Gowri Gireesh, 2026, 14:4
ISSN (Online): 2348-4098
ISSN (Print): 2395-4752

International Journal of Science, Engineering and Technology
An Open Access Journal

Human Psychology in the Era of Artificial Intelligence: Human–AI Interaction, Trust, Emotion, and Cognitive Behavior

Gowri Gireesh[1], Dr Abhilash S.Vasu[2]

[1]Department of Psychology, Fatima Mata National College, Kollam, Kerala, India Mail id:gowrigireesh2004@gmail.com (Corresponding author)

[2]Department of Electronics and Communication Engineering, NICHE, Kumaracoil, Thuckalay, Kanyakumari District, Tamil Nadu, India

Abstract- Artificial intelligence (AI) has moved from task-oriented automation toward systems that communicate with, advise, and collaborate with humans. Consequently, psychological factors such as trust, emotion, cognitive bias, perceived intelligence, anthropomorphism, and decision confidence have become central to AI adoption. This paper presents a human-centered framework connecting psychological characteristics with AI system characteristics and interaction context. A focused literature synthesis is used to examine emotion-aware AI, human–AI trust, explainability, cognitive behavior, and the emerging influence of generative AI. The paper identifies research gaps in longitudinal assessment, culturally diverse datasets, multimodal psychological measurement, and calibration of appropriate—not excessive—trust. A proposed methodology combines questionnaire-based psychological measures, controlled human–AI experiments, behavioral indicators, and machine-learning analysis. The framework can support the design of trustworthy, transparent, adaptive, and psychologically compatible AI systems for education, healthcare, workplaces, and consumer applications.

Keywords— artificial intelligence, human psychology, human–AI interaction, trust, emotion recognition, cognitive bias, explainable AI, generative AI.

I. INTRODUCTION

Artificial intelligence is increasingly embedded in everyday decision-making, communication, education, employment, and digital services. As AI systems become conversational and adaptive, the quality of interaction is no longer determined only by computational accuracy. Users form expectations about an AI system's competence, reliability, intention, warmth, and social presence. Recent psychological research emphasizes that trust in AI depends on the trustor, the AI system, and the interaction context [1], while systematic reviews identify capability, anthropomorphism, individual factors, and explainability among important predictors of trust [2].

The psychological dimension is particularly important for generative AI because conversational systems can produce fluent and socially convincing responses even when their underlying reasoning is uncertain. Excessive trust may result in automation bias and uncritical acceptance, whereas insufficient trust can cause useful systems to be ignored. Therefore, the objective should be appropriate trust calibrated to system capability, uncertainty, and task risk.

This paper develops a conceptual human-psychology–AI framework and discusses four connected dimensions: emotional response, trust and acceptance, cognitive behavior, and interaction quality. It further proposes a research methodology that can experimentally evaluate how AI characteristics influence psychological outcomes.

II. HUMAN PSYCHOLOGY AND AI: CONCEPTUAL FOUNDATION

Human psychology provides models for understanding perception, attention, learning, memory, motivation, emotion, social cognition, and

decision-making. AI provides computational mechanisms for learning patterns from data, generating predictions, recognizing signals, and producing adaptive responses. Their intersection creates human-centered AI, in which psychological characteristics become design variables rather than secondary considerations.

The proposed framework considers three layers. The first is the user layer, containing prior experience, personality-related tendencies, digital literacy, cognitive load, expectations, and risk perception. The second is the AI layer, containing accuracy, response quality, explainability, anthropomorphism, uncertainty communication, and adaptiveness. The third is the interaction-context layer, containing task criticality, time pressure, social setting, privacy expectations, and consequences of error. These layers jointly influence trust, emotion, reliance, satisfaction, and decision quality.

Table. 1. Proposed human psychology–AI conceptual framework

User Psychology	Ai Characteristics	Interaction Context	Psychological Response	Outcome
Expectations Cognitive load Experience	Accuracy Explainability Anthropomorphism	Task risk Time pressure Privacy	Trust Emotion Perceived competence	Acceptance Reliance Decision quality
Input factors	System factors	Situational factors	Mediators	Measured outcomes

III. PSYCHOLOGICAL DIMENSIONS OF HUMAN–AI INTERACTION

1. Trust and Appropriate Reliance

Trust is a central determinant of whether people accept and rely on AI. Research distinguishes human-related, AI-related, and context-related antecedents of trust [1], while systematic reviews show that trust outcomes include attitudes, usefulness perceptions, behavioral intention, and acceptance [2], [3]. Explainability and uncertainty communication can help users judge when AI should be followed and when human verification is necessary.

The key design challenge is not maximizing trust. Instead, AI should promote calibrated trust: high confidence when evidence supports reliability and reduced reliance when the system is uncertain or operating outside its competence. NIST's AI Risk Management Framework emphasizes validity, safety, security, accountability, transparency, explainability, privacy, and fairness as characteristics of trustworthy AI [4].

2. Emotion and Affective Computing

AI can estimate emotional states from text, speech, facial behavior, physiological signals, or multimodal combinations. Emotion-aware interaction may improve personalization and engagement, but emotion inference is probabilistic and can be affected by culture, context, individual differences, and noisy observations. Therefore, emotion recognition should be treated as an uncertain inference rather than a direct measurement of a person's internal state.

Generative AI is also changing the psychological relationship between users and machines. Recent reviews describe potential benefits in psychological research and personalized support while highlighting privacy, bias, and depersonalization concerns [5]. Emerging research specifically examines trust in AI-generated emotional support, demonstrating the need for careful theoretical and empirical study of these interactions [6].

3. Cognitive Bias and Decision-Making

Human users may display automation bias, anchoring, confirmation bias, overconfidence, and algorithm aversion during AI-assisted decisions. A highly fluent AI response can increase perceived competence even when factual reliability is limited. Conversely, unexplained errors can cause users to reject an otherwise useful system. Interface design

should therefore expose evidence, uncertainty, alternative options, and appropriate verification pathways rather than presenting AI output as an unquestionable answer.

4. Anthropomorphism and Social Presence

Human-like names, voices, communication styles, and conversational behavior can increase perceived social presence. Recent research indicates that anthropomorphic characteristics, particularly voice and communication style, can influence perceived human-likeness and trust [7]. However, human-like design can also encourage users to attribute intentions, emotions, or competence that the system does not actually possess. Designers should therefore balance natural interaction with clear disclosure that the user is interacting with an AI system.

IV. COMPARATIVE ANALYSIS OF AI AND PSYCHOLOGICAL FACTORS

AI capability	Psychological factor	Potential benefit	Potential risk
Explainable AI	Trust, perceived control	Improved understanding and calibrated reliance	Information overload
Emotion-aware AI	Engagement, affective response	Personalized interaction	Misclassification, privacy concerns
Generative AI	Cognitive support, creativity	Faster ideation and assistance	Over-reliance, misinformation
Anthropomorphic AI	Social presence, rapport	Natural communication	Excessive trust or attachment
Adaptive AI	Personalization, satisfaction	Context-aware assistance	Manipulation, loss of autonomy
Decision-support AI	Confidence, reliance	Improved consistency	Automation bias

V. RESEARCH GAPS

Despite rapid progress, several gaps remain. First, much of the trust literature relies on cross-sectional self-report studies; recent systematic work on AI chatbots highlights the shortage of longitudinal research [3]. Second, psychological responses can vary with culture, age, expertise, task type, and prior AI experience, yet many datasets do not adequately represent this diversity. Third, studies frequently examine trust, emotion, or decision quality separately, whereas real human–AI interaction is multimodal and dynamic. Fourth, the relationship between generative AI use and long-term cognitive behavior remains insufficiently established. Finally, standardized methods are needed to measure appropriate trust, cognitive load, emotional response, and reliance together.

VI. PROPOSED RESEARCH METHODOLOGY

A controlled human–AI interaction experiment is proposed. Participants can be divided into experimental conditions using: (i) conventional AI responses, (ii) explainable AI responses, and (iii) uncertainty-aware explainable AI responses. Each participant completes standardized decision tasks with varying levels of task difficulty and risk. Psychological variables can be measured before, during, and after interaction.

Independent variables include explainability, anthropomorphic communication, uncertainty disclosure, and task risk. Dependent variables

include trust, perceived competence, emotional response, cognitive load, reliance, satisfaction, and decision accuracy. Trust and acceptance can be measured using validated questionnaires, while behavioral indicators can include response time, frequency of AI acceptance, correction rate, and verification behavior. Optional physiological measures such as heart-rate variability or electrodermal activity can provide additional affective indicators.

The analysis can combine descriptive statistics, reliability testing, correlation analysis, ANOVA or mixed-effects models, and machine-learning methods such as random forests for identifying influential psychological and AI-related features. A final calibrated-trust model can estimate the conditions under which users should accept, verify, or reject AI recommendations.

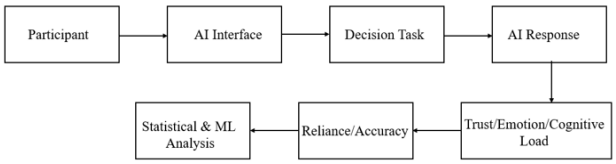

Fig. 1. Proposed experimental architecture

VII. ETHICAL AND PRIVACY CONSIDERATIONS

Psychology-oriented AI requires stronger safeguards because systems may infer sensitive information from user behavior. Data minimization, informed consent, secure storage, anonymization, transparent model limitations, and human oversight should be incorporated into research and deployment. AI should not claim certainty about emotions or psychological states when the evidence is ambiguous. The NIST AI RMF and its generative-AI profile provide useful risk-management principles for trustworthy development [4], [8].

Future Directions

Future research should investigate longitudinal human–AI relationships, culturally adaptive interaction, multimodal emotion inference, cognitive-load-aware interfaces, and personalized explainability. Another promising direction is the development of psychologically adaptive AI that changes explanation depth, interaction style, and uncertainty communication according to user expertise and task risk. Large-scale studies combining behavioral, linguistic, physiological, and interaction-log data could establish more robust models of human–AI co-adaptation.

VIII. CONCLUSION

Human psychology is becoming a fundamental component of AI design because AI systems increasingly participate in human communication and decision-making. Trust, emotion, cognitive bias, anthropomorphism, and perceived competence strongly influence how users interpret and rely on AI. This paper proposed a human-centered framework connecting user psychology, AI characteristics, and interaction context, together with a methodology for experimentally measuring psychological and behavioral outcomes. Future AI systems should not merely be accurate; they should also communicate uncertainty, support calibrated trust, protect privacy, and preserve human autonomy. Integrating psychological science with AI engineering can therefore contribute to safer, more acceptable, and more effective human–AI collaboration.

REFERENCES

1. Y. Li, B. Wu, Y. Huang, and S. Luan, "Developing trustworthy artificial intelligence: insights from research on interpersonal, human-automation, and human-AI trust," Frontiers in Psychology, vol. 15, 2024, doi: 10.3389/fpsyg.2024.1382693.
2. Q. Dang and G. Li, "Antecedents and Consequences of Human Trust in AI: Evidence from a Systematic Review," Academy of Management Proceedings, 2025, doi: 10.5465/AMPROC.2025.15921.
3. S. W. T. Ng and R. Zhang, "Trust in AI chatbots: A systematic review," Telematics and Informatics, vol. 97, 2025, Art. no. 102240, doi: 10.1016/j.tele.2025.102240.
4. E. Tabassi, "Artificial Intelligence Risk Management Framework (AI RMF 1.0)," NIST AI 100-1, National Institute of Standards and Technology, 2023, doi: 10.6028/NIST.AI.100-1.

5. M. Salah, F. Abdelfattah, and H. Al Halbusi, "The good, the bad, and the GPT: Reviewing the impact of generative artificial intelligence on psychology," Current Opinion in Psychology, vol. 59, 2024, Art. no. 101872, doi: 10.1016/j.copsyc.2024.101872.
6. "Trusting emotional support from generative artificial intelligence: a conceptual review," Computers in Human Behavior: Artificial Humans, vol. 5, 2025, Art. no. 100195, doi: 10.1016/j.chbah.2025.100195.
7. T. Franke, T. Radüntz, and C. Peifer, "To trust or not to trust a human(-like) AI—A scoping review and conjoint analyses on factors influencing anthropomorphism and trust," Zeitschrift für Arbeitswissenschaft, 2025, doi: 10.1007/s41449-025-00481-6.
8. National Institute of Standards and Technology, "Artificial Intelligence Risk Management Framework: Generative Artificial Intelligence Profile," NIST AI 600-1, 2024.
9. S. Mehrotra, C. Degachi, O. Vereschak, C. M. Jonker, and M. L. Tielman, "A Systematic Review on Fostering Appropriate Trust in Human-AI Interaction: Trends, Opportunities and Challenges," ACM Journal on Responsible Computing, vol. 1, no. 4, Art. no. 26, 2024, doi: 10.1145/3696449.
10. A. Afroogh, A. Akbari, E. Malone, M. Kargar, et al., "Trust in AI: progress, challenges, and future directions," Humanities and Social Sciences Communications, vol. 11, Art. no. 1568, 2024.
11. National Institute of Standards and Technology, "AI Risk Management Framework," NIST, 2023–2026.
12. National Institute of Standards and Technology, "Psychological Foundations of Explainability and Interpretability in Artificial Intelligence," NISTIR 8367.
13. National Institute of Standards and Technology, "Trust and Artificial Intelligence," NISTIR 8332.
14. National Institute of Standards and Technology, "Towards a Standard for Identifying and Managing Bias in Artificial Intelligence," NIST SP 1270.
15. Y. Li et al., "Human-AI trust and trustworthy AI: psychological and interaction perspectives," Frontiers in Psychology, 2024.
16. S. Mehrotra et al., "Appropriate trust in human-AI interaction: trends and challenges," ACM Journal on Responsible Computing, 2024.
17. D. Dang and G. Li, "Human trust in AI: antecedents and consequences," Academy of Management Proceedings, 2025.
18. J. Dang and L. Liu, "Public trust in artificial intelligence users," Current Opinion in Psychology, vol. 66, 2025, Art. no. 102148.