AdapTutor: Reinforcement Learning-Enhanced Personalized AI Tutoring System with Direct User Feedback Integration

Dhruv Agarwal[1]

[1]Manipal institute of technology

dhruvagarwal5018@gmail.com

Abstract

We present an Adaptive AI Tutor with Reinforcement Learning from User Feedback (RLUF) that continuously optimizes teaching strategies based on student quiz performance. Using Merlyn Mind's 12B education-specific language model enhanced with direct verifiable rewards (+1 correct/-1 incorrect), the system incorporates comprehensive user profiles into personalized prompts for individualized learning experiences.

1 Introduction

Educational technology stands at a transformative crossroads where artificial intelligence can revolutionize personalized learning, yet current AI tutoring systems remain fundamentally static and generic [1, 2]. Traditional educational AI platforms fail to adapt their teaching methods based on student performance outcomes, providing one-size-fits-all responses that ignore individual learning patterns, academic backgrounds, and cognitive preferences [2, 5].

Our project addresses this fundamental challenge by developing an Adaptive AI Tutor with Reinforcement Learning from User Feedback (RLUF) that continuously optimizes its teaching strategies based on student quiz performance and comprehensive user profiling [3, 4, 7]. Unlike existing systems that rely on pre-programmed responses, our solution leverages Merlyn Mind's education-specific language models enhanced with reinforcement learning algorithms that adapt to individual student needs in real-time [6]. The system incorporates detailed user demographics, learning preferences, and academic history directly into personalized prompts, creating truly individualized educational experiences that improve over time [1, 2, 5].

2 Problem Description

Our Adaptive AI Tutor with Reinforcement Learning from User Feedback (RLUF) creates a dynamic educational system that continuously optimizes its teaching strategies based on student quiz performance. The system leverages Merlyn Mind's education-specific 12B parameter language model enhanced with reinforcement learning algorithms that adapt to individual student needs in real-time.[6]

The system implements a simplified RLUF approach using direct verifiable rewards (+1 for correct answers, -0.1 for incorrect) to eliminate the complexity and cost of traditional reward models while maintaining learning capabilities. Students interact with the AI tutor, receive personalized explanations, take follow-up quizzes, and their performance automatically trains the system to provide better future responses. This creates a continuous improvement loop where the AI becomes increasingly effective at teaching each individual student based on their demonstrated learning patterns and comprehension levels.

3 *Current Limitations*

- Static Teaching Methods: Existing AI tutors cannot adapt their explanations or questioning strategies based on whether students actually learn from their responses, resulting in repeated ineffective teaching approaches.
- Lack of Deep Personalization: Current systems fail to integrate comprehensive student profiles (learning style, demographics, academic background) into real-time responses, providing generic rather than individually tailored educational experiences.
- No Learning from Outcomes: Traditional AI tutors cannot optimize their teaching effectiveness based on student comprehension and retention data, missing opportunities to improve educational outcomes through iterative refinement.
- High Implementation Costs: Complex reward model systems require expensive computational resources ($50,000+) and extensive development time (8-12 weeks), making advanced AI tutoring inaccessible to many educational institutions.
- Generic Content Delivery: Most AI tutors draw from general knowledge bases rather than curriculum-specific, age-appropriate educational content, leading to responses that may be technically correct but ineffective for specific learning objectives and developmental stages.

4 Technical Proposal

4.1 Technical Approach

4.1.1 *System Architecture*

Our adaptive AI tutor implements a streamlined three-component architecture: (1) Merlyn Mind's 12B education-specific language model as the foundation [6] (2) Dynamic prompt generation engine that injects comprehensive user profiles into real-time interactions (3) Direct RLUF training pipeline using verifiable quiz rewards (+1 correct, -0.1 incorrect) without complex reward models [3, 9]. The system leverages Knowledge Graph-enhanced Retrieval-Augmented Generation (KG-RAG) to ground responses in structured educational content, eliminating hallucination issues common in general-purpose models [8]. User profiles containing demographics, learning style, academic history, and performance patterns are dynamically integrated into personalized prompts, creating individualized educational experiences that adapt continuously based on quiz outcomes [1].

4.1.2 *Reinforcement Learning Implementation*

We implement Proximal Policy Optimization (PPO) with direct verifiable rewards, eliminating the need for separate reward model training [3, 7]. The system processes student interactions through a continuous feedback loop: tutor response $\rightarrow$ quiz administration $\rightarrow$ performance evaluation $\rightarrow$ reward calculation $\rightarrow$ model update [9].

4.1.3 *Personalization Framework*

The system builds comprehensive learner models incorporating: academic background, learning preferences, demographic factors, performance history, and real-time engagement metrics [1, 2]. These profiles are injected directly into model prompts, enabling context-aware responses that adapt teaching style, difficulty level, and content presentation to individual student needs without requiring separate personalization models [1, 5].

4.2 Evaluation Methods

- Quiz Performance Improvement: Pre/post assessment comparison with target 25-35% improvement in accuracy [1, 2].
- Knowledge Retention Testing: Follow-up assessments at 1-week and 1-month intervals to measure long-term learning [1, 2].
- Engagement Analytics: Session duration, return rate, and voluntary interaction frequency [1, 3].

5 Related Work

Reinforcement Learning from User Feedback (RLUF) represents a paradigm shift in AI alignment, moving from expert-annotated preferences to direct user signals [4]. Han et al. (2025) demonstrated that RLUF using binary feedback (like "love" reactions) achieves 28% improvement in positive user engagement while maintaining safety standards [3].

Intelligent Tutoring Systems (ITS) research shows significant promise for personalized education [1, 2]. Létourneau et al. (2025) found that AI-driven ITS improve K-12 learning outcomes by 25-40% through adaptive learning pathways and real-time feedback mechanisms [2]. Contemporary studies emphasize that effective personalization requires comprehensive learner modeling incorporating cognitive states, learning styles, and performance patterns [1]. Liu et al. (2024) demonstrated that adaptive prompts in ITS, grounded in educational theories, enhance problem-solving efficiency and deepen understanding through dynamic adjustment to student performance [5].

Adaptive Learning Systems utilizing AI have shown substantial educational benefits [1]. Wang et al. (2024) established that incorporating detailed user profiles in educational AI systems leads to 35% improvement in learning efficiency [1]. Recent systematic reviews highlight that AI-enabled adaptive learning systems achieve superior outcomes through continuous assessment, real-time adjustments, and multi-objective optimization balancing engagement, comprehension, and retention [1, 2]. However, current literature identifies critical gaps in systems that can learn from educational outcomes to optimize teaching effectiveness over time, representing the key innovation opportunity addressed by our proposed RLUF-enhanced tutoring system

References

[1] Artificial intelligence in education: A systematic literature review. *Expert Systems with Applications*, 2024. Covers adaptive learning, personalization, and reported outcome improvements in AIED.

[2] A systematic review of ai-driven intelligent tutoring systems (its) in k-12 education. *Systematic Review (K-12 ITS), PMC*, 2025. Synthesizes 28 studies; reports generally positive learning effects with examples of sizable pre/post gains.

[3] Reinforcement learning from user feedback. *arXiv preprint arXiv:2505.14946*, 2025. Analyzes binary production signals (e.g., Love Reactions) and reports up to 28% increases in positive user feedback.

[4] Paul F. Christiano, Jan Leike, Tom B. Brown, Miljan Martic, Shane Legg, and Dario Amodei. Deep reinforcement learning from human preferences. In *Advances in Neural Information Processing Systems*, 2017. Also available as arXiv:1706.03741.

[5] H. Liu and colleagues. The design of guiding and adaptive prompts for intelligent tutoring systems. *IEEE Transactions on Learning Technologies*, 2024. Proposes principles and adaptive prompt design for mathematical ITS.

[6] Merlyn Mind. Merlyn-education-teacher-assistant (12b) for the education domain, 2023. 12B decoder-style transformer fine-tuned from Pythia-12B for classroom assistance.

[7] Long Ouyang, Jeff Wu, Xianyi Jiang, Diogo Almeida, Carroll Wainwright, et al. Training language models to follow instructions with human feedback. In *Advances in Neural Information Processing Systems*, 2022. Also available as arXiv:2203.02155.

[8] Diego Sanmartin. Kg-rag: Bridging the gap between knowledge and creativity. *arXiv preprint arXiv:2405.12035*, 2024.

[9] John Schulman, Filip Wolski, Prafulla Dhariwal, Alec Radford, and Oleg Klimov. Proximal policy optimization algorithms. *arXiv preprint arXiv:1707.06347*, 2017.