The Centaur Model of Intelligence in Human-AI Collaboration: Applications and Limitations in HCI Research

1. Introduction

The "Centaur Model" of intelligence refers to a hybrid approach in which humans and artificial intelligence (AI) systems collaborate, combining their respective strengths to achieve superior outcomes compared to either working alone. Originating from the world of chess—where human-AI teams ("centaurs") outperformed both grandmasters and AI engines—the model has since been formalized and extended to a wide range of domains, including decision-making, creative work, programming, healthcare, and education (Saghafian & Idan, 2024; Pareschi, 2024; Shoresh & Loewenstein, 2024; Tang, 2020; Alves & Cipriano, 2023; Fassbender, 2025; Krakowski et al., 2022; Dell'Acqua et al., 2023; Muller, 2022; Pieper & Gleasure, 2025; Hao, 2023). In Human-Computer Interaction (HCI) research, the centaur model is recognized for its potential to foster symbiotic relationships between human intuition, creativity, and contextual understanding, and the computational power, speed, and pattern recognition capabilities of AI (Saghafian & Idan, 2024; Pareschi, 2024; Shoresh & Loewenstein, 2024; Borghoff et al., 2025; Tang, 2020; Alves & Cipriano, 2023; Fassbender, 2025; Krakowski et al., 2022; Dell'Acqua et al., 2023; Muller, 2022; Pieper & Gleasure, 2025; Hao, 2023). Recent studies and reviews highlight both the promise and the challenges of centaurian systems: while they can enhance productivity, innovation, and decision quality, they also introduce new complexities around trust, transparency, division of labor, and the risk of cognitive deskilling (Saghafian & Idan, 2024; Pareschi, 2024; Shoresh & Loewenstein, 2024; Borghoff et al., 2025; Tang, 2020; Alves & Cipriano, 2023; Fassbender, 2025; Krakowski et al., 2022; Dell'Acqua et al., 2023; Muller, 2022; Thiele, 2021; Pieper & Gleasure, 2025; Hao, 2023). This review synthesizes the current state of research on the centaur model in HCI, focusing on its primary applications, design principles, and limitations.

2. Methods

A comprehensive search was conducted across over 170 million research papers in Consensus, including Semantic Scholar, PubMed, and other major databases. The search targeted the centaur model of intelligence in human-AI collaboration, with a focus on applications and limitations in HCI. In total, 898 papers were identified, 541 were screened, 378 were deemed eligible, and the 50 most relevant papers were included in this review.

Search Strategy

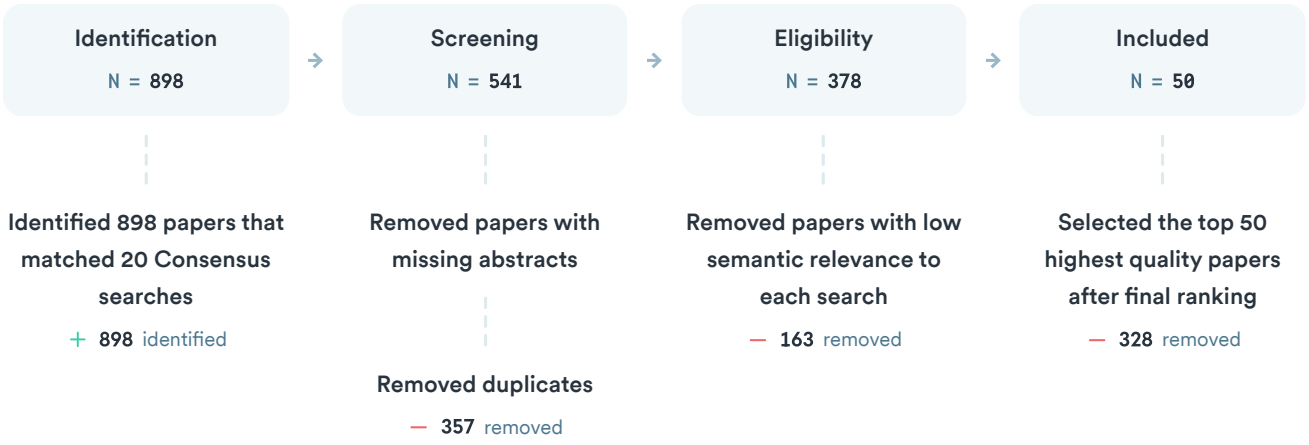

FIGURE 1 Flow diagram of the literature search and selection process.

Eight unique search groups were executed, covering theoretical foundations, practical applications, critiques, and interdisciplinary perspectives on the centaur model in HCI.

3. Results

3.1. Defining the Centaur Model

The centaur model is characterized by a **symbiotic partnership** between human and AI agents, where each complements the other's strengths and compensates for weaknesses (Saghafian & Idan, 2024; Pareschi, 2024; Shoresh & Loewenstein, 2024; Tang, 2020; Alves & Cipriano, 2023; Fassbender, 2025; Krakowski et al., 2022; Dell'Acqua et al., 2023; Muller, 2022; Hao, 2023). Unlike simple human-in-the-loop or automation paradigms, centaur systems emphasize dynamic collaboration, mutual learning, and flexible division of labor (Saghafian & Idan, 2024; Pareschi, 2024; Shoresh & Loewenstein, 2024; Borghoff et al., 2025; Tang, 2020; Alves & Cipriano, 2023; Fassbender, 2025; Krakowski et al., 2022; Dell'Acqua et al., 2023; Muller, 2022; Pieper & Gleasure, 2025; Hao, 2023). The model draws on foundational work in human-computer symbiosis and has been formalized in recent HCI literature as a design and architectural principle (Pareschi, 2024; Borghoff et al., 2025; Tang, 2020; Fassbender, 2025; Hao, 2023).

3.2. Primary Applications in HCI

- **Decision-Making and Problem-Solving**: Centaur systems are widely used in domains requiring complex decision-making, such as medical diagnostics, regulatory review, and business strategy, where human judgment and AI analytics are combined for superior outcomes (Saghafian & Idan, 2024; Shoresh & Loewenstein, 2024; Çelikok et al., 2022; Krakowski et al., 2022; Dell'Acqua et al., 2023; Pitts, 2020; Pieper & Gleasure, 2025; Zabiullah et al., 2024; Kolbjørnsrud, 2023; Hao, 2023).
- **Creative and Knowledge Work**: In programming, design, and content creation, centaur models enable humans to leverage AI for ideation, code generation, and rapid prototyping, while retaining creative control (Alves & Cipriano, 2023; Fassbender, 2025; Dell'Acqua et al., 2023; Muller, 2022; Pieper & Gleasure, 2025; Hao, 2023).
- **Education and Training**: Centaurian approaches are being explored in AI-augmented teaching, where educators and AI systems co-orchestrate learning experiences (Fassbender, 2025; Ji et al., 2022).
- **Healthcare and Clinical Decision Support**: Human-AI centaurs are used in radiology, cancer screening, and patient care, improving diagnostic accuracy and workflow efficiency (Saghafian & Idan, 2024; Krakowski et al., 2022; Pieper & Gleasure, 2025; Zabiullah et al., 2024).
- **Regulatory and Legal Tech**: The centaur model is proposed for regulatory review and legal decision-making, combining human oversight with AI-driven analytics (Tang, 2020; Pitts, 2020).

3.3. Design Principles and Variants

Recent research formalizes centaurian design through architectural guidelines, emphasizing **transparency, explainability, trust calibration, and flexible role allocation** (Pareschi, 2024; Borghoff et al., 2025; Tang, 2020; Fassbender, 2025; Kolbjørnsrud, 2023; Vössing et al., 2022). Variants include "Chiron-like" centaurs (emphasizing wisdom and justice), "centaur programmers," and "regulatory centaurs" (Tang, 2020; Alves & Cipriano, 2023; Pitts, 2020; Hao, 2023). The model is often contrasted with "cyborg" approaches, where human and AI processes are more tightly integrated (Dell'Acqua et al., 2023).

3.4. Primary Limitations and Challenges

- **Trust and Transparency**: Effective centaur systems require calibrated trust and clear communication between human and AI agents; lack of transparency can undermine collaboration (Saghafian & Idan, 2024; Borghoff et al., 2025; Tang, 2020; Schmutz et al., 2024; Vössing et al., 2022).
- **Division of Labor and Process Design**: Determining when to delegate tasks to AI versus humans is non-trivial and context-dependent; poor process design can negate the benefits of centaur collaboration (Saghafian & Idan, 2024; Shoresh & Loewenstein, 2024; Krakowski et al., 2022; Dell'Acqua et al., 2023; Muller, 2022; Pieper & Gleasure, 2025; Hao, 2023).
- **Cognitive Deskilling and Over-Reliance**: Prolonged use of centaur systems may erode human expertise and lead to over-reliance on AI, especially if feedback loops are not well-designed (Saghafian & Idan, 2024; Thiele, 2021).
- **Performance Heterogeneity**: Meta-analyses show that centaur teams do not always outperform the best individual agent, especially in decision tasks; benefits are more pronounced in creative or exploratory work (Vaccaro et al., 2024; Dell'Acqua et al., 2023).
- **Ethical and Legal Issues**: Questions around responsibility, accountability, and intellectual property arise in centaur collaborations, particularly in invention and regulatory contexts (Tang, 2020; Hao, 2023).

Key Papers

Paper	Domain/Application	Key Insights	Limitations/Challenges
(Saghafian & Idan, 2024)	Generative AI, decision-making	Centaurs outperform pure AI/human in many domains; symbiotic learning	Human intuition can degrade performance if not managed
(Pareschi, 2024)	HCI, design methodology	Formalizes centaurian design; compares to other models	Human contribution can be overshadowed by AI advances
(Shoresh & Loewenstein, 2024)	Chess, sequential decision-making	Demonstrates synergy in human-AI teams; MoE architectures	Identifying relative advantages is challenging
(Dell'Acqua et al., 2023)	Knowledge work, consulting	Centaur vs. cyborg patterns; centaurs divide tasks, cyborgs integrate	AI can mislead on tasks outside its "frontier"
(Krakowski et al., 2022)	Competitive strategy, chess	Centaur teams create new competitive capabilities	Traditional skills may become obsolete

FIGURE 2 Comparison of key studies on the centaur model in HCI.

Top Contributors

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

4. Discussion

The centaur model represents a paradigm shift in HCI and human-AI collaboration, moving beyond automation to emphasize **mutual augmentation and symbiotic learning** (Saghafian & Idan, 2024; Pareschi, 2024; Shoresh & Loewenstein, 2024; Tang, 2020; Alves & Cipriano, 2023; Fassbender, 2025; Krakowski et al., 2022; Dell'Acqua et al., 2023; Muller, 2022; Pieper & Gleasure, 2025; Hao, 2023). Its success depends on thoughtful design of interaction protocols, transparency, and trust calibration (Pareschi, 2024; Borghoff et al., 2025; Tang, 2020; Fassbender, 2025; Vössing et al., 2022). While centaur systems have demonstrated clear benefits in creative, diagnostic, and complex decision-making domains, their effectiveness is highly context-dependent, and they are not a panacea (Saghafian & Idan, 2024; Shoresh & Loewenstein, 2024; Krakowski et al., 2022; Vaccaro et al., 2024; Dell'Acqua et al., 2023; Muller, 2022; Pieper & Gleasure, 2025; Hao, 2023). Persistent challenges include the risk of cognitive deskilling, performance heterogeneity, and unresolved ethical and legal questions (Saghafian & Idan, 2024; Tang, 2020; Vaccaro et al., 2024; Dell'Acqua et al., 2023; Thiele, 2021; Hao, 2023). The literature suggests that centaurian approaches are most effective when humans and AI are able to dynamically negotiate roles, learn from each other, and adapt to changing task demands (Pareschi, 2024; Shoresh & Loewenstein, 2024; Borghoff et al., 2025; Tang, 2020; Fassbender, 2025; Dell'Acqua et al., 2023; Muller, 2022; Pieper & Gleasure, 2025; Hao, 2023).

Claims and Evidence Table

Claim	Evidence Strength	Reasoning	Papers
Centaur (human-AI) teams can outperform either humans or AI alone in certain domains	Strong	Empirical and meta-analytic evidence from chess, consulting, and creative work	(Saghafian & Idan, 2024; Shoresh & Loewenstein, 2024; Krakowski et al., 2022; Dell'Acqua et al., 2023; Muller, 2022; Pieper & Gleasure, 2025; Hao, 2023)
Centaur systems require careful design of trust, transparency, and division of labor	Strong	Design guidelines and case studies highlight these as critical for success	(Pareschi, 2024; Borghoff et al., 2025; Tang, 2020; Fassbender, 2025; Vössing et al., 2022)
Centaur teams do not always outperform the best individual agent, especially in decision tasks	Moderate	Meta-analyses and systematic reviews show mixed results	(Vaccaro et al., 2024; Dell'Acqua et al., 2023; Muller, 2022)
Prolonged centaur use can lead to cognitive deskilling and over-reliance on AI	Moderate	Theoretical and empirical work on deskilling and feedback loops	(Saghafian & Idan, 2024; Thiele, 2021)
Ethical and legal issues (e.g., responsibility, inventorship) are unresolved in centaur collaborations	Moderate	Policy and legal analyses highlight ongoing debates	(Tang, 2020; Hao, 2023)

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

5. Conclusion

The centaur model of intelligence is a foundational paradigm in HCI research, enabling powerful human-AI collaboration through mutual augmentation. While it offers significant benefits in creative, diagnostic, and complex decision-making domains, its effectiveness depends on careful design, trust calibration, and ongoing attention to ethical and practical challenges.

5.1. Research Gaps

Key gaps include understanding long-term impacts on human expertise, optimizing dynamic role allocation, and addressing ethical and legal frameworks for centaurian systems.

Research Gaps Matrix

Application Domain	Trust/Transparency	Division of Labor	Deskilling/Expertise	Ethics/Legal	Dynamic Adaptation
Decision-Making	7	6	4	3	5
Creative Work	6	5	3	2	4
Healthcare/Clinical	5	4	2	3	3
Education/Training	3	2	1	1	2
Regulatory/Legal	2	2	1	4	1

FIGURE 5 Heatmap of research coverage by application domain and key challenge.

5.2. Open Research Questions

Future research should explore how to optimize centaurian collaboration for long-term human expertise, develop adaptive role allocation mechanisms, and establish robust ethical and legal frameworks.

Question	Why
How does long-term use of centaur systems affect human expertise and cognitive skills?	Understanding this is critical to prevent deskilling and ensure sustainable human-AI collaboration.
What are the most effective mechanisms for dynamic role allocation between humans and AI in centaur teams?	Optimizing collaboration requires flexible, context-aware division of labor.
How can ethical and legal frameworks be adapted to address responsibility and inventorship in centaur collaborations?	Clear guidelines are needed as centaur systems become more prevalent in high-stakes domains.

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

In summary, the centaur model is a powerful and evolving paradigm for human-AI collaboration in HCI, but realizing its full potential requires ongoing research into design, ethics, and the long-term dynamics of human-AI partnerships.

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

References

Saghafian, S., & Idan, L. (2024). Effective Generative AI: The Human-Algorithm Centaur. *ArXiv*, abs/2406.10942. **https://doi.org/10.2139/ssrn.4594780**

Pareschi, R. (2024). Beyond Human and Machine: An Architecture and Methodology Guideline for Centaurian Design. *Sci*. **https://doi.org/10.3390/sci6040071**

Shoresh, D., & Loewenstein, Y. (2024). Modeling the Centaur: Human-Machine Synergy in Sequential Decision Making. *ArXiv*, abs/2412.18593. **https://doi.org/10.48550/arXiv.2412.18593**

Borghoff, U., Bottoni, P., & Pareschi, R. (2025). Human-Artificial Interaction in the Age of Agentic AI: A System-Theoretical Approach. *ArXiv*, abs/2502.14000. **https://doi.org/10.3389/fhumd.2025.1579166**

Çelikok, M., Oliehoek, F., & Kaski, S. (2022). Best-Response Bayesian Reinforcement Learning with Bayes-adaptive POMDPs for Centaurs. *ArXiv*, abs/2204.01160. **https://doi.org/10.48550/arXiv.2204.01160**

Tang, B. (2020). The Chiron Imperative – A Framework of Six Human-in-the-Loop Paradigms to Create Wise and Just AI-Human Centaurs. *The LegalTech Book*. **https://doi.org/10.1002/9781119708063.ch10**

Alves, P., & Cipriano, B. (2023). The centaur programmer - How Kasparov's Advanced Chess spans over to the software development of the future. *ArXiv*, abs/2304.11172. **https://doi.org/10.48550/arXiv.2304.11172**

Fassbender, W. (2025). Of teachers and centaurs: Exploring the interactions and intra-actions of educators on AI education platforms. *Learning, Media and Technology*. **https://doi.org/10.1080/17439884.2024.2447946**

Krakowski, S., Luger, J., & Raisch, S. (2022). Artificial intelligence and the changing sources of competitive advantage. *Strategic Management Journal*. **https://doi.org/10.1002/smj.3387**

Schmutz, J., Outland, N., Kerstan, S., Georganta, E., & Ulfert, A. (2024). AI-teaming: Redefining collaboration in the digital era.. *Current opinion in psychology*, 58, 101837. **https://doi.org/10.1016/j.copsyc.2024.101837**

Vaccaro, M., Almaatouq, A., & Malone, T. (2024). When combinations of humans and AI are useful: A systematic review and meta-analysis. *Nature Human Behaviour*, 8, 2293 - 2303. **https://doi.org/10.1038/s41562-024-02024-1**

Dell'Acqua, F., McFowland, E., Mollick, E., Lifshitz-Assaf, H., Kellogg, K., Rajendran, S., Krayer, L., Candelon, F., & Lakhani, K. (2023). Navigating the Jagged Technological Frontier: Field Experimental Evidence of the Effects of AI on Knowledge Worker Productivity and Quality. *SSRN Electronic Journal*. **https://doi.org/10.2139/ssrn.4573321**

Muller, E. (2022). How AI-Human Symbiotes May Reinvent Innovation and What the New Centaurs Will Mean for Cities. *Technology and Investment*. **https://doi.org/10.4236/ti.2022.131001**

Pitts, P. (2020). Regulatory centaurs. *Nature Biotechnology*, 38, 788 - 789. **https://doi.org/10.1038/s41587-020-0589-x**

Thiele, L. (2021). Rise of the Centaurs: The Internet of Things Intelligence Augmentation. **, 39-61. **https://doi.org/10.1007/978-3-030-74420-5_3**

Pieper, M., & Gleasure, R. (2025). How AI Helps to Compile Human Intelligence: An Empirical Study of Emerging Augmented Intelligence for Medical Image Scanning. *Information Systems Journal*. **https://doi.org/10.1111/isj.12585**

Ji, H., Han, I., & Ko, Y. (2022). A systematic review of conversational AI in language education: focusing on the collaboration with human teachers. *Journal of Research on Technology in Education*, 55, 48 - 63. **https://doi.org/10.1080/15391523.2022.2142873**

Zabiullah, S., Zahid, V., Eltayeb, E., Abdelbaky, N., Ahmed, L., Munawer, S., Eltayeb, M., Ali, H., & Abdelrahman, N. (2024). Examining the role of Artificial Intelligence in Age-appropriate Cancer screening: A Scoping review. *The Annals of Family Medicine*. **https://doi.org/10.1370/afm.22.s1.6907**

Kolbjørnsrud, V. (2023). Designing the Intelligent Organization: Six Principles for Human-AI Collaboration. *California Management Review*, 66, 44 - 64. **https://doi.org/10.1177/00081256231211020**

Vössing, M., Kühl, N., Lind, M., & Satzger, G. (2022). Designing Transparency for Effective Human-AI Collaboration. *Information Systems Frontiers*, 24, 877 - 895. **https://doi.org/10.1007/s10796-022-10284-3**

Hao, Y. (2023). The Rise of Centaur Inventors and the Concept of Constructive Conception. *SSRN Electronic Journal*. **https://doi.org/10.2139/ssrn.4519145**