Established ethical and practical frameworks for designing AI systems that compensate for users' cognitive or emotional challenges emphasize **human-centered design**, **privacy and bias mitigation**, and **interdisciplinary integration** of psychological, ethical, and technical principles.

1. Introduction

The rapid integration of artificial intelligence (AI) into domains such as mental health, education, and eldercare has driven the development of frameworks to ensure that AI systems ethically and effectively compensate for users' cognitive or emotional challenges. These frameworks are grounded in principles of human-centered design, transparency, privacy, and bias mitigation, and often draw from interdisciplinary theories such as emotional intelligence, affective computing, and cognitive psychology (Deckker & Sumanasekara, 2025; Thakkar et al., 2024; Samsonovich, 2020; Zhao et al., 2022; Velagaleti et al., 2024; Tretter, 2024; Beg et al., 2024; Thieme et al., 2022; Liu et al., 2024; Kundu, 2022). Practical frameworks emphasize adaptive personalization, explainability, and the integration of psychological theories (e.g., Cognitive Behavioral Therapy) to enhance user well-being and autonomy (Deckker & Sumanasekara, 2025; Benita et al., 2025; Patil et al., 2025; Kallivalappil et al., 2023; Thieme et al., 2022). Ethical frameworks stress the importance of safeguarding user privacy, ensuring informed consent, addressing algorithmic bias, and maintaining human oversight, especially in sensitive contexts like mental health care (Zhang & Wang, 2024; Thakkar et al., 2024; Tretter, 2024; Beg et al., 2024; Graham et al., 2019; Thieme et al., 2022; Liu et al., 2024). Despite significant progress, challenges remain in standardizing evaluation metrics, ensuring cultural inclusivity, and balancing technological innovation with ethical safeguards (Deckker & Sumanasekara, 2025; Thakkar et al., 2024; Tretter, 2024; Beg et al., 2024; Liu et al., 2024). This review synthesizes the most influential ethical and practical frameworks, highlighting their core principles, implementation strategies, and ongoing challenges in the design of compensatory AI systems.

2. Methods

We conducted a comprehensive search across over 170 million research papers in Consensus, including Semantic Scholar, PubMed, and other major databases. The search targeted ethical and practical frameworks for AI systems designed to compensate for cognitive or emotional challenges. In total, 1024 papers were identified, 719 were screened, 560 were deemed eligible, and the 50 most relevant papers were included in this review.

FIGURE 1 Flow diagram of the literature search and selection process.

Eight unique search groups were executed, systematically covering foundational, practical, domain-specific, and critical perspectives on ethical and practical frameworks for compensatory AI.

3. Results

3.1. Core Ethical Frameworks

Ethical frameworks for compensatory AI systems consistently emphasize user autonomy, privacy, transparency, and fairness. Key guidelines include the need for informed consent, robust data protection, and the mitigation of algorithmic bias, especially in mental health and educational applications (Deckker & Sumanasekara, 2025; Zhang & Wang, 2024; Thakkar et al., 2024; Tretter, 2024; Beg et al., 2024; Graham et al., 2019; Thieme et al., 2022; Liu et al., 2024). The importance of human oversight and the prevention of over-reliance on AI are also central themes (Zhang & Wang, 2024; Tretter, 2024; Beg et al., 2024; Thieme et al., 2022). Frameworks such as the "Ethics of Emotional AI" and "Human-Centered AI" advocate for ongoing ethical review and stakeholder engagement throughout the AI system lifecycle (Tretter, 2024; Thieme et al., 2022; Liu et al., 2024).

3.2. Practical Design Frameworks

Practical frameworks focus on adaptive personalization, explainability, and the integration of psychological and educational theories. For example, AI systems in mental health often incorporate Cognitive Behavioral Therapy (CBT) principles, emotion recognition, and adaptive feedback to support user well-being (Deckker & Sumanasekara, 2025; Benita et al., 2025; Patil et al., 2025; Kallivalappil et al., 2023; Beg et al., 2024; Shegekar et al., 2024; Thieme et al., 2022). In education, frameworks emphasize social-emotional learning (SEL), real-time emotion detection, and personalized interventions to enhance engagement and learning outcomes (Deckker & Sumanasekara, 2025; Liu et al., 2024; Bilquise et al., 2022; Zong & Yang, 2025; Shi, 2024; Moghadam et al., 2023; Darejeh et al., 2024; Zahra et al., 2025). The R-CAGE model and other user-centered architectures prioritize sustainable emotional engagement and cognitive autonomy (Choi, 2025; Samsonovich, 2020; Zhao et al., 2022; Kundu, 2022).

3.3. Interdisciplinary Integration

Many frameworks draw from interdisciplinary theories, combining affective computing, emotional intelligence, cognitive psychology, and HCI principles (Deckker & Sumanasekara, 2025; Thakkar et al., 2024; Samsonovich, 2020; Zhao et al., 2022; Velagaleti et al., 2024; Tretter, 2024; Beg et al., 2024; Thieme et al., 2022; Liu et al., 2024; Kundu, 2022). This integration enables AI systems to better recognize, interpret, and respond to user emotions and cognitive states, while also addressing ethical and practical challenges unique to each domain (Deckker & Sumanasekara, 2025; Samsonovich, 2020; Zhao et al., 2022; Velagaleti et al., 2024; Tretter, 2024; Beg et al., 2024; Thieme et al., 2022; Liu et al., 2024).

3.4. Challenges and Limitations

Despite advances, significant challenges persist. These include algorithmic bias, cross-cultural misinterpretation of emotions, privacy concerns, lack of standardized evaluation metrics, and the risk of over-reliance on AI (Deckker & Sumanasekara, 2025; Zhang & Wang, 2024; Thakkar et al., 2024; Tretter, 2024; Beg et al., 2024; Graham et al., 2019; Thieme et al., 2022; Liu et al., 2024). The need for culturally inclusive design, longitudinal validation, and hybrid human-AI scaffolding is frequently highlighted (Deckker & Sumanasekara, 2025; Thakkar et al., 2024; Tretter, 2024; Beg et al., 2024; Liu et al., 2024).

Key Papers

Paper	Framework Type	Domain	Key Principles	Notable Features
(Deckker & Sumanasekara, 2025)	Systematic review	Education, mental health	Emotion-aware, adaptive, ethical governance	Emphasizes cultural inclusivity, hybrid scaffolding
(Tretter, 2024)	Ethical analysis	Decision-support systems	Emotional capacity, bias mitigation, social discourse	Advocates for separate ethical review per AI type
(Thieme et al., 2022)	Human-centered design	Online CBT, mental health	Stakeholder engagement, balanced info, human oversight	Design sessions with clinicians, risk of over-reliance
(Beg et al., 2024)	Narrative review	Psychotherapy, mental health	Privacy, trust, human-AI relationship	Cautious integration, patient well-being focus
(Liu et al., 2024)	Editorial review	Healthcare, education	Trust, transparency, user acceptance	Multimodal systems, cross-cultural validation

FIGURE 2 Comparison of key studies on ethical and practical frameworks for compensatory AI.

Top Contributors

Type	Name	Papers
Author	Anja Thieme	(Thieme et al., 2022)
Author	A. Samsonovich	(Samsonovich, 2020)
Author	Gustavo Assunção	(Assunção et al., 2022)
Journal	*Frontiers in Psychology*	(Liu et al., 2024)
Journal	*IEEE Access*	(Nag et al., 2025; Rokhsaritalemi et al., 2023)
Journal	*Frontiers in Digital Health*	(Thakkar et al., 2024)

FIGURE 3 Authors & journals that appeared most frequently in the included papers.

4. Discussion

The literature demonstrates that established ethical and practical frameworks for compensatory AI systems are grounded in human-centered design, privacy protection, and interdisciplinary integration (Deckker & Sumanasekara, 2025; Thakkar et al., 2024; Samsonovich, 2020; Zhao et al., 2022; Tretter, 2024; Beg et al., 2024; Thieme et al., 2022; Liu et al., 2024; Kundu, 2022). High-quality evidence supports the effectiveness of adaptive, emotion-aware AI in improving user outcomes, particularly when systems are designed with ethical safeguards and stakeholder input (Deckker & Sumanasekara, 2025; Benita et al., 2025; Patil et al., 2025; Kallivalappil et al., 2023; Beg et al., 2024; Thieme et al., 2022). However, persistent challenges such as algorithmic bias, privacy risks, and the need for cultural inclusivity highlight the importance of ongoing ethical review and transparent communication (Deckker & Sumanasekara, 2025; Zhang & Wang, 2024; Thakkar et al., 2024; Tretter, 2024; Beg et al., 2024; Graham et al., 2019; Thieme et al., 2022; Liu et al., 2024). The field is moving toward more robust, context-sensitive frameworks that balance technological innovation with ethical responsibility, but further research is needed to standardize evaluation metrics and ensure equitable access.

Claims and Evidence Table

Claim	Evidence Strength	Reasoning	Papers
Human-centered, ethical frameworks are essential for compensatory AI design	Strong	Multiple reviews and empirical studies show improved outcomes and reduced risks with user-centered, ethical approaches	(Deckker & Sumanasekara, 2025; Tretter, 2024; Beg et al., 2024; Thieme et al., 2022; Liu et al., 2024)
Adaptive, emotion-aware AI enhances user well-being and engagement	Strong	Systematic reviews and case studies demonstrate positive effects in education and mental health	(Deckker & Sumanasekara, 2025; Benita et al., 2025; Patil et al., 2025; Kallivalappil et al., 2023; Beg et al., 2024; Thieme et al., 2022)
Privacy, bias, and transparency are persistent ethical challenges	Strong	Recurrent themes in reviews and ethical analyses; unresolved in many real-world deployments	(Deckker & Sumanasekara, 2025; Zhang & Wang, 2024; Thakkar et al., 2024; Tretter, 2024; Beg et al., 2024; Graham et al., 2019; Thieme et al., 2022; Liu et al., 2024)
Interdisciplinary integration improves framework robustness	Moderate	Combining psychological, ethical, and technical principles leads to more effective and inclusive AI systems	(Deckker & Sumanasekara, 2025; Thakkar et al., 2024; Samsonovich, 2020; Zhao et al., 2022; Velagaleti et al., 2024; Tretter, 2024; Beg et al., 2024; Thieme et al., 2022; Liu et al., 2024; Kundu, 2022)
Standardized evaluation metrics and cultural inclusivity are lacking	Moderate	Reviews highlight variability in assessment and limited cross-cultural validation	(Deckker & Sumanasekara, 2025; Thakkar et al., 2024; Tretter, 2024; Beg et al., 2024; Liu et al., 2024)
Over-reliance on AI and lack of human oversight can undermine outcomes	Moderate	Empirical and design studies warn of risks without balanced human-AI collaboration	(Zhang & Wang, 2024; Tretter, 2024; Beg et al., 2024; Thieme et al., 2022)

FIGURE 4 Key claims and support evidence identified in these papers.

5. Conclusion

Ethical and practical frameworks for compensatory AI systems are increasingly robust, emphasizing human-centered design, privacy, transparency, and interdisciplinary integration. While these frameworks have improved the effectiveness and safety of AI in compensating for cognitive and emotional challenges, ongoing challenges in bias mitigation, cultural inclusivity, and standardized evaluation remain.

5.1. Research Gaps

Despite progress, research gaps persist in the standardization of evaluation metrics, cross-cultural validation, and the long-term impact of compensatory AI on user autonomy and well-being.

Research Gaps Matrix

Framework Focus	Education	Mental Health	Elder Care	Cross-Cultural	Longitudinal Impact
Human-Centered Design	7	8	4	2	1
Privacy & Bias	5	7	3	1	GAP
Adaptive Personalization	6	5	2	1	GAP
Interdisciplinary Integration	4	4	2	1	GAP

FIGURE 5 Heatmap of research coverage by framework focus and application domain.

5.2. Open Research Questions

Future research should focus on standardizing evaluation metrics, ensuring cultural inclusivity, and assessing the long-term impact of compensatory AI on user autonomy and well-being.

Question	Why
How can evaluation metrics for compensatory AI systems be standardized across domains and cultures?	Standardization is crucial for comparing effectiveness, ensuring safety, and facilitating regulatory oversight across diverse applications.
What are the long-term effects of compensatory AI on user autonomy and psychological well-being?	Understanding these effects is essential to prevent over-reliance and ensure that AI supports, rather than undermines, human agency.
How can frameworks be adapted to ensure cultural inclusivity and reduce algorithmic bias in compensatory AI?	Addressing cultural and demographic diversity is vital for equitable and effective AI deployment in global contexts.

FIGURE 6 Open research questions and their significance for future work.

In summary, while established frameworks have advanced the ethical and practical design of compensatory AI, ongoing research and interdisciplinary collaboration are needed to address persistent challenges and ensure equitable, effective, and safe AI systems for all users.

These papers were sourced and synthesized using Consensus, an AI-powered search engine for research. Try it at ***https://consensus.app***

References

Assunção, G., Patrão, B., Castelo-Branco, M., & Menezes, P. (2022). An Overview of Emotion in Artificial Intelligence. *IEEE Transactions on Artificial Intelligence*, 3, 867-886. **https://doi.org/10.1109/TAI.2022.3159614**

Deckker, D., & Sumanasekara, S. (2025). SYSTEMATIC REVIEW ON AI IN EMOTIONAL INTELLIGENCE AND PSYCHOLOGICAL EDUCATION. *EPRA International Journal of Research & Development (IJRD)*. **https://doi.org/10.36713/epra21351**

Zhang, Z., & Wang, J. (2024). Can AI replace psychotherapists? Exploring the future of mental health care. *Frontiers in Psychiatry*, 15. **https://doi.org/10.3389/fpsyt.2024.1444382**

Benita, J., Jaswanth, S., Bhuvaneshwar, N., Yuvaraj, R., & Narayana, Y. (2025). Phoenix: A Conversational Agent for Emotional Well-Being and Psychological Support. *2025 International Conference on Multi-Agent Systems for Collaborative Intelligence (ICMSCI)*, 1137-1142. **https://doi.org/10.1109/ICMSCI62561.2025.10894579**

Thakkar, A., Gupta, A., & De Sousa, A. (2024). Artificial intelligence in positive mental health: a narrative review. *Frontiers in Digital Health*, 6. **https://doi.org/10.3389/fdgth.2024.1280235**

Choi, S. (2025). R-CAGE: A Structural Model for Emotion Output Design in Human-AI Interaction. **.

Liu, Y., Zhang, H., Jiang, M., Chen, J., & Wang, M. (2024). A systematic review of research on emotional artificial intelligence in English language education. *System*. **https://doi.org/10.1016/j.system.2024.103478**

Nag, P., Bhagat, A., & Priya, V. (2025). Expanding AI's Role in Healthcare Applications: A Systematic Review of Emotional and Cognitive Analysis Techniques. *IEEE Access*, 13, 69129-69160. **https://doi.org/10.1109/ACCESS.2025.3562131**

Samsonovich, A. (2020). Socially emotional brain-inspired cognitive architecture framework for artificial intelligence. *Cognitive Systems Research*, 60, 57-76. **https://doi.org/10.1016/j.cogsys.2019.12.002**

Zhao, J., Wu, M., Zhou, L., Wang, X., & Jia, J. (2022). Cognitive psychology-based artificial intelligence review. *Frontiers in Neuroscience*, 16. **https://doi.org/10.3389/fnins.2022.1024316**

Bilquise, G., Ibrahim, S., & Shaalan, K. (2022). Emotionally Intelligent Chatbots: A Systematic Literature Review. *Human Behavior and Emerging Technologies*. **https://doi.org/10.1155/2022/9601630**

Zong, Y., & Yang, L. (2025). How AI-Enhanced Social–Emotional Learning Framework Transforms EFL Students' Engagement and Emotional Well-Being. *European Journal of Education*. **https://doi.org/10.1111/ejed.12925**

Shi, L. (2024). The Integration of Advanced AI-Enabled Emotion Detection and Adaptive Learning Systems for Improved Emotional Regulation. *Journal of Educational Computing Research*, 63, 173 - 201. **https://doi.org/10.1177/07356331241296890**

Velagaleti, S., Choukaier, D., Nuthakki, R., Lamba, V., Sharma, V., & Rahul, S. (2024). Empathetic Algorithms: The Role of AI in Understanding and Enhancing Human Emotional Intelligence. *Journal of Electrical Systems*. **https://doi.org/10.52783/jes.1806**

Moghadam, T., Darejeh, A., Delaramifar, M., & Mashayekh, S. (2023). Toward an artificial intelligence-based decision framework for developing adaptive e-learning systems to impact learners' emotions. *Interactive Learning Environments*, 32, 3665 - 3685. **https://doi.org/10.1080/10494820.2023.2188398**

Patil, S., Shinde, V., & Nemade, S. (2025). Building Empathetic AI for Mental Health Support: A Human-Centered Approach Combining Prompt Engineering, Machine Learning, and Psychological Theories. *Indian Journal of Computer Science and Technology*. **https://doi.org/10.59256/indjcst.20250401029**

Darejeh, A., Moghadam, T., Delaramifar, M., & Mashayekh, S. (2024). A Framework for AI-Powered Decision Making in Developing Adaptive e-Learning Systems to Impact Learners' Emotional Responses. *2024 11th International and the 17th National Conference on E-Learning and E-Teaching (ICeLeT)*, 1-6. **https://doi.org/10.1109/ICeLeT62507.2024.10493103**

Kallivalappil, N., D'souza, K., Deshmukh, A., Kadam, C., & Sharma, N. (2023). Empath.ai: a Context-Aware Chatbot for Emotional Detection and Support. *2023 14th International Conference on Computing Communication and Networking Technologies (ICCCNT)*, 1-7. **https://doi.org/10.1109/ICCCNT56998.2023.10306584**

Tretter, M. (2024). Equipping AI-decision-support-systems with emotional capabilities? Ethical perspectives. *Frontiers in Artificial Intelligence*, 7. **https://doi.org/10.3389/frai.2024.1398395**

Rokhsaritalemi, S., Sadeghi-Niaraki, A., & Choi, S. (2023). Exploring Emotion Analysis Using Artificial Intelligence, Geospatial Information Systems, and Extended Reality for Urban Services. *IEEE Access*, 11, 92478-92495. **https://doi.org/10.1109/ACCESS.2023.3307639**

Beg, M., Verma, M., M., V., & Verma, M. (2024). Artificial Intelligence for Psychotherapy: A Review of the Current State and Future Directions. *Indian Journal of Psychological Medicine*. **https://doi.org/10.1177/02537176241260819**

Graham, S., Depp, C., Lee, E., Nebeker, C., Tu, X., Kim, H., & Jeste, D. (2019). Artificial Intelligence for Mental Health and Mental Illnesses: an Overview. *Current Psychiatry Reports*, 21. **https://doi.org/10.1007/s11920-019-1094-0**

Zahra, S., Samra, M., & Gizawi, L. (2025). Working Toward Advanced Architectural Education: Developing an AI-Based Model to Improve Emotional Intelligence in Education. *Buildings*. **https://doi.org/10.3390/buildings15030356**

Shegekar, G., Gajbhiye, S., Bhosale, G., & Adikane, S. (2024). Review Paper on AI Chatbot for Mental Health Support. *International Journal for Research in Applied Science and Engineering Technology*. **https://doi.org/10.22214/ijraset.2024.64499**

Thieme, A., Hanratty, M., Lyons, M., Palacios, J., Marques, R., Morrison, C., & Doherty, G. (2022). Designing Human-centered AI for Mental Health: Developing Clinically Relevant Applications for Online CBT Treatment. *ACM Transactions on Computer-Human Interaction*, 30, 1 - 50. **https://doi.org/10.1145/3564752**

Liu, Y., Kauttonen, J., Zhao, B., Li, X., & Peng, W. (2024). Editorial: Towards Emotion AI to next generation healthcare and education. *Frontiers in Psychology*, 15. **https://doi.org/10.3389/fpsyg.2024.1533053**

Kundu, S. (2022). HCI-Driven Emotion-Adaptive UIs for Cognitive Efficiency: Real-Time Adaptation to Fatigue, Focus, and Emotional Expression. *International Scientific Journal of Engineering and Management*. **https://doi.org/10.55041/isjem00111**