(Libyan Journal of Contemporary Academic Studies) المجلة الليبية للدراسات الأكاديمية المعاصرة
LJCAS, eISSN: 5970-3005
Volume 3, Issue 2 (Part II), 2025, Pages: 101-108

Designing Partnership: A Comprehensive Review of Human–AI Collaboration

Anwar Alhenshiri [1*], Hoda Badesh [2]
[1,2] Department of Computer Science, Faculty of Information Technology, Misurata University, Misurata, Libya
*Corresponding author: alhenshiri@it.misuratau.edu.ly

تصميم الشراكة: مراجعة شاملة للتعاون بين الإنسان والذكاء الاصطناعي

أنور أحمد الهنشيري*[1]، هدى سالم بادش [2]
[2،1] قسم علوم الحاسوب، كلية تقنية المعلومات، جامعة مصراته، مصراته ، ليبيا

Received: 07-09-2025; **Accepted:** 19-11-2025; **Published:** 06-12-2025

Abstract:

Human–AI collaboration has become a pivotal area of research at the intersection of human–computer interaction (HCI), artificial intelligence (AI), and cognitive systems. Rather than replacing human work, collaborative systems aim to integrate human judgment and contextual expertise with the computational strengths of AI to achieve shared goals. This survey provides a comprehensive overview of the field, examining foundational frameworks and models of collaboration, methods for evaluating human–AI teamwork, and applications across domains such as healthcare, education, creative industries, and transportation. Key challenges—including the calibration of trust, the design of transparent and usable explanations, and the balance between human control and AI autonomy—are analyzed in depth. The paper concludes by identifying open research questions and outlining future directions for advancing human-centered approaches to AI collaboration that enhance performance while safeguarding user agency, accountability, and ethical values.

Keywords: HCI, AI, Collaboration, User, Experience, Evaluation, Study.

الملخص :

لقد أصبح **التعاون بين الإنسان والذكاء الاصطناعي** مجالاً محورياً للبحث في تقاطع التفاعل بين الإنسان والحاسوب(HCI) ، والذكاء الاصطناعي(AI) ، والأنظمة المعرفية. وبدلاً من أن تحل هذه النظم محل العمل البشري، فإن هدفها هو دمج الحكم البشري والخبرة السياقية مع القدرات الحسابية للذكاء الاصطناعي من أجل تحقيق أهداف مشتركة. يقدم هذا الاستعراض نظرة شاملة على المجال، حيث يتناول الأطر والنماذج الأساسية للتعاون، وطرائق تقييم العمل الجماعي بين الإنسان والذكاء الاصطناعي، إضافةً إلى التطبيقات في مجالات مثل الرعاية الصحية، والتعليم، والصناعات الإبداعية، والنقل. كما يناقش التحديات الرئيسة بعمق، بما في ذلك **معايرة الثقة**، وتصميم التفسيرات الشفافة وسهلة الاستخدام، وتحقيق التوازن بين **التحكم البشري** و**استقلالية الذكاء الاصطناعي** .ويختتم البحث بتحديد أسئلة بحثية مفتوحة ورسم اتجاهات مستقبلية لتطوير مناهج تتمحور حول الإنسان في التعاون مع الذكاء الاصطناعي، بما يعزز الأداء مع الحفاظ على **وكالة المستخدم**، و**المساءلة**، و**القيم الأخلاقية**.

الكلمات المفتاحية: التفاعل بين الانسان والحاسوب، خبرة المستخدم، تقييم، دراسة، تعاون.

1. Introduction

Human–AI collaboration describes systems in which artificial intelligence (AI) and people work together—sharing tasks, exchanging information, and making joint decisions—rather than systems that merely automate human work or replace human involvement. The rapid proliferation of AI across domains such as healthcare, education, creative work, and business has shifted research and design emphasis from standalone algorithms to how AI can effectively partner with human users in real-world tasks[2, 10]. This shift foregrounds questions of usability, trust, transparency, and shared decision-making that sit at the intersection of Human–Computer Interaction (HCI), cognitive science, and machine learning.

Historically, HCI research treated interactive systems as predictable tools, grounded in classic principles of usability such as visibility and feedback. Modern AI systems, however, are adaptive, probabilistic, and often opaque, which changes the nature of interaction design [32]. Foundational work on automation reliability showed that human reliance depends not only on accuracy but also on how users form mental models and perceive system reliability in uncertain situations [4, 20]. These human factors remain central in collaborative AI: poorly calibrated trust leads to over-reliance (automation bias) or under-utilization of useful capabilities.

Recent scholarship emphasizes that technical advances alone do not guarantee effective collaboration. Survey and synthesis studies argue that success depends on human-centered design choices—clear communication of system intent and limitations, user control, and support for graceful error recovery—rather than solely on improved model performance [17, 41]. Empirical findings further show mixed outcomes: in some contexts, human–AI teams outperform either humans or AI alone, while in others, naïve combinations perform worse than the best individual contributor [7, 33]. These mixed results highlight the need for principled design frameworks and rigorous evaluation methods tailored to collaborative settings.

Despite growing interest, several gaps remain. First, there is no universally accepted taxonomy or evaluation standard for what constitutes "good" collaboration—metrics vary across studies and domains, from task accuracy to trust and cognitive workload [36]. Second, while Explainable AI (XAI) has advanced rapidly [11, 39], translating technical outputs into explanations usable by non-experts is still unresolved. Third, the recent emergence of large language models (LLMs) and other foundation models has introduced new collaborative opportunities (co-creative writing, decision support) but also risks of hallucination, misplaced confidence, and ethical concerns [6, 16]. Together, these gaps motivate a comprehensive survey that synthesizes conceptual frameworks, empirical findings, design principles, and open challenges for human–AI collaboration.

This paper provides a structured overview of human–AI collaboration research. The work (1) clarifies terminology and situates collaboration within related paradigms such as automation, assistance, and human-in-the-loop systems; (2) reviews key conceptual frameworks and interaction models, including levels of automation, mixed-initiative systems, and shared mental models; (3) synthesizes empirical findings and evaluation methods across domains; (4) discusses case studies in healthcare, education, creative tools, and transportation; and (5) identifies open research directions in evaluation standards, adaptive explanation, multimodal collaboration, and ethical governance. By integrating perspectives from HCI, AI, and human factors, this survey offers a roadmap for designing AI systems that genuinely augment human capabilities while preserving user agency and societal values.

The remainder of this paper is structured as follows. Section 2 introduces the conceptual foundations of human–AI collaboration. Section 3 reviews key frameworks and models. Section 4 discusses the core dimensions of collaboration. Section 5 outlines research methods and evaluation approaches. Section 6 surveys applications across domains. Section 7 addresses challenges and open issues. Section 8 presents future research directions. Section 9 concludes with a synthesis of insights.

2. Background and Conceptual Foundations

Human–AI collaboration builds on research in HCI, human factors, and cognitive psychology. Early studies of automation, such as Sheridan and Verplank's [40] model of **Levels of Automation (LOA),** highlighted trade-offs between efficiency and human control, showing that higher automation can reduce workload but also erode vigilance and situational awareness.

As AI systems grew more adaptive, traditional HCI approaches were insufficient. **Human-in-the-loop (HITL)** frameworks emphasized the need for oversight and feedback, framing AI as a partner rather than a replacement [1, 28]. Cognitive theories also contributed: ***distributed cognition*** views AI as part of a shared cognitive system, while ***shared mental models*** stress alignment of goals and expectations between human and AI partners [26].

More recent perspectives, such as **hybrid intelligence**, highlight co-adaptation, where humans provide intuition and ethical reasoning while AI offers scalability and pattern recognition [10]. Parallel calls for **human-centered AI** emphasize accountability, safety, and ethical principles [15, 41].

Together, these foundations frame collaboration as a dynamic partnership requiring shared understanding, trust calibration, and alignment with human values—principles that inform the frameworks reviewed in Section 3.

3. Frameworks and Models of Collaboration

A number of frameworks have been developed to conceptualize how tasks, initiative, and authority are shared between humans and AI. These models, originating in human factors and HCI, remain central to understanding collaboration today.

One of the earliest was Sheridan and Verplank's [40] ***Levels of Automation*** **(LOA)**, which outlined a continuum from full human control to complete machine autonomy. Subsequent work refined this idea, emphasizing the importance of ***adjustable autonomy***, where control can shift flexibly depending on the reliability of the system, the complexity of the task, and the state of the user [28, 43]. Although widely used, LOA frameworks tend to oversimplify modern collaboration, where initiative often moves back and forth between human and AI in more fluid ways.

The concept of ***mixed-initiative interaction*** [23] addresses this fluidity by describing systems in which both humans and AI agents can initiate actions, propose alternatives, or redirect tasks. Research applying this model to tutoring and planning systems [1] shows its potential, but also highlights challenges around timing, conflict resolution, and the risk of overwhelming users with interruptions [12]. **Figure 1** conceptually depicts a mixed-initiative interaction loop, showing how human and AI exchange proposals, negotiate actions, and adapt based on outcomes.

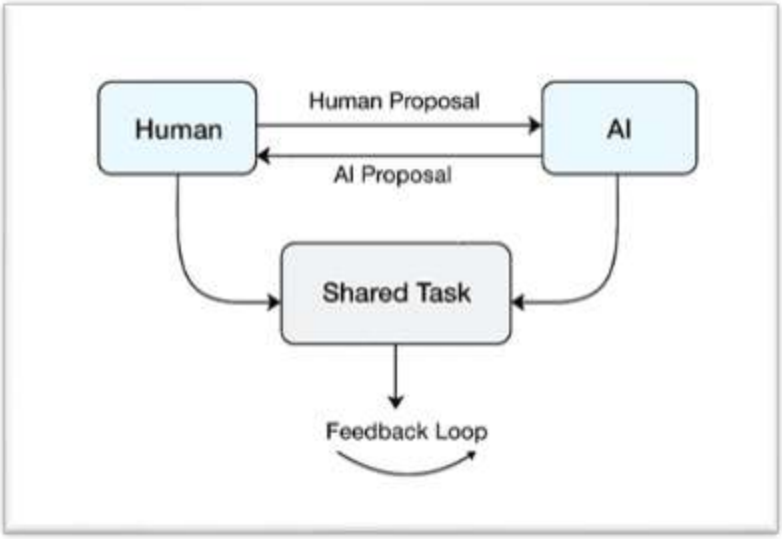

Figure 1. Human-AI Interaction Loop

Another influential perspective comes from *team cognition*. The theory of ***shared mental models*** [26] stresses the importance of humans and AI aligning their understanding of goals, roles, and task strategies. Studies in robotics suggest that shared mental models improve coordination and reduce errors, provided systems communicate their reasoning and limitations clearly [18]. In this respect, explainable AI has become essential, since well-designed explanations help users maintain accurate mental models of what the AI can and cannot do [11].

Finally, the idea of ***hybrid intelligence*** emphasizes collaboration as an evolving partnership in which humans and AI co-adapt. Humans contribute contextual reasoning, ethics, and intuition, while AI provides scalability and pattern recognition [10]. Hybrid intelligence thus frames collaboration as iterative and dynamic, rather than a static division of labor.

Taken together, these frameworks underline three recurring themes: effective collaboration requires balancing autonomy and control, enabling initiative from both sides, and fostering alignment through shared understanding. While developed in different contexts, they provide a common foundation for designing human–AI systems across domains.

4. Core Dimensions of Human–AI Collaboration

Research has identified several dimensions that consistently shape the quality of human–AI collaboration. These dimensions are interdependent, and effective system design requires balancing them in context rather than optimizing any one in isolation.

A central concern is ***trust and reliance***. Trust must be carefully calibrated: too little leads users to ignore valuable recommendations, while too much results in over-reliance and automation bias [20, 28]. Trust is not static but evolves as users interact with a system, influenced by prior experience, perceived reliability, and the stakes of the decision [17]. Recent studies confirm that users may continue to defer to AI even after observing errors, underscoring the difficulty of aligning trust with actual performance [33].

Closely related is the issue of ***transparency and explainability***. Explanations are intended to help users form accurate mental models and understand the rationale behind AI decisions. Yet research shows that explanations can sometimes mislead or overwhelm, depending on their form and context [32 35]. For this reason, scholars emphasize the need for adaptive, context-sensitive approaches that tailor explanations to the expertise and goals of users [12]. Transparency, therefore, is not simply a matter of more information but of providing the right information at the right time.

The ***balance between human control and AI autonomy*** is another defining dimension. While high autonomy can reduce workload and improve efficiency, meaningful human control is critical for safety, accountability, and acceptance [8, 29]. Many researchers advocate adjustable autonomy, where the level of control shifts depending on context and user needs [43]. In domains such as healthcare and transportation, maintaining a clear path for human intervention is essential [22].

Equally important are ***collaboration dynamics,*** which reflect lessons from team research. Successful human–AI teams require coordination, role clarity, and effective communication of intent [26]. Systems that signal uncertainty, adapt to user behavior, and negotiate initiative more smoothly tend to support higher-quality teamwork [21, 16].

Finally, ***ethics and accountability*** permeate all dimensions. Questions of responsibility become especially complex in joint decision-making, where both human and AI contribute. Scholars argue that responsibility must remain human-centered, even as AI becomes more autonomous [41]. At the same time, persistent issues of bias and fairness remind us that collaboration can reinforce inequalities unless carefully designed [31, 34]. Ethical frameworks [15] provide guidance but must be translated into concrete design and evaluation practices.

Together, these dimensions—trust, transparency, autonomy, collaboration dynamics, and ethics—capture the multifaceted nature of human–AI partnerships. They also illustrate why collaboration cannot be reduced to system accuracy alone; it is a socio-technical challenge requiring thoughtful integration of human, technical, and ethical factors.

5. Research Methods and Evaluation Approaches

Studying *human–AI collaboration* requires methods that capture both technical performance and human experience. Unlike traditional AI evaluation, which relies on accuracy or efficiency, collaborative systems must also be assessed in terms of *trust, usability, workload, and ethical alignment*. This has led to a diverse set of methodological approaches.

Controlled laboratory experiments remain the backbone of research because they allow causal effects of design choices to be isolated. By manipulating factors such as explanation style, level of automation, or user expertise, these studies reveal how collaboration outcomes change under different conditions [7, 35]. While rigorous, such experiments often lack ecological validity, since real-world collaboration unfolds in more complex environments.

To complement this, researchers increasingly rely on ***field studies and longitudinal deployments***. These capture how users adapt to AI over time, how trust evolves with repeated exposure, and how systems integrate into professional workflows. For example, Yang et al. [44] documented how clinicians gradually adjusted their reliance on diagnostic AI, while Lai et al. [27] showed how journalists negotiated the strengths and weaknesses of AI-generated news summaries. Such studies provide insights that cannot be observed in one-off lab sessions.

Surveys and self-report measures are widely used to capture subjective perceptions of trust, workload, and fairness [17]. These instruments are valuable but imperfect, since self-reports may diverge from actual behavior. To address this, researchers pair them with ***behavioral and physiological measures***, including eye-tracking, EEG, or biometric sensors, which provide real-time indicators of trust calibration and cognitive load [33, 45].

Another important strand involves ***simulation studies***, particularly in domains where experimentation carries significant risk or is logistically impractical. Multi-agent simulations and computational models allow researchers to explore scenarios that would be too costly, dangerous, or time-consuming to replicate in the real world. For example, in aviation, simulations can model pilot–AI coordination under emergency conditions, while in robotics they can capture swarm behavior and distributed task allocation [18]. Traffic systems research also benefits from simulation, enabling the study of human–AI interactions in large-scale environments without endangering safety [36]. These approaches not only provide valuable insights into system dynamics but also allow for controlled manipulation of variables, offering a powerful tool for testing hypotheses that complement laboratory and field experiments.

Increasingly, scholars adopt ***mixed-methods approaches*** that combine quantitative performance metrics with qualitative insights from interviews, observations, or think-aloud protocols. This reflects growing recognition that collaboration is not simply a matter of accuracy or efficiency but a socio-technical process shaped by human perceptions, organizational contexts, and ethical considerations. Quantitative data, such as task completion times or error rates, provide measurable outcomes, while qualitative accounts reveal how users interpret AI behavior, how they negotiate control, and how collaboration affects trust and decision-making [12, 42]. By integrating these perspectives, mixed-methods research provides a more holistic picture of human–AI collaboration, capturing not only what works but also why and under what conditions. Table 1 provides a summary of the main research methods used in human–AI collaboration studies, their strengths, and limitations.

Table 1. Main Research Methods in Human-AI Collaboration

Method	Description	Strengths	Limitations	Example Studies
Laboratory Experiments	Controlled tasks with AI	Causal inference, control	Limited ecological validity	[7, 35]
Field Studies	Real-world deployment	High realism, longitudinal	Harder to control variables	[27, 44]
Surveys/User Studies	Structured questionnaires	Scalable, captures attitudes	Self-report bias	[16]
Simulation/Modeling	Agent-based or computational	Safe, scalable, exploratory	May lack realism	[18, 36]
Cognitive/Physiological	Eye-tracking, EEG, biometrics	Rich, real-time data	Expensive, privacy concerns	[33, 45]
Mixed-Methods	Combines multiple approaches	Triangulation, deeper insights	Resource-intensive	[12, 42]

Despite these advances, evaluation remains fragmented. Different domains emphasize different outcomes—clinicians may prioritize safety, educators focus on equity, and businesses look for efficiency. Without standardized frameworks, results are hard to compare across studies. Developing multi-dimensional evaluation standards that integrate accuracy, efficiency, trust, satisfaction, and fairness remains one of the field's most pressing challenges.

6. Applications Across Domains

The principles of *human–AI collaboration* have been explored in a wide range of domains, each with distinct opportunities and challenges. Although the contexts differ, recurring issues such as trust, transparency, and accountability appear across them.

Healthcare has been one of the most prominent fields of application. AI supports tasks such as diagnostic imaging, prognosis, and treatment planning. Studies show that clinician–AI teams often outperform either partner alone. For example, McKinney et al. [30] demonstrated that AI-assisted breast cancer screening could achieve accuracy comparable to expert radiologists, while Rajpurkar et al. [37] found that collaboration between clinicians and AI improved chest X-ray interpretation. At the same time, ethical and practical challenges remain: algorithms may reproduce systemic biases, as Obermeyer et al. [34] showed in a case where Black patients' health needs were underestimated. For this reason, transparency, trust calibration, and clinician accountability remain essential [22].

In ***education***, intelligent tutoring systems and adaptive platforms offer personalized learning experiences and feedback. Such systems can reduce teacher workload and provide students with tailored support [19, 24]. However, fairness and transparency remain significant concerns, especially when AI recommendations shape learning opportunities. Teachers' roles are not replaced but redefined, as they use AI-generated insights while preserving pedagogical control [3].

In ***business and finance***, AI is used in decision-support systems for tasks ranging from risk management to customer analytics. Hybrid intelligence approaches enhance performance by combining algorithmic insights with human judgment [10]. Yet reliance on AI introduces risks of automation bias, and historical inequities in training data may be perpetuated in hiring or lending decisions [39]. As Bansal et al. [4] note, trust calibration is particularly critical in high-stakes business contexts.

Creative industries illustrate collaboration in less deterministic domains. Generative AI systems have been adopted for co-writing, music composition, and visual design. Users often treat these systems as partners that inspire new ideas, while still exercising final control over content [9, 13]. Nevertheless, concerns remain about authorship, originality, and the ethical implications of using large datasets that may embed cultural or linguistic biases [6].

Finally, ***transportation*** highlights the safety-critical nature of shared control. Autonomous vehicles must balance automation with human oversight. Endsley [14] identified risks of automation complacency, while Lu et al. [29] showed that transparency about system intent can enhance driver trust. Adaptive autonomy mechanisms, paired with clear intervention protocols, are vital to ensure safety [25].

Across these domains, a consistent pattern emerges: while AI enhances efficiency and performance, its success depends on careful integration into human workflows. Whether in medicine, classrooms, finance, creative practices, or autonomous driving, the principles of trust, transparency, fairness, and accountability remain universal.

7. Challenges and Open Issues

Although human–AI collaboration has made rapid progress, several unresolved challenges continue to limit its effectiveness. These challenges are not isolated but deeply interrelated, requiring both technical and socio-technical solutions.

A central issue is ***trust calibration***. While designers aim to build user trust, the real challenge lies in aligning it with actual system reliability. Over-trust can result in automation bias, where users defer to AI despite errors, while under-trust prevents them from using valuable recommendations. Studies show that this miscalibration is persistent: users sometimes continue to rely on AI even after seeing mistakes [33]. Large language models exacerbate the problem, as their fluent outputs can create a false sense of authority despite factual inaccuracies [16].

Another persistent concern is ***transparency and explainability***. Although explainable AI has advanced, explanations do not always lead to better decisions. Poursabzi-Sangdeh et al. [35] found that partial transparency sometimes worsens outcomes, as users misinterpret irrelevant details. Too little transparency undermines trust, while too much detail can overwhelm, creating cognitive overload. The challenge is therefore not only technical but also human-centered: explanations must be designed to match user expertise, context, and cognitive needs [12, 32].

Bias and fairness also remain pressing challenges. Algorithms trained on historical or biased data may perpetuate inequities, as demonstrated in healthcare, where an AI system underestimated the needs of Black patients [34]. In recruitment and hiring, automated systems risk replicating gender and racial disparities [38]. Addressing these issues requires fairness-aware machine learning, systematic auditing, and participatory design that incorporates diverse perspectives [5, 31].

A fourth open problem is the lack of ***standardized evaluation metrics***. Studies measure outcomes in very different ways, from accuracy and efficiency to user satisfaction and trust. This fragmentation makes it difficult to compare findings or build cumulative evidence [36]. Without integrated frameworks that combine objective and subjective measures, progress will remain uneven across domains.

Finally, ***accountability and responsibility*** present unresolved tensions. In joint decision-making, responsibility can become blurred: if an AI provides a recommendation that a human accepts, who is to blame when harm results? Scholars argue that accountability must remain with humans [41], but legal and regulatory frameworks are still adapting. The European Union's proposed AI Act (2021) represents one step toward clarifying oversight, yet implementation remains challenging [15].

In brief, human–AI collaboration faces challenges that are simultaneously technical, cognitive, and ethical. Trust calibration, transparency, fairness, evaluation, and accountability must all be addressed together if collaboration is to move from promising prototypes to reliable real-world systems.

8. Future Research Directions

Looking ahead, research on human–AI collaboration must move beyond optimizing algorithms toward designing systems that are adaptive, trustworthy, and ethically aligned. Several directions stand out as particularly promising.

One important area is ***adaptive trust calibration***. Current systems rarely account for the dynamic nature of trust, yet studies show that reliance shifts with context, stakes, and prior experience [28]. Future systems could monitor behavioral or physiological cues—such as hesitation, overrides, or gaze—to detect mis-calibrated trust and respond by adjusting explanations or confidence displays [33]. This would help mitigate over-trust in fluent but fallible systems, such as large language models.

A second priority is ***context-sensitive and personalized explainability***. Explanations are unlikely to be equally useful for all users; what benefits an expert clinician may confuse a layperson. Research is beginning to explore adaptive explanations that vary in detail or form depending on the user and task [12]. Narrative rationales, interactive explanations, and multimodal presentations offer promising avenues, though ethical safeguards will be needed to avoid manipulation [32, 39].

Future collaboration will also increasingly involve ***multimodal and embodied interaction***. Beyond text and graphics, systems will use speech, gesture, gaze, or even augmented and virtual reality to support richer teamwork [21]. Robotics and embodied AI highlight the importance of physical presence, shared context, and nonverbal cues in building effective collaboration [18].

The rise of ***large language models (LLMs)*** introduces new possibilities and risks. LLMs such as GPT-4 can act as general-purpose collaborators, assisting with brainstorming, co-writing, and decision support [13]. Yet their tendency to "hallucinate" and overstate confidence raises concerns about over-reliance [16]. Systematic research is needed on how to structure collaboration with LLMs, including mechanisms for error detection, fact-checking, and role negotiation in team settings.

Another gap lies in ***longitudinal and ecologically valid evaluation***. Most studies remain short-term or lab-based, but real collaboration unfolds over weeks or months as users adapt. Field studies in healthcare and journalism show how reliance patterns evolve over time [27, 44]. Expanding such research will be vital to designing systems that remain effective after initial novelty wears off.

Finally, ***governance and ethics*** will shape the future of collaboration. Scholars argue for embedding principles such as accountability, fairness, and explicability into system design [15]. Regulatory frameworks like the EU AI Act (2021) mark an important step, but translating policy into concrete design practices remains an open challenge. Collaborative systems will need institutional and cross-cultural approaches to ensure their alignment with social values.

To summarize, these directions suggest a shift in emphasis: from building powerful models to designing ***human-centered ecosystems*** that integrate AI into workflows responsibly and sustainably. Future progress will depend not only on technical innovation but also on interdisciplinary research that brings together computer science, HCI, psychology, and ethics.

9. Conclusion

Human–AI collaboration represents a shift from *automation as replacement* to *automation as partnership*. This survey has reviewed the conceptual foundations and frameworks that shape collaboration, examined the core dimensions of trust, transparency, autonomy, and accountability, and discussed methods for evaluating human–AI teamwork across healthcare, education, business, creative industries, and transportation. Together, these insights show that effective collaboration depends on more than accuracy: it requires calibrated trust, usable explanations, meaningful human control, and integration into human workflows.

Looking forward, future research must address open challenges by developing adaptive trust mechanisms, context-sensitive explanations, and multimodal systems, while also embedding fairness and accountability into design. The rise of large language models expands possibilities for collaboration but also amplifies risks of over-reliance and misinformation. The path ahead lies in building systems that are not only powerful but also ***human-centered, transparent, and ethically responsible***, ensuring that AI enhances rather than undermines human capabilities.

Compliance with ethical standards

Disclosure of conflict of interest

The author(s) declare that they have no conflict of interest.

References

1. Amershi, S., Cakmak, M., Knox, W. B., & Kulesza, T. (2014). Power to the people: The role of humans in interactive machine learning. *AI Magazine, 35*(4), 105–120. https://doi.org/10.1609/aimag.v35i4.2513
2. Amershi, S., Weld, D. S., Vorvoreanu, M., Fourney, A., Nushi, B., Collisson, P., ... & Horvitz, E. (2019). Guidelines for human–AI interaction. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19)* (pp. 1–13). ACM. https://doi.org/10.1145/3290605.3300233
3. Baker, R. S., & Siemens, G. (2022). Ethics and equity in learning analytics. *British Journal of Educational Technology, 53*(3), 457–472. https://doi.org/10.1111/bjet.13130
4. Bansal, G., Wu, T., Zhou, J., & Fok, R. (2021). Does the whole explain the parts? The effect of AI explanations on human decision-making. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems (CHI '21)* (pp. 1–15). ACM. https://doi.org/10.1145/3411764.3445717
5. Barocas, S., Hardt, M., & Narayanan, A. (2021). *Fairness and machine learning*. MIT Press.
6. Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the dangers of stochastic parrots: Can language models be too big? In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency (FAccT '21)* (pp. 610–623). ACM. https://doi.org/10.1145/3442188.3445922
7. Buçinca, Z., Malaya, M. B., & Gajos, K. Z. (2020). Proxy tasks and explainability: Calibrating trust in human–AI collaboration. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (CHI '20)* (pp. 1–13). ACM. https://doi.org/10.1145/3313831.3376615
8. Cummings, M. L. (2004). Automation bias in intelligent time-critical decision support systems. In *AIAA 1st Intelligent Systems Technical Conference* (pp. 6313). AIAA. https://doi.org/10.2514/6.2004-6313
9. Davis, N., et al. (2022). Creativity support with generative AI. In *Proceedings of the Creativity and Cognition Conference (C&C '22)* (pp. 43–54). ACM. https://doi.org/10.1145/3527927.3532801
10. Dellermann, D., Ebel, P., Kolbe, L. M., & vom Brocke, J. (2019). Hybrid intelligence. *Business & Information Systems Engineering, 61*(5), 637–643. https://doi.org/10.1007/s12599-019-00595-2
11. Doshi-Velez, F., & Kim, B. (2017). Towards a rigorous science of interpretable machine learning. *arXiv preprint arXiv:1702.08608*.
12. Ehsan, U., & Riedl, M. O. (2020). Human-centered explainable AI: Toward a reflective sociotechnical approach. *International Journal of Human–Computer Studies, 140*, 102428. https://doi.org/10.1016/j.ijhcs.2020.102428
13. Elsden, C., et al. (2023). Co-writing with GPT-3: Shared authorship in human–AI writing. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI '23)* (pp. 1–15). ACM. https://doi.org/10.1145/3544548.3581076
14. Endsley, M. R. (2017). From here to autonomy: Lessons learned from human–automation research. *Human Factors, 59*(1), 5–27. https://doi.org/10.1177/0018720816681350
15. Floridi, L., & Cowls, J. (2019). A unified framework of five principles for AI in society. *Harvard Data Science Review, 1*(1). https://doi.org/10.1162/99608f92.8cd550d1
16. Gilardi, F., Alizadeh, M., & Kubli, M. (2023). ChatGPT outperforms crowd-workers for text annotation tasks. *Proceedings of the National Academy of Sciences, 120*(6), e2217269120. https://doi.org/10.1073/pnas.2217269120
17. Glikson, E., & Woolley, A. W. (2020). Human trust in artificial intelligence: Review of empirical research. *Academy of Management Annals, 14*(2), 627–660. https://doi.org/10.5465/annals.2018.0057
18. Gombolay, M. C., Shah, J. A., & Stentz, A. (2017). Computational design of human–robot team interaction. *Science Robotics, 2*(9), eaam5419. https://doi.org/10.1126/scirobotics.aam5419
19. Graesser, A. C., et al. (2023). Intelligent tutoring systems: Advances and challenges. *International Journal of Artificial Intelligence in Education, 33*(1), 15–32. https://doi.org/10.1007/s40593-022-00287-y
20. Hoff, K. A., & Bashir, M. (2015). Trust in automation: Integrating empirical evidence on factors that influence trust. *Human Factors, 57*(3), 407–434. https://doi.org/10.1177/0018720814547570
21. Hoffman, G., & Breazeal, C. (2004). Collaboration in human–robot teams. In *Proceedings of the HRI Workshop on Human–Robot Collaboration* (pp. 1–8).
22. Holzinger, A., et al. (2022). Human-in-the-loop machine learning for healthcare. *Nature Reviews Methods Primers, 2*(71), 1–15. https://doi.org/10.1038/s43586-022-00131-5
23. Horvitz, E. (1999). Principles of mixed-initiative user interfaces. In *Proceedings of the 1999 CHI Conference on Human Factors in Computing Systems (CHI '99)* (pp. 159–166). ACM. https://doi.org/10.1145/302979.303030

24. Kasneci, E., et al. (2023). ChatGPT for education: Opportunities and challenges. *Learning and Instruction, 90*, 102686. https://doi.org/10.1016/j.learninstruc.2023.102686
25. Körber, M. (2019). Theoretical considerations and empirical findings on trust in automation. In *Proceedings of the 11th International Conference on Automotive User Interfaces (AutomotiveUI '19)* (pp. 1–8). ACM. https://doi.org/10.1145/3342197.3344529
26. Kozlowski, S. W. J., & Ilgen, D. R. (2006). Enhancing the effectiveness of work groups and teams. *Psychological Science in the Public Interest*, *7*(3), 77–124. https://doi.org/10.1111/j.1529-1006.2006.00030.x
27. Lai, V., Chen, C., & Tan, C. (2021). Human-in-the-loop evaluations of interpretable models. In *Proceedings of the 35th AAAI Conference on Artificial Intelligence (AAAI '21)* (pp. 11686–11695). AAAI. https://doi.org/10.1609/aaai.v35i13.17371
28. Lee, J. D., & See, K. A. (2004). Trust in automation: Designing for appropriate reliance. *Human Factors, 46*(1), 50–80. https://doi.org/10.1518/hfes.46.1.50.30392
29. Lu, Z., et al. (2022). Driver trust in autonomous vehicles: The impact of transparency and control. *Transportation Research Part F: Traffic Psychology and Behaviour, 85*, 198–210. https://doi.org/10.1016/j.trf.2022.02.015
30. McKinney, S. M., et al. (2020). International evaluation of an AI system for breast cancer screening. *Nature, 577*(7788), 89–94. https://doi.org/10.1038/s41586-019-1799-6
31. Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K., & Galstyan, A. (2021). A survey on bias and fairness in machine learning. *ACM Computing Surveys, 54*(6), 115. https://doi.org/10.1145/3457607
32. Miller, T. (2019). Explanation in artificial intelligence: Insights from the social sciences. *Artificial Intelligence*, *267*, 1–38. https://doi.org/10.1016/j.artint.2018.07.007
33. Nourani, M., et al. (2021). Calibrating trust in AI decision-making. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems (CHI '21)* (pp. 1–14). ACM. https://doi.org/10.1145/3411764.3445464
34. Obermeyer, Z., Powers, B., Vogeli, C., & Mullainathan, S. (2019). Dissecting racial bias in an algorithm used to manage health. *Science*, *366*(6464), 447–453. https://doi.org/10.1126/science.aax2342
35. Poursabzi-Sangdeh, F., et al. (2021). Manipulating and measuring model interpretability. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems (CHI '21)* (pp. 1–13). ACM. https://doi.org/10.1145/3411764.3445315
36. Rahwan, I., et al. (2019). Machine behavior. *Nature*, *568*(7753), 477–486. https://doi.org/10.1038/s41586-019-1138-y
37. Rajpurkar, P., et al. (2022). Clinician–AI collaboration improves chest X-ray interpretation. *Nature Medicine, 28*(1), 85–93. https://doi.org/10.1038/s41591-021-01624-9
38. Raghavan, M., Barocas, S., Kleinberg, J., & Levy, K. (2020). Mitigating bias in algorithmic hiring: Evaluating claims and practices. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency (FAccT '20)* (pp. 469–481). ACM. https://doi.org/10.1145/3351095.3372828
39. Ribeiro, M. T., Singh, S., & Guestrin, C. (2022). Anchors: High-precision model-agnostic explanations. In *Proceedings of the AAAI Conference on Artificial Intelligence (AAAI '22)* (pp. 7736–7744). AAAI. https://doi.org/10.1609/aaai.v36i7.20724
40. Sheridan, T. B., & Verplank, W. L. (1978). *Human and computer control of undersea teleoperators*. MIT Man-Machine Systems Laboratory.
41. Shneiderman, B. (2020). Human-centered artificial intelligence: Reliable, safe & trustworthy. *International Journal of Human–Computer Interaction, 36*(6), 495–504. https://doi.org/10.1080/10447318.2020.1741118
42. Wang, D., Yang, Q., Abdul, A., & Lim, B. Y. (2019). Designing for human–AI collaboration. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19)* (pp. 1–13). ACM. https://doi.org/10.1145/3290605.3300646
43. Wang, Y., et al. (2020). Adjustable autonomy in AI systems. In *Proceedings of the AAAI Conference on Artificial Intelligence (AAAI '20)* (pp. 13686–13693). AAAI.
44. Yang, Q., Steinfeld, A., Rosé, C., & Zimmerman, J. (2020). Investigating human–AI collaboration in healthcare. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (CHI '20)* (pp. 1–13). ACM. https://doi.org/10.1145/3313831.3376716
45. Zhou, Z., & Ji, Q. (2022). Measuring cognitive workload in human–AI interaction with physiological data. *Frontiers in Human Neuroscience, 16*, 845123. https://doi.org/10.3389/fnhum.2022.845123

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of **LJCAS** and/or the editor(s). **LJCAS** and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.