

Ứng dụng Trí tuệ nhân tạo trong xây dựng hệ thống Học tăng cường hỗ trợ dạy và học STEM

1st Nguyễn Hữu Lộc
 Khoa Công nghệ Thông tin
 Trường Đại học Sài Gòn
 TP. Hồ Chí Minh, Việt Nam
 lockbkbang@gmail.com

2nd Văn Tuấn Kiệt
 Khoa Công nghệ Thông tin
 Trường Đại học Sài Gòn
 TP. Hồ Chí Minh, Việt Nam
 vankiet27012004@gmail.com

Abstract—Trong kỷ nguyên công nghiệp 4.0, giáo dục STEM đóng vai trò then chốt trong việc đào tạo nguồn nhân lực chất lượng cao. Tuy nhiên, các phương pháp giảng dạy truyền thống và hệ thống quản lý học tập (LMS) hiện hành thường áp dụng cách tiếp cận “một kích cỡ cho tất cả”, thất bại trong việc đáp ứng nhu cầu cá nhân hóa của từng người học. Bài báo này đề xuất một hệ thống gợi ý lô trình học tập thông minh sử dụng kỹ thuật Học tăng cường (Reinforcement Learning - RL), cụ thể là thuật toán Q-learning, được tích hợp vào nền tảng Moodle qua chuẩn LTI 1.3. Hệ thống mô hình hóa quá trình học tập dưới dạng Quy trình quyết định Markov (MDP), sử dụng dữ liệu hành vi thực tế để phân cụm người học và tối ưu hóa chiến lược gợi ý. Kết quả thực nghiệm mô phỏng trên 500 vòng lặp cho thấy thuật toán giúp tăng 22.5% điểm số trung bình và giảm 51.0% số lượng kỹ năng yếu so với phương pháp truyền thống.

Index Terms—Học tăng cường, Q-learning, Cá nhân hóa học tập, Giáo dục STEM, Data-driven Simulation, MDP

I. GIỚI THIỆU

Sự phát triển mạnh mẽ của Trí tuệ nhân tạo (AI) đang định hình lại nhiều lĩnh vực, trong đó có giáo dục. Theo nghiên cứu của Frey và Osborne, khoảng 47% các công việc truyền thống có nguy cơ bị tự động hóa, đặt ra yêu cầu cấp thiết về việc trang bị các kỹ năng mới cho người lao động, đặc biệt là các kỹ năng STEM (Khoa học, Công nghệ, Kỹ thuật và Toán học) [2]. Giáo dục STEM chú trọng phát triển tư duy phản biện và khả năng giải quyết vấn đề, tuy nhiên, việc triển khai hiệu quả gặp nhiều rào cản do sự đa dạng về năng lực và tốc độ tiếp thu của học viên.

Thách thức lớn nhất hiện nay là cá nhân hóa trải nghiệm học tập (Personalized Adaptive Learning - PAL) trên quy mô lớn. Các hệ thống LMS truyền thống như Moodle, Blackboard chủ yếu đóng vai trò lưu trữ tài liệu và quản lý điểm số, thiếu khả năng phân tích hành vi để đưa ra các can thiệp kịp thời [1]. Tại Việt Nam, các nghiên cứu về ứng dụng AI trong giáo dục chủ yếu tập trung vào bài toán dự báo (prediction) - ví dụ như dự báo nguy cơ bỏ học hoặc dự đoán điểm số cuối kỳ - mà chưa chú trọng nhiều đến bài toán đưa ra khuyến nghị hành động (prescription) để cải thiện kết quả đó [11].

Để giải quyết vấn đề này, nhu cầu về một hệ thống hỗ trợ dạy và học STEM cá nhân hóa ứng dụng Học tăng cường

(Reinforcement Learning - RL) trở nên cấp thiết. RL cho phép hệ thống tự động tối ưu hóa chiến lược giảng dạy thông qua cơ chế thử-sai (trial-and-error).

Nghiên cứu này đóng góp vào lĩnh vực Cá nhân hóa học tập (Personalized Adaptive Learning) thông qua ba điểm chính:

- 1) **Đề xuất khung giải pháp thích ứng:** Xây dựng mô hình Quy trình quyết định Markov (MDP) với hàm phần thưởng đa mục tiêu, kết hợp giữa đặc điểm phân cụm người học và lý thuyết hành vi (ICAP framework).
- 2) **Quy trình Mô phỏng hướng dữ liệu (Data-driven Simulation):** Giải quyết thách thức “khởi động lạnh” (cold-start) và sự khan hiếm dữ liệu thực nghiệm bằng cách xây dựng môi trường giả lập dựa trên tham số thống kê từ dữ liệu khóa học thực tế.
- 3) **Kiểm chứng thực nghiệm:** Chứng minh hiệu quả của thuật toán Q-learning thông qua A/B testing trên tập dữ liệu mô phỏng, cho thấy sự vượt trội so với các chiến lược truyền thống về điểm số và mức độ tham gia.

Figure 1. Conceptual Comparison between Traditional vs. Adaptive Learning Approaches.

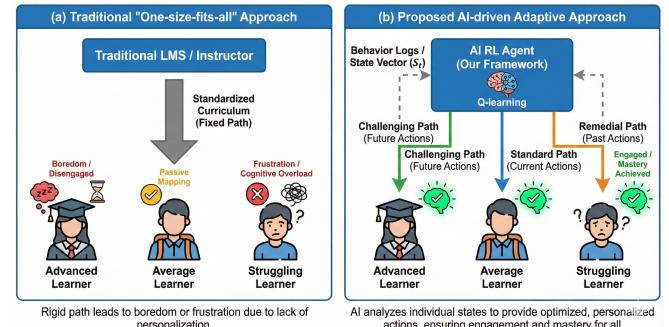


Figure 1. Comparison of learning approaches. (a) The traditional approach applies a fixed curriculum to divergent learners, leading to suboptimal engagement outcomes like boredom or frustration. (b) The proposed framework utilizes an AI Reinforcement Learning Agent to analyze individual student states from log data and recommend personalized actions (remedial, standard, or challenging), aiming for maximized engagement and mastery for all learner types.

II. PHƯƠNG PHÁP ĐỀ XUẤT

Dựa trên các hạn chế của LMS truyền thống, nghiên cứu đề xuất một khung giải pháp học tập thích ứng (Adaptive Learning Framework) sử dụng thuật toán Q-learning. Quy trình xử lý tổng thể đi từ dữ liệu hành vi thô, qua bước trích xuất đặc trưng để xây dựng không gian trạng thái, và cuối cùng là tác nhân AI đưa ra quyết định tối ưu (Hình 2).

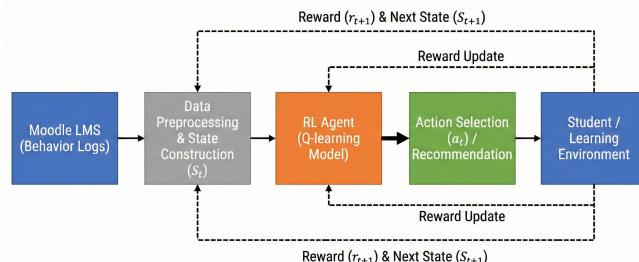


Figure 2. Tổng quan phương pháp đề xuất: Dữ liệu hành vi từ Moodle được chuyển đổi thành các trạng thái (State), qua đó Agent lựa chọn hành động (Action) tối ưu để nhận phần thưởng (Reward).

Figure 2. Tổng quan phương pháp đề xuất: Dữ liệu hành vi từ Moodle được chuyển đổi thành các trạng thái (State), qua đó Agent lựa chọn hành động (Action) tối ưu để nhận phần thưởng (Reward).

A. Mô hình hóa bài toán (Problem Formulation)

Để cá nhân hóa lộ trình học tập, chúng tôi mô hình hóa bài toán dưới dạng Quy trình Quyết định Markov (MDP), được định nghĩa bởi bộ ba $\langle S, A, R \rangle$ như sau:

1) Không gian Trạng thái (State Space - S): Tại thời điểm t , hệ thống quan sát trạng thái người học S_t . Để đảm bảo tính tổng quát, S_t được định nghĩa là một vector đặc trưng d chiều:

$$S_t = \{f_1, f_2, \dots, f_d\} \quad (1)$$

Trong nghiên cứu này, chúng tôi đề xuất bộ đặc trưng bao gồm:

- C (Cluster): Nhóm người học xác định qua phân cụm.
- M (Module): Chỉ số bài học hiện tại.
- P (Progress): Mức độ hoàn thành module.
- Sc (Score): Phân loại điểm số tích lũy.
- Ph (Phase): Giai đoạn học tập (ví dụ: theo khung ICAP).
- E (Engagement): Mức độ tương tác.

2) Không gian Hành động (Action Space - A): Dựa trên S_t , tác nhân (Agent) lựa chọn một hành động a_t từ tập hợp A gồm m hành động sư phạm khả dĩ ($A = \{a_0, a_1, \dots, a_{m-1}\}$). Các hành động này được phân loại theo trực thời gian (Quá khứ - Hiện tại - Tương lai) nhằm phục vụ các chiến lược ôn tập (Remedial) hoặc bồi dưỡng (Advanced).

3) Hàm phần thưởng (Reward Function - R): Mục tiêu của hệ thống là tối đa hóa tổng phần thưởng tích lũy. Hàm thưởng được thiết kế đa mục tiêu:

$$R_{total} = R_{base} + R_{LO} + R_{bonus} - P_{penalty} \quad (2)$$

Trong đó, R_{base} là phần thưởng cơ bản, R_{LO} dựa trên mức độ đạt chuẩn đầu ra, R_{bonus} cho các chuỗi hành vi tốt, và $P_{penalty}$ là điểm phạt để hạn chế hành vi kém hiệu quả.

B. Quy trình Xử lý dữ liệu và Phân cụm

Dữ liệu log thô từ LMS thường chứa nhiều và không cấu trúc. Trước khi đưa vào mô hình RL, dữ liệu cần được tiền xử lý và chuẩn hóa. Thuật toán K-means được áp dụng để phân chia người học thành K cụm (Clusters) có đặc điểm hành vi tương đồng. Việc xác định giá trị K tối ưu được thực hiện thông qua phương pháp Elbow và chỉ số Silhouette. Giá trị Cluster ID sau đó trở thành một thành phần quan trọng trong vector trạng thái S_t .

C. Thuật toán Q-learning

Hệ thống sử dụng thuật toán Q-learning để cập nhật bảng giá trị Q (Q-table) theo công thức Bellman:

$$Q(s, a) \leftarrow Q(s, a) + \alpha [r + \gamma \max_{a'} Q(s', a') - Q(s, a)] \quad (3)$$

Trong đó α là tốc độ học và γ là hệ số chiết khấu. Chiến lược ϵ -greedy được áp dụng để cân bằng giữa khám phá (Exploration) và khai thác (Exploitation).

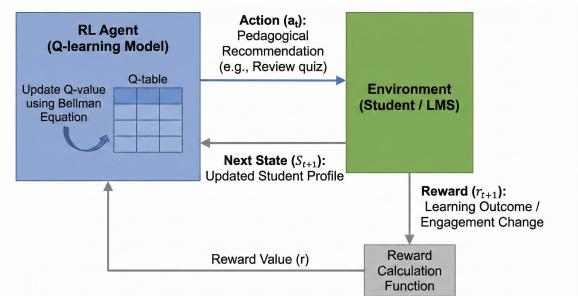


Figure 3. Sơ đồ chi tiết luồng hoạt động của thuật toán Q-learning trong việc đưa ra gợi ý sư phạm.

Figure 3. Sơ đồ chi tiết luồng hoạt động của thuật toán Q-learning trong việc đưa ra gợi ý sư phạm.

D. Khung giải thích mô hình (Explainability Framework)

Để giải quyết tính "hộp đen" của bảng Q-table và tăng sự tin cậy sư phạm, nghiên cứu tích hợp phương pháp SHAP (SHapley Additive exPlanations) [6].

1) Cơ sở toán học: Giá trị SHAP ϕ_i đo lường đóng góp biên của đặc trưng i vào giá trị Q dự đoán, dựa trên lý thuyết trò chơi hợp tác:

$$\phi_i = \sum_{S \subseteq F \setminus \{i\}} \frac{|S|!(|F| - |S| - 1)!}{|F|!} [f(S \cup \{i\}) - f(S)] \quad (4)$$

trong đó F là tập hợp đặc trưng và f là hàm tra cứu Q-table. Tính chất cộng tính giúp phân rã giá trị Q thành tổng các đóng góp: $Q(s, a^*) = \phi_0 + \sum \phi_i(s)$.

2) Cơ chế triển khai: Do không gian trạng thái rời rạc, chúng tôi sử dụng KernelExplainer để xấp xỉ giá trị Shapley. Quy trình bao gồm: (1) Xây dựng hàm dự đoán từ Q-table đã hội tụ; (2) Lấy mẫu nền (Background sampling) từ tập trạng thái quan sát được; và (3) Tính toán trọng số đóng góp cho từng đặc trưng trong vector trạng thái S_t .

E. Khung mô phỏng hướng dữ liệu (Data-driven Simulation Framework)

Để giải quyết bài toán "khởi động lạnh" (Cold-start) và đảm bảo tính hội tụ của thuật toán trước khi triển khai thực tế, chúng tôi đề xuất quy trình huấn luyện hai giai đoạn dựa trên khái niệm "Digital Twin" (Bản sao số).

1) *Mô hình hóa Xác suất chuyển đổi* (Transition Probability Modeling): Dựa trên dữ liệu lịch sử $\mathcal{D}_{history}$, hệ thống xây dựng một Ma trận xác suất chuyển đổi (Transition Probability Matrix - TPM) \mathcal{P} . Với mỗi cặp trạng thái - hành động (s, a) , xác suất người học chuyển sang trạng thái tiếp theo s' được ước lượng bởi:

$$P(s'|s, a) = \frac{\text{count}(s, a, s')}{\sum_{s^*} \text{count}(s, a, s^*)} \quad (5)$$

Ma trận \mathcal{P} này đóng vai trò là "môi trường nền" (Baseline Environment), phản ánh thói quen và phản ứng tự nhiên của người học trong quá khứ.

2) *Cơ chế Huấn luyện và Đánh giá* (Training & Evaluation Pipeline): Quy trình được thực hiện theo chu trình khép kín:

- 1) **Khởi tạo (Initialization):** Xây dựng các Tác nhân Người học (Student Agents) với các đặc trưng hành vi được tham số hóa từ $\mathcal{D}_{history}$.
- 2) **Huấn luyện (Training):** Agent tương tác với các Tác nhân Người học trong môi trường giả lập để tối ưu bảng Q-table.
- 3) **Đối sánh (Comparison):** So sánh chiến lược tối ưu π^* của Q-learning với chiến lược ngẫu nhiên có tham số (Parametric Policy) mô phỏng lại hành vi lịch sử \mathcal{P} .

III. THỰC NGHIỆM VÀ KẾT QUẢ

A. Thiết lập Môi trường Giả lập (Simulation Setup)

1) *Mô hình hóa Phong cách học tập:* Các tác nhân ảo (Virtual Agents) không hành động ngẫu nhiên mà sở hữu các phong cách học tập phi tuyến tính. Phân phối phong cách được thiết lập dựa trên tham số thực tế:

- **Linear Learner (70%):** Tuân thủ lô trình tuần tự truyền thống.
- **Practice-first (10%):** Ưu tiên thực hiện bài tập/quiz trước khi xem lý thuyết.
- **Video/Reading-first (20%):** Ưu tiên tiêu thụ nội dung thụ động trước khi tương tác.

2) *Cấu hình Ngẫu nhiên và Tham số Mô phỏng* (Simulation Settings & Stochasticity): Để đảm bảo tính khách quan và khả năng tái lập (reproducibility) của thực nghiệm, môi trường mô phỏng được thiết lập với các tham số chi tiết về nhiễu (noise) và quy mô mẫu như sau:

1) **Quy mô và Tái lập:** Quá trình huấn luyện được thực hiện trên tổng số 500 episodes, tương ứng với 500 tác nhân người học ảo (Student Agents) được khởi tạo ngẫu nhiên. Hạt giống ngẫu nhiên (Random Seed) được cố định ở giá trị 42 để đảm bảo sự nhất quán giữa các lần chạy thử nghiệm.

2) Mô hình Nhiễu (Noise Modeling): Mô phỏng tích hợp tính ngẫu nhiên để phản ánh sự biến thiên trong hành vi thực tế của sinh viên.

- **Biến thiên Điểm số (σ):** Điểm số đạt được sau mỗi hành động không cố định mà chịu tác động của nhiễu phân phối đều (Uniform Noise). Công thức tính điểm thực tế S_{real} được định nghĩa:

$$S_{real} = \text{clip}(S_{base} + \mathcal{U}(-\sigma_c, \sigma_c), 0, 1) \quad (6)$$

Trong đó σ_c là độ biến thiên đặc trưng cho từng cụm: Nhóm Yếu có độ biến động cao nhất ($\sigma = 0.18$), tiếp theo là Trung bình ($\sigma = 0.10$) và Thấp nhất ở nhóm Giỏi ($\sigma = 0.05$).

- **Biến thiên Thời gian:** Thời gian hoàn thành một hành động được lấy mẫu ngẫu nhiên trong khoảng từ 5 đến 30 phút để mô phỏng sự chênh lệch về tốc độ xử lý thông tin: $T \sim \mathcal{U}(5, 30)$ [cite: 1441].

3) Điều kiện dừng (Termination): Mỗi episode kết thúc khi tác nhân người học hoàn thành toàn bộ $N = 6$ module của khóa học hoặc khi đạt giới hạn bước tối đa (max steps = 100) để ngăn chặn các vòng lặp vô tận trong giai đoạn đầu của quá trình thăm dò (exploration) [cite: 1183, 1516].

Table I
TỔNG HỢP THAM SỐ CẤU HÌNH MÔ PHỎNG

Tham số	Mô tả	Giá trị
$N_{episodes}$	Số lượng vòng lặp huấn luyện	500
<i>Seed</i>	Hạt giống ngẫu nhiên	42
$N_{modules}$	Số module trong khóa học	6
$P_{success}$	Xác suất thành công cơ sở (Weak/Med/Strong)	0.72 / 0.78 / 0.90
c_{learn}	Tốc độ học tập (Weak/Med/Strong)	0.22 / 0.32 / 0.30

B. Thiết lập thực nghiệm (Experimental Setup)

1) *Dữ liệu huấn luyện* (Dataset Ground Truth): Nghiên cứu sử dụng bộ dữ liệu chuẩn Moodle Log & Grades [18]. Trong số các khóa học có sẵn, chúng tôi lựa chọn **Khóa học ID 670** làm cơ sở để xây dựng môi trường mô phỏng (Data-driven Environment).

Quyết định này dựa trên sự cân bằng lý tưởng của dữ liệu: Khóa học ghi nhận 13,995 điểm tương tác từ 23 sinh viên với phân phối điểm số chuẩn ($\mu = 7.64, \sigma = 2.95$). Điều này khắc phục được hạn chế dữ liệu bị lệch (skewed data) thường thấy ở các khóa học khác (ví dụ Course ID 42 có Mean=1.07), giúp thuật toán phân biệt rõ ràng chiến lược học tập giữa các nhóm sinh viên (Giỏi, Khá, Yếu).

2) *Tiền xử lý và Trích xuất đặc trưng:* Để giải quyết bài toán bùng nổ không gian trạng thái, chúng tôi áp dụng kỹ thuật **Lọc tương quan (Correlation Filtering)** với ngưỡng 0.95 trên 114 đặc trưng hành vi thô ban đầu. Kết quả đã thu gọn không gian đầu vào xuống còn **15 đặc trưng cốt lõi** (Hình 4), loại bỏ hoàn toàn hiện tượng đa cộng tuyến.

Sau bước trích xuất, thuật toán K-means được áp dụng với $K = 6$ cụm. Đáng chú ý, cụm dữ liệu của giảng viên đã

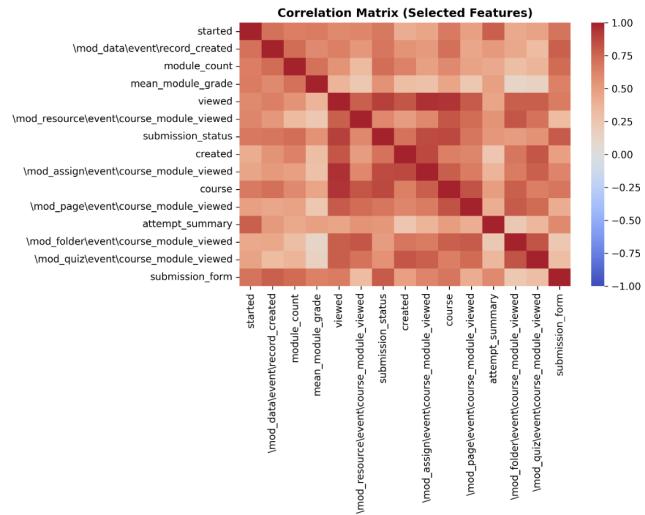


Figure 4. Ma trận tương quan Pearson giữa 15 đặc trưng hành vi cốt lõi. Các cặp biến có $|r| > 0.95$ đã được loại bỏ, đảm bảo không có đa công tuyến trong mô hình Q-learning.

được phát hiện và loại bỏ khỏi tập huấn luyện để đảm bảo tính chính xác cho Agent.

3) *Tham số mô hình (Model Parameters)*: Trong quá trình thực nghiệm, các tham số mô hình được thiết lập chi tiết để đảm bảo sự hội tụ của thuật toán:

- **Không gian trạng thái (S):** Trạng thái S_t được định nghĩa là vector 6 chiều phản ánh ngữ cảnh người học:

$$S_t = (C, M, P, Sc, Ph, E) \quad (7)$$

Trong đó: C là cụm người học ($K = 6$); M là chỉ số module; P, Sc là mức tiến độ và điểm số (rời rạc hóa 4 mức); Ph là giai đoạn học tập (0: Pre, 1: Active, 2: Reflective); và E là mức độ tương tác (Low, Med, High). Kích thước không gian là $5 \times 6 \times 4 \times 4 \times 3 \times 3 \approx 4,320$ trạng thái.

- **Không gian hành động (A):** Gồm $m = 15$ hành động được sàng lọc qua bộ lọc ICAP và Pareto ($> 1\%$ tần suất). Để tối ưu hiển thị, các hành động được nhóm theo ngữ cảnh thời gian như trình bày tại Bảng II.
- **Tham số Q-learning:** Tốc độ học $\alpha = 0.1$, hệ số chiết khấu $\gamma = 0.95$, và chiến lược ϵ -greedy giảm dần.

C. Độ đo đánh giá (Evaluation Metrics)

Hiệu quả của mô hình được đánh giá qua các chỉ số:

- **Tổng phần thưởng tích lũy (Total Reward):** Đo lường mức độ hội tụ của Agent.
- **Điểm số trung bình (Average Score):** Điểm kết thúc khóa học (thang 10).
- **Số kỹ năng yếu (Weak Skills Count):** Số lượng Chuẩn đầu ra (LO) có độ thông thạo < 0.5 .

D. Quy trình Sinh dữ liệu và Phân tích Hội tụ

1) *Quy trình Sinh dữ liệu (Simulation Loop):* Quy trình mô phỏng được thực hiện trên quy mô 500 vòng lặp

Table II
KHÔNG GIAN HÀNH ĐỘNG HOÀN CHỈNH (15 ACTIONS)

Nhóm	ID	Mã hành động	Ý nghĩa sự phạm
PAST (Ôn tập)	0	view_assign_past	Xem lại yêu cầu cũ
	1	view_content_past	Ôn lại bài giảng cũ
	2	attempt_quiz_past	Làm lại trắc nghiệm cũ
	3	review_quiz_past	Phân tích lỗi sai cũ
	4	post_forum_past	Thảo luận chủ đề cũ
CURRENT (Hiện tại)	5	view_assign_curr	Xem yêu cầu bài mới
	6	view_content_curr	Học nội dung tuần này
	7	submit_assign_curr	Nộp bài tập lớn
	8	attempt_quiz_curr	Làm bài kiểm tra
	9	submit_quiz_curr	Nộp bài lấy điểm
	10	review_quiz_curr	Xem kết quả vừa nộp
FUTURE (Chuẩn bị)	11	post_forum_curr	Thảo luận bài hiện tại
	12	view_content_fut	Xem trước bài mới
	13	attempt_quiz_fut	Thử sức bài tương lai
	14	post_forum_fut	Tìm hiểu chủ đề sắp tới

(episodes). Trong mỗi episode, hệ thống khởi tạo 100 tác nhân ảo (Virtual Agents) với phân phối năng lực mô phỏng lớp học thực tế: 20% Yếu, 60% Trung bình, và 20% Giỏi. Tổng cộng, mô hình đã huấn luyện trên tương tác của 50,000 lượt sinh viên ảo.

2) *Phân tích Hội tụ và Độ bao phủ:* Kết quả huấn luyện cho thấy dung lượng Q-table đạt xấp xỉ 219KB. Trong không gian lý thuyết 4,320 trạng thái, tác nhân đã khám phá và tối ưu hóa được 802 trạng thái cốt lõi (Core States). Đây là các trạng thái bao phủ hầu hết các kịch bản học tập thực tế, trong khi các trạng thái hiếm (Rare States) được xử lý thông qua cơ chế tổng quát hóa của Epsilon trong giai đoạn đầu.

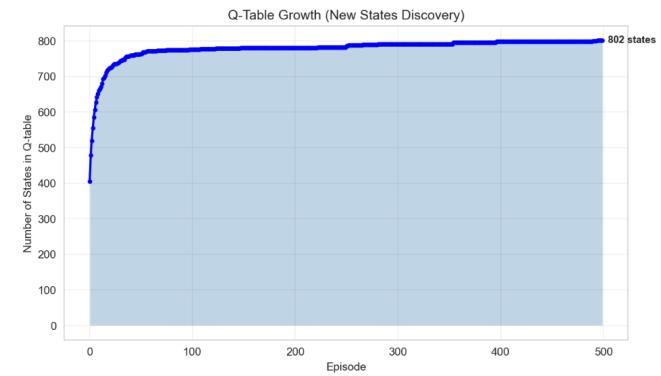


Figure 5. Biểu đồ hội tụ phần thưởng tích lũy (Cumulative Reward) qua 500 episodes. Sự ổn định bắt đầu xuất hiện rõ rệt sau episode 350.

Biểu đồ tại Hình 5 và Hình 6 minh họa mối tương quan nghịch biến giữa giá trị Epsilon và Phần thưởng tích lũy. Khi ϵ giảm dần về 0.01 (Giai đoạn khai thác), phần thưởng trung bình tăng trưởng ổn định, chứng tỏ tác nhân đã học được chiến lược tối ưu để tối đa hóa kết quả học tập.

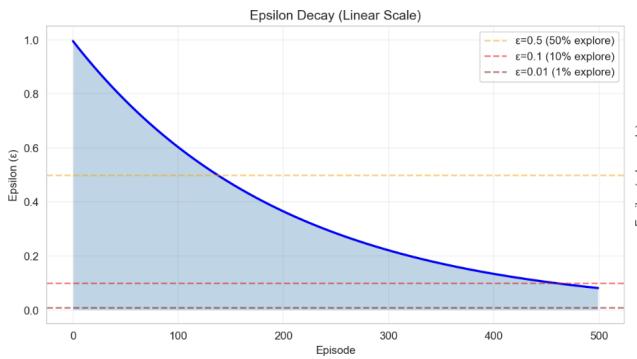


Figure 6. Chiến lược giảm dần số Epsilon (ϵ) qua 3 giai đoạn: Khám phá, Chuyển đổi và Khai thác.

E. Kết quả so sánh (A/B Testing)

Quá trình huấn luyện diễn ra qua 500 episodes với tham số $\alpha = 0.1$, $\gamma = 0.95$. Kết quả định lượng tổng hợp tại Bảng III và chi tiết từng phân cụm tại Hình 7 cho thấy sự vượt trội rõ rệt của thuật toán Q-learning.

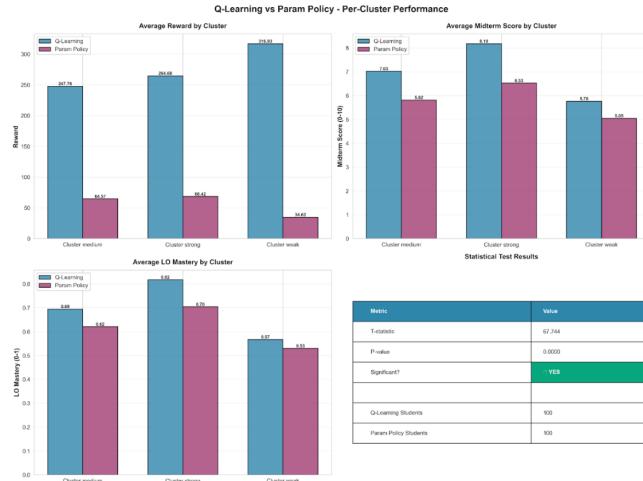


Figure 7. So sánh hiệu suất giữa Q-Learning và Param Policy trên các chỉ số: Phần thưởng (Reward), Điểm số (Score), và Độ thành thạo (LO Mastery).

Table III
KẾT QUẢ SO SÁNH HIỆU SUẤT TRUNG BÌNH

Chỉ số	Đối chứng	Q-learning	Cải thiện
Tổng phần thưởng	88.4	389.6	+340.8%
Điểm TB (thang 10)	6.25	7.66	+22.5%
Số kỹ năng yếu	3.02	1.48	-51.0%

F. Phân tích độ quan trọng đặc trưng

Phân tích SHAP (Hình 8) trên mô hình đã huấn luyện chỉ ra rằng *Module ID* và *Engagement* là hai yếu tố ảnh hưởng

nhất. Sự phân tán cao của giá trị SHAP đối với Engagement ($\sigma^2 = 995.79$) khẳng định rằng tác nhân AI đã học được cách đánh giá ngữ cảnh linh hoạt thay vì áp đặt luật cứng nhắc.

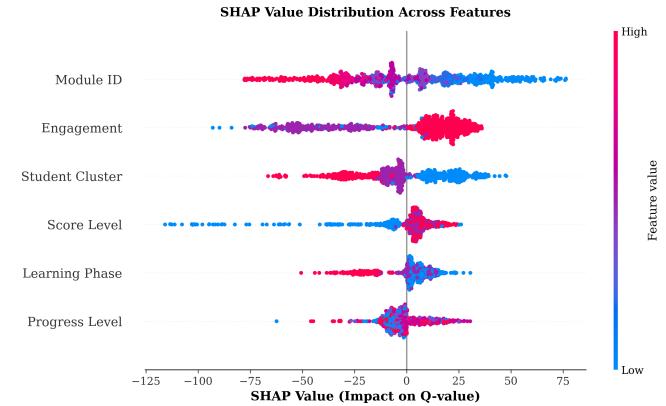


Figure 8. Phân phối giá trị SHAP từ thực nghiệm. Module ID và Engagement có ảnh hưởng rộng nhất đến quyết định của Agent.

G. Phân tích tác động sư phạm

Dựa trên biểu đồ trực quan tại Hình 7, hệ thống thích ứng tốt với từng nhóm người học:

- Nhóm Yếu (Weak):** Được gọi ý nhiều hành động ôn tập (Remedial), dẫn đến mức tăng trưởng phần thưởng cao nhất và giảm mạnh số lỗ hổng kiến thức (-51%) [4].
- Nhóm Giỏi (Strong):** Đạt điểm số tuyệt đối cao nhất (8.18/10) nhờ các gợi ý mang tính thách thức (Advanced).

Kiểm định T-test độc lập xác nhận sự khác biệt có ý nghĩa thống kê ($p < 0.001$) với kích thước ảnh hưởng lớn (Cohen's $d = 6.78$).

IV. KẾT LUẬN VÀ HƯỚNG PHÁT TRIỂN

Bài báo đã trình bày một giải pháp toàn diện để cá nhân hóa giáo dục STEM thông qua việc ứng dụng Học tăng cường. Các đóng góp chính bao gồm: (1) Kiến trúc hệ thống mở dựa trên Microservices và LTI 1.3; (2) Quy trình mô hình hóa dữ liệu người học chi tiết; và (3) Thuật toán Q-learning với cơ chế thưởng thích ứng.

Kết quả thực nghiệm cho thấy hệ thống không chỉ cải thiện điểm số (+22.5%) mà quan trọng hơn là giúp lấp đầy các lỗ hổng kiến thức cho sinh viên yếu (-51% số kỹ năng yếu), hiện thực hóa mục tiêu “không ai bị bỏ lại phía sau”.

Tuy nhiên, nghiên cứu vẫn còn hạn chế khi chưa được triển khai trên lớp học thực tế (Live Deployment) và không gian trạng thái bị giới hạn bởi phương pháp rời rạc hóa.

Hướng phát triển trong tương lai bao gồm:

- Deep Reinforcement Learning (DRL):** Áp dụng mạng nơ-ron sâu (DQN, PPO) để xử lý không gian trạng thái liên tục và phức tạp hơn [5].

- **Triển khai thực tế:** Tích hợp hệ thống vào các khóa học STEM tại trường đại học để thu thập dữ liệu phản hồi thực và tinh chỉnh mô hình.
- **Federated Learning:** Nghiên cứu cơ chế học tập liên kết để bảo vệ quyền riêng tư dữ liệu người học khi triển khai trên nhiều cơ sở giáo dục [4].

LỜI CẢM ƠN

Nhóm tác giả xin chân thành cảm ơn TS. Đỗ Như Tài đã hướng dẫn tận tình. Nghiên cứu được thực hiện tại Khoa Công nghệ Thông tin, Trường Đại học Sài Gòn.

REFERENCES

- [1] E. du Plooy et al., "Personalized adaptive learning in higher education: A scoping review of key characteristics and impact on academic performance and engagement," *Heliyon*, vol. 10, no. 21, p. e39630, 2024. [cite: 721]
- [2] C. B. Frey and M. A. Osborne, "The future of employment: How susceptible are jobs to computerisation?," *Tech. Forecast. Soc. Change*, vol. 114, pp. 254–280, 2017. [cite: 120]
- [3] A. Riedmann, P. Schaper, and B. Lugrin, "Reinforcement Learning in Education: A Systematic Literature Review," *Int. J. Artif. Intell. Educ.*, 2025. DOI: 10.1007/s40593-025-00494-6. [cite: 724]
- [4] I. Gligoreia et al., "Adaptive Learning Using Artificial Intelligence in e-Learning: A Literature Review," *Educ. Sci.*, vol. 13, no. 12, 2023. [cite: 729]
- [5] R. S. Sutton and A. G. Barto, *Reinforcement Learning: An Introduction*. MIT Press, 2018. [cite: 460]
- [6] S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," arXiv:1705.07874, 2017. DOI: 10.48550/arXiv.1705.07874. [cite: 1705.07874]
- [7] S. Wu et al., "A Comprehensive Exploration of Personalized Learning in Smart Education: From Student Modeling to Personalized Recommendations," arXiv:2402.01666, 2024. [cite: 728]
- [8] W. Villegas-Ch et al., "Adaptive intelligent tutoring systems for STEM education: analysis of the learning impact and effectiveness of personalized feedback," *Int. J. Educ. Technol. High. Educ.*, 2025. [cite: 730]
- [9] V. Aleven et al., "Adaptive Learning Technologies," in *Handbook of Learning Analytics*, 2016. [cite: 726]
- [10] P. L. Nguyen, "Vietnam's STEM Education Landscape: Evolution, Challenges, and Policy Interventions," *Vietnam J. Educ.*, vol. 8, no. 2, pp. 177–189, 2024. [cite: 723]
- [11] T. B. Thuan, "Ứng dụng machine learning dự báo sinh viên diện cảnh báo học tập tại trường đại học kinh tế Huế," *Hue Uni. Journal of Science*, 2022. [cite: 736]
- [12] L. H. Sang, N. T. Hai, T. T. Dien, and N. T. Nghe, "Dự báo kết quả học tập bằng kỹ thuật học sâu với mạng nơ-ron đa tầng," *Can Tho Univ. J. Sci.*, vol. 56, 2020. DOI: 10.22144/ctu.jvn.2020.049. [cite: 737]
- [13] "Ứng dụng AI trong thiết kế Khóa học trực tuyến tại Khoa Công nghệ số và Kỹ thuật Trường Đại học Đồng Tháp," *Tạp chí Thiết bị Giáo dục*, 2024. [cite: 735]
- [14] IMS Global, "LTI 1.3 Implementation Guide," [Online]. Available: <https://www.imsglobal.org/spec/lti/v1p3>. [cite: 733]
- [15] Kong Gateway Developer Documentation. [Online]. Available: <https://developer.konghq.com/index/gateway/>. [cite: 733]
- [16] MoodleDev, "Update LTI tool provider feature to support 1.3." [Online]. Available: <https://moodledev.io/general/releases/4.0>. [cite: 725]
- [17] R. W. Bybee, *The Case for STEM Education: Challenges and Opportunities*. NSTA Press, 2013. [cite: 738]
- [18] M. Sneiders, "Moodle Grades and Action Logs Dataset," Kaggle, 2021. [Online]. Available: <https://www.kaggle.com/datasets/martinssneiders/moodle-grades-and-action-logs>.