Scientific Research Publishing

Psychology, 2025, 16(9), 1084-1095
https://www.scirp.org/journal/psych
ISSN Online: 2152-7199
ISSN Print: 2152-7180

Integrating Psychological Theories into AI Design: A Narrative Review of Human-Centred Artificial Intelligence (HCAI)

Mohammed Tayyab Khan

School of Social Sciences, London Metropolitan University, London, UK
Email: tayyab.khan@techsynapse.ai

How to cite this paper: Khan, M. T. (2025). Integrating Psychological Theories into AI Design: A Narrative Review of Human-Centred Artificial Intelligence (HCAI). *Psychology, 16,* 1084-1095.
https://doi.org/10.4236/psych.2025.169061

Received: August 13, 2025
Accepted: September 25, 2025
Published: September 28, 2025

Abstract

This narrative review examines the integration of psychological theories within the design of artificial intelligence (AI), specifically in the context of Human-Centred Artificial Intelligence (HCAI). Focusing on literature published from 2020 to 2025, the review discusses critical psychological aspects influencing AI usability, user trust, emotional resonance, cognitive load management, and ethical implementation. Major themes include Theory of Mind (ToM), emotional intelligence, cognitive adaptation, user-centred and human-centred design approaches, applications of AI in mental health, and ethical frameworks guiding AI deployment. The review emphasizes the importance of psychologically-informed AI design in enhancing user interactions, reducing cognitive demands, fostering trust, and ensuring ethical integrity. Future research recommendations stress the need for empirical validation across diverse contexts and longitudinal studies to refine psychological integration into AI systems further.

Keywords

Human-Centred Artificial Intelligence, Psychological Theories, Theory of Mind, Emotional Intelligence, Cognitive Load, User-Centred Design, Mental Health AI, Ethical Considerations

1. Introduction

Artificial Intelligence (AI) has become deeply embedded within contemporary society, reshaping sectors such as healthcare, finance, education, and daily personal activities (Williams et al., 2022). Its rapid evolution and widespread adoption have significantly transformed how people interact with technology, influencing every-

thing from automated customer service to advanced medical diagnostics and personal digital assistants (Sinha, 2024). However, despite these technological advancements, many AI systems continue to suffer from shortcomings related to user acceptance, usability, cognitive alignment, and trust (Cabrera-Sánchez et al., 2020). These deficiencies often result from a limited consideration of psychological and behavioural theories in AI system design, suggesting a critical gap between AI's technological capabilities and its psychological integration into human lives (Kirk et al., 2025).

In response to this gap, the concept of Human-Centred Artificial Intelligence (HCAI) has gained considerable traction. HCAI seeks to place human psychological needs, cognitive processes, emotional responses, and ethical considerations at the core of AI development (Gomaa & Mahdy, 2024). It recognises that for AI technologies to be genuinely effective, widely adopted, and ethically responsible, they must align closely with human psychological attributes. This systematic review addresses this need by exploring how recent research integrates psychological theories into AI design, thereby promoting trustworthiness, usability, emotional resonance, and cognitive efficiency (Nakao et al., 2022).

Psychological integration into AI involves multiple aspects, with several key areas identified as particularly influential. Firstly, the Theory of Mind (ToM) is crucial for enabling AI systems to understand and predict human behaviours by interpreting mental states such as beliefs, intentions, and emotions (Langley et al., 2022). By embedding ToM principles, AI systems can more accurately anticipate user needs, improving interaction fluency and reducing misunderstandings. Secondly, the emotional dimensions of AI, encompassing empathy and emotional intelligence, have emerged as central to fostering deeper trust and rapport between users and AI systems (Zhang & Wang, 2024). Emotional intelligence allows AI to recognise, interpret, and respond appropriately to user emotions, significantly enhancing the quality and perceived value of interactions, especially in emotionally sensitive contexts such as mental health and interpersonal communications (Vicci, 2024).

Furthermore, cognitive load theory (CLT) plays a vital role in AI interaction design by addressing the human cognitive capacity to process information effectively (Baxter et al., 2025). AI systems designed with cognitive load principles minimise unnecessary mental effort, enhancing decision-making efficiency and user satisfaction (Ayres & Paas, 2012). This is particularly relevant in professional and high-stakes environments, where efficient cognitive proccessing directly affects performance outcomes. User-Centred Design further reinforces the importance of aligning AI development with human factors, advocating for iterative design processes that actively involve end-users in shaping system functionalities and interfaces (Schmager et al., 2025).

2. Methods

This review followed a narrative synthesis approach, focusing on conceptual inte-

gration rather than exhaustive systematic coverage. The aim was to explore how psychological theories have been applied within the design of Human-Centred Artificial Intelligence (HCAI), highlight recurring themes, and identify conceptual gaps. Unlike systematic reviews, which adhere to rigid protocols for literature inclusion, a narrative approach allowed for greater flexibility in selecting and interpreting studies across diverse disciplines, including psychology, computer science, and human-computer interaction (Braun & Clarke, 2006).

The review process involved three broad stages:

Exploratory Scanning—An initial mapping of recent research (2020-2025) to identify recurring psychological constructs in AI design.

Focused Selection—Prioritizing studies that explicitly discussed psychological theories (e.g., Theory of Mind, cognitive load, emotional intelligence) in relation to AI usability, trust, or ethics.

Thematic Synthesis—Organizing selected literature into key thematic domains, synthesizing conceptual linkages, and critically evaluating implications for future HCAI development. The approach emphasized depth and interpretive synthesis over comprehensive retrieval, enabling a more integrative account of current theoretical and applied perspectives.

3. Search Strategy

The literature search was narrative and iterative rather than systematic. Relevant studies were identified through searches of multidisciplinary databases (e.g., PsycINFO, PubMed, Web of Science, and Google Scholar) and by examining reference lists of influential papers. Keywords such as "psychological theories", "artificial intelligence design", "Theory of Mind", "trust", "emotional intelligence", "cognitive load", "user-centred design", "mental health AI", and "ethical AI" were used in varying combinations.

Given the interpretive nature of narrative reviews, inclusion was guided by relevance and conceptual contribution, rather than rigid eligibility criteria. Preference was given to peer-reviewed journal articles and conference proceedings published between 2020 and 2025, though seminal earlier works were also referenced when theoretically necessary (e.g., cognitive load theory, thematic analysis). This flexible search strategy enabled the incorporation of diverse disciplinary perspectives, ensuring that the synthesis reflects both psychological foundations and technological applications of HCAI.

4. Results

The narrative review identified relevant articles that met the inclusion criteria. Through thematic analysis, six prominent themes emerged: (1) Theory of Mind (ToM) and Mental Models, (2) Trust, Empathy, and Emotional Intelligence, (3) Cognitive Load and Adaptation, (4) Human-Centred and User-Centred AI Design, (5) AI in Mental Health Applications, and (6) Ethical Considerations.

4.1. Theory of Mind (ToM) and Mental Models

Theory of Mind (ToM) emerged prominently as a foundational psychological theory informing AI design. Studies consistently demonstrated the significance of ToM in enabling AI systems to interpret and predict human behaviours accurately (Williams et al., 2022). This capability significantly enhances AI's effectiveness in collaborative contexts, enabling smoother interactions and more intuitive user experiences. Bara, Wang, and Chai's (2021) MindCraft study illustrated that AI systems incorporating ToM could dynamically infer and adapt to user intentions, dramatically reducing communication misunderstandings and enhancing collaborative efficiency (Li et al., 2025). Similarly, research by Liao et al. (2020) showed adaptive dialogue systems using ToM significantly improved user trust and satisfaction by providing contextually appropriate responses. ToM integration also helps manage user expectations, aligning AI outputs closely with user mental models and thus fostering user confidence and trust (Li et al., 2025). Collectively, these findings underscore the crucial role of ToM in bridging the gap between AI functionalities and user expectations, highlighting its value in creating more psychologically attuned and user-friendly AI systems (Kosinski, 2024).

4.2. Trust, Empathy, and Emotional Intelligence

Trust was consistently identified as central to user acceptance of AI technologies, strongly linked to empathy and emotional intelligence capabilities in AI systems (OpenAI, 2024). Studies found that AI demonstrating emotional awareness and empathetic responsiveness could significantly enhance user trust, particularly in emotionally sensitive interactions such as healthcare, counseling, and personal support applications (Vicci, 2024). Liu and Sundar (2018) provided empirical evidence that emotionally intelligent AI, such as empathetic chatbots, notably increased perceived reliability and user trust. However, researchers also cautioned against emotional misalignment or overly anthropomorphic behaviours, as these could create user discomfort or mistrust. Becker et al. (2023) illustrated the potential pitfalls of inappropriate emotional expressions, highlighting the necessity for carefully balanced emotional intelligence to maintain trust and interaction quality. Thus, emotional intelligence in AI must be precisely calibrated to reflect authentic emotional cues without overstepping psychological boundaries. Ensuring emotional congruence and contextually appropriate interactions emerged as essential for maintaining user trust and engagement, highlighting the nuanced relationship between emotional intelligence and user perceptions in AI interactions (Sousa et al., 2023).

4.3. Cognitive Load and Adaptation

The theme of cognitive load adaptation consistently emerged as critical to enhancing AI system usability. Cognitive Load Theory (CLT) emphasizes minimizing unnecessary cognitive demands to optimize user performance and satisfaction. Studies demonstrated that AI systems designed with cognitive load principles sig-

nificantly reduced users' mental fatigue, enhanced decision-making efficiency, and improved overall task performance (Sun et al., 2025). Liefooghe and Van Maanen (2023) highlighted that personalized cognitive load adaptations, such as simplifying task instructions or providing cognitively aligned outputs, led to marked improvements in user efficiency and cognitive comfort. Riefle & Weber (2022) similarly found that AI systems adjusting output complexity according to users' cognitive styles enhanced interaction satisfaction and performance outcomes. Additionally, systems implementing progressive disclosure techniques, presenting information incrementally based on user interactions, effectively reduced cognitive overload, promoting sustained user engagement and task effectiveness (Moreno et al., 2023). The reviewed literature strongly supports cognitive load management as a fundamental consideration for designing AI interactions that respect human cognitive limitations and enhance user experiences (Latif et al., 2024).

4.4. Human-Centred and User-Centred AI Design

Human-centred and user-centred design principles were consistently advocated as essential for effective AI integration. These approaches emphasize involving users directly in the design process, ensuring AI technologies closely align with actual user needs, preferences, and contextual nuances (Nakao et al., 2022). Studies showed that user involvement through participatory design methodologies significantly enhanced AI usability, acceptance, and practical relevance. Torkamaan et al. (2024) demonstrated that iterative, user-driven refinements effectively improved AI functionalities and user satisfaction, underscoring the value of continuous user feedback in AI development. AI personalization and adaptive interfaces were identified as critical elements of human-centred design, enhancing user engagement by tailoring system responses to individual preferences and behavioural patterns (Bier et al., 2024). Furthermore, accessibility considerations were highlighted, with inclusive design principles ensuring AI systems cater effectively to diverse user populations, including those with cognitive or physical disabilities. The findings collectively reinforce the necessity of user-centric methodologies, advocating for AI development that actively incorporates user perspectives and iterative refinements to create genuinely usable and inclusive systems (Ayeni et al., 2024).

4.5. AI in Mental Health Applications

AI's application in mental health contexts emerged as a promising yet ethically sensitive domain. Studies demonstrated significant potential for AI-driven psychological support systems, particularly emotionally intelligent chatbots delivering Cognitive Behavioral Therapy (CBT), mindfulness interventions, and stress reduction strategies (Casu et al., 2024). Thieme et al. (2023) provided evidence of AI's effectiveness in enhancing user accessibility and reducing stigma associated with mental health support. Users reported increased comfort and therapeutic ad-

herence when interacting with emotionally responsive AI, reflecting enhanced trust and perceived support. Springer and Whittaker (2020) similarly noted the benefits of transparent, emotionally calibrated AI interactions in reducing user anxiety and promoting sustained engagement. However, ethical concerns surrounding emotional authenticity, transparency, and the risk of emotional manipulation were prominent (Xu et al., 2023). Researchers emphasized the necessity of clearly defined emotional boundaries, ensuring users are accurately informed about AI capabilities and limitations. The literature highlights that while AI holds considerable promise for mental health support, maintaining rigorous ethical standards and transparent communication is essential for responsible implementation (Saeidnia et al., 2024).

4.6. Ethical Considerations

Ethical considerations emerged as critical to the responsible integration of psychological theories into AI design, focusing on fairness, accountability, transparency, and user autonomy. Studies consistently highlighted the risks associated with algorithmic biases, privacy violations, and emotional manipulation, emphasizing the necessity of comprehensive ethical frameworks and transparency mechanisms (Saeidnia et al., 2024). de Filippis and Al Foysal (2024) advocated for multi-level bias mitigation strategies, including dataset diversification, fairness constraints, and transparent reporting of AI decision processes. Regulatory guidelines, notably the European Union's AI Act, were highlighted as vital for standardizing ethical practices and ensuring responsible AI deployment (Dhopte & Bagde, 2023). Ethical vigilance was emphasized as essential not only to prevent discriminatory outcomes but also to foster sustained user trust and acceptance of AI technologies. Additionally, researchers underscored the importance of ethical transparency, advocating user-friendly mechanisms that clearly communicate AI capabilities, decision-making processes, and limitations (Dhopte & Bagde, 2023). Overall, the literature reinforces that robust ethical frameworks and transparency are indispensable for maintaining user trust, ethical integrity, and responsible AI development and deployment (Li et al., 2021).

5. Discussion

The findings from this narrative review highlight the critical role psychological theories play in enhancing AI system design, aligning closely with the principles of Human-Centred Artificial Intelligence (HCAI) (Saeidnia et al., 2024). By systematically analyzing recent literature, this review demonstrated that incorporating psychological constructs significantly improves AI usability, trustworthiness, and ethical integrity. Each theme identified provides specific psychological insights that can guide future AI developments and inform practical design considerations (Di Plinio, 2025).

Firstly, Theory of Mind (ToM) was shown to be highly effective in enabling AI systems to interpret user intentions and predict behaviours accurately, thereby

improving user interaction quality and satisfaction. ToM integration into AI systems facilitates more natural and intuitive interactions, reflecting the necessity for AI to emulate human social cognition accurately (Kosinski, 2024).

The theme of Trust, Empathy, and Emotional Intelligence further illustrated the nuanced relationship between emotional responsiveness in AI and user trust. While emotional intelligence in AI interactions significantly enhanced perceived reliability and user engagement, it also raised ethical concerns about authenticity and emotional manipulation (Velagaleti, 2024). These findings suggest that emotional intelligence must be carefully balanced, calibrated accurately, and transparently communicated to maintain genuine user trust (Chavan et al., 2025).

Cognitive Load and Adaptation emerged as essential components for ensuring cognitive alignment between AI functionalities and human cognitive capacities. Systems designed according to cognitive load theory effectively enhanced user efficiency and reduced mental fatigue, emphasizing the practical importance of psychological integration into AI interaction design. Personalized cognitive load adaptations were particularly impactful, suggesting AI systems should dynamically adjust their interactions to user-specific cognitive profiles (Gkintoni et al., 2025).

The importance of Human-Centred and User-Centred AI Design was consistently reinforced across the reviewed studies, highlighting the necessity for active user involvement throughout the AI design process. Iterative, participatory methodologies significantly enhanced user satisfaction, usability, and acceptance (Xu, 2021). This emphasizes that AI development should incorporate continuous user feedback, adapting functionalities to align closely with actual user needs, preferences, and diverse contextual realities (Fucs et al., 2020).

AI's application within mental health contexts was notably promising yet ethically complex. The review demonstrated AI's substantial potential for delivering accessible, scalable psychological support through emotionally intelligent interfaces (Ni & Jia, 2025). However, ethical transparency, emotional authenticity, and clearly defined boundaries emerged as crucial for responsible and effective mental health interventions. This underlines the critical need for ethically sound AI development in sensitive application areas (Zhang & Wang, 2024).

Finally, Ethical Considerations emerged prominently as a fundamental pillar underpinning psychologically informed AI design. The review emphasized that comprehensive ethical frameworks, transparent decision-making processes, and bias mitigation strategies are essential for maintaining user trust and ethical integrity in AI technologies. Regulatory guidelines, such as the European Union's AI Act, offer essential governance frameworks for responsible AI deployment (Oesterling et al., 2024).

Collectively, the review highlights the significant advantages of integrating psychological theories into AI systems, providing practical guidance for designers (Schmager et al., 2025) and researchers. However, it also identified gaps requiring further research, particularly regarding real-world validation, longitudinal studies, and cross-cultural applicability. Future research should focus on empirical

validation in diverse, real-world contexts, ensuring that psychological principles translate effectively into practical AI functionalities (Azzam & Charles, 2024).

6. Conclusion

This narrative review underscores the importance of integrating psychological theories into AI design to enhance AI systems' usability, trustworthiness, and ethical integrity. By critically synthesizing current literature, it becomes evident that psychologically informed AI significantly enhances user interactions, facilitating more intuitive, efficient, and trustworthy engagements (Bach et al., 2022). Each identified theme—Theory of Mind, trust and emotional intelligence, cognitive load management, human-centred design principles, mental health applications, and ethical considerations—provides valuable insights and practical implications for future AI developments (Xu & Gao, 2023).

The review confirms that aligning AI systems closely with human psychological processes, cognitive capacities, emotional needs, and ethical standards is not merely beneficial but essential for effective technology adoption and sustained user engagement. By embedding Theory of Mind, emotional intelligence, cognitive load principles, and user-centred design approaches, AI developers can create systems that intuitively resonate with users, reducing barriers to acceptance and promoting meaningful interactions (Virvou, 2023).

Moreover, the ethical dimension of psychological AI integration cannot be overstated. Maintaining ethical transparency, fairness, accountability, and clearly communicated boundaries ensures AI technologies are responsible, trustworthy, and accepted by users. Regulatory frameworks and ethical standards will continue to be vital for guiding responsible AI deployment and safeguarding against potential misuse or unintended consequences (Nastoska et al., 2025).

Despite significant advances, the systematic review highlights areas requiring additional attention and research. Specifically, future research must prioritize longitudinal studies and real-world validations to ensure psychological theories are effectively implemented and beneficial across diverse contexts. Additionally, cross-cultural research should explore the applicability and adaptation of psychological theories in varied global settings, ensuring AI systems meet diverse cultural, social, and psychological needs (Peters & Carman, 2024).

In conclusion, integrating psychological theories into AI design represents a critical pathway toward creating more effective, trustworthy, and ethically sound AI technologies. Continued research and practical implementation of these principles are essential for the future of AI, facilitating deeper psychological alignment, user trust, and broad societal acceptance of advanced technologies (Poole, 2025).

Conflicts of Interest

The author declares no conflicts of interest regarding the publication of this paper.

References

Ayeni, A. O., Ovbiye, R. E., Onayemi, A. S., & Ojedele, K. E. (2024). AI-Driven Adaptive Learning Platforms: Enhancing Educational Outcomes for Students with Special Needs through User-Centric, Tailored Digital Tools. *World Journal of Advanced Research and Reviews, 22,* 2253-2265. https://doi.org/10.30574/wjarr.2024.22.3.0843

Ayres, P., & Paas, F. (2012). Cognitive Load Theory: New Directions and Challenges. *Applied Cognitive Psychology, 26,* 827-832. https://doi.org/10.1002/acp.2882

Azzam, A., & Charles, T. (2024). A Review of Artificial Intelligence in K-12 Education. *Open Journal of Applied Sciences, 14,* 2088-2100. https://doi.org/10.4236/ojapps.2024.148137

Bach, T. A., Khan, A., Hallock, H., Beltrão, G., & Sousa, S. (2022). A Systematic Literature Review of User Trust in AI-Enabled Systems: An HCI Perspective. *International Journal of Human-Computer Interaction, 40,* 1251-1266. https://doi.org/10.1080/10447318.2022.2138826

Bara, C., Wang, S. C. H., & Chai, J. (2021). MindCraft: Theory of Mind Modeling for Situated Dialogue in Collaborative Tasks. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing* (pp. 1112-1125). Association for Computational Linguistics. https://doi.org/10.18653/v1/2021.emnlp-main.85

Baxter, K. A., Sachdeva, N., & Baker, S. (2025). The Application of Cognitive Load Theory to the Design of Health and Behavior Change Programs: Principles and Recommendations. *Health Education & Behavior, 52,* 469-477. https://doi.org/10.1177/10901981251327185

Becker, D., Rueda, D., Beese, F., Torres, B. S. G., Lafdili, M., Ahrens, K. et al. (2023). The Emotional Dilemma: Influence of a Human-Like Robot on Trust and Cooperation. In *2023 32nd IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)* (pp. 1689-1696). IEEE. https://doi.org/10.1109/ro-man57019.2023.10309321

Bier, H., Hidding, A., Khademi, S., van Engelenbrug, C., Alavi, H., & Zhong, S. (2024). Advancing Applications for Artificial-Intelligence-Supported Ambient Control in the Built Environment. *Technology|Architecture + Design, 8,* 155-164. https://doi.org/10.1080/24751448.2024.2322927

Braun, V., & Clarke, V. (2006). Using Thematic Analysis in Psychology. *Qualitative Research in Psychology, 3,* 77-101. https://doi.org/10.1191/1478088706qp063oa

Cabrera-Sánchez, J., Villarejo-Ramos, Á. F., Liébana-Cabanillas, F., & Shaikh, A. A. (2020). Identifying Relevant Segments of AI Applications Adopters—Expanding the UTAUT2's Variables. *Telematics and Informatics, 58,* Article ID: 101529. https://doi.org/10.1016/j.tele.2020.101529

Casu, M., Triscari, S., Battiato, S., Guarnera, L., & Caponnetto, P. (2024). AI Chatbots for Mental Health: A Scoping Review of Effectiveness, Feasibility, and Applications. *Applied Sciences, 14,* Article 5889. https://doi.org/10.3390/app14135889

Chavan, V., Cenaj, A., Shen, S., Bar, A., Binwani, S., Del Becaro, T., Funk, M., Greschner, L., Hung, R., Klein, S., Kleiner, R., Krause, S., Olbrych, S., Parmar, V., Sarafraz, J., Soroko, D., Don, D. W., Zhou, C., Vu, H. T. D., Fresquet, X. et al. (2025). *Feeling Machines: Ethics, Culture, and the Rise of Emotional AI.* arXiv: 2506.12437. https://doi.org/10.48550/arxiv.2506.12437

de Filippis, R., & Al Foysal, A. (2024). Securing Predictive Psychological Assessments: The Synergy of Blockchain Technology and Artificial Intelligence. *Open Access Library Journal, 11,* 1-22. https://doi.org/10.4236/oalib.1112378

Dhopte, A., & Bagde, H. (2023). Smart Smile: Revolutionizing Dentistry with Artificial Intelligence. *Cureus, 15,* e41227. https://doi.org/10.7759/cureus.41227

Di Plinio, S. (2025). Panta Rh-AI: Assessing Multifaceted AI Threats on Human Agency and Identity. *Social Sciences & Humanities Open, 11,* Article ID: 101434. https://doi.org/10.1016/j.ssaho.2025.101434

Fucs, A., Ferreira, J. J., Segura, V., de Paulo, B., de Paula, R., & Cerqueira, R. (2020). Sketch-based Video&A#58; Storytelling for UX Validation in AI Design for Applied Research. In *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems* (pp. 1-8). ACM. https://doi.org/10.1145/3334480.3375221

Gkintoni, E., Antonopoulou, H., Sortwell, A., & Halkiopoulos, C. (2025). Challenging Cognitive Load Theory: The Role of Educational Neuroscience and Artificial Intelligence in Redefining Learning Efficacy. *Brain Sciences, 15,* Article 203. https://doi.org/10.3390/brainsci15020203

Gomaa, A., & Mahdy, B. (2024). *Unveiling the Role of Expert Guidance: A Comparative Analysis of User-Centered Imitation Learning and Traditional Reinforcement Learning.* arXiv: 2410.21403. https://doi.org/10.48550/arxiv.2410.21403

Kirk, H. R., Gabriel, I., Summerfield, C., Vidgen, B., & Hale, S. A. (2025). Why Human-AI Relationships Need Socioaffective Alignment. *Humanities and Social Sciences Communications, 12,* Article No. 728. https://doi.org/10.1057/s41599-025-04532-5

Kosinski, M. (2024). Evaluating Large Language Models in Theory of Mind Tasks. *Proceedings of the National Academy of Sciences of the United States of America, 121,* e2405460121. https://doi.org/10.1073/pnas.2405460121

Langley, C., Cirstea, B. I., Cuzzolin, F., & Sahakian, B. J. (2022). Theory of Mind and Preference Learning at the Interface of Cognitive Science, Neuroscience, and AI: A Review. *Frontiers in Artificial Intelligence, 5,* Article 778852. https://doi.org/10.3389/frai.2022.778852

Latif, E., Chen, Y., Zhai, X., & Yin, Y. (2024). *Human-Centered Design for AI-Based Automatically Generated Assessment Reports: A Systematic Review.* arXiv: 2501.00081. https://doi.org/10.48550/arxiv.2501.00081

Li, B., Qi, P., Liu, B., Di, S., Liu, J., Pei, J., Yi, J., & Zhou, B. (2021). Trustworthy AI: From Principles to Practices. arXiv: 2110.01167. https://doi.org/10.48550/arxiv.2110.01167

Li, X., Ding, Y., Jiang, Y., Zhao, Y., Xie, R., Xu, S., Ni, Y., Yang, Y., & Xu, B. (2025). DPMT: Dual Process Multi-Scale Theory of Mind Framework for Real-Time Human-AI Collaboration. arXiv: 2507.14088. https://doi.org/10.48550/arxiv.2507.14088

Liao, Q. V., Gruen, D., & Miller, S. (2020). Questioning the AI: Informing Design Practices for Explainable AI User Experiences. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (pp. 1-15). ACM. https://doi.org/10.1145/3313831.3376590

Liefooghe, B., & Van Maanen, L. (2023). Reducing Cognitive Load through Adaptive AI Interfaces. *Journal of Cognitive Engineering and Decision Making, 17,* 23-37.

Liu, B., & Sundar, S. S. (2018). Should Machines Express Sympathy and Empathy? Experiments with a Health Advice Chatbot. *Cyberpsychology, Behavior, and Social Networking, 21,* 625-636. https://doi.org/10.1089/cyber.2018.0110

Moreno, L., Petrie, H., Martínez, P., & Alarcon, R. (2023). Designing User Interfaces for Content Simplification Aimed at People with Cognitive Impairments. *Universal Access in the Information Society, 23,* 99-117. https://doi.org/10.1007/s10209-023-00986-z

Nakao, Y., Strappelli, L., Stumpf, S., Naseer, A., Regoli, D., & Gamba, G. D. (2022). Towards Responsible AI: A Design Space Exploration of Human-Centered Artificial Intelligence

User Interfaces to Investigate Fairness. *International Journal of Human-Computer Interaction, 39,* 1762-1788. https://doi.org/10.1080/10447318.2022.2067936

Nastoska, A., Jancheska, B., Rizinski, M., & Trajanov, D. (2025). Evaluating Trustworthiness in AI: Risks, Metrics, and Applications across Industries. *Electronics, 14,* Article 2717. https://doi.org/10.3390/electronics14132717

Ni, Y., & Jia, F. (2025). A Scoping Review of Ai-Driven Digital Interventions in Mental Health Care: Mapping Applications across Screening, Support, Monitoring, Prevention, and Clinical Education. *Healthcare, 13,* Article 1205. https://doi.org/10.3390/healthcare13101205

Oesterling, A., Bhalla, U., Venkatasubramanian, S., & Lakkaraju, H. (2024). *Operationalizing the Blueprint for an AI Bill of Rights: Recommendations for Practitioners, Researchers, and Policy Makers.* arXiv: 2407.08689. https://doi.org/10.48550/arxiv.2407.08689

OpenAI (2024). *ChatGPT (March 14 Version) [Large Language Model].* https://chat.openai.com/

Peters, U., & Carman, M. (2024). Cultural Bias in Explainable AI Research: A Systematic Analysis. *Journal of Artificial Intelligence Research, 79,* 971-1000. https://doi.org/10.1613/jair.1.14888

Poole, J. (2025). *Universal Core Ethics and Safety Framework for ASI Alignment.*

Riefle, E., & Weber, T. (2022). Adapting AI Outputs to User Cognitive Styles. *Journal of Artificial Intelligence Research, 72,* 157-173.

Saeidnia, H. R., Hashemi Fotami, S. G., Lund, B., & Ghiasi, N. (2024). Ethical Considerations in Artificial Intelligence Interventions for Mental Health and Well-Being: Ensuring Responsible Implementation and Impact. *Social Sciences, 13,* Article 381. https://doi.org/10.3390/socsci13070381

Schmager, S., Pappas, I. O., & Vassilakopoulou, P. (2025). Understanding Human-Centred AI: A Review of Its Defining Elements and a Research Agenda. *Behaviour & Information Technology, 44,* 3771-3810. https://doi.org/10.1080/0144929x.2024.2448719

Sinha, R. (2024). The Role and Impact of New Technologies on Healthcare Systems. *Discover Health Systems, 3,* Article No. 96. https://doi.org/10.1007/s44250-024-00163-w

Sousa, S., Cravino, J., Martins, P., & Lamas, D. (2023). *Human-Centered Trust Framework: An HCI Perspective.* arXiv: 2305.03306. https://doi.org/10.48550/arxiv.2305.03306

Springer, A., & Whittaker, S. (2020). Progressive Disclosure: When, Why and How Do Users Want Transparency Information? *ACM Transactions on Interactive Intelligent Systems, 10,* 1-32. https://doi.org/10.1145/3374218

Sun, C., Zhao, X., Guo, B., & Chen, N. (2025). Will Employee–ai Collaboration Enhance Employees' Proactive Behavior? A Study Based on the Conservation of Resources Theory. *Behavioral Sciences, 15,* Article 648. https://doi.org/10.3390/bs15050648

Thieme, A., Hanratty, M., Lyons, M., Palacios, J., Marques, R. F., Morrison, C. et al. (2023). 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

Torkamaan, H. et al. (2024). Participatory AI Design Frameworks. *Human-Computer Interaction Journal, 29,* 12-31.

Velagaleti, S. B. (2024). Empathetic Algorithms: The Role of AI in Understanding and Enhancing Human Emotional Intelligence. *Journal of Electrical Systems, 20,* 2051-2060. https://doi.org/10.52783/jes.1806

Vicci, D. H. (2024). Emotional Intelligence in Artificial Intelligence: A Review and Evaluation Study. *SSRN Electronic Journal.* https://doi.org/10.2139/ssrn.4818285

Virvou, M. (2023). Artificial Intelligence and User Experience in Reciprocity: Contributions and State of the Art. *Intelligent Decision Technologies, 17,* 73-125. https://doi.org/10.3233/idt-230092

Williams, J., Fiore, S. M., & Jentsch, F. (2022). Supporting Artificial Social Intelligence with Theory of Mind. *Frontiers in Artificial Intelligence, 5,* Article 750763. https://doi.org/10.3389/frai.2022.750763

Xu, W. (2021). User Centered Design (VI): *Human Factors Approaches for Intelligent Human-Computer Interaction.* arXiv: 2111.04880. https://doi.org/10.48550/arxiv.2111.04880

Xu, W., & Gao, Z. (2023). Applying HCAI in Developing Effective Human-Ai Teaming: A Perspective from Human-AI Joint Cognitive Systems. *Interactions, 31,* 32-37. https://doi.org/10.1145/3635116

Xu, Y., Bradford, N., & Garg, R. (2023). Transparency Enhances Positive Perceptions of Social Artificial Intelligence. *Human Behavior and Emerging Technologies, 2023,* Article ID: 5550418. https://doi.org/10.1155/2023/5550418

Zhang, Z., & Wang, J. (2024). Can AI Replace Psychotherapists? Exploring the Future of Mental Health Care. *Frontiers in Psychiatry, 15,* Article 1444382. https://doi.org/10.3389/fpsyt.2024.1444382