frontiers | Frontiers in Computer Science

TYPE Perspective
PUBLISHED 14 January 2026
DOI 10.3389/fcomp.2025.1692728

OPEN ACCESS

EDITED BY
Annahita Nezami,
Kepler Space Institute, United States

REVIEWED BY
Yen-Hung Hu,
Norfolk State University, United States

*CORRESPONDENCE
Vanessa Farsadaki
✉ vfarsadaki@sestrategies.org

RECEIVED 06 September 2025
REVISED 04 November 2025
ACCEPTED 25 November 2025
PUBLISHED 14 January 2026

CITATION
Farsadaki V and Griffy-Brown C (2026) AI affective computing and behavioral health. *Front. Comput. Sci.* 7:1692728. doi: 10.3389/fcomp.2025.1692728

AI affective computing and behavioral health

Vanessa Farsadaki[1,2,3]* and Charla Griffy-Brown[1]

[1]Thunderbird School of Global Management, Arizona State University, Phoenix, AZ, United States, [2]Space Exploration Strategies, Houston, TX, United States, [3]Kepler Space University, Bradenton, FL, United States

The integration of artificial intelligence (AI) and affective computing into behavioral health is transforming how mental wellbeing is assessed, monitored, and treated. As emotional and cognitive states can now be inferred through facial expressions, vocal tone, physiological signals, and behavioral cues, new technological paradigms are emerging to complement traditional approaches to mental healthcare. This is especially relevant in the wake of rising global mental health challenges, where access, personalization, and real-time feedback are essential components of effective care. Affective computing, a multidisciplinary field at the intersection of psychology, computer science, and cognitive science, seeks to enable machines to recognize, interpret, and respond to human emotions. When coupled with AI-driven data analytics and virtual reality (VR), it offers powerful tools for enhancing self-awareness, supporting clinical diagnostics, and delivering immersive therapeutic interventions. This paper explores how AI and affective computing can be leveraged across the behavioral health spectrum, from early detection and remote monitoring to therapy delivery and outcome prediction, with a particular emphasis on virtual environments as mediators of emotionally adaptive systems. We aim to review current innovations, examine their psychological validity, discuss ethical implications, and propose a research framework for advancing human-centered AI in behavioral health. Through this lens, we highlight the potential of emotionally intelligent systems not only to augment clinical practice but also to empower users in managing their mental wellbeing in real time.

KEYWORDS

affective computing, behavioral health, artificial intelligence, virtual reality (VR), digital phenotyping

1 Introduction

Behavioral health, encompassing mental health, emotional wellbeing, and substance use, remains one of the most complex and urgent domains in global healthcare (World Health Organization, 2022). Today, nearly one in every eight people worldwide lives with a mental health disorder, yet the majority do not receive adequate care (Ritchie and Roser, 2018). Access remains fragmented and uneven, stigma persists despite increasing awareness, and the cost of effective treatment continues to rise (Patel et al., 2018). The traditional boundaries of therapy, diagnosis, and support are being tested by a landscape marked by chronic provider shortages, economic strain, and social isolation, a reality brought into stark relief by recent global crises such as the COVID-19 pandemic (Moreno et al., 2020).

In this climate, the need for innovation in behavioral health is not just timely; it is imperative. Artificial intelligence (AI) and affective computing are transformative forces at the intersection of psychology, technology, and data science (Picard, 1997; Calvo et al., 2015). AI enables machines to learn from massive datasets, detect subtle patterns, and make predictions at scale (Esteva et al., 2019). Affective computing brings a critical human dimension: the capacity for machines to recognize, interpret, and respond to emotional states through cues such as facial expression, vocal tone, physiological responses, and behavior (Dzedzickis et al., 2020).

This technological synergy marks a profound shift in how we think about mental wellbeing and intervention. Traditional care models often rely on self-reporting and scheduled encounters, while AI and affective computing offer the promise of real-time, continuous insights. These systems can detect emotional changes before they escalate, personalize interventions, and reach people where and when they need it most (Torous et al., 2021). These technologies do not just augment the clinical toolkit; they have the potential to democratize behavioral health, breaking through barriers of access, stigma, and resource constraints (Kumar et al., 2022).

Further amplifying this transformation is the integration of virtual reality (VR) and immersive digital environments, which can mediate emotionally adaptive systems in entirely new ways (Freeman et al., 2017). Imagine a therapeutic platform that recognizes your emotional state and adapts in real time, offering tailored support, feedback, and immersive simulations that foster self-awareness and resilience (Fitzpatrick et al., 2017). By fusing AI, VR, and affective computing, we are moving closer to a model of behavioral health that is both highly personalized and universally accessible (Lindner et al., 2019).

The urgency is clear: global behavioral health cannot wait for incremental change. The fusion of AI and affective computing is not just a technological leap; it is a moral and practical imperative to build a future where mental wellbeing is within reach for all, powered by emotionally intelligent systems that understand, adapt, and empower (Topol, 2019).

2 Foundations of affective computing

Affective computing is the study and development of systems and devices that can recognize, interpret, process, and simulate human emotions. It draws on psychology, computer science, and cognitive neuroscience to help machines interact with people in ways that feel more natural and empathetic (Picard, 1997; Calvo et al., 2015). Core concepts include emotion detection, sentiment analysis, and multimodal sensing, which together form the foundation of emotionally intelligent artificial intelligence (AI) (Poria et al., 2017).

Emotion detection involves identifying human feelings through observable cues such as facial expressions, vocal tone, and physiological changes (Soleymani et al., 2017). Sentiment analysis, often applied to text, uses natural language processing (NLP) to determine the emotional tone and intent behind words (Medhat et al., 2014). Multimodal sensing combines visual, auditory, and biometric data to form a more complete picture of a person's emotional state, with the goal of approaching the nuance of human perception (Zeng et al., 2009).

Examples of affective computing technologies include facial expression recognition systems that detect micro-expressions and subtle facial muscle movements, NLP algorithms that identify emotional content in speech and text, voice analytics tools that analyze pitch, tone, and prosody, and wearable biosensors that monitor signals such as heart rate, skin conductance, and respiration (Picard, 1997; Calvo et al., 2015). When used together, these modalities can generate real-time emotional profiles with a high degree of detail.

Despite advances, the field faces significant challenges. Many algorithms are trained on datasets lacking demographic diversity, which can lead to biased results and reduced accuracy for certain groups (Buolamwini and Gebru, 2018). Systems that perform well in laboratory conditions often struggle in real-world environments, where uncontrolled variables make prediction harder (McDuff et al., 2019). Cultural differences in emotional expression can also cause misinterpretation (Jack et al., 2012). To ensure affective computing technologies are both accurate and ethical, ongoing efforts must focus on reducing bias, improving inclusivity, increasing transparency, and designing systems that work across diverse contexts (Calvo et al., 2015; McDuff et al., 2019).

3 Applications in behavioral health

3.1 Early detection and monitoring

Artificial intelligence and affective computing have dramatically expanded our capacity for early detection and monitoring of behavioral health conditions. Passive data collection—gathering information from everyday device use without user intervention—has emerged as a powerful strategy for identifying early signs of depression, anxiety, and post-traumatic stress disorder (PTSD) (Torous et al., 2021). For example, algorithms can analyze patterns in smartphone sensor data, such as sleep disruptions, changes in social interaction, or reduced mobility, which often precede the onset or worsening of mental health symptoms.

Beyond sensors, AI chatbots now engage users in natural conversation to unobtrusively assess mood, thought patterns, and risk factors. Woebot, for instance, is a digital mental health agent that monitors user input for signs of depression or anxiety, providing support and escalating care if risk is detected (Fitzpatrick et al., 2017). Passive mobile sensing platforms like Mindstrong Health utilize smartphone usage data and keystroke dynamics to detect cognitive changes in real time, enabling clinicians to monitor patients remotely and intervene early (Onnela and Rauch, 2016).

3.2 Therapeutic interventions

The convergence of AI and virtual reality (VR) is redefining the therapeutic landscape. Emotionally adaptive VR environments, powered by affective computing, can sense a user's emotional state through physiological sensors or facial analysis and adjust therapeutic content accordingly (Freeman et al., 2017). For example, in exposure therapy for anxiety or PTSD, the virtual environment can be modified in real time to gradually introduce stressors or provide calming stimuli, maximizing both safety and efficacy (Lindner et al., 2019).

AI-enhanced cognitive behavioral therapy (CBT) extends beyond scripted interventions. Advanced systems, such as Wysa or Tess, use machine learning to deliver CBT in a conversational format, dynamically adapting guidance to user responses (Inkster et al., 2018). In addition, VR-based therapies increasingly incorporate biofeedback—providing users with real-time physiological information such as heart rate variability or skin conductance—to facilitate emotional regulation and accelerate therapeutic progress (Repetto et al., 2019).

3.3 Personalized behavioral coaching

Personalized behavioral coaching represents the next frontier in digital mental health. AI companions, designed to offer real-time emotional support, leverage advances in natural language processing and sentiment analysis to detect user needs and respond empathetically (Laranjo et al., 2018). These companions provide ongoing encouragement, suggest coping strategies, and can even escalate to human intervention when necessary.

Reinforcement learning allows these systems to refine their support over time, learning which interventions are most effective for each individual and adjusting recommendations dynamically (Hochreiter and Schmidhuber, 1997; Mohr et al., 2017). Integration with digital phenotyping—the continuous collection of behavioral data from personal devices—enables hyper-personalized interventions, matching support to each user's unique psychological profile and moment-to-moment context (Insel, 2017).

By combining affective computing, AI analytics, and immersive environments, these innovations are making behavioral health more proactive, personalized, and accessible than ever before.

4 Virtual environments as emotionally adaptive systems

4.1 The role of VR and immersive technology

Virtual reality (VR) and related immersive technologies are transforming the way behavioral health professionals deliver care, offering unprecedented opportunities to assess, treat, and support individuals in emotionally responsive ways. By creating fully controlled, interactive digital environments, VR allows clinicians to simulate real-world scenarios and observe user responses in a safe, repeatable manner (Freeman et al., 2017). This capability is particularly valuable for exposure therapy, where gradual, controlled exposure to feared stimuli can be tailored to each individual's needs and progress (Maples-Keller et al., 2017).

Moreover, VR's immersive qualities facilitate deep engagement, offering an experience that is not only multisensory but also emotionally evocative (Parsons and Rizzo, 2008). This makes VR an ideal platform for behavioral health interventions that require both active participation and emotional processing.

4.2 Personalization and user engagement

Affective computing further amplifies the impact of VR by enabling environments and avatars to detect and respond to users' emotional states in real time. Emotionally adaptive systems can monitor physiological data (such as heart rate or galvanic skin response), vocal tone, and facial expressions to personalize content, pacing, and difficulty level (Dzedzickis et al., 2020; Wiederhold and Riva, 2019). For example, a VR environment designed for social anxiety might scale the complexity of a virtual crowd based on the user's measured stress levels, providing optimal exposure while preventing overwhelm (Lindner et al., 2019).

Personalization is key to engagement and therapeutic effectiveness. Studies show that adaptive VR environments, which adjust in response to a user's affective signals, are perceived as more supportive and result in higher adherence to therapeutic protocols (Riva et al., 2019).

4.3 Case studies and real-world examples

4.3.1 Exposure therapy

One of the most validated uses of VR in behavioral health is for exposure therapy in anxiety and PTSD. In a clinical trial, VR-based exposure therapy demonstrated effectiveness for veterans with PTSD, providing controlled and repeatable traumatic scene re-creations while allowing real-time monitoring of user distress (Maples-Keller et al., 2017; Rizzo et al., 2015). The emotional adaptability of these environments enables clinicians to titrate exposure with unprecedented precision.

4.3.2 Stress reduction and resilience training

Immersive VR programs are also being used for stress reduction and resilience building. For example, VR mindfulness and relaxation environments, which respond to physiological feedback, have shown significant reductions in self-reported stress and physiological arousal among users (Annerstedt et al., 2013; Wiederhold et al., 2020). These systems often leverage soothing nature scenes or guided meditations that adjust according to real-time biometric data.

4.3.3 Emotionally responsive avatars and environments

Avatars in VR can be designed to mirror and respond to a user's facial expressions, posture, and tone of voice, enhancing feelings of social presence and empathy (de Melo et al., 2019). Such emotionally responsive avatars have been used to facilitate social skills training in individuals with autism and to provide supportive coaching for those with depression or anxiety (Georgescu et al., 2014).

4.3.4 VR-based emotion elicitation for clinical assessment

VR is increasingly recognized as a powerful tool for emotion elicitation and assessment in research and clinical settings. By placing individuals in controlled, immersive scenarios, clinicians can observe authentic emotional and behavioral reactions that might be difficult to replicate in traditional settings (Parsons and Rizzo, 2008; Shiban et al., 2015). This approach enables more nuanced assessment of emotional regulation, reactivity, and coping strategies.

Overall, VR and immersive technologies, when paired with affective computing, offer not only new modes of therapy but also rich opportunities for assessment, engagement, and empowerment in behavioral health.

5 Psychological validity and effectiveness

5.1 Evidence from recent studies

The psychological validity and clinical effectiveness of AI-driven affective computing and immersive technologies in behavioral health have been increasingly supported by a growing body of empirical research. Randomized controlled trials have demonstrated that AI-powered chatbots and virtual agents can reduce symptoms of depression and anxiety with effect sizes comparable to traditional low-intensity interventions (Fitzpatrick et al., 2017; Inkster et al., 2018). Similarly, VR-based exposure therapy has been shown to be as effective, if not superior, to in-person exposure therapy for a variety of anxiety disorders and PTSD (Carl et al., 2019; Maples-Keller et al., 2017).

Mobile sensing and digital phenotyping approaches, which leverage passive data from smartphones and wearables, have been validated as reliable tools for early detection and monitoring of mental health status. These technologies can predict clinical deterioration days or even weeks before conventional assessments, increasing opportunities for early intervention (Torous et al., 2021).

Furthermore, emotionally adaptive environments, where VR content or chatbot responses change dynamically in response to user affect, have been shown to improve engagement, reduce dropout rates, and enhance perceived support during therapy (Riva et al., 2019; Lindner et al., 2019). These findings highlight the promise of emotionally intelligent systems in not only delivering effective care but also in fostering therapeutic alliance and trust.

5.2 Strengths and limitations

Among the strengths of AI and affective computing in behavioral health is scalability. Digital interventions can be delivered across geographies, reducing barriers of access, cost, and stigma that have traditionally limited care (Kumar et al., 2022). These systems also support continuous, real-time monitoring, enabling truly personalized interventions and the capacity to reach high-risk individuals outside clinical settings (Insel, 2017).

However, limitations remain. Psychological validity is contingent on the quality and representativeness of training data; many AI systems risk perpetuating bias if not carefully designed and validated across diverse populations (Obermeyer et al., 2019). While digital tools can enhance engagement, they may not be suitable for everyone, individual differences in technology acceptance, digital literacy, and trust must be considered (Mohr et al., 2017). Additionally, some studies report that automated interventions are most effective when integrated with some form of human support or oversight, rather than being fully autonomous (Laranjo et al., 2018).

Another challenge is the potential for over-reliance on quantitative data, which may not fully capture the nuanced, contextual factors that influence mental health (Harari et al., 2016). Ethical concerns around privacy, data security, and informed consent also require robust regulatory frameworks and ongoing vigilance (Moreno et al., 2020).

5.3 Real-world outcomes

Deployment at scale shows that emotionally aware, AI-enabled tools can produce measurable improvements in symptoms and functioning across diverse populations. AI chatbots such as Woebot and Wysa report reductions in depressive and anxiety symptoms comparable to low-intensity psychological interventions, with strong engagement in real-world use (Fitzpatrick et al., 2017; Inkster et al., 2018). VR exposure programs have translated into clinically meaningful and durable gains for PTSD and anxiety in health system settings, including veteran care, where controlled stimulus delivery and repeatability matter (Rizzo et al., 2015; Maples-Keller et al., 2017; Carl et al., 2019). Digital phenotyping has enabled earlier detection of deterioration through continuous behavioral signals, which creates new windows for timely outreach and risk mitigation in daily life contexts, not only in clinic visits (Onnela and Rauch, 2016; Torous et al., 2021).

These outcomes are driven by three reinforcing mechanisms. First, reach and immediacy: 24/7 availability lowers access barriers and stigma while supporting just-in-time interventions. Second, personalization: affect-sensitive systems adapt content and pacing to the user's state, which strengthens alliance and adherence. Third, continuous observation: passive sensing and repeated measures detect small changes that accumulate into clinically relevant signals. Benefits are moderated by factors such as severity, digital literacy, and preference for human contact, and they depend on privacy, safety, and equity safeguards that sustain trust (Riva et al., 2019; Lindner et al., 2019; Price and Cohen, 2019; Obermeyer et al., 2019).

In practice, the strongest and safest gains appear when these tools are embedded in stepped-care pathways that include human oversight, clear escalation protocols, and transparent data practices. Real-world effectiveness therefore hinges on thoughtful integration into clinical workflows, ongoing monitoring of equity and bias, and participatory feedback that keeps systems aligned with user goals and values (Topol, 2019; Price and Cohen, 2019; Obermeyer et al., 2019).

6 Affective computing in behavioral health: ethical and practical challenges

Table 1 shows that integrating affective computing into behavioral health offers transformative potential, but it also brings significant ethical and practical challenges. Because these systems handle deeply sensitive emotional data and mental

TABLE 1 Real-world behavioral outcomes from AI and affective computing interventions.

Intervention type	Representative outcomes	Core mechanisms behind outcomes	Noted limitations and caveats
AI chatbots for CBT and coaching (e.g., Woebot, Wysa)	Reductions in depressive and anxiety symptoms; strong short-term engagement and daily use adherence (Fitzpatrick et al., 2017; Inkster et al., 2018)	On-demand access, low stigma context, structured CBT micro-skills, sentiment-aware dialogue that tailors prompts	Lower effect in severe or crisis cases without human backup; reliance on self-report; need for escalation pathways and safeguarding
VR exposure and skills training	Symptom improvement in PTSD and anxiety disorders comparable to, and sometimes exceeding, in-person exposure; durable gains in follow-up (Rizzo et al., 2015; Maples-Keller et al., 2017; Carl et al., 2019)	Immersive realism with controllable intensity, graded exposure, affect-adaptive pacing, high presence that supports emotional processing	Cost and workflow integration; motion sickness for a subset; therapist training required; equity concerns for hardware access
Digital phenotyping and passive sensing	Earlier detection of deterioration and risk states; opportunities for proactive outreach and relapse prevention (Onnela and Rauch, 2016; Torous et al., 2021)	Continuous behavioral signals, deviation detection from personal baselines, aggregation across modalities for robust patterns	Privacy and consent management; potential bias if data are not representative; false positives without clinical context
Emotion-adaptive environments and agents	Improved engagement, lower dropout, stronger perceived support and alliance in guided programs (Riva et al., 2019; Lindner et al., 2019)	Real-time adjustment to arousal and affect, personalized pacing and content, feedback loops that reinforce self-efficacy	Generalizability across cultures and contexts; need for transparent models to maintain trust; validation beyond lab settings

health indicators, privacy, consent, and algorithmic fairness must be central design priorities (Shen et al., 2020; Vinciarelli and Mohammadi, 2014).

Privacy is one of the most pressing concerns in emotional AI. Emotionally responsive systems often require continuous data collection from personal devices, wearable biometric sensors, and even social interactions (McDuff et al., 2019). Safeguarding this information involves ensuring informed consent, collecting only the minimum necessary data, and giving users control over how their information is stored and shared (Martínez-Miranda and Aldea, 2005). Behavioral health data is especially sensitive, and privacy-by-design approaches are essential. These should include strong encryption, secure storage systems, and compliance with frameworks such as the Health Insurance Portability and Accountability Act (HIPAA) in the United States and the General Data Protection Regulation (GDPR) in the European Union (Hao et al., 2021).

Bias and fairness in emotion recognition systems represent another serious challenge. Many current AI models are trained on datasets that fail to adequately represent different ages, races, cultural backgrounds, or genders (Buolamwini and Gebru, 2018). This lack of diversity can lead to misinterpretation of emotional signals or inappropriate system responses—an especially dangerous risk in mental health contexts where errors can affect diagnosis or care (Howard and Borenstein, 2018). Addressing this problem requires inclusive data collection, algorithmic transparency, and ongoing monitoring for disparate impacts (Shen et al., 2020).

Human-AI interaction and trust are equally critical for adoption in behavioral health. For these systems to be effective, users must view them as credible, empathetic, and respectful of their emotional states (Lisetti et al., 2013). Designing with explainability in mind—so users understand how and why the system responds in a particular way—can help foster a therapeutic alliance (Abd-Alrazaq et al., 2019). Without trust, even the most advanced AI tools will fail to achieve meaningful engagement.

Finally, ethical deployment requires interdisciplinary oversight. Clinicians, technologists, ethicists, and patients should collaborate to set standards for transparency, accountability, and consent (Hao et al., 2021). This collaboration not only mitigates risks but also ensures that affective computing in behavioral health aligns with human values and contributes positively to mental health outcomes (Howard and Borenstein, 2018).

7 A framework for human-centered AI in behavioral health

7.1 Principles of design and deployment

The preceding sections surveyed current affective computing and AI applications in behavioral health, identifying their strengths, limitations, and real-world outcomes. Building on this review, Section 7 proposes a human-centered deployment framework to ensure that emotionally intelligent behavioral health systems are safe, ethical, equitable, and clinically meaningful. The principles outlined here were selected based on their recurrence across major digital mental health guidelines (e.g., Topol, 2019; Calvo et al., 2015), WHO and APA standards for responsible digital care, and AI ethics frameworks emphasizing trust, transparency, and inclusion as prerequisites for therapeutic adoption and effectiveness.

Each principle reflects a critical requirement for behavioral health: safety and privacy protect users from harm; empathy and engagement ensure adherence; transparency and explainability foster trust; inclusivity and equity address biases and disparities; and empowerment and co-design sustain long-term user relevance. These principles are interdependent and collectively serve as guardrails for implementation. Their practical effectiveness

TABLE 2 Mapping human-centered principles to their justification, enabling technologies, policy alignment, and implementation pathways.

Principle	Why it matters in behavioral health	Example enabling technologies	Supporting policy/ethical frameworks	Example implementation pathway
Safety and privacy	Affective data and behavioral signals are deeply personal; breaches erode trust and increase harm (Price and Cohen, 2019)	End-to-end encryption, federated learning, on-device affect detection	GDPR, HIPAA, privacy-by-design standards	Secure AI-based emotion tracking in CBT apps with opt-in consent and local processing
Empathy and engagement	Therapeutic success depends on user trust, emotional containment, and perceived relational support (Picard, 1997; Fiske et al., 2019)	Sentiment-aware NLP, affect-responsive avatars, adaptive VR coaching	WHO engagement guidelines	Emotionally adaptive chatbot delivering CBT with tone-matching responses
Transparency and explainability	Users and clinicians must understand how emotional inferences and decisions are made to avoid distrust or misinterpretation (Obermeyer et al., 2019)	Explainable AI (XAI) models, interpretability dashboards	EU AI Act, AI clinical audit standards	Dashboard showing how stress levels are inferred from heart rate and voice patterns
Inclusivity and equity	Bias in emotion recognition can reinforce disparities and harm marginalized populations (Buolamwini and Gebru, 2018)	Culturally diverse training datasets, bias detection algorithms	IEEE AI ethics standards; equity-first design	Facial emotion recognition model retrained with multi-ethnic datasets
Empowerment and co-design	Users are more likely to sustain engagement when they co-shape tools aligned with their therapeutic narratives (Sanders and Stappers, 2014)	Participatory design interfaces, customizable therapeutic journeys	Patient-centered care frameworks	Co-created VR anxiety exposure scenes with user-defined triggers and pacing

depends on translating each principle into specific technical, clinical, and policy strategies that can be operationalized in real-world deployments. These human-centered principles, along with their ethical justification, enabling technologies, and policy alignment, are mapped in Table 2.

Operationalizing these principles requires cross-functional alignment among AI developers, behavioral clinicians, ethicists, policymakers, and end-users. For instance, an affective VR therapy platform designed under this framework would incorporate encryption (safety), adaptive avatar responsiveness (empathy), user-facing emotional feedback explanations (transparency), validated cross-cultural emotion models (equity), and iterative patient feedback loops (co-design). In this way, the framework provides both a conceptual foundation and a deployment roadmap toward responsible and effective affective AI in behavioral health.

8 Research gaps and future directions

Despite rapid advances, current applications of affective computing and AI in behavioral health face several unresolved gaps that limit their long-term efficacy, safety, scalability, and equitable adoption. These gaps were identified based on recurring limitations observed in existing systems (Sections 3–6), implementation challenges (Section 5.3), and unmet needs for human-centered alignment (Section 7.1). They fall into four main categories: technological challenges, clinical integration gaps, policy and regulatory limitations, and ethical/social considerations. Addressing these gaps is essential for moving from experimental success toward safe, effective, and sustainable real-world deployment at scale. Key research gaps across technology, clinical integration, policy, and ethics, along with their relative priority, are outlined in Table 3.

8.1 Technology gaps

8.1.1 Longitudinal validation remains limited

Most current evaluations are short-term, which restricts understanding of how emotional AI performs in prolonged use across mental health trajectories (Lindner et al., 2019).

8.1.2 Multimodal data fusion is still underdeveloped

Effective affective computing requires integrating voice, facial expression, behavior, and physiological biosignals, yet many current systems rely on single modalities, which limits emotional sensitivity and resilience in real-world conditions (Dzedzickis et al., 2020; Harari et al., 2016).

8.1.3 Context-aware emotional modeling is insufficient

Emotion recognition systems often lack situational awareness and fail to differentiate between clinically relevant emotional states and benign affective fluctuations, reducing diagnostic precision.

TABLE 3 Key research gaps, their significance, and priority levels for advancing affective computing and AI applications in behavioral health.

Gap category	Specific gap	Why it matters	Priority level
Technology	Lack of longitudinal validation	Short-term studies cannot confirm sustained effectiveness or safety	High
Technology	Insufficient multimodal emotional integration	Single-modality detection reduces accuracy and robustness	High
Technology	Poor context sensitivity	Risk of false alerts and reduced clinical relevance	Medium
Clinical integration	Low interoperability with healthcare systems	Limits clinician uptake and workflow adoption	High
Clinical integration	Underdeveloped hybrid AI-human care models	Fully automated interventions may reduce therapeutic alliance	High
Clinical integration	Lack of standard escalation protocols	Unclear responsibility in high-risk mental health situations	Medium
Policy and regulation	Absence of emotional AI approval pathways	Slows safe adoption in clinical environments	High
Policy and regulation	No standardized fairness benchmarks	Leads to unequal harm and legal uncertainty	Medium
Ethics and society	Risk of emotional surveillance and manipulation	Threatens autonomy and user trust	High
Ethics and society	Persistent bias in emotion detection	Misdiagnosis risk for marginalized groups	High
Ethics and society	Digital exclusion and inequality	Widening mental health access gap	Medium

8.2 Clinical integration gaps

8.2.1 Limited interoperability with existing healthcare systems

AI-generated emotional insights are often siloed, lacking seamless integration into electronic health records and clinical workflows (Topol, 2019).

8.2.2 Inadequate human-AI hybrid care models

Fully automated mental health support may not be equally effective for all users; hybrid models that define when and how clinicians intervene remain underdeveloped (Mohr et al., 2017).

8.2.3 Unclear frameworks for clinical responsibility and escalation

There is still limited guidance on how clinical accountability should be managed when AI-driven emotional inferences are used in patient care.

8.3 Policy and regulatory gaps

8.3.1 Lack of standardized regulatory pathways for emotional AI

Most existing frameworks (e.g., GDPR, HIPAA) address data privacy but not affective inference risk or emotional manipulation.

8.3.2 No universal benchmarks for assessing accuracy and fairness in emotion recognition

Regulatory bodies lack validated performance thresholds for approving affective computing tools, especially those used in high-stakes behavioral health decision-making.

8.4 Ethical and societal gaps

8.4.1 Emotional surveillance concerns

Prolonged monitoring of affective signals raises questions about user autonomy, emotional manipulation, and consent boundaries (Shen et al., 2020).

8.4.2 Bias and cultural misinterpretation persist

Even improved datasets struggle to fully capture diverse affective norms, risking misclassification and inequitable treatment (Buolamwini and Gebru, 2018).

8.4.3 Digital exclusion risks

Populations with low digital literacy, limited connectivity, or mistrust in AI may be further marginalized if emotionally adaptive care becomes standard without inclusive strategies.

Addressing these gaps requires coordinated collaboration between developers, behavioral clinicians, ethicists, regulatory agencies, and affected communities. Future research must focus on building context-aware, culturally adaptive emotional models; validating long-term effectiveness in diverse populations; constructing clinically actionable AI-human collaboration protocols; and developing policy frameworks that ensure fairness, transparency, and accountability. Only through a multidisciplinary and inclusive approach can AI-driven affective computing evolve into a trusted pillar of behavioral healthcare.

9 Discussion

The integration of artificial intelligence and affective computing into behavioral health is redefining the landscape of mental wellbeing and care. These technologies, when thoughtfully designed and ethically deployed, have the power to dramatically improve behavioral health outcomes on a global scale. From early detection and real-time monitoring to adaptive therapeutic interventions and personalized coaching, AI-driven systems are enhancing access, personalization, and efficacy in ways that traditional approaches alone could never achieve (Fitzpatrick et al., 2017; Torous et al., 2021).

Virtual reality and immersive environments further amplify this potential, offering emotionally responsive systems that foster engagement, resilience, and deep personal insight (Freeman et al., 2017; Riva et al., 2019). As a result, the promise of emotionally intelligent, human-centered digital care is no longer a distant vision—it is already beginning to transform how we diagnose, treat, and support mental health for millions worldwide.

Yet, realizing the full potential of these innovations requires more than technological advancement. It demands genuine, sustained interdisciplinary collaboration. Clinicians, technologists, ethicists, designers, researchers, and patients themselves must work together to ensure that these tools are safe, effective, transparent, and equitable (Topol, 2019; Moreno et al., 2020). Only by bridging the worlds of clinical wisdom, ethical rigor, and human-centered design can we build emotionally intelligent systems that are truly fit for purpose.

The call to action is clear: we must break down silos and foster active partnerships across disciplines and sectors. Together, we can shape a future where behavioral health support is accessible, adaptive, and empowering for all—delivered not just by machines, but by emotionally aware, human-guided AI that honors the complexity and dignity of every individual.

Data availability statement

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

Author contributions

VF: Investigation, Writing – review & editing, Conceptualization, Writing – original draft. CG-B: Writing – original draft.

Funding

The author(s) declare that no financial support was received for the research and/or publication of this article.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The author(s) declare that no Gen AI was used in the creation of this manuscript.

Any alternative text (alt text) provided alongside figures in this article has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Abd-Alrazaq, A., Alajlani, M., Alalwan, A., Bewick, B. M., Gardner, P., and Househ, M. (2019). An overview of the features of chatbots in mental health: a scoping review. *Int. J. Med. Inform.* 132:103978. doi: 10.1016/j.ijmedinf.2019.103978

Annerstedt, M., Johansson, M., Ivarsson, J., Hultén, S., and Östergren, P.-O. (2013). Inducing physiological stress recovery with sounds of nature in a virtual reality forest-Results from a pilot study. *Physiol. Behav.* 118, 240–250. doi: 10.1016/j.physbeh.2013.05.023

Buolamwini, J., and Gebru, T. (2018). Gender shades: Intersectional accuracy disparities in commercial gender classification. *arXiv [Preprint]*. doi: 10.48550/arXiv.1810.00471

Calvo, R. A., D'Mello, S., Gratch, J., and Kappas, A. (2015). *The Oxford Handbook of Affective Computing*. Oxford: Oxford University Press. doi: 10.1093/oxfordhb/9780199942237.001.0001

Carl, E., Stein, A. T., Levihn-Coon, A., Pogue, J. R., Rothbaum, B., Emmelkamp, P., et al. (2019). Virtual reality exposure therapy for anxiety and related disorders: a meta-analysis of randomized controlled trials. *J. Anxiety Disord.* 61, 27–36. doi: 10.1016/j.janxdis.2018.08.003

de Melo, C. M., Gratch, J., and Carnevale, P. J. (2019). Humans versus virtual agents: The effect of emotion expressions on people's decision making. *IEEE Trans. Affect. Comput.* 10, 291–305. doi: 10.1109/TAFFC.2017.2769121

Dzedzickis, A., Tamosiunaite, M., and Maskeliunas, R. (2020). Human emotion recognition: Review of sensors and methods. *Sensors* 20:592. doi: 10.3390/s20030592

Esteva, A., Robicquet, A., Ramsundar, B., Kuleshov, V., DePristo, M., Chou, K., et al. (2019). A guide to deep learning in healthcare. *Nat. Med.* 25, 24–29. doi: 10.1038/s41591-018-0316-z

Fiske, A., Henningsen, P., and Buyx, A. (2019). Your robot therapist will see you now: Ethical implications of embodied artificial intelligence in mental health care. *J. Med. Internet Res.* 21:e13216. doi: 10.2196/13216

Fitzpatrick, K. K., Darcy, A., and Vierhile, M. (2017). Delivering cognitive behavior therapy to young adults with symptoms of depression and anxiety using a fully automated conversational agent (Woebot): a randomized controlled trial. *JMIR Mental Health* 4:e19. doi: 10.2196/mental.7785

Freeman, D., Reeve, S., Robinson, A., Ehlers, A., Clark, D., Spanlang, B., et al. (2017). Virtual reality in the assessment, understanding, and treatment of mental health disorders. *Psychol. Med.* 47, 2393–2400. doi: 10.1017/S003329171700040X

Georgescu, A. L., Kuzmanovic, B., Santos, N. S., Tepest, R., Bente, G., and Tittgemeyer, M. (2014). Perceiving nonverbal behavior: Neural correlates of processing movement fluency and contingency in dyadic interactions. *Hum. Brain Mapp.* 35, 1362–1378. doi: 10.1002/hbm.22259

Hao, Q., Zhang, L., and Wen, F. (2021). Privacy protection in affective computing: challenges and future directions. *IEEE Trans. Affect. Comput.* 12, 609–623. doi: 10.1109/TAFFC.2019.2909096

Harari, G. M., Lane, N. D., Wang, R., Crosier, B. S., Campbell, A. T., and Gosling, S. D. (2016). Using smartphones to collect behavioral data in psychological science: opportunities, practical considerations, and challenges. *Perspect. Psychol. Sci.* 11, 838–854. doi: 10.1177/1745691616650285

Hochreiter, S., and Schmidhuber, J. (1997). Long short-term memory. *Neural Comput.* 9, 1735–1780. doi: 10.1162/neco.1997.9.8.1735

Howard, A., and Borenstein, J. (2018). The ugly truth about ourselves and our robot creations: the problem of bias and social inequity. *Sci. Eng. Ethics* 24, 1521–1536. doi: 10.1007/s11948-017-9975-2

Inkster, B., Sarda, S., and Subramanian, V. (2018). An empathy-driven, conversational artificial intelligence agent (Wysa) for digital mental wellbeing: real-world data evaluation. *JMIR mHealth uHealth* 6:e12106. doi: 10.2196/12106

Insel, T. R. (2017). Digital phenotyping: technology for a new science of behavior. *JAMA* 318, 1215–1216. doi: 10.1001/jama.2017.11295

Jack, R. E., Garrod, O. G. B., Yu, H., Caldara, R., and Schyns, P. G. (2012). Facial expressions of emotion are not culturally universal. *Proc. Nat. Acad. Sci.* 109, 7241–7244. doi: 10.1073/pnas.1200155109

Kumar, S., Nilsen, W. J., Abernethy, A., Atienza, A., Patrick, K., Pavel, M., et al. (2022). Mobile health technology evaluation: the mHealth evidence workshop. *Am. J. Prev. Med.* 45, 228–236. doi: 10.1016/j.amepre.2013.03.017

Laranjo, L., Dunn, A. G., Tong, H. L., Kocaballi, A. B., Chen, J., Bashir, R., et al. (2018). Conversational agents in healthcare: a systematic review. *J. Am. Med. Inform. Assoc.* 25, 1248–1258. doi: 10.1093/jamia/ocy072

Lindner, P., Miloff, A., Hamilton, W., Reuterskiöld, L., Andersson, G., Powers, M. B., et al. (2019). Creating state of the art, next-generation virtual reality exposure therapies for anxiety disorders using consumer hardware platforms: design considerations and future directions. *Cogn. Behav. Ther.* 48, 404–420. doi: 10.1080/16506073.2017.1280843

Lisetti, C. L., Amini, R., Yasavur, U., and Rishe, N. (2013). I can help you change! An empathic virtual agent delivers behavior change health interventions. *ACM Trans. Manag. Inf. Syst.* 4, 1–28. doi: 10.1145/2544103

Maples-Keller, J. L., Bunnell, B. E., Kim, S. J., and Rothbaum, B. O. (2017). The use of virtual reality technology in the treatment of anxiety and other psychiatric disorders. *Harv. Rev. Psychiatry* 25, 103–113. doi: 10.1097/HRP.0000000000000138

Martínez-Miranda, J., and Aldea, A. (2005). Emotions in human and artificial intelligence. *Comput. Hum. Behav.* 21, 323–341. doi: 10.1016/j.chb.2004.02.01

McDuff, D., Mahmoud, A., Mavadati, S., Amr, M., and Kaliouby, R. E. (2019). "AFFDEX SDK: a cross-platform real-time multi-face expression recognition toolkit," in *Proceedings of the 2019 International Conference on Multimodal Interaction,* 414–415.

Medhat, W., Hassan, A., and Korashy, H. (2014). Sentiment analysis algorithms and applications: a survey. *Ain Shams Eng. J.* 5, 1093–1113. doi: 10.1016/j.asej.2014.04.011

Mohr, D. C., Zhang, M., and Schueller, S. M. (2017). Personal sensing: Understanding mental health using ubiquitous sensors and machine learning. *Annu. Rev. Clin. Psychol.* 13, 23–47. doi: 10.1146/annurev-clinpsy-032816-044949

Moreno, C., Wykes, T., Galderisi, S., Nordentoft, M., Crossley, N., Jones, N., et al. (2020). How mental health care should change as a consequence of the COVID-19 pandemic. *Lancet Psychiat.* 7, 813–824. doi: 10.1016/S2215-0366(20)30307-2

Obermeyer, Z., Powers, B., Vogeli, C., and Mullainathan, S. (2019). Dissecting racial bias in an algorithm used to manage the health of populations. *Science* 366, 447–453. doi: 10.1126/science.aax2342

Onnela, J. P., and Rauch, S. L. (2016). Harnessing smartphone-based digital phenotyping to enhance behavioral and mental health. *Neuropsychopharmacology* 41, 1691–1696. doi: 10.1038/npp.2016.7

Parsons, T. D., and Rizzo, A. A. (2008). Affective outcomes of virtual reality exposure therapy for anxiety and specific phobias: a meta-analysis. *J. Behav. Ther. Exp. Psychiatry* 39, 250–261. doi: 10.1016/j.jbtep.2007.07.007

Patel, V., Saxena, S., Lund, C., Thornicroft, G., Baingana, F., Bolton, P., et al. (2018). The Lancet Commission on global mental health and sustainable development. *Lancet* 392, 1553–1598. doi: 10.1016/S0140-6736(18)31612-X

Picard, R. W. (1997). *Affective Computing*. London: MIT Press. doi: 10.7551/mitpress/1140.001.0001

Poria, S., Cambria, E., Bajpai, R., and Hussain, A. (2017). A review of affective computing: From unimodal analysis to multimodal fusion. *Inf. Fusion* 37, 98–125. doi: 10.1016/j.inffus.2017.02.003

Price, W. N., and Cohen, I. G. (2019). Privacy in the age of medical big data. *Nat. Med.* 25, 37–43. doi: 10.1038/s41591-018-0272-7

Repetto, C., Serino, S., Macedonia, M., and Riva, G. (2019). Virtual reality as an embodied tool to enhance episodic memory in elderly. *Front. Psychol.* 10:2622. doi: 10.3389/fpsyg.2019.02622

Ritchie, H., and Roser, M. (2018). *Mental health.* Our World in Data. Available online at: https://ourworldindata.org/mental-health (Accessed March 20, 2025).

Riva, G., Wiederhold, B. K., and Mantovani, F. (2019). Neuroscience of virtual reality: From virtual exposure to embodied medicine. *Cyberpsychol. Behav. Soc. Netw.* 22, 82–96. doi: 10.1089/cyber.2017.29099.gri

Rizzo, A. S., Koenig, S. T., and Talbot, T. B. (2015). Virtual reality as a tool for delivering PTSD exposure therapy and stress resilience training. *Milit. Behav. Health* 3, 254–264. doi: 10.1080/21635781.2015.1044092

Sanders, E. B. N., and Stappers, P. J. (2014). Probes, toolkits and prototypes: three approaches to making in codesigning. *CoDesign* 10, 5–14. doi: 10.1080/15710882.2014.888183

Shen, J., Nam, W., and Wang, J. (2020). Ethics of artificial intelligence in affective computing: a systematic literature review. *Comput. Human Behav.* 115:106612. doi: 10.1016/j.chb.2020.106612

Shiban, Y., Pauli, P., and Mühlberger, A. (2015). Effect of multiple context exposure on renewal in spider phobia. *Behav. Res. Ther.* 68, 22–29. doi: 10.1016/j.brat.2015.02.005

Soleymani, M., Asghari-Esfeden, S., Fu, Y., and Pantic, M. (2017). Analysis of EEG signals and facial expressions for continuous emotion detection. *IEEE Trans. Affect. Comput.* 7, 17–28. doi: 10.1109/TAFFC.2015.2436926

Topol, E. (2019). *Deep Medicine: How Artificial Intelligence Can Make Healthcare Human Again.* Hachette UK: Basic Books.

Torous, J., Larsen, M. E., Depp, C., Cosco, T. D., Barnett, I., Nock, M. K., et al. (2021). Smartphones, sensors, and machine learning to advance real-time prediction and interventions for suicide prevention: a review of current progress and next steps. *Curr. Psychiatry Rep.* 20:51. doi: 10.1007/s11920-018-0914-y

Vinciarelli, A., and Mohammadi, G. (2014). A survey of personality computing. *IEEE Trans. Affect. Comput.* 5, 273–291. doi: 10.1109/TAFFC.2014.2330816

Wiederhold, B. K., Miller, I. T., and Wiederhold, M. D. (2020). Using virtual reality to mobilize health care: Mobile virtual reality for mental health treatment. *Cyberpsychol. Behav. Soc. Netw.* 23, 385–389. doi: 10.1089/cyber.2020.29189.bkw

Wiederhold, B. K., and Riva, G. (2019). Virtual reality therapy: emerging topics and future challenges. *Cyberpsychol. Behav. Soc. Netw.* 22, 3–6. doi: 10.1089/cyber.2018.29136.bkw

World Health Organization (2022). *World Mental Health Report: Transforming Mental Health For All.* Geneva: World Health Organization.

Zeng, Z., Pantic, M., Roisman, G. I., and Huang, T. S. (2009). A survey of affect recognition methods: audio, visual, and spontaneous expressions. *IEEE Trans. Pattern Anal. Mach. Intell.* 31, 39–58. doi: 10.1109/TPAMI.2008.52