Journal of Digital Educational Technology
2026, 6(2), ep2621
e-ISSN: 2752-5503
https://www.jdet.net/
Review Article
OPEN ACCESS

Human-AI collaboration for ethical, transparent, and equitable educational practices

Diwakar Singh [1*]

[1] Jerome Cheese Company, Jerome, Idaho, 83338, UNITED STATES
*Corresponding Author: dsndri33@gmail.com

Citation: Singh, D. (2026). Human-AI collaboration for ethical, transparent, and equitable educational practices. *Journal of Digital Educational Technology, 6*(2), Article ep2621. https://doi.org/10.29333/jdet/19286

ARTICLE INFO

Received: 29 Mar. 2026

Accepted: 15 Aug. 2026

ABSTRACT

Artificial intelligence (AI) and immersive technologies are reshaping education by enabling adaptive, personalized, and experiential learning environments that enhance engagement, instructional effectiveness, and administrative efficiency. This review explores Human-in-the-loop (HITL) frameworks that integrate human expertise with AI to enhance trust, interpretability, collaborative decision-making, and continuous feedback. AI applications in teacher professional development facilitate personalized training, competency growth, and ethical governance, emphasizing human-centered frameworks and structured professional learning. HITL approaches extend to reality training, knowledge graph validation, emotion recognition, and large language model integration, combining automated outputs with human oversight to ensure contextual appropriateness, fairness, and ethical compliance. Reinforcement learning augmented with human guidance enhances performance, robustness, and reliability, demonstrating scalable, interpretable, and human-aligned solutions. Challenges persist in explainability, bias mitigation, interface design, and the underexplored role of agentic AI, highlighting the need for continuous human involvement in AI-assisted educational systems. Empirical and conceptual evidence underscores that human-AI collaboration improves engagement and ethical alignment while enabling scalable, inclusive, and responsible AI incorporation across diverse educational and operational domains. This study proposed Human-AI Adaptive Synergy Theory (HAAST) for continuously co-adapting through reciprocal learning, feedback, and shared decisions. The study developed a Human-Centered Adaptive Intelligence Framework (HCAIF) that provides an operational structure for designing, evaluating, and governing such systems. The review synthesizes insights from educational, industrial, and technological applications, emphasizing that future directions should include designing transparent and interactive HITL systems. Collectively, integrating human intelligence with AI fosters resilient, inclusive, and adaptive learning ecosystems while modernizing intelligent educational and operational systems.

Keywords: human-in-the-loop, artificial intelligence, adaptive learning, educational technology, immersive learning, large language models

INTRODUCTION

Educational systems are experiencing profound advancements as artificial intelligence (AI) and immersive technologies reshape teaching, learning, and assessment practices. Adaptive learning systems, data-driven analytics, and intelligent platforms facilitate personalized instruction, real-time feedback, and competency-based evaluation. Simultaneously, virtual and augmented reality environments foster experiential learning, engagement, and practical skill development, enhancing critical thinking and student-centered pedagogy. AI tools further support teacher professional development, institutional efficiency, and automated administrative processes, enabling educators to focus on instructional quality while leveraging intelligent systems for continuous monitoring, assessment, and decision-making. However, these technological advancements necessitate new competencies, ethical awareness, and trust-building to ensure effective and responsible assimilation within educational contexts.

AI in Education (AIED) enhances teaching, learning, and educational decision-making through intelligent systems. It enables personalized support and improved instructional strategies. However, the limited empirical evidence indicates the need for rigorous evaluation. The successful incorporation of AIED requires balancing technological innovation with pedagogical needs, human expertise, and transparent practices. Future AIED development should emphasize interdisciplinary collaboration, transparent governance, and human-centered practices to ensure durable and equitable educational progress (Hwang et al., 2020). Machine learning

(ML) and analytics help identify student strengths, predict learning needs, and provide timely feedback, improving engagement and outcomes. AI also supports administration, assessment, and game-based learning, though challenges remain in privacy, equity, and teacher competencies (Susanti et al., 2025). Immersive technologies, including virtual and augmented reality, complement AI by creating interactive, experiential learning experiences. AI systems adapt content and feedback to individual learners, enhancing motivation, retention, and engagement. Virtual reality and augmented reality simulate real-world scenarios, fostering critical thinking and practical skills while enriching pedagogical approaches; however, challenges include digital literacy, privacy, and resource demands (Wang and Huang, 2025).

Recent advancements in AI are supporting student learning and informing teaching decisions. Within this evolution, human involvement remains essential for designing and implementing AI systems, particularly through closed-loop processes involving data recording, pattern detection, and adaptive interventions (Ninaus and Sailer, 2022). Learning analytics and AI have emerged as key fields for collecting and analyzing learning data to guide educational decisions. However, their adoption in authentic contexts remains constrained by insufficient attention to stakeholder needs, limited pedagogical contextualization, and low trust (Topali et al., 2025). Concurrently, rapid technological integration presents universities with opportunities and challenges, requiring educators to address diverse student needs while using AI to enhance learning experiences, encourage active engagement, and reduce academic boredom (Kottara and Asonitou, 2025). A large-scale analysis reveals established themes including intelligent tutoring systems, learning analytics, and natural language processing, alongside growing interest in generative AI (GenAI) and large language models (LLMs) (Feng et al., 2025). Importantly, AI's educational improvement requires uniquely human capabilities, emphasizing critical thinking, creativity, problem-solving, and metacognitive skills through Learning Mastery and Knowledge Mastery frameworks (Luckin, 2025).

AI unification in learning and development enhances personalization, analytics, and institutional effectiveness across higher education and workplaces. Adaptive models, such as self-directed, blended, and experiential learning, benefit from AI, while psychological factors like resilience, growth mindset, and emotional intelligence influence outcomes. Ethical governance, inclusivity, and human-centered frameworks ensure responsible AI adoption, preserving agency and enhancing lifelong learning and organizational resilience (Sharma and Singh, 2026). Human-AI collaboration emphasizes augmentation, combining computational power and pattern recognition with human judgment and contextual insight. Interactive ML interfaces facilitate iterative guidance, transparency, and trust. Effective human-AI partnerships enhance decision-making and performance, highlighting the importance of interface design and collaborative frameworks to empower users and optimize outcomes across education, organizational, and societal contexts (Saha et al., 2023).

Human-in-the-loop (HITL) frameworks emphasize collaboration between human expertise and AI capabilities, promoting interpretability and ethical oversight. By integrating educators' judgment, pedagogical knowledge, and real-time feedback with automated systems, HITL approaches enhance emotion recognition and language model deployment, ensuring contextual appropriateness and reliability. Applications span autonomous navigation, robotics, cybersecurity, and adaptive forecasting, demonstrating improved efficiency, trust, and scalability across domains. Reinforcement learning (RL), Bayesian inference, and interactive supervision illustrate how human-guided AI optimizes decision-making and performance while maintaining safety and alignment with human values. Collectively, embedding human oversight in AI-driven education enables personalized, inclusive, and responsible learning by combining computational intelligence with human judgment, pedagogy, and ethics. AI and immersive technologies enhance engagement, experiential learning, and human-AI collaboration, while addressing equity, ethics, and teacher competencies ensures stable and effective educational systems that maximize human potential and organizational resilience. Despite extensive research on AI and immersive technologies in education, gaps remain in understanding the full consolidation of HITL frameworks across adaptive learning, teacher development, reinforcement learning, and emotion recognition. Limited empirical studies are available on ethical alignment, trust calibration, interface design, scalability, and multi-domain generalization, highlighting the need for structured, human-centered, and context-aware implementation strategies.

The objective of this review is to comprehensively examine the alignment of AI and immersive technologies in education, with particular focus on HITL frameworks, adaptive learning models, and human-AI collaboration. It aims to evaluate applications in teacher development, personalized learning, RL, and emotion recognition while identifying ethical, pedagogical, and technological considerations. The study seeks to synthesize advances in human-centered AI, large language model (LLM) assimilation, and interactive intelligent systems, highlighting mechanisms for trust, transparency, interpretability, and fairness. By analyzing frameworks across education, autonomous systems, and industrial applications, the review identifies strategies that enhance engagement, reliability, and performance, ensuring scalable, responsible, and effective deployment of AI while preserving human judgment, agency, and ethical oversight in dynamic learning and professional environments.

METHODOLOGY

This study adopted a structured review methodology to explore recent advancements in educational practices. The review synthesized empirical and theoretical evidence for resilient and inclusive education. The methodology followed a process involving clearly defined inclusion and exclusion criteria, comprehensive literature searching, critical evaluation of eligible studies, thematic synthesis, and structured reporting.

Identify the Review's Purpose

The review was designed to evaluate the current state of human-centered digital advancement in education, identify emerging research trends, examine ethical and technological developments, and highlight future research directions supporting responsible AI unification in educational environments.

Set up Review Protocol

A review protocol was established before conducting the study to ensure transparency and reproducibility. The protocol defined the review objectives, search strategy, inclusion and exclusion criteria, selected databases, and qualitative synthesis procedures.

Searching and Screening of the Literature

A comprehensive literature search was conducted using Google Scholar, Scopus, Web of Science, PubMed, ScienceDirect, Springer, Taylor & Francis, Wiley, and MDPI databases. Boolean operators combined keywords including "Human-Centered Digital Innovation," "Human-in-the-Loop," "Artificial Intelligence," "Adaptive Learning," "Educational Technology," "Explainable AI," "Ethical AI," "Immersive Learning," and "Large Language Models." Only peer-reviewed journal articles published in English were considered. Non-English publications, conference abstracts, abstract-only papers, editorials, opinion articles, and studies lacking direct relevance were excluded. Titles, abstracts, and full texts were systematically screened to identify studies meeting the eligibility criteria. The methodological quality of the included studies was assessed using the Mixed Methods Appraisal Tool (MMAT), version 2018. The MMAT was selected because the review incorporated diverse forms of evidence, including qualitative, quantitative, mixed-methods, and nonrandomized studies. A total of 128 studies were identified across databases, with 15 duplicates removed. Of the remaining studies, 113 were screened, 92 underwent title and abstract screening, and 14 were excluded for irrelevance or lack of peer review. Subsequently, 78 full-text studies were assessed for eligibility, 9 were excluded due to insufficient details, lack of focus, or duplicate findings, resulting in 69 studies included in the final synthesis.

Data Extraction

Relevant information from eligible studies published between 2020 and 2026 was systematically extracted and organized into thematic categories addressing AI-driven educational technologies, human-centered design principles, ethical and transparent AI practices, adaptive learning systems, and equitable educational implementation. This structured coding facilitated consistent comparison and interpretation across studies.

Synthesis of Studies

A qualitative thematic analysis was applied to synthesize the selected literature, identifying recurring themes, technological advancements, implementation challenges, and research gaps.

Writing the Review

The findings were integrated into a structured review that presents current developments, practical implications, and future opportunities in global education through ethical, transparent, and self-sustaining technological practices (see **Figure 1**).

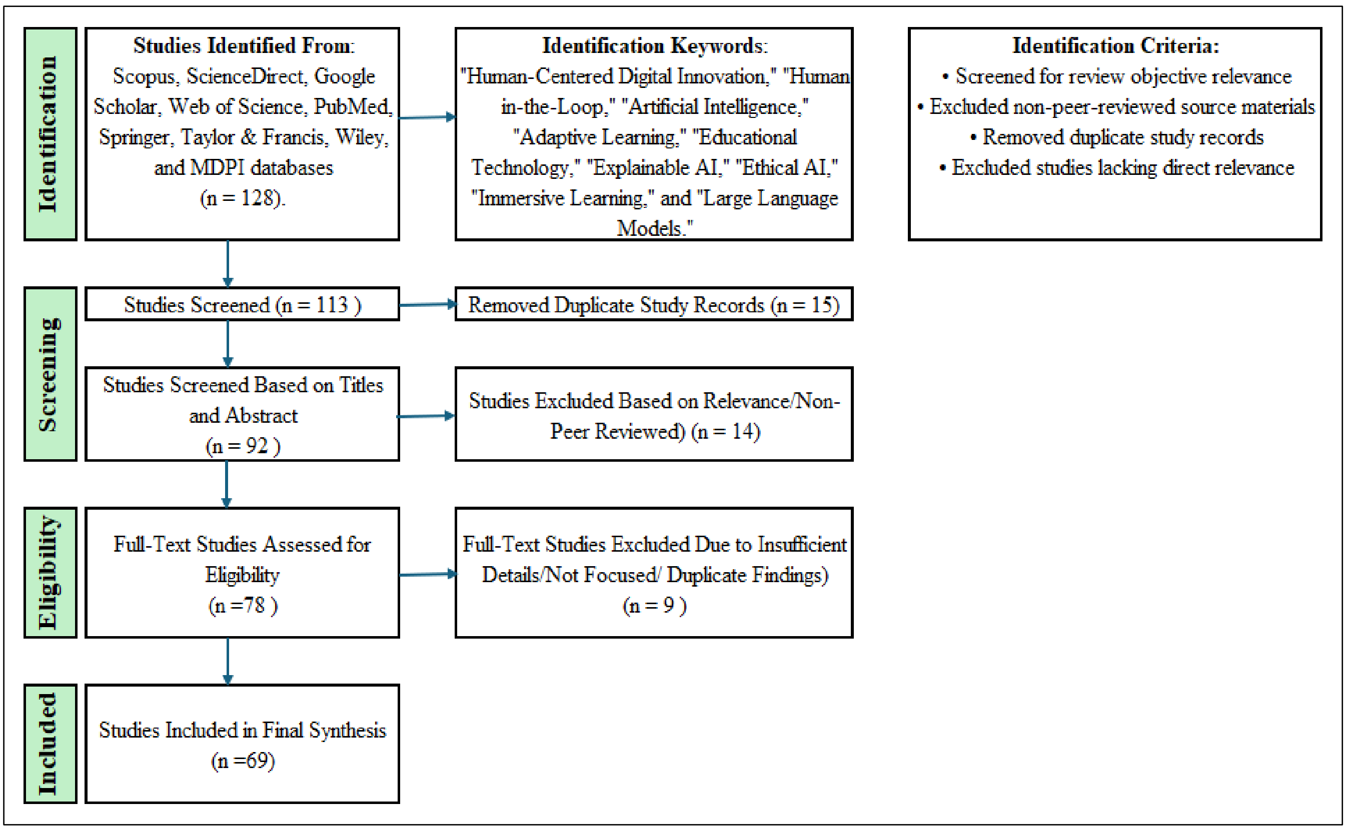

Figure 1. PRISMA flow diagram of the literature search and study selection process (Source: Author's own elaboration)

HUMAN-IN-THE-LOOP AI FOR EDUCATION AND PROFESSIONAL DEVELOPMENT

This section explores the role of HITL-AI in advancing education and professional development by integrating human expertise with AI to support learner-centered educational practices.

Human-AI Collaboration and Pedagogical Integration to Co-Adaptive Education

AI-enhanced education facilitates determining technological autonomy that can be balanced with human judgment. Ninaus and Sailer (2022) argued that rather than allowing AI to independently generate recommendations or implement interventions, hybrid systems should provide transparent AI-generated information that enables teachers and learners to retain decision-making agency. This approach acknowledges that both AI predictions and human judgments are fallible and that errors may have substantially different consequences depending on the educational context. Their iterative framework integrates human and data-driven selection of learner data, ML-based pattern detection, and adaptive recommendations or interventions, with subsequent learning outcomes feeding back into the system. This model positions AI as a co-adaptive partner rather than an autonomous authority, although its effectiveness depends on contextual sensitivity and continuous human oversight (Ninaus and Sailer, 2022).

The growing emphasis on Human-Centered Learning Analytics (HCLA) and Human-Centered AI (HCAI) reflects recognition that technological effectiveness cannot be separated from stakeholder needs and pedagogical contexts. Topali et al. (2025) found that stakeholder involvement is increasingly incorporated into solution design, but remains limited during ideation, prototyping, and evaluation. Similarly, Kottara and Asonitou (2025) demonstrated that thoughtfully designed AI-supported courses and adaptive learning strategies can enhance student engagement, motivation, and academic achievement while reducing boredom. Nevertheless, these findings suggest that personalization alone does not guarantee meaningful learning; AI must complement rather than replace teacher agency and pedagogical interaction. The limited empirical evaluation reported by Topali et al. (2025) therefore represents a significant gap, particularly concerning whether human-centered approaches translate into sustained adoption, user satisfaction, and improved educational outcomes.

Recent AIED research further indicates a transition from predominantly technical advancement toward human-centered and co-adaptive paradigms. Feng et al. (2025) identified LLMs, GenAI, multimodal learning analytics, and human-AI collaboration as emerging frontiers, while emphasizing personalization, self-regulated learning, feedback, assessment, motivation, ethics, and equity. However, the field's strong technical orientation and reliance on keyword metadata highlight the need for classroom-based, experimental, and qualitative research. Complementing this perspective, Luckin (2025) argued that AI can personalize learning, support teacher development, and automate administrative work, but requires educators to preserve human interaction, promote AI literacy and digital citizenship, and address privacy and ethical concerns. Collectively, these studies indicate that responsible AI integration requires moving beyond technological capability toward systems that preserve human agency, strengthen higher-order thinking, support equitable participation, and align AI perspectives with meaningful pedagogical objectives.

Trends and Transformative Potential of Human-in-the-Loop AI in Education

AI is reshaping education, but human involvement remains essential for ethical, equitable, and effective implementation. A Delphi study highlighted three critical areas: privacy-conscious AI use, transparent and trustworthy algorithms, and fairness in deployment. Collaboration among educators, policymakers, and researchers ensures AI enhances learning while maintaining human agency (Ifenthaler et al., 2024). Another major theme concerns the balance between AI-enabled personalization and human pedagogical judgment. AI-driven tools such as virtual dissections, adaptive feedback, and personalized learning pathways increase engagement, comprehension, and retention. AI has the potential to optimize instructional practices; however, its implementation should be balanced with traditional teaching approaches, experiential learning, and ethical oversight to avoid excessive dependence on technology and preserve meaningful human interaction (Joseph et al., 2025). Similarly, Wang (2021) argued that AI enhances educational leadership by improving analytical efficiency through data-driven insights while supporting, rather than replacing, human decision-making. Effective leadership requires balancing empirical evidence with ethical reasoning and contextual awareness. A human-centered approach, supported by transparent governance and continuous evaluation, ensures responsible AI adoption across diverse educational settings (Wang, 2021). This position complements the human-AI hybrid model proposed by Holstein et al. (2020), in which AI supports instructional decisions through data-driven insights while teachers provide contextual, pedagogical, and socio-emotional expertise. Overall, the studies complement one another, but they provide different forms of evidence. These studies converge on a human-centered vision of AI in education, emphasizing that technological effectiveness depends on ethical governance, explainability, and sustained human expertise. While AI strengthens decision-making, personalization, and instructional efficiency, limited empirical validation of hybrid frameworks underscores the need for rigorous evidence and evaluation of its practical effectiveness to ensure equitable, context-sensitive, and pedagogically meaningful implementation.

Human-in-the-Loop AI in Education for Enhanced Collaboration and Pedagogical Applications

The distribution and structure of human-AI roles as a collaborative rather than substitutive paradigm is one of the prominent themes. HITL and AI in education through an Entity-Relationship (ER) lens suggests underexplored human roles in educational AI systems (Memarian and Doleck, 2024). This observation is complemented by Ranade et al. (2025), who demonstrated that prompt engineering for AI-assisted writing benefits from HITL incorporation, emphasizing context, audience, purpose, and constraints. Using rhetorical situation theory, input prompts can be designed to guide outputs more ethically and effectively. User-centered HITL approaches empower professionals, improve collaboration, and highlight the need for broader research across genres, frameworks, and diverse applications (Ranade et al., 2025).

A second theme is related to the concerns about the uniformly distributed educational benefits and limitations of generative AI. GenAI tools, including ChatGPT, are drivers of research growth, primarily in higher education. Findings reveal cognitive and personal benefits, such as motivation and autonomy, but social, teacher-centered, and affective gains are underexplored. Ethical use, digital literacy, and pedagogical integrity remain key challenges for effective human-centered AI consolidation (Garzón et al., 2025). GenAI in education supports personalized learning while complementing human expertise. AI adapts content to individual learners, tracking progress and enhancing engagement, while teachers provide guidance and foster critical thinking. Successful implementation requires attention to privacy, fairness, and inclusivity, ensuring AI enhances rather than replaces human judgment in educational decision-making and learning design (Varakantham and Tio, 2025). Practical evidence further demonstrates how continuous human involvement can translate these principles into educational practice. The deployment of a Socratic GenAI tutor in undergraduate statistics courses demonstrates practical HITL incorporation. Collaboration with instructors enabled personalized instruction, iterative refinement, and ethical oversight. Evaluation shows improved student engagement and learning outcomes, highlighting the necessity of continuous human involvement in design, monitoring, and pedagogical guidance for effective AI-supported education (Qiu et al., 2025). These studies position HITL-AI as a collaborative paradigm that combines AI efficiency with human expertise to strengthen educational quality, personalization, and ethical governance. While evidence demonstrates improved engagement, writing, and adaptive learning, limited attention to diverse human roles, socio-emotional outcomes, and long-term validation highlights important directions for future research and implementation.

Human-in-the-Loop AI in Adaptive Learning and Knowledge Systems

HITL-AI plays a critical role in dynamic educational systems by integrating human expertise with intelligent algorithms to personalize instruction, enhance knowledge systems, and ensure responsive, learner-centered educational environments. Extended Reality (XR) training benefits from integrating HITL supervision, multimodal sensing, transparent AI explanations, and adjustable autonomy. The framework delivers interpretable and contestable scenario recommendations, enhancing task performance, user engagement, and trust. By integrating personalized guidance with human supervision, the composite design preserves human agency, supports situational awareness, and strengthens decision-making in safety-critical training environments (Pretolesi et al., 2026). A further theme concerns assessment, professional competency, and ethical governance. AI-driven assessment in Ibero-American higher education highlights teacher competencies and ethical safeguards. A function-by-purpose taxonomy differentiates pre-AI baselines from AI-specific mechanisms, emphasizing explainability, data stewardship, and clear assistance boundaries. Competency clusters address feedback literacy, fairness, rubric validation, integrity, and platform orchestration. Professional development is needed to ensure AI complements rather than replaces human judgment in assessment workflows (Fajardo-Ramos et al., 2025).

SPARC, a human-centered framework for learning spatial concepts, leverages human feedback and parametric similarity measures to improve learning efficiency, accuracy, and adaptability. Experiments reveal how different user characteristics influence outcomes and system adaptability. The approach generates multiple explanation types, enhancing user understanding and trust, and sets a foundation for human-aligned AI reasoning in complex spatial domains (Young et al., 2025). Adapting ML models to dynamic real-world data requires human-guided autonomy. Task-Agnostic Domain-Incremental Learning (TADIL), TADILER, CAMA, and KC-Agent frameworks enhance accuracy, continual learning, and safety. Human expertise supervises AI adaptation across domains like medical imaging, autonomous driving, and finance, achieving significant performance gains with minimal computational overhead (Bravo Rocca, 2025).

AI-driven leadership development integrates HITL approaches to preserve relational learning while enhancing personalized feedback and content sequencing. The concept of taxonomical leapfrogging allows AI to elevate leadership training, balancing task- and relationship-oriented skills. Implementing human-aware AI frameworks ensures ethical practice, quality assurance, and scalable adaptation to modern leadership contexts in both higher education and industry (Jenkins and Khanna, 2025). Across these studies, HITL-AI emerges as a valuable approach for adaptive learning and knowledge systems by combining algorithmic capabilities with human oversight, feedback, and ethical judgment. Although applications demonstrate improved personalization, trust, and performance across domains, challenges remain regarding scalability, competency development, explainability, and maintaining human agency in increasingly autonomous educational technologies.

AI-driven Teacher Development for Skill Enhancement and Professional Training

AI impacts teacher professional development by enabling personalized learning, skill advancement, and ethical practices. The findings revealed a moderate positive correlation between AI-based learning tools and training success, while regression analysis revealed limited predictive effects on teaching effectiveness, highlighting the need for complementary pedagogical strategies (Shezad et al., 2025). This finding introduces an important qualification to the generally positive evidence. In contrast, Sharma et al. (2025b) reported that AI improves efficiency, fairness, and strategic decision-making. Personalized goal setting, real-time feedback, and predictive analytics help identify talent while reducing bias. Adaptive learning and immersive technologies enhance employee engagement and retention, supporting diversity, equity, and inclusion initiatives. Effective implementation requires human-centered oversight, upskilling, and leadership frameworks to maximize benefits (Sharma et al., 2025b).

A second theme concerns teacher trust in AI-EdTech for adoption, especially in K-12 education. Experiments with AI-Grader show that explaining AI decision-making and emphasizing complementarity rather than replacement increases confidence. Professional development programs focusing on theoretical understanding, flexibility, and agency help educators integrate AI tools effectively, ensuring alignment with classroom needs and ethical practices (Nazaretsky et al., 2022). This finding complements the broader human-centered arguments of Sharma et al. (2025a), which concluded that AI improves by automating routine tasks, enabling data-driven decisions, and enhancing employee experience through personalized training and performance feedback. While offering substantial efficiency and engagement gains, ethical implementation, change management, and data governance remain crucial. Human-centered approaches balance technological excellence with inclusivity, trust, and long-term sustainability in AI-driven workplaces (Sharma et al., 2025a).

Similarly, the combination of AI with the Internet of Things (AIoT) increases demand for AI competencies among teachers. AIoT professional development program demonstrated significant improvements in participants' knowledge, skills, and attitudes toward AI applications. Structured training empowers educators to teach AI concepts effectively, bridging gaps in understanding and supporting technology-driven learning environments (Kitcharoen et al., 2024). These studies demonstrate that AI-driven professional development enhances teacher competencies, personalized learning, and decision-making through adaptive feedback and analytics. However, successful implementation depends on sustained training, trust-building, ethical governance, and human-centered strategies. Evidence suggests AI should augment educators' capabilities rather than replace professional expertise, ensuring equitable technology assimilation.

TRUSTWORTHY, ETHICAL, AND HUMAN-CENTERED AI

This section explores the principles and practices required to develop trustworthy, ethical, and human-centered AI systems that promote responsible incorporation within educational contexts.

Ethical, Transparent, and Trust-Calibrated Human-AI Collaboration in Education

AI in education presents transformative learning opportunities but entails multi-dimensional ethical risks. Technological threats include privacy breaches, algorithmic bias, and computational errors. Educational risks involve homogenized learning, teacher-student alienation, and academic misconduct. Societal risks include widening digital divides and accountability gaps. Mitigation strategies emphasize standardization, ethical guidelines, and empirical validation for responsible deployment (Zhu et al., 2025). This broader perspective complements Explainable AI (XAI), which enhances trust, transparency, and fairness in AI-supported learning, enabling adaptive, human-centered instruction. Personalized explanations respond to student feedback, improving engagement and comprehension. However, challenges include algorithmic bias, limited interpretability, privacy, and equitable access. Stakeholder collaboration, continuous evaluation, and ethical design ensure socially responsible AI unification that promotes inclusive, effective, and context-sensitive education (Shafik, 2024). This is supported by another study, which emphasizes that XAI is critical for educational contexts where explanations must be pedagogically meaningful. The study traces XAI development from early interpretability models to advanced frameworks, noting gaps in cognitive alignment, ethical considerations, and user engagement. Future research should enhance pedagogical utility while preserving algorithmic transparency and supporting student learning (Rachha and Seyam, 2023).

To balance under-reliance and over-reliance on AI systems, trust calibration is essential. Interventions that align user trust with system capabilities can enhance human-AI interaction. Although measurement limitations and study design affect result interpretation, effective trust calibration improves appropriate reliance on automation, providing valuable insights for future research to strengthen performance, safety, reliability, and decision-making in AI-supported educational and operational environments (Wischnewski et al., 2023). HITL Intelligent Testing (HITL-IT) incorporates human factors into AI-based software testing, accounting for cognitive attributes, domain knowledge, and intuitive judgment. The three-layer framework- human-factor modeling, AI testing core, and interactive feedback loop supports iterative refinement through suggestion, challenge, and relearning cycles. Preliminary evaluations indicate reduced error rates, improved safety, and enhanced reliability, demonstrating that HITL-IT balances automation efficiency with human insight to create resilient and trustworthy intelligent testing systems (Xue and Wang, 2025). Similarly, an HITL learning framework for bi-manual robot manipulation combines interactive teaching with autonomous execution. Through compliant bilateral teleoperation, the system enables precise human interventions, richer task-relevant demonstrations, and corrective feedback. Evaluation in simulation and real-world experiments shows faster skill acquisition, higher task performance, and greater flexibility for novel tasks, demonstrating that interactive human-robot collaboration effectively bridges human expertise with autonomous policy learning for complex manipulation scenarios (Wu et al., 2025).

Collaborative human-AI teams (CHAI-T) depend on dynamic trust, evolving with tasks, goals, and interaction processes. The CHAI-T framework incorporates relational trust, co-adaptation, and iterative calibration, highlighting the importance of human-like team processes. By capturing trust's dynamic nature, CHAI-T supports effective collaboration, safety, and advancement, offering actionable guidance for designing trustworthy human-AI teamwork in diverse environments (McGrath et al., 2025). Integrating these findings indicates that trustworthy AI in education depends on combining HITL methodologies, explainable systems, and ethical governance. Although human oversight enhances reliability and adaptive performance, unresolved issues related to scalability, bias, privacy, and equitable access require further investigation.

Human-Centered Ethical AI Integration in Education

Responsible AI adoption in education requires a comprehensive approach integrating fairness, ethics, and human-centered governance. Memarian and Doleck (2023) identified that while fairness dominates higher education AI research, accountability and transparency remain insufficiently examined, revealing conceptual and methodological gaps in responsible AI frameworks. Similarly, Fu and Weng (2024) expanded this perspective by emphasizing five interconnected principles-fairness and equity, privacy and security, non-maleficence and beneficence, agency and autonomy, and intelligibility-as essential foundations for human-centered AI. Together, these findings suggest that ethical AI implementation extends beyond technical performance and requires meaningful collaboration among educators, students, policymakers, and developers. However, both studies highlight limitations related to fragmented definitions, limited empirical evidence, and inadequate longitudinal research, restricting the ability to evaluate AI governance practices across diverse educational settings.

Further synthesis reveals that user perspectives and emerging GenAI challenges require stronger ethical frameworks and institutional responses. An et al. (2024) critically argued that existing AI ethics models often overlook user agency, learning analytics, and algorithmic decision-making, emphasizing the need for stakeholder-informed approaches. Extending this discussion, García-López and Trujillo-Liñán (2025) demonstrated that GenAI introduces additional concerns related to privacy, bias, inequality, learner autonomy, and regulatory limitations. Although proposed strategies such as ethics committees, digital literacy development, and human oversight provide valuable directions, their practical effectiveness requires further validation. Future research should prioritize empirical evaluations, context-sensitive governance models, and interdisciplinary collaboration to ensure AI systems promote equitable access, preserve learner autonomy, and strengthen educational outcomes while minimizing social and ethical risks.

Human-Centered Design and Robot Collaboration

AI development across education and mental health highlights the importance of human intelligence and artificial wisdom for responsible AI aggregation, leading to advances in adaptive collaboration, emotionally intelligent systems, and human-AI decision-making models. Emphasis is placed on ethical design, trust, and inclusivity to ensure AI augments human judgment and fosters equitable, resilient socio-technical ecosystems (Singh, 2026). Building on this perspective, HCAI requires structured methodologies for practical implementation, addressing ethical, social, and user-centric considerations. The HCAI-MF framework provides a five-component approach, including hierarchical requirement structuring, method taxonomies, process workflows, interdisciplinary collaboration, and multilevel design paradigms. A case study demonstrated improved stakeholder engagement and alignment with human values, while challenges in coordination and adaptation are discussed (Xu et al., 2026).

Robotics research provides more direct empirical evidence for dynamic human-AI control allocation. HITL Task and Motion Planning (HITL-TAMP) combines human guidance with automated TAMP to efficiently train robots in complex, contact-rich tasks. By dynamically allocating control between human operators and robotic systems, the framework enables rapid collection of high-quality demonstrations and effective imitation learning. Experimental results demonstrate superior task success, scalability, and reduced training time, illustrating practical benefits for real-world robotic manipulation (Mandlekar et al., 2023). The Sirius framework further advances human-robot collaboration through a division-of-labor approach, where robots autonomously handle routine decisions, and humans intervene in complex scenarios. Weighted behavioral cloning optimizes policy performance based on human trust, improving real-world success rates by 27% and memory efficiency by 85%. Sirius enables safe, efficient robot deployment while minimizing human cognitive load and enhancing shared autonomy in dynamic environments (Liu et al., 2025). These studies demonstrate that human-centered and HITL frameworks are essential for developing adaptive, ethical, and trustworthy AI systems across education, mental health, and robotics. While advances in collaboration, autonomy, and performance are evident, challenges remain in coordination, scalability, and value alignment. Future research should strengthen interdisciplinary approaches that preserve human agency while optimizing AI capabilities.

ADVANCED HUMAN-AI INTERACTION AND INTELLIGENT SYSTEMS

This section evaluates emerging developments in advanced Human-AI interaction and intelligent systems, focusing on how human expertise can be integrated with AI capabilities to create collaborative and context-aware technologies. It explores HITL approaches across educational practices, large language models, emotion recognition, and RL, emphasizing their role in enhancing personalization, real-time support, and effective human-centered decision-making.

Advancing Human-AI Interaction and Human Expertise in Intelligent Systems

Advancing human-AI interaction in education involves designing intelligent systems that effectively combine AI capabilities with human expertise to support teaching, learning, and decision-making. By integrating educators' contextual knowledge, ethical judgment, and pedagogical experience with AI-driven analytics, these systems can enhance learning personalization, improve instructional strategies, and promote more effective educational outcomes. However, the conceptual and empirical advances in human-AI interaction have identified four central challenges:

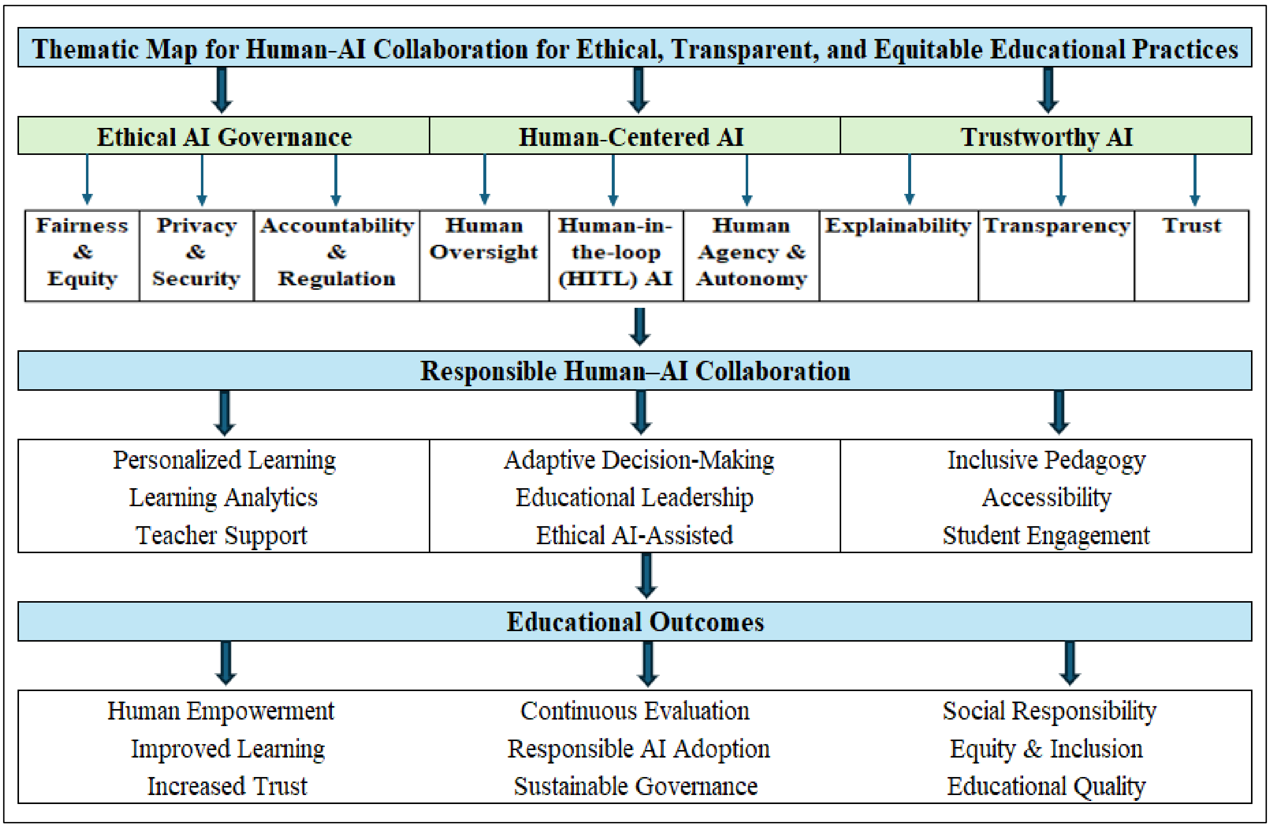

Figure 2. Conceptual framework of Human-AI collaboration in educational practices (Source: Author's own elaboration)

conversation, collaboration, explainability, and accountability, alongside fairness and bias. While these domains dominate current scholarship, agentic AI and interface design remain underexplored. The framework outlines research priorities to guide responsible, theory-driven human-AI collaboration (Abedin et al., 2022).

The importance of interface design and usability complements this broader human-centered perspective. Evolving web technologies, mobile platforms, and AI systems demand Human-Computer Interaction strategies that accommodate continuous updates and personalization. Grounded in User-Centered Design and Interaction Design principles, the proposed framework integrates usability heuristics with empirical validation, offering structured guidance for developing consistent, efficient, and user-aligned intelligent interfaces across dynamic digital environments (Costa, 2022). Evidence from human-centered robot learning research further strengthens this argument and emphasizes empowering novice users through improved communication and interpretability in Learning from Demonstration systems. A three-dimensional augmented reality taxonomy and targeted UI Interventions enhance information exchange, while competitive interaction styles increase engagement and efficiency. Findings highlight the importance of adaptable, context-sensitive design for inclusive human-robot collaboration (Phaijit, 2025). These studies highlight that effective human-AI interaction depends on integrating human expertise with explainable and user-centered intelligent systems. While frameworks addressing collaboration, accountability, and interface design provide valuable foundations, gaps remain in agentic AI, usability, and inclusive interaction. Future developments should prioritize interpretability, ethical alignment, and continuous human engagement to ensure responsible AI evolution. Figure 2 presents the conceptual framework of Human-AI collaboration in educational practices, illustrating the key dimensions, relationships, and processes that support ethical, human-centered, and effective AI alignment in teaching, learning, and educational decision-making.

Human-in-the-Loop AI for Student-Centered Educational Practices

AI-supported curriculum adaptation enables flexible and personalized instruction; however, evidence remains limited regarding long-term learning impacts (Karataş et al., 2025). Explainable HITL approaches further address concerns, although current XAI methods often prioritize performance explanations over provenance and user interpretability (Zhang et al., 2025). Student-driven HITL frameworks integrating feedback loops, prompt engineering, and Retrieval-Augmented Generation (RAG) demonstrate potential for enhancing personalized learning, learner agency, and comprehension, particularly in STEM education (Tarun et al., 2025). Similarly, Human-AI Cooperation (HAC) frameworks highlight the complementary strengths of teachers and AI in adaptive decision-making and instructional responsiveness (Chen, 2022). Table 1 summarizes key outcomes of HITL-AI applications in education, highlighting the application domains, principal findings, and reported limitations across the reviewed studies.

Broader HITL research indicates that human expertise improves AI reliability and interpretability (Kumar et al., 2024), whereas HITL testing frameworks for LLM-based

Table 1. Key outcomes of human-in-the-loop AI applications in education

Author (Year)	HITL AI Application	Key Findings	Limitations
Karataş et al. (2025)	GenAI-supported curriculum adaptation	AI enables flexible, personalized, student-centered curriculum design and improves engagement.	Limited evidence on long-term educational outcomes and implementation challenges.
Zhang et al. (2025)	Explainable HITL pipelines for knowledge graph construction	Human oversight improves the transparency, fairness, accountability, and interpretability of AI systems.	Current XAI approaches emphasize performance explanations over provenance and user understanding.
Tarun et al. (2025)	HITL GenAI for personalized learning using feedback loops and RAG	Student feedback improves AI responses, engagement, confidence, comprehension, and learning outcomes.	Preliminary studies; limited validation across diverse learners and subjects.
Chen (2022)	Human-AI cooperation (hybrid intelligence) in education	Combining human expertise and AI enhances adaptive decision-making and instructional responsiveness.	Requires stronger empirical validation and theoretical refinement.
Kumar et al. (2024)	General HITL machine learning systems, including education	Human expertise improves AI accuracy, robustness, transparency, and trustworthiness through iterative feedback.	Challenges include bias, data quality, interface design, and accountability.
Kathiresan (2025)	HITL testing framework for LLM-integrated applications	Human evaluation combined with automated testing detects inconsistencies and reliability issues.	Focuses mainly on software testing; limited educational application evidence.
Molenaar (2022)	Hybrid human-AI intelligence for teaching and learning	AI augmentation supports teacher-learner collaboration, self-regulated learning, and future educational innovation.	Conceptual framework requiring further empirical evaluation.
Shin et al. (2025)	Interactive machine learning (IML) with human feedback mechanisms	Personalized feedback loops improve user engagement, satisfaction, and AI model refinement.	Small sample size and non-education-specific application context.

applications demonstrate improved detection of hallucination issues (Kathiresan, 2025). Hybrid intelligence models emphasize stakeholder collaboration and balanced AI augmentation rather than the replacement of educators, yet require practical implementation studies (Molenaar, 2022). Furthermore, adaptive feedback mechanisms enhance user engagement, but effectiveness depends on context and personalization, requiring broader educational validation (Shin et al., 2025). Overall, HITL AI offers substantial promise, but rigorous longitudinal and cross-context research is needed.

Human-in-the-Loop Combination with Large Language Models

HITL incorporation with LLMs in education combines automated language generation with human oversight to ensure accuracy, pedagogical relevance, and ethical use in learning environments. Educators review, guide, and refine AI-generated explanations, feedback, or content, improving reliability while aligning outputs with curriculum goals and students' learning needs. Combining Belief-Desire-Intention (BDI) agents with LLMs enhances human-robot interaction. LLMs translate user commands into actionable goals while BDI agents ensure verifiable reasoning, safety, and plan management. Experiments in robot navigation show improved accuracy, explainability, and minimal computational overhead, highlighting adaptable applications for complex systems (Frering et al., 2025). The literature also demonstrates that integrating LLMs into HITL frameworks improves functional safety risk analysis under ISO 12100. Four industrial case studies-motorized gates, autonomous vehicles, weaving machines, and rotary presses-demonstrate that iterative human refinement corrects hallucinations, enhances hazard detection, and achieves agreement with expert ground truth. Time efficiency, usability, and accuracy ratings confirm the practical value of structured HITL deployment (Iyenghar, 2025).

LLMAuditor exemplifies HITL auditing for LLMs, addressing bias and inconsistencies. By generating probes with HIL-validated templates and a dual-LLM setup, human effort is minimized while ensuring reliability. Case studies confirm the detection of knowledge inconsistencies and provide scalable, transparent auditing solutions. This demonstrates that human-augmented LLM evaluation improves quality assurance rigor (Amirizaniani et al., 2025). HITL-LLM workflows improve knowledge graph validation, achieving up to 87% precision in large-scale KG construction. Hybrid approaches selectively combine automated validators with human oversight to enhance accuracy and reduce expert workload. Experiments in the Computer Science Knowledge Graph show human-AI collaboration improves F1 scores, balancing efficiency, precision, and recall while mitigating LLM limitations (Tsaneva et al., 2025).

Despite advances in large language models, human expertise remains indispensable in post-training refinement, bias mitigation, and domain adaptation. Specialists ensure reliability across low-resource languages, multimodal contexts, and human-robot applications. While LLMs assist annotation and generation, they lack judgment and contextual reasoning, necessitating sustained human oversight for trustworthy deployment (Eckert et al., 2025). These studies confirm that HITL assimilation strengthens LLM reliability by combining generative capabilities with human judgment, validation, and domain expertise. While applications in education, robotics, auditing, and knowledge systems show improved accuracy, safety, and interpretability, challenges remain in scalability, workload management, and contextual reasoning. Future research should optimize human oversight strategies for trustworthy, efficient LLM deployment.

Human-in-the-Loop Approaches for Enhanced Emotion Recognition

HITL frameworks enhance emotion recognition in education by combining AI with human expertise to interpret students' emotional cues more accurately during learning activities. Teachers can review and refine AI-generated insights, ensuring that emotional assessments are contextually appropriate and used to support responsive, empathetic, and personalized learning environments. Emotion recognition from facial expressions is crucial for applications in HCI, mental health, and robotics. This study introduces a Vision Transformer enhanced with an HITL framework to improve accuracy and generalization. By incorporating human feedback selectively through confidence-based interventions, the system corrects mispredictions and refines decision boundaries efficiently (Hasan et al., 2025). HITL annotation frameworks combine automated predictions with human judgment to improve emotion estimation. Reliable model outputs foster high trust and efficiency, whereas deliberate errors increase cognitive load and variability. Cognitive framing influences perception, highlighting psychological factors in human-AI collaboration for annotation tasks (Subramanya et al., 2025).

HITL emotion annotation emphasized collaborative human-machine interactions. Results indicate that dependable machine outputs, alongside human oversight, enhance annotation accuracy, engagement, and reliability. Negative cognitive framing can bias perceptions, making models appear more accurate, demonstrating the need for balancing automation with contextual human input (Yadnakudige Subramanya et al., 2025). Human-in-the-Loop Cyber–Physical Systems (HiTLCPS) frameworks extend HITL concepts to cyber-physical systems, integrating humans in data acquisition, inference, and actuation. Case studies show improved responsiveness, usability, and trust. Challenges include limited participation, engagement sustainability, and privacy. Emerging tools like smartphones and unobtrusive sensors enable human-centered systems for scalable, ethical, and effective implementations (Fernandes et al., 2025). These studies indicate that HITL emotion recognition frameworks improve AI accuracy, interpretability, and user trust by integrating computational analysis with human contextual judgment. Although adaptive feedback and human oversight enhance reliability, challenges remain regarding cognitive bias, privacy, engagement, and scalability.

Advances in Human-in-the-Loop Reinforcement Learning for Real-Time Guidance

Human-AI collaboration in educational practices integrates human judgment with adaptive AI guidance. Advances in HITL reinforcement learning (HITL RL) extend this approach, enabling real-time, personalized feedback and interactive decision support that further enhances learner engagement and outcomes. HITL RL incorporates human feedback to optimize agent behavior, as seen in RLHF applications like ChatGPT. Structured phases of human involvement - development, learning, evaluation, and deployment enhance alignment, performance, and transparency. XAI techniques improve interpretability for diverse users, while promising strategies advance effective human-agent collaboration in autonomous systems (Retzlaff et al., 2024). HITL and RL with Auxiliary Task (HIAT) address sample imbalance in HIRL through an auxiliary task for policy discrimination. Adaptive weighting (AWHIAT) prioritizes this task early in training, improving sample efficiency and policy performance. Evaluations on MuJoCo benchmarks show 34.3% improved efficiency and 7.02% higher reward. The framework reduces reliance on expert guidance while enhancing learning outcomes (Niu et al., 2025). Post-training adaptive optimization improves time-series forecasting without retraining models. Output transformations correct bias dynamically using reinforcement learning, bandits, or evolutionary strategies. HITL feedback via natural-language guidance enhances interpretability. Evaluations across electricity, weather, and traffic datasets show consistent accuracy improvements with low computational cost, enabling scalable and reliable forecasting deployment (Tiomoko et al., 2025).

A hierarchical HITL deep reinforcement learning (DRL) framework combines self-learning, imitation, and transfer learning with structured human input through reward shaping, action guidance, and demonstrations. Experiments using the Cogment platform in UAV defense scenarios show accelerated convergence, improved policy quality, and robust responses to adversarial conditions, illustrating scalable, real-world multi-agent applicability of HITL RL (Arabneydi et al., 2025). A shared autonomy framework supports HITL policy alignment in cognitively demanding robotic operations by treating human operators and autonomous agents as collaborative decision-makers. The framework combines hierarchical short- and long-term adaptation through Bayesian inference and deep RL to enhance coordination, adaptability, and decision-making performance. Multiple interaction channels facilitate seamless perception-decision-action coordination. Experimental evaluation on physical pick-and-place tasks demonstrates improved system performance, robust policy alignment, and enhanced user satisfaction, highlighting the value of holistic co-design for high-level robotic collaboration (Yousefi et al., 2025). These studies indicate that HITL RL enhances AI systems by integrating human feedback, guidance, and expertise into learning processes. While evidence highlights improvements in efficiency, alignment, interpretability, and decision-making, challenges remain in scalability, feedback quality, and interaction complexity.

DISCUSSION AND IMPLICATIONS

This section highlights key findings on human-AI collaboration in education, while also revealing persistent research gaps and implementation challenges related to trust, governance, accessibility, and responsible consolidation. The discussion that follows examines the broader implications of human-centered AI for educational systems and explores prospects for developing learning environments with human values.

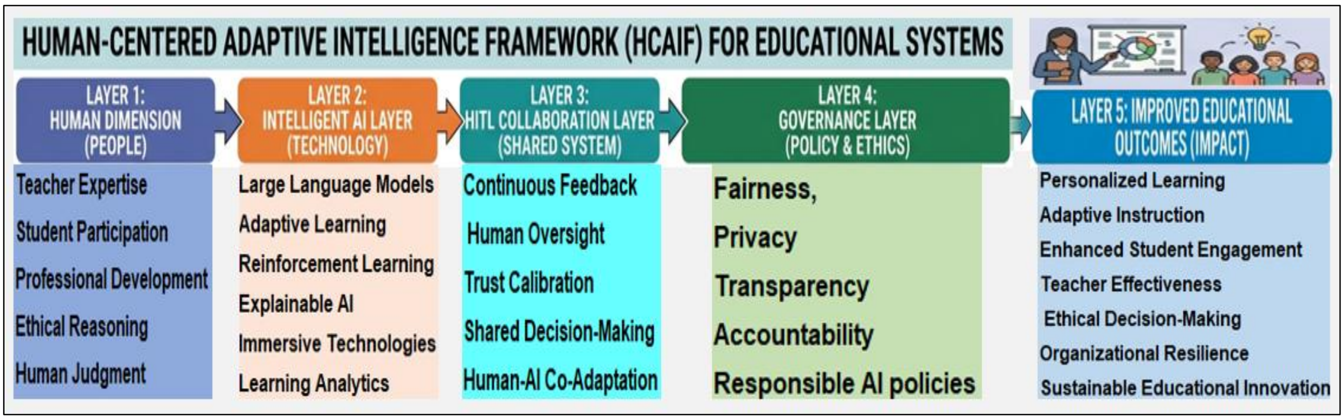

Figure 3. Schematic overview of Human-Centered Adaptive Intelligence Framework (HCAIF) (Source: Author's own elaboration)

Key Findings of the Study

Proposal of Human-AI Adaptive Synergy Theory (HAAST)

Based on the reviewed studies, Human-AI Adaptive Synergy Theory (HAAST) can be proposed for long-term educational advancement through the continuous co-adaptation of human expertise and AI within collaborative learning environments. Rather than positioning AI as an autonomous instructor or merely a decision-support system, the theory conceptualizes humans and AI as complementary co-intelligence partners that continuously learn from one another through iterative feedback, contextual adaptation, and shared decision-making. Effective human-AI collaboration depends on the dynamic interaction of five interrelated dimensions: human expertise encompassing pedagogical, ethical, and contextual knowledge; adaptive AI capabilities including learning analytics, personalization, and predictive intelligence; continuous HITL feedback; explainable and trustworthy AI that promotes transparency and accountability; and ethical governance that ensures fairness and responsible implementation.

The theory further predicts that strengthening these interconnected dimensions enhances learner engagement, instructional quality, trust in AI-supported decisions, fairness, and long-term educational sustainability, ultimately fostering resilient, inclusive, and human-centered educational ecosystems capable of adapting to evolving technological and pedagogical needs. HAAST is a conceptual theory that conceptualizes human expertise and AI capabilities as dynamically and bidirectionally co-adaptive. Unlike conventional HITL models, where humans primarily oversee, validate, or provide feedback on AI systems, HAAST positions humans and AI as complementary co-intelligence partners engaged in continuous reciprocal learning, contextual adaptation, iterative feedback, and shared decision-making. The theory emphasizes continuous feedback loops in which AI outputs are adaptively refined through human contextual knowledge, while human judgments and decision-making are simultaneously enhanced through AI-generated insights.

Development of a Human-Centered Adaptive Intelligence Framework (HCAIF)

This study develops a Human-Centered Adaptive Intelligence Framework (HCAIF) that provides a comprehensive analytical framework for evaluating and designing Human-AI educational systems through five interconnected layers. The first, the Human Dimension, encompasses teacher expertise, student participation, professional development, ethical reasoning, and human judgment. The second, the Intelligent AI Layer, incorporates large language models, adaptive learning, reinforcement learning, explainable AI, immersive technologies, and learning analytics. The third, the HITL Collaboration Layer, emphasizes continuous feedback, human oversight, trust calibration, shared decision-making, and human-AI co-adaptation. The fourth, the Governance Layer, focuses on fairness, privacy, transparency, accountability, and responsible AI policies. These layers collectively contribute to the fifth layer with improved educational outcomes, including personalized learning, adaptive instruction, enhanced student engagement, teacher effectiveness, ethical decision-making, organizational resilience, and persistent educational improvement.

HCAIF provides a practical foundation for evaluating, designing, researching, and governing future human-centered AI-enabled educational systems. HCAIF translates the theoretical principles of HAAST into a practical, multi-layer framework comprising the human dimension, intelligent AI layer, HITL collaboration layer, governance layer, and educational outcomes. Compared with existing frameworks, HCAIF highlights its distinctive integration of adaptive intelligence, continuous human-AI co-adaptation, trust calibration, governance, and measurable educational outcomes. Overall, HAAST provides the theoretical explanation of how human-AI synergy evolves dynamically, whereas HCAIF provides an operational structure for designing, evaluating, and governing such systems. **Figure 3** presents a schematic overview of the HCAIF, illustrating the interconnected components, processes, and stakeholder interactions that support adaptive, ethical, transparent, and human-centered AI assimilation in educational systems.

Identified Research Gaps and Challenges in Implementing Human-Centered AI for Educational Systems

Human-AI collaboration in education represents a significant shift toward personalized and intelligent learning ecosystems by integrating computational capabilities with human expertise. Evidence across HITL frameworks, immersive technologies, reinforcement learning, and LLM applications indicates that AI can enhance instructional personalization, decision-making, feedback quality, and learner engagement. However, critical evaluation reveals that technological advancement alone does not guarantee educational effectiveness. Successful implementation depends on pedagogical alignment, educator involvement, ethical governance, and maintaining human agency. Hybrid intelligence models demonstrate potential by combining AI efficiency with human contextual understanding, enabling trustworthy and adaptable educational systems. Nevertheless, current research remains fragmented, with limited empirical validation across diverse educational contexts and insufficient investigation of long-term outcomes.

Despite promising developments, implementing human-centered AI introduces complex technological, ethical, and organizational challenges. Data privacy, algorithmic bias, unequal access, limited digital literacy, and inadequate educator preparation remain persistent barriers. Furthermore, HITL systems require effective mechanisms for human engagement, feedback quality control, cognitive workload management, and transparency. The incorporation of large language models introduces additional concerns regarding hallucinations, explainability, and curriculum relevance, requiring continuous human monitoring and refinement. While current frameworks emphasize collaboration, gaps remain in understanding social-emotional impacts, teacher-centered outcomes, and scalable implementation strategies across varied educational environments.

Implications of Human-Centered AI in Education

The unification of AI and immersive technologies in education has practical implications for policy, curriculum design, professional development, and institutional governance. Rather than demonstrating that AI automatically produces better educational outcomes, the reviewed studies show that by embedding HITL frameworks, organizations can enhance ethical oversight, transparency, and stakeholder engagement while promoting personalized feedback and hybrid intelligence. These approaches enable educators to guide AI outputs, ensuring pedagogical alignment, fairness, and accountability, while fostering student engagement, trust, and inclusive access. The findings have implications for institutional AI governance and policy and have direct implications for curriculum, assessment, and instructional design. Teacher professional development should become an explicit component of institutional AI implementation. The findings support the design of HITL mechanisms that allocate human intervention strategically rather than continuously. The evidence on explainability and trust should be treated as functional requirements of educational AI rather than optional technical features. HITL RL and LLM-guided workflows improve decision-making, operational efficiency, and interpretability in high-stakes domains, including robotics, industry, and cyber-physical systems. Emotion recognition, immersive simulations, and interactive learning platforms benefit from structured human-AI collaboration, enhancing context-aware instruction and empathetic learning interventions. Collectively, these frameworks inform institutional strategies, professional training programs, and technology governance, supporting scalable, responsible, and stable AI adoption. The evidence supports a cautious approach to scaling AI across educational and operational environments. Educational institutions and organizations can leverage HITL methodologies to optimize system reliability, maintain ethical standards, and empower both educators and learners, creating accountable and human-centered ecosystems that align technological progress with pedagogical objectives, workforce readiness, and long-term organizational resilience.

Prospects for Adaptive and Ethical Educational Systems

Prospects for HITL-AI in education and complex systems emphasize ethical and human-centered advancement. Emerging intelligent tutoring systems, immersive simulations, and hybrid learning platforms can further personalize instruction, enhance collaborative decision-making, and strengthen teacher competencies. Advancements in multimodal sensing, context-aware personalization, and explainable AI will enable scalable, interpretable, and trustworthy human-AI partnerships. Hierarchical, multi-agent HITL-RL frameworks, combined with diffusion-based policies, post-training optimization, and natural-language feedback, can result in enhanced efficiency, robustness, and real-world applicability across education and Industry 4.0. Integrating agentic AI, multilingual LLMs, adaptive augmented reality interfaces, and unobtrusive human feedback can expand inclusivity, engagement, and equitable access while preserving human agency. Interdisciplinary collaboration, theory-driven validation, and human-aware interface design can improve usability, accountability, and ethical governance.

The reviewed evidence identifies several specific research directions for advancing HITL-AI in education. Future research should establish whether AI-supported personalization produces sustained educational benefits. Continuous research into adaptive HITL-LLM frameworks, knowledge graph validation, and automated human-feedback prioritization can refine AI outputs and ensure pedagogical alignment. Future research should strengthen human-AI educational systems through interdisciplinary investigations of adaptive co-learning, where humans and AI continuously influence and improve each other's capabilities. Longitudinal studies are needed to examine trust development, usability, learning outcomes, and professional advancement across diverse educational contexts. Research should further explore human-centered agentic AI by defining appropriate autonomy, intervention mechanisms, and ethical responsibilities. Integrating emotional intelligence, affective computing, and multi-agent collaboration can expand HITL systems beyond cognitive support toward socially responsive learning environments. Future frameworks should incorporate adaptive ethical governance, explainable AI, transparent decision-making, and equitable deployment strategies. Research should develop adaptive mechanisms for determining when human intervention is necessary and should investigate trust as a dynamic outcome rather than assuming that greater trust is always desirable. Advancing HITL approaches requires evidence-based design, continuous evaluation, and human expertise to ensure AI remains a supportive partner that enhances, rather than replaces, educators while promoting inclusive and self-sustaining educational ecosystems. Future research should broaden evaluation beyond cognitive and performance outcomes. Further exploration of social, affective, and teacher-centered outcomes alongside immersive and GenAI applications can balance technological efficiency with ethical, human-guided decision-making, supporting high-performance and inclusive educational and operational environments.

CONCLUSION

This review synthesizes the untapped value of AI and immersive technologies in education, emphasizing the critical role of human-centered and HITL frameworks. By incorporating adaptive learning systems, interactive experiences, and collaborative human-AI approaches, educational environments can achieve personalized, engaging, and equitable learning outcomes. Effective implementation requires ethical safeguards, transparency, and continuous professional development to ensure AI enhances, rather than replaces, human expertise. Across diverse educational and organizational contexts, AI-driven tools, ranging from adaptive feedback and virtual simulations to generative models, demonstrate improved engagement, comprehension, and performance. However, challenges including algorithmic bias, privacy concerns, educator competencies, and infrastructure limitations must be addressed through responsible governance, interdisciplinary collaboration, and scalable design strategies. HITL reinforcement learning and explainable AI frameworks further support ethical, trustworthy, and context-aware intelligent systems, demonstrating applicability in education.

This review highlights that durable AI integration in education depends on balancing technological improvement with pedagogical alignment, ethical oversight, and human agency. Embedding human judgment, iterative feedback, and trust calibration ensures adaptive, interpretable, and socially responsible AI deployment. This study proposed HAAST for human-AI co-adaptation, positioning humans and AI as complementary co-intelligence partners learning through feedback, contextual adaptation, and shared decisions. This bidirectional synergy extends HITL beyond oversight toward learning and educational advancement. The study also developed an HCAIF framework that translates HAAST into five interconnected layers for human-centered educational systems. It supports design, evaluation, research, and governance, connecting theory with practice effectively. Future research should focus on enhancing social, emotional, and cognitive dimensions of learning, expanding scalable human-AI frameworks, and promoting interdisciplinary collaboration among educators, policymakers, and technologists. By fostering inclusive, transparent, and accountable AI-assisted systems, HITL approaches can strengthen learner engagement, optimize educational outcomes, and support organizational resilience. Ultimately, combining intelligent technologies with human expertise creates adaptive, ethical, and effective educational ecosystems that support learners and professionals while advancing responsible AI innovation in evolving digital learning environments.

Author notes: The statements, opinions, and data presented in this publication are solely those of the author and are based on the analysis of previously published research. They do not reflect the views or positions of the author's affiliated institution.

Funding: No external funding is received for this article.

Ethics declaration: The author stated that ethical approval was not required for this study because it was a review of existing literature. No personal or sensitive data was collected.

AI statement: The author stated that, after the research was completed, the AI tool ChatGPT was used to improve the language quality and readability of the final manuscript.

Declaration of interest: The author declared no competing interest.

Availability of data and materials: All data generated or analyzed during this study are available for sharing when appropriate request is directed to the author.

REFERENCES

Abedin, B., Meske, C., Junglas, I., Rabhi, F., & Motahari-Nezhad, H. R. (2022). Designing and managing human-AI interactions. *Information Systems Frontiers, 24*(3), 691-697. https://doi.org/10.1007/s10796-022-10313-1

Amirizaniani, M., Lavergne, A., Snell Okada, E., Chadha, A., Roosta, T., & Shah, C. (2025). Developing a framework for auditing large language models using human-in-the-loop. In *Proceedings of the 2025 Annual International ACM SIGIR Conference on Research and Development in Information Retrieval in the Asia Pacific Region* (pp. 64-74). Association for Computing Machinery. https://doi.org/10.1145/3767695.3769514

An, Q., Yang, J., Xu, X., Zhang, Y., & Zhang, H. (2024). Decoding AI ethics from users' lens in education: A systematic review. *Heliyon, 10*(20), Article e39357. https://doi.org/10.1016/j.heliyon.2024.e39357

Arabneydi, J., Islam, S., Das, S., Gottipati, S. K., Duguay, W., Mars, C., Taylor, M. E., Guzdial, M., Fagette, A., & Zerouali, Y. (2025). A systematic approach to design real-world human-in-the-loop deep reinforcement learning: Salient features, challenges and trade-offs. *arXiv*. https://doi.org/10.48550/arXiv.2504.17006

Bravo Rocca, G. J. (2025). *Human-on-the-loop continual learning: Data, knowledge and agents for model adaptation* [Doctoral dissertation]. TDX. https://www.tdx.cat/handle/10803/695722

Chen, F. (2022). Human-AI cooperation in education: Human in loop and teaching as leadership. *Journal of Educational Technology and Innovation, 2*(1). https://media.sciltp.com/articles/2509001474/2509001474.pdf

Costa, A. F. M. (2022). *Human-AI interaction design: Usability guidelines for AI systems interfaces design* [Master's thesis]. Repositório Científico Lusófona. https://repositorio.ensinolusofona.pt/handle/10437.1/13135

Eckert, M., Bonial, C., Bonn, J., Hwang, J. D., & Madabushi, H. T. (2025). Keeping humans in the loop: The role of human skills in large language model–based natural language applications. *Annual Review of Linguistics, 12*. https://doi.org/10.1146/annurev-linguistics-041824-061723

Fajardo-Ramos, D. C., Chiappe, A., & Mella-Norambuena, J. (2025). Human-in-the-loop assessment with AI: Implications for teacher education in Ibero-American universities. *Frontiers in Education, 10*, Article 1710992. https://doi.org/10.3389/feduc.2025.1710992

Feng, S., Zhang, H., & Gašević, D. (2025). Mapping the evolution of AI in education: Toward a co-adaptive and human-centered paradigm. *Computers and Education: Artificial Intelligence*, Article 100513. https://doi.org/10.1016/j.caeai.2025.100513

Fernandes, J. M., Rivadeneira, J. E., Rodrigues, A., Boavida, F., & Silva, J. S. (2025). People 4.0—A model for human-in-the-loop CPS-based systems. *Computer Standards & Interfaces, 91*, Article 103895. https://doi.org/10.1016/j.csi.2024.103895

Frering, L., Steinbauer-Wagner, G., & Holzinger, A. (2025). Integrating belief-desire-intention agents with large language models for reliable human–robot interaction and explainable artificial intelligence. *Engineering Applications of Artificial Intelligence, 141*, Article 109771. https://doi.org/10.1016/j.engappai.2024.109771

Fu, Y., & Weng, Z. (2024). Navigating the ethical terrain of AI in education: A systematic review on framing responsible human-centered AI practices. *Computers and Education: Artificial Intelligence, 7*, Article 100306. https://doi.org/10.1016/j.caeai.2024.100306

García-López, I. M., & Trujillo-Liñán, L. (2025). Ethical and regulatory challenges of generative AI in education: A systematic review. *Frontiers in Education, 10*, Article 1565938. https://doi.org/10.3389/feduc.2025.1565938

Garzón, J., Patiño, E., & Marulanda, C. (2025). Systematic review of artificial intelligence in education: Trends, benefits, and challenges. *Multimodal Technologies and Interaction, 9*(8), Article 84. https://doi.org/10.3390/mti9080084

Hasan, M., Shuvo, N. I., Akter, A., Tamim, F. S., Ahmed, N., Mia, N., & Alam, S. (2025). Enhancing face-to-emotion recognition with vision transformer and human-in-the-loop approach. In *2025 International Conference on Quantum Photonics, Artificial Intelligence, and Networking (QPAIN)* (pp. 1-6). IEEE. https://doi.org/10.1109/QPAIN66474.2025.11171658

Holstein, K., Aleven, V., & Rummel, N. (2020). A conceptual framework for human–AI hybrid adaptivity in education. In *International conference on artificial intelligence in education* (pp. 240-254). Springer. https://doi.org/10.1007/978-3-030-52237-7_20

Hwang, G. J., Xie, H., Wah, B. W., & Gašević, D. (2020). Vision, challenges, roles, and research issues of artificial intelligence in education. *Computers and Education: Artificial Intelligence, 1*, Article 100001. https://doi.org/10.1016/j.caeai.2020.100001

Ifenthaler, D., Majumdar, R., Gorissen, P., Judge, M., Mishra, S., Raffaghelli, J., & Shimada, A. (2024). Artificial intelligence in education: Implications for policymakers, researchers, and practitioners. *Technology, Knowledge and Learning, 29*(4), 1693-1710. https://doi.org/10.1007/s10758-024-09747-0

Iyenghar, P. (2025). Clever Hans in the loop? A critical examination of ChatGPT in a human-in-the-loop framework for machinery functional safety risk analysis. *Eng, 6*(2), 31. https://doi.org/10.3390/eng6020031

Jenkins, D., & Khanna, G. (2025). AI-enhanced training, education, & development: Exploration and insights into generative AI's role in leadership learning. *Journal of Leadership Studies, 18*(4), 81-97. https://doi.org/10.1002/jls.70004

Joseph, T. S., Gowrie, S., Montalbano, M. J., Bandelow, S., Clunes, M., Dumont, A. S., Iwanaga, J., Tubbs, R. S., & Loukas, M. (2025). The roles of artificial intelligence in teaching anatomy: A systematic review. *Clinical Anatomy, 38*(5), 552-567. https://doi.org/10.1002/ca.24272

Karataş, F., Eriçok, B., & Tanrikulu, L. (2025). Reshaping curriculum adaptation in the age of artificial intelligence: Mapping teachers' AI-driven curriculum adaptation patterns. *British Educational Research Journal, 51*(1), 154-180. https://doi.org/10.1002/berj.4068

Kathiresan, G. (2025). *Human-in-the-loop testing for LLM-integrated software: A quality engineering framework for trust and safety* [Preprint]. TechRxiv. https://doi.org/10.36227/techrxiv.174702077.78864934/v1

Kitcharoen, P., Howimanporn, S., & Chookaew, S. (2024). Enhancing teachers' AI competencies through artificial intelligence of things professional development training. *International Journal of Interactive Mobile Technologies, 18*(2). https://doi.org/10.3991/ijim.v18i02.46613

Kottara, C., & Asonitou, S. (2025). The impact of human-centered artificial intelligence in higher education. *European Journal of Education Studies, 12*(2). https://doi.org/10.46827/ejes.v12i2.5796

Kumar, S., Datta, S., Singh, V., Datta, D., Singh, S. K., & Sharma, R. (2024). Applications, challenges, and future directions of human-in-the-loop learning. *IEEE Access, 12*, 75735-75760. https://doi.org/10.1109/ACCESS.2024.3401547

Liu, H., Nasiriany, S., Zhang, L., Bao, Z., & Zhu, Y. (2025). Robot learning on the job: Human-in-the-loop autonomy and learning during deployment. *The International Journal of Robotics Research, 44*(10-11), 1727-1742. https://doi.org/10.1177/02783649241273901

Luckin, R. (2025). Nurturing human intelligence in the age of AI: Rethinking education for the future. *Development and Learning in Organizations: An International Journal, 39*(1), 1-4. https://doi.org/10.1108/DLO-04-2024-0108

Mandlekar, A., Garrett, C. R., Xu, D., & Fox, D. (2023). Human-in-the-loop task and motion planning for imitation learning. In *Proceedings of the Conference on Robot Learning* (pp. 3030-3060). PMLR. https://proceedings.mlr.press/v229/mandlekar23b.html

McGrath, M. J., Duenser, A., Lacey, J., & Paris, C. (2025). Collaborative human-AI trust (CHAI-T): A process framework for active management of trust in human-AI collaboration. *Computers in Human Behavior: Artificial Humans*, Article 100200. https://doi.org/10.1016/j.chbah.2025.100200

Memarian, B., & Doleck, T. (2023). Fairness, accountability, transparency, and ethics (FATE) in artificial intelligence (AI) and higher education: A systematic review. *Computers and Education: Artificial Intelligence, 5*, Article 100152. https://doi.org/10.1016/j.caeai.2023.100152

Memarian, B., & Doleck, T. (2024). Human-in-the-loop in artificial intelligence in education: A review and entity-relationship (ER) analysis. *Computers in Human Behavior: Artificial Humans, 2*(1), Article 100053. https://doi.org/10.1016/j.chbah.2024.100053

Molenaar, I. (2022). Towards hybrid human-AI learning technologies. *European Journal of Education, 57*(4), 632-645. https://doi.org/10.1111/ejed.12527

Nazaretsky, T., Ariely, M., Cukurova, M., & Alexandron, G. (2022). Teachers' trust in AI-powered educational technology and a professional development program to improve it. *British Journal of Educational Technology, 53*(4), 914-931. https://doi.org/10.1111/bjet.13232

Ninaus, M., & Sailer, M. (2022). Closing the loop—The human role in artificial intelligence for education. *Frontiers in Psychology, 13*, Article 956798. https://doi.org/10.3389/fpsyg.2022.956798

Niu, B., Luo, B., Cui, Y., Xu, X., Zhao, Y., & Feng, Y. (2025). HIAT: Human-in-the-loop reinforcement learning with auxiliary task. *Artificial Intelligence Review, 58*(9), 273. https://doi.org/10.1007/s10462-025-11260-4

Phaijit, L. (2025). *Human-centred approaches to human-in-the-loop robot learning* [Doctoral dissertation, UNSW Sydney]. https://doi.org/10.26190/unsworks/31227

Pretolesi, D., Regal, G., Schrom-Feiertag, H., & Tscheligi, M. (2026). Adaptive realities: Human-in-the-loop AI for trustworthy XR training in safety-critical domains. *Multimodal Technologies and Interaction, 10*(1), Article 11. https://doi.org/10.3390/mti10010011

Qiu, W., Thway, M., Lai, J. W., & Lim, F. S. (2025). GenAI for teaching and learning: A human-in-the-loop approach. In *Companion proceedings of the 15th International Conference on Learning Analytics & Knowledge (LAK25)* (pp. 3-7). https://joellaiwj.com/wp-content/uploads/2025/04/lak25-companionproceedings.pdf

Rachha, A., & Seyam, M. (2023). Explainable AI in education: Current trends, challenges, and opportunities. In *SoutheastCon 2023* (pp. 232-239). IEEE. https://doi.org/10.1109/SoutheastCon51012.2023.10115140

Ranade, N., Saravia, M., & Johri, A. (2025). Using rhetorical strategies to design prompts: A human-in-the-loop approach to make AI useful. *AI & Society, 40*(2), 711-732. https://doi.org/10.1007/s00146-024-01905-3

Retzlaff, C. O., Das, S., Wayllace, C., Mousavi, P., Afshari, M., Yang, T., Saranti, A., Angerschmid, A., Taylor, M. E., & Holzinger, A. (2024). Human-in-the-loop reinforcement learning: A survey and position on requirements, challenges, and opportunities. *Journal of Artificial Intelligence Research, 79*, 359-415. https://doi.org/10.1613/jair.1.15348

Saha, G. C., Kumar, S., Kumar, A., Saha, H., Lakshmi, T. K., & Bhat, N. (2023). Human-AI collaboration: Exploring interfaces for interactive machine learning. *Tuijin Jishu/Journal of Propulsion Technology, 44*(2).

Shafik, W. (2024). Towards trustworthy and explainable AI educational systems. In *Explainable AI for education: Recent trends and challenges* (pp. 17-41). Springer. https://doi.org/10.1007/978-3-031-72410-7_2

Sharma, R., & Singh, D. (2026). Positive attitude—Cultivating optimism in uncertain times. In *[Book title/publisher information requires verification]* (pp. 94-108). https://doi.org/10.48001/978-81-988770-7-9-8

Sharma, R., Singh, D., & Gupta, K. (2025a). Impact of artificial intelligence on HR efficiency and employee experience in Indian IT firms. *Journal of Information Systems Engineering and Management, 10*(53s), 747-752. https://www.jisem-journal.com/index.php/journal/article/view/10968/5077

Sharma, R., Singh, D., & Gupta, K. (2025b). AI in employee lifecycle management: Current practices and future prospects. *International Journal of Advanced Research in Science, Communication and Technology, 5*(8), 612-619. https://doi.org/10.48175/IJARSCT-26968

Shezad, F., Goswami, R., Shaheen, A., & Khan, I. (2025). AI in teacher training and professional development: A tool for continuous learning and skill enhancement. *Review of Education, Administration & Law, 8*(1), 179-191. https://doi.org/10.47067/real.v8i1.416

Shin, H., Park, J., Yu, J., Kim, J., Kim, H. Y., & Oh, C. (2025). Looping in: Exploring feedback strategies to motivate human engagement in interactive machine learning. *International Journal of Human–Computer Interaction, 41*(14), 8666-8683. https://doi.org/10.1080/10447318.2024.2413293

Singh, D. (2026). Advanced human-AI synergy for effective collaboration: Ethical implications and future prospects. *Inteletica, 3*(5), 24-38. https://doi.org/10.4114/inteletica.vol3iss5pp24-38

Subramanya, S. Y., Watanabe, K., Dengel, A., & Ishimaru, S. (2025). Human-in-the-loop annotation for image-based engagement estimation: Assessing the impact of model reliability on annotation accuracy. In M. Kurosu & A. Hashizume (Eds.), *Human-computer interaction* (Lecture Notes in Computer Science, Vol. 15770, pp. 169-186). Springer. https://doi.org/10.1007/978-3-031-93864-1_12

Susanti, W., Widi, R., Nasution, T., Johan, J., & Verawardina, U. (2025). The role of artificial intelegensia technology in improving the quality of education. *Journal of Applied Business and Technology, 6*(1), 11-15. https://doi.org/10.35145/jabt.v6i1.178

Tarun, B., Du, H., Kannan, D., & Gehringer, E. F. (2025). Human-in-the-loop systems for adaptive learning using generative AI. In *2025 IEEE Frontiers in Education Conference (FIE)* (pp. 1-7). IEEE. https://doi.org/10.1109/FIE63693.2025.11328658

Tiomoko, M., Cherkaoui, H., Paolo, G., Yili, Z., Meng, Y., Keli, Z., & Ali, H. T. (2025). Human in the loop adaptive optimization for improved time series forecasting. *arXiv*. https://doi.org/10.48550/arXiv.2505.15354

Topali, P., Ortega-Arranz, A., Rodríguez-Triana, M. J., Er, E., Khalil, M., & Akçapınar, G. (2025). Designing human-centered learning analytics and artificial intelligence in education solutions: A systematic literature review. *Behaviour & Information Technology, 44*(5), 1071-1098. https://doi.org/10.1080/0144929X.2024.2345295

Tsaneva, S., Dessì, D., Osborne, F., & Sabou, M. (2025). Knowledge graph validation by integrating LLMs and human-in-the-loop. *Information Processing & Management, 62*(5), Article 104145. https://doi.org/10.1016/j.ipm.2025.104145

Varakantham, P., & Tio, S. (2025). Reimagining education: Keeping the human in the loop. *Asian Management Insights, 12*(1), 12-19. https://ink.library.smu.edu.sg/ami/288

Wang, D., & Huang, X. (2025). Transforming education through artificial intelligence and immersive technologies: Enhancing learning experiences. *Interactive Learning Environments, 33*(7), 4546-4565. https://doi.org/10.1080/10494820.2025.2465451

Wang, Y. (2021). Artificial intelligence in educational leadership: A symbiotic role of human-artificial intelligence decision-making. *Journal of Educational Administration, 59*(3), 256-270. https://doi.org/10.1108/JEA-10-2020-0216

Wischnewski, M., Krämer, N., & Müller, E. (2023). Measuring and understanding trust calibrations for automated systems: A survey of the state-of-the-art and future directions. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (pp. 1-16). Association for Computing Machinery. https://doi.org/10.1145/3544548.3581197

Wu, P., Shentu, Y., Liao, Q., Jin, D., Guo, M., Sreenath, K., Lin, X., & Abbeel, P. (2025). Robocopilot: Human-in-the-loop interactive imitation learning for robot manipulation. *arXiv*. https://doi.org/10.48550/arXiv.2503.07771

Xu, W., Gao, Z., & Dainoff, M. J. (2026). An HCAI methodological framework: Putting it into action to enable human-centered AI. *IEEE Transactions on Human-Machine Systems*. https://doi.org/10.1109/THMS.2025.3631590

Xue, W., & Wang, Y. (2025). Human-in-the-loop intelligent testing for safety-critical software. In *2025 IEEE 36th International Symposium on Software Reliability Engineering Workshops (ISSREW)* (pp. 171-172). IEEE. https://doi.org/10.1109/ISSREW67781.2025.00068

Young, B., Anderson, D. T., Keller, J., Petry, F., & Michael, C. J. (2025). SPARC: A human-in-the-loop framework for learning and explaining spatial concepts. *Information, 16*(4), 252. https://doi.org/10.3390/info16040252

Yousefi, E., Chen, M., & Sharf, I. (2025). Shared autonomy policy fine-tuning and alignment for robotic tasks. *The International Journal of Robotics Research*. https://doi.org/10.1177/02783649241312699

Zhang, B., Meroño-Peñuela, A., & Simperl, E. (2025). Towards explainable automated knowledge engineering with human-in-the-loop. *Semantic Web, 16*(5), Article 22104968251382171. https://doi.org/10.1177/22104968251382171

Zhu, H., Sun, Y., & Yang, J. (2025). Towards responsible artificial intelligence in education: A systematic review on identifying and mitigating ethical risks. *Humanities and Social Sciences Communications, 12*(1), 1-14. https://doi.org/10.1057/s41599-025-05252-6