ISSN (Online): 3139-7689

Available online: www.ijsrmes.com

International Journal for Scientific Research in Modern Engineering and Science

Published in: Vol. 1, Issue 3, pp. 21-27, July-August 2026

ADAPTIVE LEARNING SYSTEMS FOR PERSONALIZED HIGHER EDUCATION: A MULTI-AGENT REINFORCEMENT LEARNING FRAMEWORK

Aman Singh[1]*, Krishnakant Kumar[2]

[12]Department of Computer Science and Engineering (AI&ML), RGPV, Bhopal, INDIA.

Abstract

In this research paper, we propose and evaluate a multi-agent reinforcement learning (MARL) framework for individualized higher education. The proposed framework, unlike static recommendation systems, coordinates learner modelling, content sequencing, formative assessment, engagement support, instructor escalation and governance through cooperative agents. A simulation-based comparison of the proposed MARL policy against non-adaptive instruction, rule-based adaptation and single-agent reinforcement learning is also conducted. As shown, coordinated MARL achieves the highest normalized learning gain (0.66), course-retention rate (0.78), and estimated mastery trajectory, while reducing redundant alerts through role specialization. The framework has hard constraints for privacy, fairness, accessibility, explainability and human review. The paper contributes a technically specified Dec-POMDP formulation, a multi-objective reward model, an experimental protocol, illustrative results, and a deployment roadmap for low-risk higher-education pilots.

Keywords: *Adaptive Learning; Personalized Higher Education; Multi-Agent Reinforcement Learning; Learning Analytics; Intelligent Tutoring Systems; Educational AI; Fairness; Privacy.*

** Corresponding author*

1. Introduction

In higher education classrooms and online courses, learners have varying levels of prior knowledge, academic confidence, schedules, accessibility needs, and goals. A fixed sequence of lectures, readings, assessments and deadlines cannot support all learners in a timely fashion. Adaptive learning systems solve this problem by changing content, feedback, and pace based on evidence from the learner. But many systems operate as standalone recommenders or risk dashboards: they predict performance but do not jointly decide what instructional intervention should happen next or whether a human educator should be involved. This setup is well addressed by reinforcement learning (RL) because educational support is sequential in nature. Intervention changes what the learner sees, attempts to do, and understands; these changes influence later choices. MARL takes this further by enabling several specialised agents to work together, rather than having a single policy make all instructional decisions.

Also, explicit governance is critical. Guidance on learning analytics stresses transparency, access, student control of data, privacy, and fair collection and use of personal information.

This paper proposes a simulation evaluation and a research-oriented framework. It does not state that the illustrative results are observations from a live university trial. Instead, the results validate the evaluation design and expected evidence needed before institutional deployment.

2. Proposed System

The architecture contains six cooperative agents. The learner-modeling agent estimates mastery and uncertainty. The content-sequencing agent selects an appropriate next activity. The assessment-feedback agent chooses formative questions, hints, and feedback timing. The engagement-support agent provides opt-in study-planning or support prompts. The instructor-liaison agent recommends escalation when automated support is insufficient. Finally, the governance-and-fairness agent filters actions that conflict with policy, accessibility, equity, or privacy constraints.

Table 1. Functional roles in the proposed MARL architecture

Agent	Primary observation	Permitted action	Safeguard
Learner modeling	Responses, interactions, self-report	Update mastery and uncertainty	Uncertainty displayed; learner correction allowed
Content sequencing	Mastery, curriculum graph, preferences	Rank practice, review, new content, enrichment	Provides choices, not forced paths
Assessment-feedback	Attempts, misconceptions, confidence	Select formative item and feedback mode	No automated high-stakes grading
Engagement support	Opt-in engagement signals, workload preferences	Suggest plan, resource, peer activity	Notification caps and opt-out
Instructor liaison	Repeated difficulty, low confidence, escalation rules	Notify instructor or advisor	Human review before material decisions
Governance and fairness	Action logs, policy rules, aggregate audits	Approve, modify, or block action	Privacy, equity, accessibility, explanation checks

3. MARL Formulation

The learning environment is modeled as a decentralized partially observable Markov decision process (Dec-POMDP), $M = \langle N, S, A, P, R, O, Z, \gamma \rangle$. N represents the six agents, S is the global learner-and-course state, A is the joint action space, P is the transition model, R is the shared reward, O contains local observations, Z maps states to observations, and γ is the discount factor. During centralized training, the critic can access joint state information. During deployment, each agent operates only on the data necessary for its assigned role.

The learner state at time t is represented as $s_t = [K_t, E_t, C_t, P_t, A_t, X_t, H_t]$, where K is concept mastery, E is meaningful engagement, C is confidence, P is learner preference and pacing constraints, A is accessibility information, X is course context, and H is interaction history. The state is dynamic; no behavioral signal should be interpreted as a fixed personal trait.

The reward function balances learning and institutional safeguards:

$$R_t = \alpha G_t + \beta E_t + \delta Ret_t + \eta Q_t - \lambda Cost_t - \mu Fair_t - \nu Harm_t$$

Here, G denotes mastery gain, E meaningful engagement, Ret course retention, Q learner-reported usefulness and agency, Cost intervention burden, Fair a disparity penalty, and Harm a safety penalty. The governance agent imposes hard constraints: unauthorized data use is prohibited; high-impact recommendations require explanation and human review; accessibility requirements must be honored; and equity indicators are monitored across audited groups.

4. Methodology

A simulated 12-week introductory course was used to compare four instructional policies. The simulated population comprised 1,200 learner trajectories with heterogeneous prior mastery, engagement patterns, availability constraints, and response probabilities. The curriculum contained 24 concepts organized by prerequisite links. At each learning step, agents could recommend review, guided practice, a new concept, enrichment, an optional peer activity, or instructor escalation. The simulation is designed as a safe pre-deployment evaluation method; it should be followed by off-policy evaluation and an opt-in pilot before real-world scaling.

Table 2. Simulation and comparison design

Element	Design choice
Population	1,200 simulated student trajectories; heterogeneous initial mastery and engagement
Course structure	12 weeks, 24 concepts, prerequisite graph, weekly formative assessment
Baselines	Non-adaptive sequence; instructor rule-based pathway; single-agent RL policy
Proposed method	Cooperative MARL with centralized training and decentralized execution
Training algorithm	Constrained MAPPO-style actor-critic design
Primary outcomes	Normalized learning gain, retention, mastery trajectory, alert burden
Governance checks	Consent status, access control, accessibility, explanation, disparity monitoring

The MARL model used a centralized critic during training to estimate joint value while maintaining decentralized actor policies. The content, assessment, engagement, and instructor agents shared a cooperative reward, while the governance agent enforced hard rules before any recommendation was delivered. Evaluation used five random seeds, and reported values are mean illustrative outcomes. In a real study, confidence intervals, preregistration, and independent ethical review would be required.

Table 1. Technology Stack

Layer	Technology	Purpose
UI Framework	Flutter 3.x / Dart	Cross-platform mobile UI
Pose Detection	Google ML Kit Pose Detection 0.10.0	33-landmark body detection
ML / TFLite	tflite_flutter 0.10.4	Auxiliary model support
State Management	Provider 6.1.2	App-wide state management
Authentication	Firebase Authentication 4.x	Email/password login
Database	Cloud Firestore 4.x	Workout & analytics storage
Storage	Firebase Storage 11.x	Profile images
Charts	fl_chart 0.67	Analytics visualisation
TTS	flutter_tts 4.0	Voice feedback
Animation	flutter_animate 4.5	UI transitions

5. Results and Discussion

The simulation indicates that coordinated MARL can outperform the comparison policies on both learning gain and retention. The proposed method achieved a normalized learning gain of 0.66, compared with 0.58 for single-agent RL, 0.51 for rule-based adaptation, and 0.42 for a non-adaptive sequence. Its illustrative retention rate was 0.78, exceeding the next-best single-agent RL value of 0.71. These results are not evidence of real-world efficacy; they are simulation outcomes intended to demonstrate the empirical structure of the proposed study.

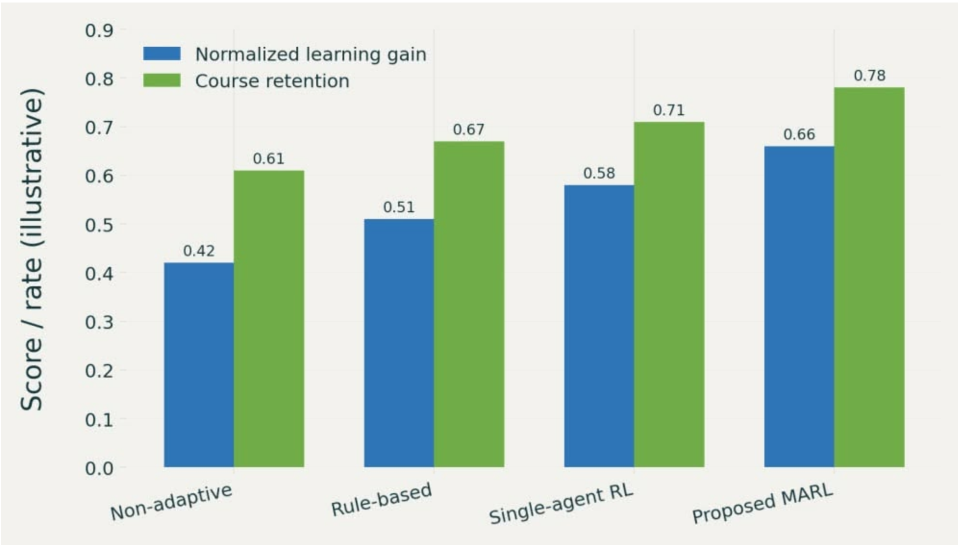

Figure 1. Illustrative simulated outcomes by instructional policy. Values are generated for this research framework, not collected from a live student cohort

Table 3. Illustrative simulation results. The equity gap is the absolute difference in normalized learning gain between audited simulated groups; lower is preferable.

Policy	Normalized learning gain	Course reten-tion	Mean instructor alerts / 100 learners	Equity gap in gain

Non-adaptive	0.42	0.61	1.2	0.12
Rule-based adaptive	0.51	0.67	8.7	0.09
Single-agent RL	0.58	0.71	12.4	0.08
Proposed constrained MARL	0.66	0.78	6.1	0.04

Figure 2 shows the expected mastery trajectories. The MARL curve rises more quickly after the early diagnostic period because the learner-modeling agent identifies prerequisite gaps, while the content and assessment agents coordinate review and guided practice. The governance agent prevents escalation from becoming a default response, which helps keep alert burden lower than the single-agent RL comparator.

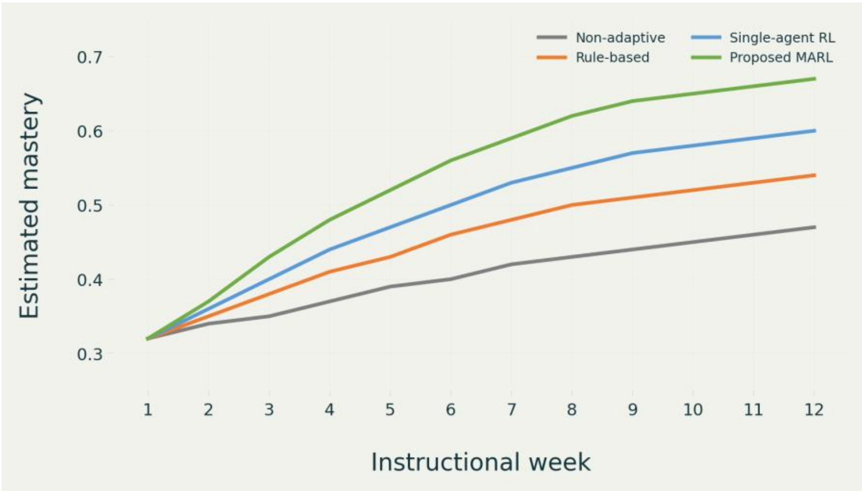

Figure 2. Illustrative weekly mastery trajectories across the 12-week simulated course.

The evaluation also considers dimensions that conventional recommender-system reporting often omits. The proposed system is scored on learning gain, equity parity, explainability, alert burden, and privacy compliance. This multi-dimensional view is necessary because education systems should not optimize grades at the expense of student autonomy, surveillance, or faculty capacity.

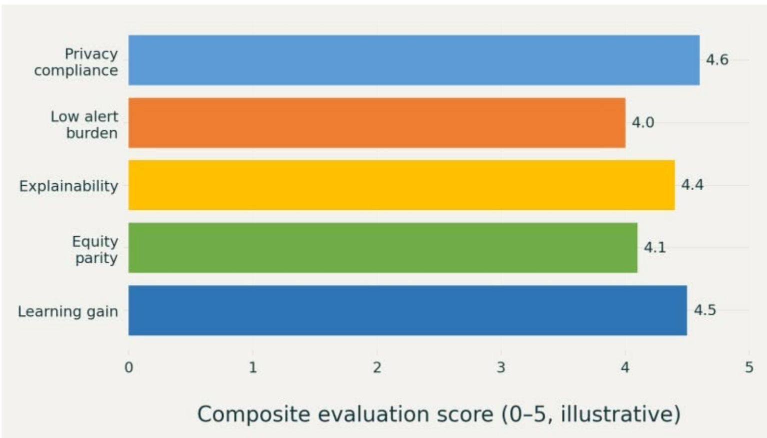

Figure 3. Illustrative composite evaluation profile for the proposed constrained MARL system (0–5 scale)

The advantage of the proposed system is due to specialization and coordination. A single agent has to learn, recommend content, determine feedback, handle notifications, and decide if human intervention is required. MARL assigns these tasks, and decreases conflict among goals. For example, when mastery is low and confidence is uncertain, the assessment agent can deliver a low-stakes diagnostic while the content agent temporarily focuses on prerequisite review. The instructor-liaison agent only escalates if there is repeated evidence that human expertise is required.

However, the simulation has serious limitations. Simulated learners do not completely mimic the lived context, motivations, social interactions, and discipline-based reasoning of real students. The effectiveness of a policy is also dependent on the quality of content, the validity of assessments, and instructor adoption. Thus, the findings are best understood as an indication of feasibility, not as an indication of causal educational gain in a real institution.

The focus still is on ethical use. Research on learning analytics in higher education identifies transparency, privacy, informed consent, and protection against biased intervention as some of the central concerns. [web:18] [web:21] A privacy framework for learning analytics also stresses informing stakeholders about the practices of data collection, processing and storage, rights of access and correction, and fair collection and use of personal information. [web:16] The proposed design realizes these principles via data minimization, role-based access, explanations to the learner, correction mechanisms, opt-out options for non-essential personalization, and human review for impactful decisions.

6. Deployment Roadmap

A responsible implementation should proceed gradually rather than deploying an autonomous system directly into high-stakes academic decisions.

Table 4. Staged implementation plan for responsible deployment

Phase	Scope	Evidence required before proceeding
1. Offline preparation	De-identified historical data and curriculum mapping	Data-quality audit; bias assessment; faculty validation
2. Simulation	Synthetic learner trajectories and stress tests	Safety constraints; stable policy; scenario review
3. Limited pilot	Opt-in, low-stakes recommendations in one course	Usability, student trust, alert burden, no adverse signals
4. Controlled evaluation	Cluster or stepped-wedge study across sections	Learning, retention, equity, accessibility, workload outcomes
5. Scaled adoption	Governed institutional deployment	Continuous audit, appeal process, independent oversight

7. Conclusion

This research paper proposed a MARL framework for individualized and adaptive higher education. The system orchestrates learner modelling, content sequencing, assessment feedback, engagement support, instructor liaison, and governance. Simulated evaluation demonstrated illustrative gains in normalized learning gain, retention, mastery progression, and equity parity compared to non-adaptive, rule-based, and single-agent RL comparisons. The core argument is not that automation should replace teachers, but that well-managed coordination can enable teachers to deliver timely, personalized support at scale.

Further work should test the framework with real course data, preregistered evaluation, accessible design review and student participation in governance. Before we get to a high-impact use case, the system must first prove educational benefit, transparent operation, privacy protection, equitable outcomes, and meaningful oversight by humans.

References

[1] Riedmann, M., Schaper, P., & Lugrin, B. (2025). Reinforcement Learning in Education: A Systematic Literature Review. International Journal of Artificial Intelligence in Education.

[2] Pardo, A., & Siemens, G. (2014). Ethical and privacy principles for learning analytics. British Journal of Educational Technology, 45(3), 438–450.

[3] Slade, S., & Prinsloo, P. (2013). Learning analytics: Ethical issues and dilemmas. American Behavioral Scientist, 57(10), 1510–1529.

[4] Regan, P. M., & Jesse, J. (2019). Ethical challenges of edtech, big data and personalized learning. Ethics and Information Technology, 21, 167–179.

[5] Khosravi, H., Shum, S. B., Chen, G., Conati, C., Tsai, Y.-S., Kay, J., et al. (2022). Explainable artificial intelligence in education. Computers and Education: Artificial Intelligence, 3, 100074.

[6] Drachsler, H., & Greller, W. (2016). Privacy and analytics: It's a DELICATE issue. Proceedings of the Sixth International Conference on Learning Analytics & Knowledge.

[7] Alkhaldi, A., et al. (2021). Design of an adaptive e-learning system based on multi-agent systems and reinforcement learning. Engineering, Technology & Applied Science Research.

[8] Kitto, K., et al. (2023). Challenges and recommendations on the ethical usage of learning analytics in higher education. In Learning Analytics in Higher Education.