/journal.pone.0313029

[53] I. Dominguez-Catena, D. Paternain, and M. Galar, "Metrics for dataset demographic bias in facial expression recognition," *arXiv preprint*, arXiv:2303.15889, 2024. [Online]. Available: https://arxiv.org/abs/2303.15889

[54] G. Benitez-Garcia, T. Nakamura, and M. Kaneko, "Facial expression recognition with Fourier descriptors," *J. Signal Inf. Process.*, vol. 8, no. 3, 2017. doi:10.4236/jsip.2017.83009

[55] R. Pusztahelyi and I. Stefán, "Social robots and data protection," *Acta Univ. Sapientiae, Legal Studies*, vol. 11, no. 1, pp. 95–118, 2022. doi:10.47745/AUSLEG.2022.11.1.06

[56] E. Schwitzgebel, "AI systems must not mislead about sentience," *Patterns*, vol. 4, no. 8, p. 100818, 2023. doi:10.1016/j.patter.2023.100818

[57] M. S. Farahani and G. Ghasemi, "Artificial intelligence and inequality," *Qeios*, 2024. doi:10.32388/7HWUZ2

[58] A. Hagerty and I. Rubinov, "Global AI ethics: Review of social impacts," *arXiv preprint*, arXiv:1907.07892, 2019. [Online]. Available: https://arxiv.org/abs/1907.07892.

[59] P. Choudhury, R. T. Allen, and M. G. Endres, "ML for pattern discovery in management research," *Strat. Manag. J.*, vol. 42, no. 1, pp. 30–57, 2021. doi:10.1002/smj.3215.