LEVERAGING REINFORCEMENT LEARNING FROM HUMAN FEEDBACK (RLHF) FOR ADAPTIVE LEARNING IN STEM HIGHER EDUCATION: FUTURE DIRECTIONS FOR AFRICA

***Dr. Clement Orver Igyu**, +2348134507985, igyuclement@gmail.com,
Department of Science and Mathematics Education,
Rev. Fr. Moses Orshio Adasu University, Makurdi, Benue State, Nigeria

Prof. Nicholas Akise Ada, FSTAN,
Department of Science and Mathematics Education,
Rev. Fr. Moses Orshio Adasu University, Makurdi, Benue State, Nigeria.

Abstract

Artificial intelligence (AI) systems that incorporate human feedback into reinforcement learning (RL) are transforming adaptive control and personalized learning. Reinforcement learning from human feedback (RLHF) enables Large Language Models and other AI assistants to align with human preferences and educational goals, rather than relying solely on engineered reward functions. In science, technology, engineering, and mathematics (STEM) education, RLHF provides a framework for building adaptive tutoring systems, intelligent laboratories, and robotics platforms that respond to teachers' and students' evaluations in real time. Drawing on recent research in adaptive control, human-AI collaboration, and educational technology, this article situates RLHF at the intersection of computational accuracy and human involvement, yet identifies a major gap in the context of education in Africa. It combines evidence from current sources to explore conceptual bases, integration methods, and pedagogical uses, while addressing ethical and operational challenges such as feedback quality, algorithmic bias, data privacy, teacher oversight, explainability, and fair implementation. The discussion shows how RLHF can support adaptive and ethical learning systems that enhance rather than replace human expertise. It is recommended that future research focus on developing robust, transparent, and human-centered RLHF systems through long-term, cross-cultural studies, improved teacher–AI collaboration, and standardized evaluation frameworks to enhance the effectiveness, fairness, and sustainability of adaptive STEM education in Africa.

Keywords: reinforcement learning, human feedback, artificial intelligence, STEM, higher education, Africa

Introduction

Artificial intelligence (AI) is now deeply embedded within the infrastructures of higher education, especially in science, technology, engineering, and mathematics (STEM) disciplines that depend heavily on iterative experimentation, computational modelling, and simulation. In these domains, AI supports a wide range of functions, from intelligent tutoring systems and automated grading to adaptive laboratory simulations and robotics education. However, genuine digital transformation in learning does not emerge merely from automation or efficiency gains; it depends on *alignment* - the capacity of AI systems to learn *from* and *with* human agents in a shared context of meaning and evaluation. Alignment ensures that the outputs of AI, including GenAIs like ChatGPT, Claude, Geminis, and specialized AI systems,

correspond not only to quantitative performance metrics but also to the qualitative values, goals, and expectations of the human stakeholders who design, teach, and learn with them.

Recent empirical studies demonstrate that students respond more positively to AI tools that visibly acknowledge and incorporate their feedback. Some studies on students' perceptions of AI in higher education found that perceived adaptability and transparency were the strongest predictors of satisfaction and trust, and that visible responsiveness in AI-based learning tools reinforced learners' persistence and intrinsic motivation, especially in technical subjects where iterative problem-solving is key (Grájeda et al., 2024; Igyu et al., 2025; Sasikala & Ravichandran, 2024). These findings, including studies in the context of higher education in Africa, show that AI systems perceived as learning companions rather than opaque evaluators encourage students to engage more deeply in reflecting and adjusting their learning behaviour in response to system feedback (Ebele & Agbade**;** 2020; Effiong, & Agbade, 2016; Ihekoronye et al. 2020). Consequently, in the evolving ecology of higher education, *adaptive feedback loops* among learners, educators, and intelligent systems are becoming central to effective teaching and learning.

Reinforcement Learning from Human Feedback (RLHF) embodies this principle of alignment by formalising a framework through which AI systems can learn from human preferences rather than pre-coded mathematical functions. Traditional reinforcement learning (RL) relies on reward signals that are explicitly engineered to maximise measurable outcomes such as accuracy, efficiency, or performance. While powerful, such reward functions often fail to capture the nuanced social and pedagogical values inherent in human learning environments. RLHF replaces these static reward formulations with dynamic, qualitative feedback provided by human students, instructors, or domain experts who evaluate the desirability or correctness of the system's actions. As Kaufmann, Weng, Bengs, and Hüllermeier (2023) explain, this approach "learns from human feedback instead of relying on an engineered reward function" (p. 3), enabling adaptive systems to approximate human judgment through iterative preference modelling (Meremikwu et al.2022; Ibu et al.2019; Adie et al. 2022).

One of the most important advances in reinforcement learning was showing that human preferences can effectively replace manually crafted reward functions. Christiano et al. (2017) demonstrated that reinforcement learning agents trained through pairwise human preference comparisons developed complex behaviors that closely matched human expectations, even when explicit reward functions were hard to define or unavailable. This represented a major shift from traditional reinforcement learning to preference-based optimization, in which human judgment plays a key role in the learning process. Building on this, Ouyang et al. (2022) introduced InstructGPT, demonstrating that large language models refined with reinforcement learning from human feedback produced responses that users consistently preferred over those from larger models trained only with standard language modeling. These important studies established RLHF as a practical approach to aligning artificial intelligence with human intentions, values, and expectations, providing a strong theoretical foundation for educational uses where quality, fairness, and contextual judgment are crucial.

In educational contexts, RLHF allows AI tutors, chatbots, and learning analytics systems to interpret teacher corrections, peer assessments, or student preferences as guidance for future behaviour. For instance, a virtual chemistry tutor could receive teacher feedback on the clarity of its explanations, or students' satisfaction ratings on problem difficulty, and adjust future interactions accordingly. Over time, the model internalises these patterns to approximate the pedagogical logic of expert instruction. This human-in-the-loop process transforms learning environments from static repositories of content into *interactive*

ecosystems where feedback serves as both a pedagogical and computational driver of improvement.

The practical significance of RLHF extends well beyond educational theory into engineering and control science, where the stability of feedback mechanisms has long been a central concern. Hafner and Riedmiller (2011) observe that any feedback-driven control system must balance *adaptation* to new information with stability against oscillation or drift. In an RLHF context, the same principle applies: a model that reacts too strongly to every piece of feedback risks erratic and inconsistent behaviour, while a model that adapts too slowly may fail to meet user expectations. This tension is particularly pronounced in higher-education settings, where human feedback can be inconsistent, subjective, or delayed. Moreira et al. (2020), in their study of human–robot interaction, show that excessive sensitivity to noisy feedback can induce instability in system performance, reinforcing the need for "moderated adaptation" techniques such as signal smoothing, weighted learning, or confidence-based reward filtering (Ibu et al. 2019; Igyu et al. 2022; Adie, et al. 2026; Meremikwu et al. 2022)

These control-theoretic insights are directly relevant to the design of educational RLHF systems that must reconcile diverse student opinions without undermining coherence or fairness. Moreover, the principles of feedback stability and adaptation echo core concepts of STEM pedagogy. In engineering education, feedback control represents the balance between responsiveness and overshoot, a metaphor that mirrors the learning process, too little feedback stifles growth, too much destabilises it. By grounding RLHF in this technical tradition, educators can develop adaptive systems that both respond to learner needs and maintain structured progression. The interconnection between reinforcement learning and control theory thus bridges computational and educational disciplines, showing how stability analysis can inform the development of intelligent tutoring systems and laboratory simulations that "learn" in step with human cognition.

The ethical dimension of RLHF is equally crucial. As Liu (2023) argues, reinforcement learning systems must sustain *value alignment* by embedding transparency, accountability, and interpretability within their architectures to ensure that machine learning remains within human normative boundaries (p. 12). In the absence of such safeguards, adaptive systems risk reproducing bias, amplifying inequity, or eroding user agency. Ethical alignment entails not only compliance with data protection and fairness standards but also a commitment to making the adaptation process *understandable* to users. Transparency about how human input modifies system behaviour cultivates trust and reinforces a sense of shared ownership over learning outcomes (Igyu et al, 2025; Ogunode, & Agbade, 2023; Olofu et al. 2019; Opara et al.2020; Agbade & Ironbar, 2017).

In STEM education, these principles are particularly relevant given the high stakes of performance evaluation and the rapid pace of technological change. Students must be assured that their feedback is not only heard but used responsibly; teachers must remain confident that algorithmic decisions align with pedagogical intent; and institutions must ensure that adaptation processes respect academic integrity. As Klimova and Pikhart (2025) observe, perceived transparency in AI systems correlates positively with both academic well-being and student confidence. Thus, ethical stewardship in RLHF is not a secondary consideration but a precondition for sustainable adoption.

This was succinctly echoed by Pope Leo IX (2026) in his *Magnifica Humanitas* - "Magnificent Humanity," issued on 25 May 2026. The whole document is about putting the human person at the center of AI governance. RLHF is literally designing how the model learns from human feedback, which translates to asking humans what counts as a good, dignified, helpful answer. For universities in Africa, that's an argument for developing their own RLHF models with local teachers and students, rather than importing models trained

only on US or EU data. Most African universities aren't running full RLHF labs yet, like Stanford, but there's active piloting, research, and adoption happening right now — and Africa has unique reasons to care about RLHF. Africa isn't necessarily lagging behind in RLHF. The work is happening in language, EdTech, and research. The next step is to fund small, course-level RLHF pilots at 5-10 universities and share the datasets. That's how you go from "ChatGPT knows calculus" to "An AI that teaches calculus the way students in Accra, Bloemfontein, and Nairobi actually learn it." (UNESCO, 2021).

These perspectives illustrate that RLHF unites three critical dimensions of educational innovation: technical adaptation, pedagogical effectiveness, and ethical stewardship. By integrating insights from control theory, cognitive science, and human–computer interaction, it offers a framework for designing AI systems that evolve through human partnership rather than human replacement. The remainder of this paper elaborates on these dimensions, exploring how RLHF can support adaptive, trustworthy, and human-centred systems in STEM teaching and learning. The conceptual narrative also makes the argument for future directions for education in Africa by leveraging on RLHF within the discipline.

Conceptual and Theoretical Foundations

Human preference learning builds on traditional reinforcement learning by acknowledging that many real-world goals cannot be sufficiently captured by fixed mathematical reward functions. Instead, human evaluative judgments are gathered and converted into reward signals that steer agent behavior. Interactive reinforcement learning enables agents to accelerate learning by incorporating immediate human feedback during task performance, reducing exploration time and improving policy quality. Likewise, preference-based reinforcement learning is especially useful in domains where subjective evaluation, uncertainty, and multiple acceptable solutions are common. Educational environments exemplify these traits because effective teaching relies not only on correct answers but also on clarity of explanation, student engagement, motivation, creativity, and contextual relevance (Knox & Stone, 2009; Wirth et al., 2017).

Classical RL formalises learning as a sequence of states, actions, and rewards. However, in domains involving social or cognitive evaluation, such as teaching, design, or laboratory practice, the desired outcome cannot be expressed as a single scalar reward. RLHF introduces preference-based feedback: humans compare or rate outcomes, and the algorithm constructs an internal reward model predicting what humans would prefer. Over time, the system internalises the evaluative logic that underpins expert judgement. This architecture parallels socio-constructivist theories of learning. Feedback in education functions not merely as evaluation but as scaffolding (Vygotsky's zone of proximal development). In RLHF, the agent similarly adjusts within a "zone of optimal adaptation," guided by human input. Gaspar-Figueiredo et al. (2025) demonstrate that personalised reinforcement agents calibrated by user satisfaction achieve statistically higher engagement, illustrating how computational and pedagogical feedback mechanisms converge.

In control engineering, reinforcement learning can be interpreted as a data-driven form of adaptive control. Hafner and Riedmiller (2011) identify core challenges, delayed responses, noise, and nonlinear dynamics that are mirrored in human feedback systems. Zohuri (2024) synthesises similar insights, reporting that hybrid controllers integrating human corrective signals with machine-learning estimators achieved faster convergence and reduced error variance. Translating these findings into STEM education in Africa implies that AI tutors or laboratories must incorporate smoothing functions and conservative update rules to prevent over-adaptation to anomalous student feedback. The principle of stability thus underpins both technical and pedagogical success. Just as engineers tune control gains,

educators and system designers must calibrate the influence of each feedback instance to sustain consistent learning trajectories.

Integration Mechanisms and System Architecture

The quality of human feedback depends on how easily users can provide it. used binary gesture-based signals in human-robot learning; educational platforms extend this to star ratings, short comments, or pairwise preferences. Also, simplicity and visibility of feedback pathways significantly raise acceptance (Grájeda et al., 2024; Moreira et al., 2020). In higher education, effective design also requires accessibility, ensuring that students from diverse backgrounds can engage with feedback without cognitive overload.

Advances in RLHF have further refined reward modeling through scalable alignment techniques that reduce annotation costs while enhancing model robustness. Successful implementation relies on maintaining consistency between reward estimation and policy updates while preventing reward hacking and preference overfitting. In STEM education, these insights suggest that adaptive learning systems should include periodic teacher validation, confidence-based feedback weighting, and iterative recalibration of reward models to ensure instructional reliability and pedagogical validity (Casper et al., 2023; Lambert et al., 2024). Once feedback is collected, algorithms train a predictive model that maps observed states and actions to human-evaluated desirability. Active-learning techniques request new feedback selectively when uncertainty is highest, reducing user fatigue. Zohuri (2024) demonstrated that blending human corrections with real-time sensor data yields reward models that are both accurate and interpretable, a finding applicable to intelligent laboratories that fuse performance metrics with subjective ratings.

The policy-learning phase updates decision rules to maximise expected human satisfaction. Hafner and Riedmiller (2011) recommend incremental updates and gain scheduling to prevent instability. Liu (2023) and Gaspar-Figueiredo et al. (2025) advocate "human-in-the-loop checkpoints," in which instructors review proposed policy changes before deployment, thereby maintaining transparency and pedagogical coherence. In educational AI, such oversight may take the form of periodic instructor validation or automated audit trails showing how feedback reshapes model parameters.

Pedagogical Applications in STEM Teaching and Learning

STEM disciplines provide fertile ground for RLHF because their epistemology already revolves around hypothesis, experimentation, and revision, precisely the iterative cycle that reinforcement learning formalizes. Reinforcement learning from human feedback (RLHF) enables intelligent tutoring systems (ITS) to personalise instruction dynamically by continually adapting to learners' cognitive states, performance, and expressed preferences. Evidence shows that mathematics and programming tutors that employ adaptive reinforcement mechanisms achieve higher achievement gains than static or rule-based systems. These systems refine instructional sequencing, hint delivery, and difficulty scaling based on human evaluative input and the effectiveness of such tutors on learners' perception of reciprocal dialogue*,* the sense that the AI listens, interprets, and adjusts in real time (Grájeda et al., 2024; Sasikala & Ravichandran, 2024; Villegas-Ch et al., 2025). Al-Zahrani and Alasmari (2024) further highlight that when teachers retain supervisory roles, adaptive tutors preserve pedagogical authenticity while reducing cognitive load. This triadic relationship, student, AI, and teacher, forms an adaptive feedback ecosystem where each interaction strengthens both engagement and conceptual understanding. Consequently, RLHF-based ITS exemplifies the fusion of computational precision with human empathy in STEM pedagogy.

The development of intelligent tutoring systems has shifted from rule-based instructional models to adaptive systems capable of personalized learning. Well-designed intelligent tutoring systems often match the effectiveness of expert human tutors by providing immediate, personalized feedback that promotes conceptual understanding. Intelligent tutors should act as collaborative learning partners, capable of monitoring learner progress, identifying misconceptions, and dynamically adjusting instructional strategies (VanLehn, 2011; Woolf, 2010). RLHF enhances these capabilities by allowing tutoring systems to continuously improve instructional decisions based on teacher evaluations and learner preferences, rather than relying solely on predefined pedagogical rules. As a result, RLHF-based tutoring systems represent a significant advancement in educational AI, combining the consistency of machine intelligence with the contextual awareness of human instructional expertise.

In robotics and control engineering education, RLHF provides a concrete bridge between theoretical modelling and experiential learning. Students often struggle to visualise how mathematical control laws translate into physical behaviour; integrating human-corrective feedback within reinforcement learning helps make that link explicit. Moreira et al. (2020) demonstrated that robots trained with human evaluative signals, such as gestures, approvals, or corrections, achieved smoother trajectory control and fewer training iterations than purely automated reinforcement agents. Similarly, Hafner and Riedmiller (2011) and Zohuri (2024) review confirm that hybrid human–machine feedback loops yield superior stability and faster convergence across dynamic environments. When applied pedagogically, these findings mean that students not only interact with robots but also actively shape their learning curves, deepening their conceptual grasp of feedback control, system identification, and error correction. Such human-in-the-loop robotics platforms epitomise the guided discovery approach, transforming abstract control theory into tangible, iterative exploration.

Virtual laboratories and simulation environments have become indispensable in STEM education, enabling experimentation beyond physical and logistical constraints. By embedding RLHF, these platforms can modulate variables such as difficulty, pacing, and explanation depth in response to learner feedback. Grájeda et al. (2024) observed that transparent adaptation significantly enhances user engagement and time-on-task, as learners perceive their input to directly influence experimental outcomes. Complementarily, Gaspar-Figueiredo, Pereira, and Costa (2025) found that personalised reinforcement agents adjusting interface complexity according to satisfaction ratings produced statistically higher learning satisfaction and retention. The Davis (2024) review also notes that explainability and real-time adaptation improve student confidence in virtual experimentation. These studies demonstrate that RLHF-driven simulations foster authentic inquiry-based learning: students hypothesize, test, observe adaptive responses, and refine their understanding iteratively. For resource-constrained institutions, like those found in most African higher institutions, such virtual ecosystems can also democratise access to laboratory experiences, providing scalable, cost-effective alternatives that maintain scientific rigour and pedagogical richness.

These laboratories also provide learning experiences comparable to traditional laboratories by promoting active experimentation, reflection, and iterative problem-solving. Such laboratories are particularly effective when learners receive timely feedback that enables them to modify experimental procedures and test alternative hypotheses (De Jong, Linn, & Zacharia, 2013; Ma & Nickerson, 2006). Incorporating RLHF into these environments introduces an additional adaptive dimension whereby learner evaluations continuously inform system behaviour, enabling simulations to personalise explanations, adjust task complexity, and recommend experimental pathways that reflect both learner performance and instructor expectations.

Automated assessment systems increasingly employ AI to grade assignments, evaluate problem-solving steps, and provide feedback. RLHF enhances these systems by turning teacher and student corrections into continuous training data for model refinement. Over successive iterations, the grading algorithm internalises human standards for creativity, partial credit, and conceptual reasoning rather than rigidly applying formulaic rubrics. This process reflects Hafner and Riedmiller's (2011) principle of closed-loop correction, the continual minimisation of error through iterative feedback. Al-Zahrani and Alasmari (2024) emphasise that such adaptive assessment tools reduce marking time while improving consistency and transparency. Furthermore, Sasikala and Ravichandran (2024) note that when feedback loops are visible, students gain metacognitive awareness of evaluation criteria, fostering fairness and self-regulation. Hence, RLHF transforms automated assessment from a static evaluative mechanism into a participatory learning process, aligning analytics with human judgment and institutional standards while maintaining reliability and efficiency.

Zeng et al. (2022) introduce the concept of curriculum reinforcement learning, in which human oversight modulates the order and difficulty of learning tasks to sustain optimal challenge levels. In STEM education, where learning progressions often build cumulatively, RLHF-based systems employing curriculum strategies can personalise trajectories that adjust to each learner's evolving mastery. By analysing patterns of error, engagement, and feedback, such systems decide when to introduce advanced concepts or revisit foundational topics. This adaptive sequencing reflects the self-regulatory dynamics of the scientific method: hypothesis, experimentation, evaluation, and refinement applied to pedagogy. When combined with instructor validation (Liu, 2023; Grájeda et al., 2024), curriculum adaptation ensures that automation complements rather than supplants professional judgement. The result is a dynamic, ethically guided system that sustains cognitive engagement, reduces frustration, and fosters a sense of progression aligned with each student's pace and potential.

Challenges, Ethical Dimensions, and Human-Centred Design

Despite RLHF's promise, its implementation in adaptive learning and control systems raises complex challenges spanning technical, ethical, and institutional domains. Human feedback is inherently subjective. Inconsistent signals, delays, contradictions, or biases can destabilize control systems, leading to oscillatory policies. The same principle applies to education: students' perceptions of difficulty or fairness vary widely, and overfitting to one learner's input may distort group outcomes. Note that such inconsistencies may also affect students' emotional well-being if adaptation appears arbitrary or opaque (Hafner & Riedmiller, 2011; Klimova & Pikhart, 2025). Robust RLHF design, therefore, requires aggregation mechanisms that normalise feedback across multiple users and iterations. Gaspar-Figueiredo et al. (2025) propose weighting models that balance individual preferences with collective satisfaction, mitigating bias and ensuring fairness, an approach transferable to large STEM classes with diverse ability levels, which can be expressed particularly in African higher education.

Because feedback often involves identifiable behavioural data, privacy remains a central ethical concern. Grájeda et al. (2024) link student trust directly to perceptions of data protection. Liu (2023) recommends transparent disclosure of how feedback modifies system behaviour and what data are retained. In academic settings, compliance with institutional research ethics and data-protection regulations must be integral to system design rather than retrofitted after deployment. RLHF systems risk undermining educators if automation replaces rather than augments professional judgement; teachers should act as meta-feedback providers, overseeing algorithmic adaptation to ensure that learning remains anchored in curricular goals. Similarly, professional development in AI literacy is essential so that instructors understand how reinforcement learning models learn from their inputs (Al-Zahrani

& Alasmari, 2024; Sasikala & Ravichandran, 2024). The shift should thus be from teacher versus machine to teacher guiding machine.

Davis (2024) warns that unequal access to computational infrastructure can amplify educational inequities, such that institutions lacking stable connectivity or hardware may be excluded from RLHF-enhanced pedagogy, which gives thought to the African educational landscape. Addressing this requires open-source toolkits, scalable cloud implementations, and capacity building. Without equitable distribution, technological sophistication risks reproducing the same structural imbalances that educational innovation aims to solve.

Ethical Alignment and Explainability

Liu's (2023) concept of value alignment underscores the need for explainable models. When learners and teachers can visualise how feedback reshapes AI behaviour, they perceive the system as trustworthy. Zohuri (2024) supports this claim by showing that hybrid controllers with interpretable decision layers maintained user confidence in the face of uncertainty. In STEM classrooms, visual dashboards illustrating adaptation paths can reinforce this transparency, aligning with ethical standards of informed participation.

Reinforcement learning from human feedback (RLHF) is a socio-technical contract that binds human judgment and algorithmic optimization in a relationship of mutual accountability. In this view, transparency in how human input modifies system behaviour is not optional but foundational to the legitimacy of learning systems. By making the feedback loop visible and interpretable, RLHF ensures that the human actor remains both participant and overseer in the machine's learning process. Reciprocal visibility is essential to prevent "algorithmic drift," a condition in which machine learning models adapt autonomously in ways misaligned with human intent or pedagogical values. The review amplifies this argument by emphasizing that explainable algorithms form the ethical backbone of responsible AI, particularly in higher education, where automated decisions directly affect assessment, grading, and opportunities (Davis, 2024; Liu, 2023). Transparency, in this sense, extends beyond technical explainability to encompass communicative clarity, so that students and educators must understand *why* and *how* the system responds to their feedback.

This interpretability transforms abstract optimisation into an intelligible dialogue between human intention and computational reasoning. Empirical findings by Klimova and Pikhart (2025) support this perspective, demonstrating that perceived transparency and agency within AI-enhanced classrooms are positively correlated with students' psychological well-being, motivation, and trust. Their study shows that learners who understand the adaptive logic of AI tools report reduced anxiety, greater satisfaction, and a stronger sense of co-authorship in their learning journeys. Together, these studies position transparency as both a cognitive and affective necessity: it safeguards fairness, reduces epistemic opacity, and fosters emotional safety. Consequently, ethical alignment in RLHF can be defined as the continuous reconciliation between machine optimisation and human values, ensuring that the algorithmic pursuit of efficiency remains grounded in the moral imperatives of education. This alignment requires that designers and educators embed mechanisms for interpretability, accountability, and participatory oversight throughout the RLHF process from the capture of human feedback to the updating of policies and the deployment of adaptive content. When such measures are in place, RLHF functions not as an impersonal optimiser but as a dialogical system in which adaptation mirrors reflection, and technological intelligence becomes an extension of the shared human commitment to transparency, learning, and trust.

Provisions from internationally recognised policy and regulatory bodies emphasise the importance of ethical governance in AI-enabled education. The UNESCO (2021) Recommendation on the Ethics of Artificial Intelligence stresses that AI systems should be transparent, accountable, inclusive, and subject to meaningful human oversight throughout

their lifecycle. Likewise, the OECD (2019) Principles on Artificial Intelligence promote human-centred values, robustness, safety, explainability, and responsible governance as essential requirements for the deployment of trustworthy AI. These principles closely align with the goals of RLHF because human feedback offers a clear mechanism for incorporating ethical values, pedagogical expectations, and contextual factors directly into algorithmic optimization. Therefore, educational RLHF systems should be designed not only to maximize learning outcomes but also to maintain learner autonomy, equity, and institutional accountability, bearing in mind the limitations of the African higher education space.

Discussion and Future Research Directions

The synthesis of engineering and educational literature reveals that RLHF embodies a broader epistemic shift, from automation toward co-operative intelligence. It unites the precision of machine optimisation with the contextual awareness of human judgement. The convergence of engineering control theory and educational psychology highlights how principles of system stability underpin sustainable learning adaptation. Findings demonstrate that the same balance between responsiveness and stability that governs robotic control systems also determines the success of adaptive learning environments. When these principles are transposed to STEM pedagogy, they reveal the importance of bounded adaptivity, which is a learning system that is responsive enough to personalise instruction without sacrificing conceptual coherence. The review strengthens this analogy by showing that hybrid control systems that integrate human feedback with machine estimation achieve smoother convergence and lower error variance (Hafner & Riedmiller, 2011; Moreira et al., 2020; Zohuri, 2024).

Feedback is central to both human learning and machine adaptation, functioning as a scaffold that supports growth through guided correction and reflection. In educational contexts, this principle is mirrored in RLHF systems, where human feedback shapes the AI's learning trajectory much like teacher guidance supports student development. Perceived responsiveness in AI systems enhances engagement and fosters deeper participation and links transparent adaptation to students' academic well-being and motivation. User satisfaction, as measured by a synthetic index of AI responsiveness, indicates that the visibility of system adjustments significantly strengthens learner trust (Grájeda et al., 2024; Klimova & Pikhart, 2025; Sasikala & Ravichandran, 2024). These studies suggest that RLHF can serve as a metacognitive scaffold, enabling students to perceive feedback not as an evaluation but as a co-constructed process. As learners observe how their input directly shapes AI behaviour, they internalise a feedback-oriented mindset, turning reflection into active participation and developing greater self-efficacy in managing their learning processes.

Furthermore, Dwivedi et al. (2023) argue that the future of educational AI is not about replacing educators but about creating collaborative intelligence where humans and AI work together in teaching, assessment, and knowledge building. Similarly, Kasneci et al. (2023) warn that although large language models present unprecedented opportunities for personalized learning, their educational value relies heavily on human oversight, transparent evaluation, and responsible integration into teaching practices.

Traditional reinforcement learning models optimise for quantifiable outcomes such as accuracy or efficiency, but education demands a more nuanced reward structure that captures persistence, curiosity, and collaboration. Gaspar-Figueiredo et al. (2025) introduce an innovative model in which reinforcement agents adjust rewards based on user satisfaction, demonstrating how human emotions and preferences can serve as valid indicators of progress. When integrated into educational AI, such human-centred reward modelling could reshape how learning success is defined and measured. Rather than rewarding only correct answers, systems could reinforce productive struggle, sustained engagement, or creative

problem-solving. This aligns closely with Liu's (2023) argument that human values must anchor algorithmic optimisation to prevent purely instrumental adaptation. Embedding moral and emotional dimensions into the reward model transforms RLHF from a mechanistic optimisation tool into an ethical framework for learning. Consequently, future educational AI should incorporate dynamic reward models that reflect diverse human outcomes, promoting both cognitive and affective development.

STEM education in Africa provides a natural environment for the evolution of symbiotic intelligence, a paradigm in which humans and AI systems co-learn through continuous, reciprocal feedback. Its iterative experimentation and emphasis on collaboration parallel the feedback loops inherent in reinforcement learning. Tarun, Du, Kannan, and Gehringer (2025) envision this human–AI partnership as a form of generative co-learning, where both entities refine their models of understanding through sustained interaction. RLHF serves as the enabling mechanism for such symbiosis, allowing machines to adjust based on human evaluation while humans refine their knowledge through AI-generated insights. In laboratory and simulation contexts, this could produce adaptive environments that evolve alongside users, reflecting collective intelligence rather than unilateral automation. By integrating feedback from both students and instructors, RLHF-based systems could become genuine partners in inquiry, mirroring the collaborative ethos of African STEM disciplines. This shift from automation to co-evolution represents the next frontier in educational AI design.

Research Directions for STEM Education in Africa

Future investigations should:

1. Multimodal feedback systems: Future systems should combine different types of feedback, like text, emotions, and behavioural cues, to understand learners more completely. This would allow AI tutors to recognise confusion or engagement and respond with more personalised support.
2. Hierarchical control frameworks: Large-scale adaptive learning requires structured control. Hierarchical frameworks can coordinate individual, group, and institutional adaptations, keeping systems stable and fair even when feedback varies widely.
3. Longitudinal studies: More long-term research is needed to determine how RLHF affects students' learning retention and confidence over time, moving beyond short-term performance measures.
4. Teacher–AI co-regulation: Studies should explore how teachers and AI systems share control in guiding learning, ensuring that AI supports rather than replaces human decision-making within ethical boundaries.
5. Adaptive-control benchmarks: There is a need to establish standard evaluation metrics, such as stability, responsiveness, and efficiency, to measure and compare RLHF system performance consistently.
6. Cross-cultural implementation: Research should investigate how RLHF systems function across diverse educational and cultural contexts to promote fairness, inclusivity, and equal access to adaptive technologies.

These directions tend to frame RLHF not merely as a computational technique but as a socio-technical methodology for cultivating reflective, participatory, and resilient learning ecosystems.

Conclusion

Reinforcement learning from human feedback (RLHF) represents a convergence of adaptive control theory, human – computer interaction, and pedagogy. It replaces static automation

with continuous dialogue between humans and machines. When applied to STEM education, RLHF mirrors the scientific method itself: observe, hypothesize, test, and refine. The evidence reviewed demonstrates that systems integrating human evaluation achieve not only higher technical stability but also greater learner trust and motivation.

Nevertheless, effective implementation requires meticulous design: ensuring feedback quality, safeguarding privacy, maintaining teacher oversight, and addressing infrastructural and developmental inequalities that exist in Africa. Ethical alignment and explainability must remain foundational. If these conditions are met, RLHF will enable a future of co-operative intelligence, where adaptation and control coexist, and learning becomes a shared journey between human intuition and machine precision.

References

Adie, EB, Inah, LI ; Ibu, PN & Anditung, PA (2022). Effect of concept mapping strategy on the academic achievement of post basic students' In Mathematics In Obudu Education Area Of Cross River State, Nigeria. *Inter-Disciplinary Journal of Science Education* (IJ-SED) 4 (1), 151-158.

Adie, EB, Ntah, HE & Anditung, PA (2026). Qualitative study on the application of mathematical principles in architectural design in Nigeria. *Zamfara International Journal of Education*, 6(1), 143-149.

Agbade**,** OP & Ironbar, V. E. (2017). Training the adult facilitator for improved productivity in the 21st century: A case study of the Federal Capital Territory, Abuja. *International Journal of Continuing Education and Development Studies* (IJCEDE). 4 (1):165-170.

Al-Zahrani, A. M., & Alasmari, A. M. (2024). *Artificial intelligence and personalised learning in higher education: Opportunities and challenges. Social Sciences, 13*(502).

Casper, S., et al. (2023). *Open problems and fundamental limitations of reinforcement learning from human feedback*. arXiv. https://arxiv.org/abs/2307.15217

Christiano, P. F., Leike, J., Brown, T., Martic, M., Legg, S., & Amodei, D. (2017). *Deep reinforcement learning from human preferences*. *Advances in Neural Information Processing Systems, 30*.

Davis, A. J. (2024). AI rising in higher education: opportunities, risks, and limitations. *Asian Education and Development Studies, 13* (4), 307-319. *https://doi.org/10.1108/AEDS-01-2024-0017*

de Jong, T., Linn, M. C., & Zacharia, Z. C. (2013). *Physical and virtual laboratories in science and engineering education*. *Science, 340*(6130), 305–308. https://doi.org/10.1126/science.1230579

Dwivedi, Y. K., et al. (2023). *So what if ChatGPT wrote it? Multidisciplinary perspectives on opportunities, challenges and implications of generative conversational AI for research, practice and policy*. *International Journal of Information Management, 71*, 102642. https://doi.org/10.1016/j.ijinfomgt.2023.102642

Ebele, FU. & Agbade, OP. (2020). Sustaining learning activities in tertiary institutions in Nigeria amidst COVID-19 pandemic lockdown: The perspective of e-learning strategy. Uyo: Benchmark Educational Services. pages 17-33. ISBN: 978-978-984-508-8.

Effiong, L. V. & Agbade, OP (2016). Relationship between teachers' induction and brain-drain in the teaching profession in Abuja Municipal Area Council. *Abuja Journal of Education and Management Sciences* (ABIJEMS). 4 (1): 122-128.

Gaspar-Figueiredo, J., Pereira, A., & Costa, R. (2025). *Personalised reinforcement agents and user-centred adaptation in human–AI interaction. arXiv:2504.20782v1.*

Grájeda, J. A., Maza, A., Álvarez, J., & Tordera, A. (2024). *Assessing student-perceived impact of using artificial intelligence tools: Construction of a synthetic index of application in higher education. Heliyon, 10*(8).

Hafner, R., & Riedmiller, M. (2011). *Reinforcement learning in feedback control: Challenges and benchmarks from technical process control. Machine Learning, 84*(1–2), 137–169.

Ibu, P. N., Adie, E. B. & Andortan, J. A. (2019). Home environmental variables and mathematics study habits among secondary school three (JSS3) students in Calabar Education Zone of Cross River State, Nigeria. *Prestige Journal of Education* 2 (2), 189 – 197.

Ibu, P. N., Adie, E. B., & Okaba, L. A. (2019). Science teachers' perception of corruption in secondary schools in Calabar Municipality, Cross River State, Nigeria. *Interdisciplinary Journal of Science Education* (IJ – SED) 1 (1), 107 – 119.

Igyu, CO, Adie, EB, Ogah, HA & Anditung PA (2022). Interactive effect of variation teaching strategy in the algebraic motivation, understanding and performance in basic 5 in Benue State, Nigeria. *Vunoklang Multidisciplinary Journal of Science and Technology,* 10(2), 65-71).

Igyu, C. O., Abakpa, B. O., Humbe, B. B. & Iyornum, Z. N. (2025). Perceptions of AI use and risks among science and mathematics education students. *BSU Journal of Science, Mathematics and Computer Education, 5(1).*

Ihekoronye, E.O., Agbade, OP & Opara, J.C. (2020). Inducting novice teachers for improved job performance in public secondary schools in Abuja Municipal Area Council of the Federal Capital Territory, Abuja. *Benue State University Journal of Educational Management*. 2(1), 29-38.

Kasneci, E., et al. (2023). *ChatGPT for good? On opportunities and challenges of large language models for education*. *Learning and Individual Differences, 103*, 102274. https://doi.org/10.1016/j.lindif.2023.102274

Kaufmann, E., Weng, P., Bengs, V., & Hüllermeier, E. (2023). *A survey of reinforcement learning from human feedback. arXiv preprint.*

Klimova, B., & Pikhart, M. (2025). *Exploring the effects of artificial intelligence on student and academic well-being in higher education: A mini-review. Humanities and Social Sciences Communications, 12*(34).

Knox, W. B., & Stone, P. (2009). *Interactively shaping agents via human reinforcement: The TAMER framework*. Proceedings of the Fifth International Conference on Knowledge Capture.

Lambert, N., et al. (2024). *Reinforcement Learning from Human Feedback.* https://rlhfbook.com

Liu, J. (2023). *Transforming human interactions with AI via reinforcement learning with human feedback (RLHF).* Unpublished manuscript.

Ma, J., & Nickerson, J. V. (2006). *Hands-on, simulated, and remote laboratories: A comparative literature review*. *ACM Computing Surveys, 38*(3). https://doi.org/10.1145/1132960.1132961

Meremikwu, AN, Benimpuye, EB, Inah, LI, Okri, JA & Tawo CN (2022). Influence of school feeding programme on academic achievement of primary 3 pupils in mathematics in Akwa Ibom State, Nigeria. *Inter-Disciplinary Journal of Science Education* (IJ-SED) 4 (1), 82-88.

Meremikwu, AN, Benimpuye, AE, Idoko, LI, Arikpo OJ &Tawo, CN (2022). Classroom routine activities as predictors of students' academic achievement in mathematics in Calabar Metropolis of Cross River State, Nigeria. Inter-disciplinary Journal of Science Education, 4(1), 60-70.

Moreira, I. A., Ferreira, M. J., Lourenço, M., & Gonçalves, J. (2020). *Deep reinforcement learning with interactive feedback in a human–robot environment. Applied Sciences, 10*(5574).\

OECD. (2019). *OECD Principles on Artificial Intelligence*. https://oecd.ai/en/ai-principles

Ogunode, NJ & Agbade, OP (2023). Impact of subsidy removal on school administration, teachers' job performance and students' academic performance in secondary schools in Nigeria. *International Journal on Integrated Education,* 6(9), 19-24.

Olofu, PA, Ihekoronye, E.O. & Opara, J.C. (2019). The role of organizational structure in facilitating effective management of the school system in the 21st century in Nigeria. *International Journal of Continuing Education and Development Studies* (IJCEDE), 5 (1):153-163.

Opara, J.C., Olofu, PA & Ihekoronye, E.O. (2020). Assessment of parents' perception on the impact of covid-19 pandemic school closure on secondary school students in Gwagwalada Area Council, Abuja. *Benue State University Journal of Educational Management*. 2(1), 93-102.

Ouyang, L., et al. (2022). *Training language models to follow instructions with human feedback*. *Advances in Neural Information Processing Systems, 35*.

Pope Leo XIV. (2026). *Magnifica Humanitas: On safeguarding the human person in the time of artificial intelligence*. Vatican Publishing House.

Sasikala, P., & Ravichandran, K. S. (2024). *Study on the impact of artificial intelligence on higher education. Procedia Computer Science, 240*, 105862.

Tarun, T., Du, J., Kannan, V., & Gehringer, E. (2025). *Human-in-the-loop systems for adaptive learning using generative AI. arXiv preprint.*

UNESCO. (2021). *Recommendation on the Ethics of Artificial Intelligence*. https://unesdoc.unesco.org

VanLehn, K. (2011). *The relative effectiveness of human tutoring, intelligent tutoring systems, and other tutoring systems*. *Educational Psychologist, 46*(4), 197–221. https://doi.org/10.1080/00461520.2011.611369

Villegas-Ch, W., Buenano-Fernandez, D., Navarro, A.M. (2025). Adaptive intelligent tutoring systems for STEM education: analysis of the learning impact and effectiveness of personalized feedback. *Smart Learn. Environ. 12, (41).* https://doi.org/10.1186/s40561-025-00389-y

Wirth, C., Akrour, R., Neumann, G., & Fürnkranz, J. (2017). *A survey of preference-based reinforcement learning methods*. *Journal of Machine Learning Research, 18*, 1–46.

Woolf, B. P. (2010). *Building Intelligent Interactive Tutors, Student-Centered Strategies for Revolutionizing E-Learning*. Morgan Kaufmann. https://www.researchgate.net/publication/232322117_Building_Intelligent_Interactive_Tutors_Student-Centered_Strategies_for_Revolutionizing_E-Learning

Zeng, R., Duan, X., Li, M., Ferrara, E., Pinto, L., Kuo, C.-C., & Nikolaidis, S. (2022). *Human decision makings on curriculum reinforcement learning with difficulty adjustment. arXiv preprint.*

Zohuri. B. (2024). Artificial Intelligence and Machine Learning Driven Adaptive Control Applications. *J Mat Sci Eng Technol. 2(4): 1-4. DOI: doi.org/10.61440/JMSET.2024.v2.27*