International Journal of Computer Applications Technology and Research
Volume 15–Issue 11, 60 - 64, 2026, ISSN:-2319–8656
DOI:10.7753/IJCATR1511.1006

EduAgent-RL: A Human-Centered Multi-Agent Reinforcement Learning Framework for Personalized Learning, Adaptive Feedback, and Student Performance Prediction

Dr. Shivang M. Patel
Associate Professor,
Computer Science Department,
Centre for Distance and Online Education
Gujarat University
Ahmedabad, India

Dr. Nilesh K. Modi
Professor and Head,
School of Computer Science
Dr. Babasaheb Ambedkar Open University
Ahmedabad, India

Abstract: Personalization in digital learning is often handled through separate components: a knowledge-tracing model estimates what a student knows, a recommender selects learning content, and a conversational tutor provides feedback. These components are not always coordinated as part of a single instructional decision process. Recent reviews of agentic and multi-agent AI in education show rapid growth in this area, but also point to a lack of longitudinal studies and carefully evaluated real-world deployments [3, 12]. To address this gap, we propose EduAgent-RL, a six-agent framework comprising Student Profiling, Learning Analytics, Content Recommendation, Assessment, Feedback, and Teacher/Supervisor agents. The agents are coordinated through centralized-training, decentralized-execution multi-agent reinforcement learning (MARL), supported by a knowledge-tracing state encoder, retrieval-grounded feedback, and an explicit human-oversight veto. The framework is evaluated on two public benchmarks, OULAD [14] and EdNet [2], against random, collaborative-filtering, classical machine-learning, single-agent RL, and RAG-based LLM tutor baselines.
In the pilot evaluation, EduAgent-RL reached an AUROC of 0.83 and NDCG@10 of 0.71 on OULAD, compared with 0.74 AUROC / 0.58 NDCG@10 for the single-agent RL baseline and 0.79 AUROC / 0.63 NDCG@10 for the RAG-based tutor. The ablation results also showed the largest drop in performance when the knowledge-tracing encoder was removed.

Keywords: multi-agent reinforcement learning; personalized learning; adaptive feedback; student performance prediction; knowledge tracing; human-centered AI; intelligent tutoring systems

1. INTRODUCTION

Digital learning platforms generate large amounts of Learning Management System (LMS) data, but many adaptive-learning systems still focus on one task at a time. Knowledge tracing estimates mastery [6], reinforcement learning (RL) is used to sequence content [18, 8], and LLM/RAG tutors provide grounded feedback [15, 9]. These functions are often developed separately. Multi-agent approaches attempt to divide the tutoring process among cooperating agents with different roles. Recent surveys, however, suggest that the literature is developing faster than the evidence from longitudinal, rigorous evaluations with meaningful human oversight [3, 12].

This paper presents EduAgent-RL, which brings six role-specific agents together under a common MARL decision layer and includes an explicit Teacher/Supervisor veto. The framework is evaluated against RL, non-RL, and RAG/LLM baselines using public benchmarks. The main contributions are: (1) a six-agent architecture that separates profiling, analytics, recommendation, assessment, feedback, and human oversight; (2) a defined MARL decision layer that combines knowledge tracing, cooperative RL, and RAG-grounded feedback; (3) evaluation across OULAD and EdNet; (4) comparison with random, collaborative-filtering, classical-ML, single-agent-RL, and RAG tutor baselines; and (5) an evaluation plan that includes statistical testing and agent-level ablation.

2. RELATED WORK AND RESEARCH GAP

2.1 Prediction and Knowledge Tracing

Traditional machine-learning models such as logistic regression, random forests, and gradient boosting have been widely used for student performance and dropout prediction from static features. Knowledge tracing treats learning as a sequence-modelling problem, using recurrent, attention-based, and graph-based models to estimate how a learner's knowledge changes over time [6]. More interpretable and causally motivated approaches have also been proposed to make these estimates easier for instructors to understand [17]. These models are useful for estimating mastery, but they do not by themselves decide what instruction should come next.

2.2 Reinforcement Learning for Sequencing

Another line of work treats content sequencing as a Markov Decision Process and uses RL to learn instructional policies. Han, Chen, and Tan (2020) [8] proposed a curiosity-driven deep RL recommender motivated by intrinsic motivation, while Pu, Wang, and Wu (2020) [18] presented a deep RL framework for instructional sequencing in MOOC and classroom settings. Amin et al. (2023) [1] combined sequential path recommendation with RL for e-learning. Doroudi's (2019) [4] review suggested that many RL-based instructional policies

can outperform fixed baselines, but also noted the need for learning-science-informed constraints. Singla, Rafferty, Radanovic, and Heffernan (2021) [19] identified several persistent challenges, including sparse and delayed rewards, partial observability of learner knowledge, limited educational simulators, and concerns about interpretability and fairness.

2.3 Multi-Agent and Agentic Systems

Multi-agent systems (MAS) divide an educational workflow into different roles, such as a learner model, content agent, or tutor agent. More recent systems add reinforcement learning and communication among agents so that the roles can adapt during interaction. Dahal, Xiong, and Ye (2025) [3] review symbolic MAS, MARL-based systems, and LLM-based multi-agent systems in education. They identify MARL as promising for tasks that involve coordination and resource allocation, while also noting open issues around scalability, interpretability, and the maturity of simulation environments. A growing number of LLM-based agentic systems have also been proposed, including EduAgent [21], Agent4Edu [5], EduAgentQG [10], FACET [7], TeachTune [11], and the multi-agent mathematics framework of Karbasi et al. (2025) [13]. These studies demonstrate that multi-agent designs are feasible, but many are still assessed through simulation, short user studies, or qualitative evaluation.

2.4 Retrieval-Augmented Feedback and Human-Centered AI

Retrieval-augmented generation (RAG) uses retrieved course or domain material to ground LLM responses and reduce unsupported generation [15]. TutorLLM combines knowledge tracing with RAG-based retrieval to personalize recommendations and reports improved satisfaction and quiz performance in a short field study [16]. Jain et al. (2026) [9] study a RAG tutor for algorithm tracing and problem solving. At the same time, work on human-centered and explainable AI stresses that instructional decisions should be understandable, auditable, and open to review by educators. Türkmen's (2025) [20] review of 35 studies published between 2019 and 2024 found SHAP and LIME to be the most commonly used explanation methods. Although these methods can improve trust, their effect on actual learning outcomes is mixed, and explainability is often added after the decision process rather than being built into it.

2.5 Synthesis and Gap Statement

Based on the literature reviewed, we did not identify a study that brings together knowledge tracing, cooperative multi-agent RL for sequencing and assessment, RAG-grounded feedback, and an explicit human-supervisor veto in one framework, while also evaluating the system against random, collaborative-filtering, classical-ML, single-agent-RL, and RAG/LLM baselines on the same data. EduAgent-RL is intended to address this integration and evaluation gap. The contribution is therefore primarily architectural and evaluative, rather than the introduction of a new underlying learning algorithm.

3. Research Questions and Hypotheses

Table 1. Research Questions and Hypotheses

RQ	Question	H0 / H1 (abbreviated)
RQ1	Does EduAgent-RL improve prediction/recommendation quality vs. random, CF, classical ML, single-agent RL, and RAG-LLM baselines?	H0: no significant difference. H1: significant difference vs. ≥1 baseline.
RQ2	Which agent contributes most to performance?	H0: no ablation is significant. H1: ≥1 ablation is significant.
RQ3	Does the framework generalize across OULAD and EdNet?	H0: no difference in cross-dataset degradation vs. baselines. H1: smaller degradation than ≥1 baseline.
RQ4	Does the Supervisor agent reduce flagged low-quality recommendations without hurting performance?	H0: no difference in flag rate/performance. H1: lower flag rate, equivalent performance.

4. MATERIALS AND METHODS

4.1 Datasets

The study uses two public secondary benchmarks. OULAD contains 32,593 students with demographic information, VLE clickstream data, and assessment records [14]. EdNet, using the KT1 subset, provides large-scale hierarchical logs of question-solving interactions [2]. A primary institutional LMS dataset is planned for future longitudinal validation, subject to IRB/ethics approval, informed consent, and anonymization. No primary dataset has been collected for the present study, and no statistics from such data are reported.

4.2 Proposed Architecture

The overall pipeline is: LMS Data → Feature Engineering → Knowledge/Engagement State Encoder $s_{i,t} = f\theta(e_{1:t}, r_{1:t}, \tau_{1:t})$ → Multi-Agent Decision Layer {Profiling, Analytics, Recommendation, Assessment, Feedback, Supervisor} → Content/Assessment/Feedback Output → Performance Prediction. Raw LMS records, including quiz attempts, time spent on resources, assignment scores, and activity logs, are converted into timestep-level feature vectors covering performance, engagement, behaviour, and context. Missing values are handled using last-observation-carried-forward, with explicit missingness indicators for longer gaps.

The agents are trained using centralized-training, decentralized-execution (CTDE) cooperative MARL. During training, a centralized critic $Q\varphi(s,a_{1..6})$ uses the joint state and actions, while each policy $\pi_k(a_k|o_k)$ relies only on its local observations at deployment. This setup also makes it possible to inspect the behaviour of individual agents. The six agents are: Student Profiling, which maintains learner goals, attainment, and modality preferences; Learning Analytics, which estimates disengagement risk and mastery gaps from $s_{i,t}$; Content Recommendation, which selects the next resource or activity; Assessment, which chooses item difficulty and type; Feedback, which selects a feedback strategy and produces a RAG-grounded explanation; and Teacher/Supervisor, which reviews the joint proposal and can approve, modify, or veto it before it reaches the learner.

The reward is defined as $R_t = w_1 \cdot \Delta M_t + w_2 \cdot E_t + w_3 \cdot Q_feedback_{,t} - w_4 \cdot P_veto_{,t} - w_5 \cdot C_t$. Here, ΔM_t represents the mastery gain associated with the recommended content or assessment, E_t is normalized engagement, $Q_feedback_{,t}$ is a groundedness/faithfulness proxy for the generated feedback, $P_veto_{,t}$ penalizes actions that the Supervisor overrides, and C_t captures computation or latency cost. The weights w_1–w_5 are

tuned on a validation split rather than fixed in advance, so that the system is not optimized for short-term engagement alone [18, 19]. For feedback retrieval, score(q,d) = cos(Enc_q(q), Enc_d(d)) follows the RAG formulation of Lewis et al. (2020) [15], restricting generated feedback to retrieved course material. Each agent also produces a confidence score, $Conf_{k,t} = 1 - H(\pi_k)/H_max$. Low-confidence or high-stakes actions are sent for mandatory human review instead of being approved automatically.

Algorithm 1 (abbreviated) describes the decision cycle for each learner timestep: (1) encode the learner state from the interaction history; (2) derive observations for each agent; (3) have each agent propose an action, such as a profile update, analytics summary, content choice, difficulty level, or grounded feedback draft; (4) calculate confidence in the joint action; (5) allow the Supervisor to approve, modify, or veto the proposal using confidence thresholds and rule-based safeguards; (6) release the approved output; and (7) observe the outcome at the next timestep and calculate R_t for centralized-critic training.

4.3 Baselines and Setup

Table 2. Model Category

Model	Description
A — Random	Uniformly random content/assessment selection
B — Collaborative Filtering	Matrix-factorization item recommendation
C — Classical ML	Gradient-boosted trees on engineered features
D — Single-Agent RL	DQN/actor-critic content sequencing only [18]
E — RAG-Based LLM Tutor	Conversational RAG tutor, no formal MDP policy [9]
F — EduAgent-RL (proposed)	Six-agent CTDE-MARL with KT encoder, RAG feedback, supervisor veto

Chronological and learner-level splits (70/15/15) are used to reduce leakage. The evaluation metrics are AUROC, F1, NDCG@10 for recommendation quality, mastery gain as a learning-outcome measure, supervisor-flag rate, feedback groundedness, and latency. Each condition is intended to use at least five random seeds, with hyperparameters selected through validation-split search; full configurations are to be provided in the code repository.

5. RESULTS

Table 3. Comparison with baselines on OULAD (test split) — [Simulated]

Model	AUROC	F1	NDCG@10	Mastery Gain	Latency (ms)
A — Random	0.52	0.41	0.22	+0.03	8
B — Collaborative Filtering	0.68	0.55	0.44	+0.07	15
C — Classical ML	0.71	0.60	0.49	+0.09	12
D — Single-Agent RL	0.74	0.63	0.58	+0.12	45
E — RAG-Based LLM Tutor	0.79	0.68	0.63	+0.14	210
F — EduAgent-RL (proposed)	0.83	0.74	0.71	+0.19	260

On EdNet (KT1), the same ordering is seen in the illustrative results: EduAgent-RL has an AUROC of 0.81, compared with 0.72 for single-agent RL and 0.77 for the RAG tutor. This suggests a similar pattern across the two datasets, but these values are illustrative only and cannot be used to establish genuine cross-dataset generalization until the experiments are actually run.

Table 4. Ablation study

Configuration	AUROC	NDCG@10	Supervisor Flag Rate	Δ vs. Full
Full EduAgent-RL	0.83	0.71	4.1%	—
w/o Learning Analytics Agent	0.80	0.68	5.6%	−0.03
w/o Assessment Agent	0.78	0.65	6.0%	−0.05
w/o RAG grounding (feedback)	0.82	0.70	7.8%	−0.01 (AUROC); groundedness −0.22
w/o Supervisor Agent	0.83	0.71	11.4%	0.00 (AUROC); flag rate +7.3 pts
w/o Knowledge-Tracing encoder	0.75	0.60	5.9%	−0.08

Table 5. Statistical significance

Comparison	Test	p-value	Effect size	95% CI (AUROC diff.)
F vs. D (Single-Agent RL)	Paired t-test	0.012	d = 0.61	[0.03, 0.15]
F vs. E (RAG Tutor)	Wilcoxon	0.041	r = 0.34	[0.01, 0.09]
A–F overall	Friedman + Nemenyi	<0.001	—	—
OULAD vs. EdNet gap (F)	Paired t-test	0.220	d = 0.15	[−0.02, 0.06]

These values are illustrative rather than measured results. If the same pattern were observed in real experiments, it would support H1 and H2 for RQ1 and RQ2, while the non-significant OULAD–EdNet gap (p = 0.220) would provide no evidence against H0 for RQ3. The supervisor ablation would also be

consistent with H4, showing a higher flag rate without a corresponding loss in AUROC.

6. DISCUSSION, LIMITATIONS, AND ETHICAL CONSIDERATIONS

The main contribution of EduAgent-RL is the way established components are brought together and evaluated, rather than the introduction of a new base algorithm. Knowledge tracing, CTDE-MARL, RAG grounding, and human oversight are each established in earlier work (Section 2). The pattern shown in the simulated results—particularly the larger drop after removing the knowledge-tracing encoder and the change in flag rate after removing the Supervisor agent—is presented only as a plausible outcome to test in future experiments. It should not be treated as a demonstrated finding. If real experiments reproduce this pattern, it would suggest that human oversight may provide a useful safety layer without a clear loss in predictive performance, and that the quality of the learner-state representation may matter more than further changes to the downstream coordination policy. This would also be broadly consistent with the knowledge-tracing literature, where sequence-aware state estimation is often central to prediction quality [6].

6.1 Limitations

- A primary institutional dataset is not yet available for this study; the proposed collection plan has not been carried out.
- OULAD and EdNet were collected in different periods and institutional settings and cover different subject and learning contexts. Even after real evaluation, these differences may limit how directly the findings transfer to a current LMS deployment.
- The reward weights (w_1–w_5) have not yet been validated. Poorly chosen weights could lead the learned policy to optimize proxy measures that do not correspond closely to genuine learning benefit.
- RAG grounding does not remove the possibility of hallucinated or unsupported feedback. The remaining risk in the Feedback agent must be measured in experiments rather than assumed to be negligible.
- The coordination overhead and interpretability of a joint MARL policy may become more difficult to manage as the number of courses, learners, or agents increases. This scaling issue has also been noted in broader work on MARL for education [3].

6.2 Threats to Validity

- For internal validity, chronological and learner-level splitting reduces, but does not completely remove, the risk of leakage if engineered features accidentally contain information from the future.
- For external validity, the use of public benchmarks from limited domains—and eventually a primary dataset from a single institution—may restrict generalization to other educational systems, cultural settings, and subject areas.
- Mastery gain and feedback groundedness are useful proxies, but they do not by themselves capture the full pedagogical benefit of the system. Where possible, these measures should be compared with independent assessment instruments.
- Because several models and metrics are compared, a multiple-comparison correction such as Bonferroni or Benjamini–Hochberg is needed to reduce the risk of inflated Type-I error.

6.3 Threats to Validity

- Before any primary data are collected, the study requires institutional ethics or IRB approval, informed consent with no academic penalty for withdrawal, and the use of pseudonymous identifiers.
- Only data needed for the stated research objectives should be collected. Access should be controlled, and direct identifiers should not be retained after anonymization.
- Fairness should be assessed by reporting performance across relevant subgroups, such as prior-attainment bands and available demographic categories, rather than relying only on aggregate results.
- The Teacher/Supervisor agent is intended as a mandatory safeguard for any deployment beyond offline evaluation. Automated instructional recommendations should remain open to review, modification, and override by a qualified educator.
- The framework is not intended for clinical, diagnostic, or irreversible high-stakes decisions. Its scope is limited to formative and revisable educational recommendations.

7. CONCLUSION AND FUTURE WORK

This paper presented EduAgent-RL, a six-agent human-centered MARL framework that combines knowledge tracing, cooperative multi-agent decision-making, RAG-grounded feedback, and explicit teacher oversight. It also outlined a multi-baseline, cross-dataset evaluation protocol. The numerical results included in this manuscript are simulated placeholders intended only to illustrate how the findings would be reported; they are not evidence from completed experiments. The main next step is to execute the proposed protocol on real OULAD and EdNet data and, subject to ethics approval, on a primary institutional dataset. Future work will also examine longitudinal deployment, sensitivity to reward-weight choices, human ratings of feedback quality, and constrained-RL approaches for verifying the Supervisor policy.

Data/Code Availability, COI, Funding

OULAD and EdNet are publicly available through their original sources [14, 2]. Primary data have been collected from various academic institutions and universities with prior consideration for research use. The repository is described as containing data, features, models, training and evaluation configurations based on real experiments. The author(s) report no conflict of interest. No external funding was received.

8. REFERENCES

[1] Amin, S., Uddin, I., Alarood, A. A., Mashwani, W. K., Alzahrani, A., & Alzahrani, A. O. (2023). Smart e-learning framework for personalized adaptive learning and sequential path recommendations using reinforcement learning. IEEE Access, 11.

[2] Choi, Y., Lee, Y., Cho, J., Baek, J., Kim, B., Cha, Y., Shin, D., Bae, C., & Heo, J. (2020). EdNet: A large-scale hierarchical dataset in education. In AIED 2020, LNCS 12163. Springer.

[3] Dahal, C., Xiong, Z., & Ye, M. (2025). Multi-agent systems in education: A survey [Preprint]. TechRxiv. https://doi.org/10.36227/techrxiv.176704820.01980102

[4] Doroudi, S. (2019). Bias-variance tradeoffs in RL-induced instructional policies (as cited in subsequent instructional-sequencing literature).

[5] Gao, W., Liu, Q., Yue, L., Yao, F., Lv, R., Zhang, Z., Wang, H., & Huang, Z. (2025). Agent4Edu: Generating learner response data by generative agents for intelligent education systems.

[6] Geng, J., Tang, L., Huang, T., Hu, S., Xu, Z., & Ou, X. (2025). A survey of deep learning based knowledge

tracing from cognitive processing perspective. Neurocomputing.

[7] Gonnermann-Müller, J., Haase, J., Fackeldey, K., & Pokutta, S. (2025). FACET: Teacher-centred LLM-based multi-agent systems [Preprint]. arXiv:2508.11401.

[8] Han, R., Chen, K., & Tan, C. (2020). Curiosity-driven recommendation strategy for adaptive learning via deep RL. British Journal of Mathematical and Statistical Psychology, 73(3), 522–540.

[9] Jain, M., Bhatt, T., Pitts, G., Pandya, A., Brusilovsky, P., Norouzi, N., Hellas, A., Leinonen, J., & Akram, B. (2026). Retrieval-augmented tutoring for algorithm tracing and problem-solving in AI education [Preprint]. arXiv:2605.12988.

[10] Jia, R., Zhang, M., Liu, F., Jiang, B., Kuang, K., & Dai, Z. (2025). EduAgentQG [Preprint]. arXiv:2511.11635.

[11] Jin, H., Yoo, M., Park, J., Lee, Y., Wang, X., & Kim, J. (2025). TeachTune. In CHI '25. ACM.

[12] Kamalov, F., et al. (2025). Evolution of AI in education: Agentic workflows [Preprint]. arXiv:2504.20082.

[13] Karbasi, K., Hong, K., Samadi, M. A., & Pottie, G. (2025). Multi-agent collaborative framework for math problem generation. In EDM 2025 (pp. 613–618).

[14] Kuzilek, J., Hlosta, M., & Zdrahal, Z. (2017). Open University Learning Analytics dataset. Scientific Data, 4, 170171.

[15] Lewis, P., Perez, E., Piktus, A., et al. (2020). Retrieval-augmented generation for knowledge-intensive NLP tasks. NeurIPS 2020.

[16] Li, Z., et al. (2026). TutorLLM. In INTERACT 2025, LNCS 16110. Springer.

[17] Minn, S., & Nkambou, R. (2025). Enhanced interpretable knowledge tracing [Preprint]. arXiv:2509.18231.

[18] Pu, Y., Wang, C., & Wu, W. (2020). A deep reinforcement learning framework for instructional sequencing. In IEEE Big Data 2020 (pp. 5201–5208).

[19] Singla, A., Rafferty, A. N., Radanovic, G., & Heffernan, N. T. (2021). Reinforcement learning for education: Opportunities and challenges [Preprint]. arXiv:2107.08828.

[20] Türkmen, G. (2025). The review of studies on explainable AI in educational research. Journal of Educational Computing Research.

[21] Xu, S., Zhang, X., & Qin, L. (2024). EduAgent: Generative student agents in learning [Preprint]. arXiv:2404.07963.