# 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 lộ 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, Microservices, LTI 1.3.**

## 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 sư phạm 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, với cơ chế học thử-sai (trial-and-error), cho phép hệ thống tự động khám phá và tối ưu hóa chiến lược giảng dạy dựa trên phản hồi liên tục từ người học.

Nghiên cứu này đề xuất xây dựng một hệ thống gợi ý thông minh tích hợp vào Moodle LMS, với các đóng góp chính sau:

1) Đề xuất kiến trúc Microservices tích hợp qua chuẩn LTI 1.3, đảm bảo khả năng mở rộng và tương thích với nhiều nền tảng LMS.
2) Xây dựng quy trình xử lý dữ liệu và phân cụm người học để giải quyết bài toán không gian trạng thái trong RL.
3) Thiết kế và tối ưu hóa thuật toán Q-learning với hàm phần thưởng đa mục tiêu, thích ứng với đặc điểm của từng nhóm người học.
4) Kiểm chứng hiệu quả của hệ thống thông qua mô phỏng so sánh (A/B testing) với dữ liệu tham số hóa từ thực tế.

## II. Tổng quan nghiên cứu

### A. Hệ thống học tập thích ứng (Adaptive Learning)

Các hệ thống PAL điều chỉnh lộ trình học tập dựa trên nhu cầu riêng biệt của sinh viên. Theo báo cáo tổng quan của du Plooy và cộng sự (2024), khảo sát 69 công trình nghiên cứu, 59% số nghiên cứu ghi nhận sự cải thiện về kết quả học tập và 36% chỉ ra sự gia tăng mức độ tham gia khi áp dụng PAL [1]. Các nền tảng như Moodle và McGraw-Hill's Connect LearnSmart là những môi trường phổ biến để triển khai các giải pháp này.

### B. Học tăng cường trong giáo dục

Trong những năm gần đây, Học tăng cường (RL) đã nổi lên như một phương pháp hiệu quả để xây dựng các gia sư thông minh (Intelligent Tutoring Systems). Theo Riedmann (2025), xu hướng sử dụng RL trong giáo dục tăng mạnh từ năm 2018, đặc biệt trong lĩnh vực STEM [3]. Về mặt thuật toán, Q-learning và Deep Q-Network (DQN) là phổ biến nhất nhờ tính linh hoạt của phương pháp Model-free, cho phép hệ thống học chiến lược tối ưu mà không cần mô hình hóa chính xác quy luật phức tạp của hành vi con người [5].

Tuy nhiên, Riedmann cũng chỉ ra hạn chế lớn của các nghiên cứu hiện tại là sự phụ thuộc vào dữ liệu mô phỏng và thiếu các triển khai thực tế tại các quốc gia đang phát triển [3].

### C. Tình hình nghiên cứu tại Việt Nam

Tại Việt Nam, xu hướng ứng dụng công nghệ trong giáo dục đang có sự chuyển dịch mạnh mẽ từ số hóa bài giảng đơn thuần sang khai thác dữ liệu thông minh (Educational Data Mining). Các nghiên cứu trong giai đoạn 2020-2024 chủ yếu tập trung vào hai nhóm vấn đề cốt lõi:

*1) Các mô hình dự báo và cảnh báo học vụ:* Đây là hướng nghiên cứu chiếm ưu thế nhờ tận dụng nguồn dữ liệu điểm số sẵn có tại các trường đại học. Nghiên cứu của Trần Bá Thuấn (2022) đã giải quyết bài toán dữ liệu mất cân bằng để dự báo sớm nguy cơ bị cảnh báo học vụ. Tác giả đã thực nghiệm so sánh các thuật toán như kNN, Decision Tree và Naïve Bayes, kết quả cho thấy Random Forest đạt hiệu suất cao nhất nhờ khả năng giảm thiểu tình trạng học lệch (overfitting)[cite: 187, 188].

Tại Đại học Cần Thơ, Lưu Hoài Sang và cộng sự (2020) đã đề xuất phương pháp dự báo kết quả học phần sử dụng mạng nơ-ron đa tầng (Multi-layer Perceptron - MLP). Nghiên cứu chỉ ra rằng việc áp dụng Deep Learning với cơ chế Dropout giúp mô hình xử lý tốt các mối quan hệ phi tuyến tính trong dữ liệu giáo dục, đạt độ lỗi RMSE thấp hơn so với các phương pháp thống kê truyền thống[cite: 189, 190].

Tuy nhiên, hạn chế chung của các mô hình này là tính chất "dự báo tĩnh" (Static Prediction). Hệ thống chỉ đưa ra kết quả dự đoán (ví dụ: rớt môn) nhưng chưa tự động đề xuất được các hành động can thiệp cụ thể (Prescriptive Analytics) theo thời gian thực để giúp người học cải thiện tình hình[cite: 191, 192].

*2) Hệ thống gợi ý và Tích hợp LMS:* Về bài toán gợi ý, các nghiên cứu trong nước thường áp dụng kỹ thuật Lọc cộng tác (Collaborative Filtering). Tuy nhiên, phương pháp này phụ thuộc lớn vào lịch sử tương đồng giữa các người dùng, do đó dễ gặp vấn đề "khởi động lạnh" (Cold-start) khi áp dụng cho khóa học mới hoặc sinh viên mới[cite: 193, 195].

Gần đây, một số nghiên cứu như của Phạm Huệ Minh (2024) đã bắt đầu tích hợp AI tạo sinh (như ChatGPT) vào Moodle để hỗ trợ tra cứu thông tin[cite: 197]. Mặc dù vậy, đa số các giải pháp này được triển khai dưới dạng Plugin cài trực tiếp lên mã nguồn LMS (kiến trúc Monolithic). Cách tiếp cận này thiếu tính linh hoạt khi mở rộng quy mô và chưa tận dụng được các chuẩn kết nối độc lập như LTI để tách biệt dữ liệu xử lý AI khỏi dữ liệu quản lý đào tạo[cite: 198, 199].

*3) Khoảng trống nghiên cứu:* Từ phân tích trên, có thể thấy tại Việt Nam vẫn thiếu vắng các giải pháp kết hợp đồng thời ba yếu tố: (1) Thuật toán Học tăng cường (RL) để tạo ra gợi ý động; (2) Kiến trúc Microservices để đảm bảo khả năng mở rộng; và (3) Chuẩn LTI 1.3 để tích hợp liền mạch. Đề tài này được thực hiện nhằm lấp đầy khoảng trống đó[cite: 204, 205].

## III. Kiến trúc hệ thống và Xử lý dữ liệu

### A. Kiến trúc Microservices và Công nghệ nền tảng

*1) Cơ sở lựa chọn kiến trúc:* Đa số các giải pháp tích hợp AI vào LMS hiện nay tại Việt Nam thường được triển khai dưới dạng Plugin cài đặt trực tiếp lên mã nguồn (Monolithic). Mặc dù đơn giản, cách tiếp cận này bộc lộ nhiều hạn chế về khả năng mở rộng và rủi ro bảo mật. Để khắc phục, hệ thống được thiết kế theo kiến trúc Microservices, tách biệt hoàn toàn module xử lý AI khỏi hệ thống quản lý đào tạo.

*2) Các thành phần cốt lõi:* Hệ thống được phân rã thành các dịch vụ độc lập, giao tiếp qua RESTful API và được quản lý tập trung bởi Kong API Gateway (như mô tả tại Hình 1):

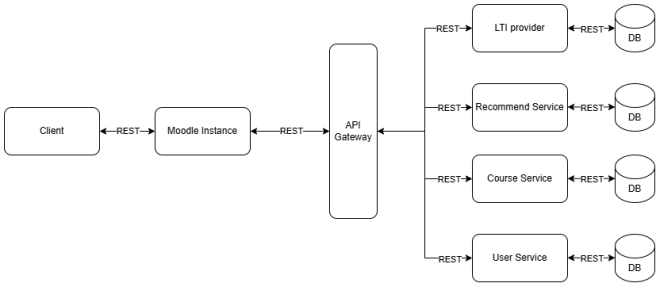


Figure 1. Sơ đồ kiến trúc Microservices và luồng dữ liệu tích hợp.

- **LTI Integration Service:** Đóng vai trò là cổng giao tiếp an toàn, thực hiện quy trình bắt tay (handshake) và xác thực OAuth 2.0 với Moodle theo chuẩn LTI 1.3.
- **Recommend Service:** Trung tâm trí tuệ của hệ thống, vận hành thuật toán Q-learning để tính toán và trả về danh sách gợi ý tối ưu.
- **User & Course Service:** Quản lý đồng bộ metadata cấu trúc khóa học và hồ sơ người dùng từ LMS.
- **Frontend:** Ứng dụng ReactJS hiển thị Dashboard tương tác.

Dữ liệu được lưu trữ phân tán theo mô hình "Database-per-service". Trong đó, MongoDB được sử dụng cho Recommend Service để xử lý dữ liệu log hành vi phi cấu trúc và trạng thái người học thay đổi liên tục.

*3) Cơ chế lưu trữ dữ liệu:* Hệ thống áp dụng mô hình "Database-per-service" để đảm bảo tính lỏng lẻo (loose coupling). Trong đó, MongoDB được lựa chọn làm cơ sở dữ liệu chính cho Recommend Service. Với đặc thù dữ liệu log hành vi phi cấu trúc và trạng thái người học thay đổi liên tục theo thời gian thực (Real-time State Tracking), cơ sở dữ liệu NoSQL như MongoDB cho phép truy xuất và ghi nhận vector trạng thái nhanh chóng hơn so với các RDBMS truyền thống[cite: 325].

### B. Quy trình Xử lý dữ liệu và Phân cụm

Nghiên cứu sử dụng bộ dữ liệu mở Moodle Log & Grades [18], bao gồm hơn 1.2 triệu dòng nhật ký hành vi thô. Sau quá trình tiền xử lý loại bỏ các sự kiện hệ thống tự động (như *webservice_function_called*), chúng tôi đã sàng lọc và

lựa chọn **Khóa học ID 670** làm đối tượng huấn luyện mô hình (Ground Truth).

Quyết định lựa chọn này dựa trên sự so sánh đối chiếu về chất lượng dữ liệu. Trong khi các khóa học khác (như Course ID 42) có mật độ tương tác cao nhưng dữ liệu điểm số bị lệch (Mean=1.07, chủ yếu là điểm 0), Course 670 thể hiện sự cân bằng lý tưởng cho bài toán Học tăng cường:

- **Dữ liệu hành vi phong phú:** Ghi nhận 13,995 điểm dữ liệu tương tác từ 23 sinh viên, đảm bảo độ dài chuỗi hành động đủ lớn cho mô hình Markov.
- **Phân phối điểm số chuẩn:** Với điểm trung bình $\mu = 7.64$ và độ lệch chuẩn $\sigma = 2.95$, phổ điểm có độ phân tán tốt, cho phép thuật toán phân biệt rõ ràng chiến lược học tập giữa các nhóm sinh viên (Giỏi, Khá, Yếu).

*1) Tiền xử lý và Phân tích đặc trưng (EDA):* Dữ liệu log thô chứa nhiều nhiễu từ các sự kiện hệ thống tự động (như *webservice_function_called*), vì vậy các sự kiện này được loại bỏ và thời gian được chuẩn hóa trước khi trích xuất đặc trưng. Từ 114 đặc trưng hành vi ban đầu, chúng tôi áp dụng **Lọc tương quan (Correlation Filtering)** với ngưỡng 0.95 để loại bỏ đa cộng tuyến, thu gọn còn **15 đặc trưng cốt lõi** (ví dụ: *module_count*, *mean_grade*, *time_spent*). Ma trận tương quan ở Hình 2 cho thấy không gian đặc trưng đã được tinh gọn, giúp tránh bùng nổ bảng Q-table. Bộ đặc trưng này là đầu vào cho bước phân cụm được trình bày ở mục tiếp theo.

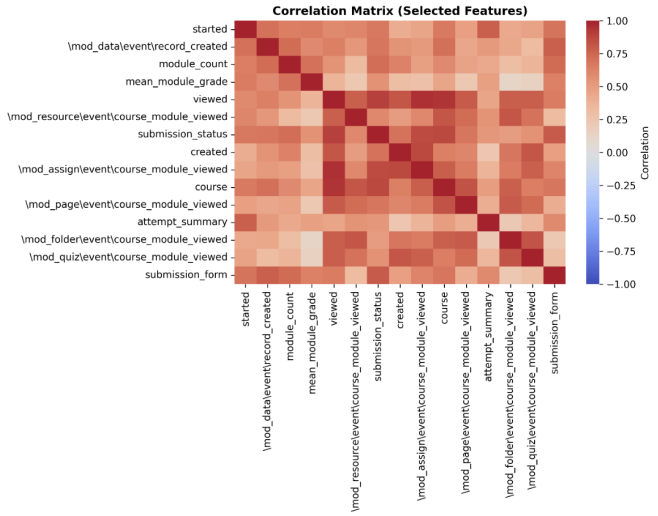


Figure 2. Ma trận tương quan giữa 15 đặc trưng hành vi cốt lõi sau khi loại bỏ các biến đa cộng tuyến.

*2) Phân cụm người học (Clustering):* Việc cá nhân hóa đòi hỏi hệ thống phải nhận diện được đặc điểm của từng nhóm đối tượng. Chúng tôi sử dụng thuật toán K-means để phân nhóm sinh viên dựa trên 15 đặc trưng đã trích xuất. Số lượng cụm tối ưu $K = 6$ được xác định thông qua phương pháp Elbow và chỉ số Silhouette [4]. Tuy nhiên, qua phân tích chi tiết, Cluster 3 được xác định là chứa dữ liệu log của tài khoản giảng viên (có pattern tương tác quản trị và chỉnh sửa nội dung khóa học), do đó đã được loại bỏ khỏi tập dữ liệu huấn luyện. Các nhóm người học hợp lệ được định danh bao gồm:

- **Cluster 0 (Nhóm cần hỗ trợ - 34.8%):** Tương tác thấp, thụ động, điểm số thấp.
- **Cluster 1 (Nhóm tự giác):** Thường xuyên theo dõi tiến độ và bảng điểm.
- **Cluster 2 (Nhóm chủ động - 30.4%):** Tương tác cao, nộp bài đầy đủ.
- **Cluster 4 (Nhóm nghiên cứu):** Tập trung vào việc xem và tải tài liệu.
- **Cluster 5 (Nhóm thành tích):** Quan tâm đến huy hiệu và xếp hạng [4].

Thông tin Cluster ID này sẽ được đưa vào vector trạng thái của thuật toán Q-learning.

## IV. Mô hình hóa bài toán Q-learning

Bài toán gợi ý lộ trình học tập được mô hình hóa dưới dạng Quy trình Quyết định Markov (MDP), bao gồm bộ ba $< S, A, R >$.

### A. Không gian trạng thái (State Space)

Trạng thái $S_t$ tại thời điểm $t$ đại diện cho tình trạng học tập hiện tại của sinh viên, được định nghĩa là một vector 6 chiều:

$$S_t = (C, M, P, Sc, Ph, E) \tag{1}$$

Trong đó:

- $C$ (Cluster): Nhóm người học $(0 \dots 5)$ từ kết quả phân cụm.
- $M$ (Module): Chỉ số bài học hiện tại trong lộ trình tuần tự.
- $P$ (Progress): Mức độ hoàn thành module, được rời rạc hóa thành 4 mức (0.25, 0.5, 0.75, 1.0).
- $Sc$ (Score): Điểm số tích lũy, chia thành 4 mức tương ứng (Yếu, TB, Khá, Giỏi).
- $Ph$ (Phase): Giai đoạn học tập (0: Pre-learning, 1: Active-learning, 2: Reflective-learning).
- $E$ (Engagement): Mức độ tương tác (Thấp, TB, Cao), tính toán dựa trên trọng số hành động theo khung ICAP [4].

Kích thước không gian trạng thái là $5 \times 6 \times 4 \times 4 \times 3 \times 3 \approx 4,320$ trạng thái, đảm bảo tính khả thi cho việc hội tụ của bảng Q-table.

### B. Không gian hành động (Action Space)

Trong mô hình MDP, không gian hành động $A$ đại diện cho tập hợp các tương tác sư phạm mà hệ thống có thể đề xuất. Việc thiết kế $A$ không đơn thuần là liệt kê các sự kiện nhật ký (logs), mà là quá trình sàng lọc và mở rộng có chủ đích.

*1) Quy trình sàng lọc hành động cốt lõi:* Hệ thống Moodle ghi nhận hàng trăm loại sự kiện khác nhau. Để chọn ra các hành động có ý nghĩa, chúng tôi áp dụng 3 tiêu chí sàng lọc:

1) **Tác động học tập (ICAP Framework):** Chỉ giữ lại các hành động thuộc nhóm *Interactive*, *Constructive*,

và *Active*. Loại bỏ các hành động quản trị hệ thống (*Passive/System*).

2) **Tần suất dữ liệu (Pareto Principle):** Loại bỏ các hành động hiếm gặp ($< 1\%$ tổng log) để đảm bảo mô hình hội tụ.
3) **Tác động chuyển đổi trạng thái:** Ưu tiên các hành động gây biến động lớn lên vector trạng thái (ví dụ: *submit_quiz* làm thay đổi điểm số và tiến độ).

Kết quả sàng lọc thu được **7 hành động cốt lõi**: *view_assignment, view_content, attempt_quiz, submit_quiz, review_quiz, submit_assignment, post_forum*.

*2) Mở rộng ngữ cảnh thời gian (Contextual Expansion):* Một hệ thống thông minh cần xác định không chỉ "làm gì" mà còn "làm ở đâu" (trong quá khứ, hiện tại hay tương lai). Do đó, 7 hành động cốt lõi được nhân rộng theo 3 ngữ cảnh thời gian:

- **Past (Quá khứ):** Phục vụ mục đích ôn tập (Review) và lấp lỗ hổng kiến thức.
- **Current (Hiện tại):** Thực hiện nhiệm vụ (Execution) đúng tiến độ.
- **Future (Tương lai):** Chuẩn bị bài mới (Preparation) và kích thích tò mò.

Tuy nhiên, không phải mọi tổ hợp đều hợp lệ (ví dụ: không thể nộp bài tập tương lai khi chưa mở). Sau khi áp dụng ma trận lọc logic sư phạm, không gian hành động cuối cùng bao gồm **15 hành động** được đánh chỉ số từ 0 đến 14 như trình bày tại Bảng I.

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

| Nhóm | ID | Mã hành động | Ý nghĩa sư 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 |
| | 11 | post_forum_curr | Thảo luận bài hiện tại |
| FUTURE (Chuẩn bị) | 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 |

Việc thiết kế không gian hành động phân theo thời gian này giúp Agent có khả năng đưa ra các chiến lược linh hoạt: *Remedial Strategy* (gợi ý Past khi điểm thấp), *Progressive Strategy* (gợi ý Current để hoàn thành tiến độ), và *Anticipatory Strategy* (gợi ý Future cho sinh viên khá giỏi).

### C. Hàm phần thưởng (Reward Function)

Hàm thưởng là thành phần quan trọng nhất định hướng hành vi của tác nhân. Chúng tôi thiết kế hàm thưởng tổng hợp $R_{total}$ bao gồm 4 thành phần:

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

*1) Phần thưởng cơ bản thích ứng ($R_{base}$):* Áp dụng chiến lược Cluster-Adaptive Reward để điều chỉnh động lực. Nhóm sinh viên Yếu (Weak Cluster) nhận được điểm thưởng cao hơn (+10) khi hoàn thành một hành động so với nhóm Trung bình (+7) và nhóm Giỏi (+5). Cơ chế này nhằm tạo động lực ngoại sinh, khuyến khích nhóm yếu duy trì tương tác thay vì bỏ cuộc [5].

*2) Phần thưởng dựa trên Chuẩn đầu ra ($R_{LO}$):* Được tính dựa trên mức độ cải thiện năng lực ($\Delta Mastery$) của sinh viên đối với các Chuẩn đầu ra (LO).

$$R_{LO} = \sum(\Delta Mastery_i \times W_{midterm,i} \times K_{cluster}) \quad (3)$$

Trong đó:

- $\Delta Mastery = \alpha \times (Score - M_t)$ là mức tăng trưởng kiến thức thực tế, với hệ số học $\alpha$ được cấu hình riêng (Nhóm Yếu $\alpha = 0.4$, Nhóm Giỏi $\alpha = 0.2$).
- $W_{midterm}$ là trọng số quan trọng của LO trong bài thi.
- $K_{cluster}$ là hệ số khích lệ tâm lý (Weak: 1.5, Medium: 1.2, Strong: 1.0).

*3) Phần thưởng bổ sung ($R_{bonus}$):* Thành phần này ($R_{bonus} = R_{phase} + R_{seq}$) tích hợp các nguyên tắc sư phạm vào quá trình học:

- **Phần thưởng Giai đoạn ($R_{phase}$):** Thưởng thêm nếu hành động của sinh viên khớp với giai đoạn học tập hiện tại (ví dụ: đang ở giai đoạn *Active Learning* mà thực hiện *attempt_quiz*).
- **Phần thưởng Chuỗi hành động ($R_{seq}$):** Khuyến khích các chuỗi hành động hợp lý (Beneficial Sequences) như: *view_content* $\rightarrow$ *attempt_quiz* (Tiếp thu $\rightarrow$ Thực hành), hoặc *submit_quiz* $\rightarrow$ *review_quiz* (Thực hành $\rightarrow$ Phản tư).

*4) Điểm phạt thất bại ($P_{penalty}$):* Để duy trì chất lượng học tập, hệ thống áp dụng điểm phạt khi kết quả không đạt yêu cầu. Mức phạt được cá nhân hóa để tránh gây áp lực ngược: Nhóm Yếu bị phạt thấp (-1.0) để giữ động lực, trong khi Nhóm Giỏi chịu mức phạt cao hơn (-2.0) để yêu cầu sự chỉn chu.

### D. Thuật toán Q-learning

Hệ thống sử dụng thuật toán Q-learning tiêu chuẩn để cập nhật bảng giá trị Q:

$$Q(s,a) \leftarrow Q(s,a) + \alpha[r + \gamma \max_{a'} Q(s',a') - Q(s,a)] \quad (4)$$

Với tốc độ học $\alpha = 0.1$ và hệ số chiết khấu $\gamma = 0.95$. Chiến lược lựa chọn hành động là $\epsilon$-greedy, với $\epsilon$ giảm dần từ 1.0 xuống 0.01 sau 400 episodes để cân bằng giữa khám phá và khai thác [5].

### *E. Khung giải thích mô hình (Explainability Framework)*

Để giải quyết bản chất "hộp đen" của bảng Q-table và tăng tính minh bạch cho các quyết định gợi ý, nghiên cứu áp dụng phương pháp **SHAP (SHapley Additive exPlanations)** [6] - một kỹ thuật giải thích mô hình dựa trên lý thuyết trò chơi hợp tác.

*1) Cơ sở toán học:* Giá trị SHAP $\phi_i$ đo lường mức độ đóng góp của đặc trưng $i$ vào giá trị Q dự đoán, được định nghĩa qua giá trị Shapley [6]:

$$\phi_i = \sum_{S \subseteq F \setminus \{i\}} \frac{|S|!(|F| - |S| - 1)!}{|F|!} [f(S \cup \{i\}) - f(S)] \quad (5)$$

trong đó $F$ là tập hợp toàn bộ các đặc trưng trạng thái, $S$ là tập con các đặc trưng, và $f$ là hàm tra cứu Q-table. SHAP đảm bảo tính chất cộng tính (additivity):

$$Q(s, a^*) = \phi_0 + \sum_{i=1}^{6} \phi_i(s) \quad (6)$$

với $\phi_0$ là giá trị kỳ vọng cơ sở và $a^*$ là hành động tối ưu.

*2) Triển khai kỹ thuật:* Do bảng Q-table là một hàm rời rạc không khả vi, chúng tôi sử dụng **KernelExplainer** - một phương pháp model-agnostic không yêu cầu gradient. Quy trình thực hiện gồm 3 bước:

1) **Xây dựng hàm dự đoán:** Tạo wrapper function $f(s) = \max_a Q(s, a)$ ánh xạ từ vector trạng thái 6 chiều sang giá trị Q tối đa.
2) **Lấy mẫu nền (Background sampling):** Chọn ngẫu nhiên 100 trạng thái từ Q-table làm baseline để tính toán hiệu ứng biên (marginal effect).
3) **Tính toán SHAP values:** KernelExplainer xấp xỉ giá trị Shapley bằng cách thực hiện hồi quy tuyến tính có trọng số trên không gian các tập con đặc trưng (feature coalitions). Với $2^6 = 64$ tập con có thể, thuật toán hoàn tất trong thời gian $O(N \cdot 2^F)$ với $N$ là số mẫu kiểm tra.

*3) Phân tích độ quan trọng đặc trưng:* Từ SHAP values, chúng tôi tính toán hai chỉ số chính:

- **Mean Absolute SHAP**: $I_i = \frac{1}{N} \sum_{j=1}^{N} |\phi_i(s_j)|$ - đo mức độ ảnh hưởng trung bình.
- **SHAP Variance**: $V_i = \text{Var}(\phi_i)$ - đo tính nhất quán của ảnh hưởng.

Kết quả định lượng được tổng hợp tại Bảng II và trực quan hóa tại Hình 3. Phân tích cho thấy *Module ID* (mean |SHAP| = 28.32) và *Engagement* (26.53) là hai yếu tố quan trọng nhất ảnh hưởng đến quyết định gợi ý của Agent.

Đặc biệt, biểu đồ beeswarm (Hình 3) tiết lộ các pattern quan trọng về tác động phi tuyến của các đặc trưng. *Engagement* thể hiện phương sai cao nhất ($\sigma^2 = 995.79$), với phân phối SHAP values trải rộng từ -50 đến +50, chứng tỏ tác động của yếu tố này mang tính ngữ cảnh cao (context-dependent): engagement cao có thể tăng Q-value mạnh (màu đỏ, phía dương) khi sinh viên đang ở module phù hợp, nhưng cũng có thể giảm Q-value (màu xanh, phía âm) nếu tương tác không đúng hướng.

Ngược lại, *Progress Level* có mean |SHAP| thấp nhất (7.42) và phương sai nhỏ (95.96), với các điểm dữ liệu tập trung quanh giá trị 0 trên biểu đồ, xác nhận rằng tiến độ hoàn thành đơn thuần ít ảnh hưởng đến chiến lược. Điều này hỗ trợ cho giả thuyết sư phạm: "chất lượng tương tác" (engagement quality) quan trọng hơn "số lượng hoàn thành" (completion quantity) trong việc dự báo kết quả học tập.

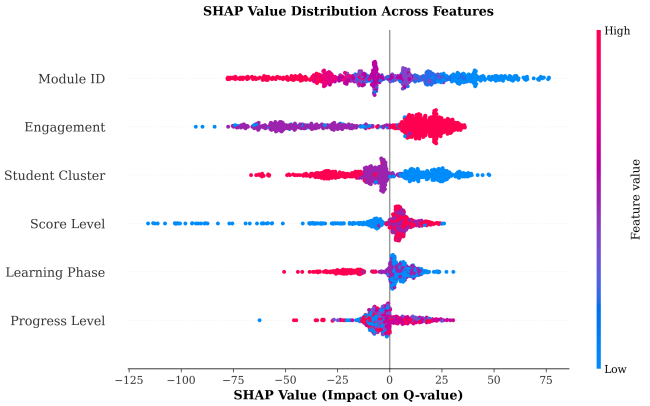


Figure 3. Phân phối SHAP values trên 802 trạng thái từ Q-table. Mỗi điểm biểu diễn SHAP value của một state, màu sắc thể hiện giá trị feature (đỏ = cao, xanh = thấp). Trục X cho thấy mức độ ảnh hưởng lên Q-value (dương = tăng, âm = giảm). Engagement và Module ID có impact range rộng nhất, trong khi Progress Level tập trung quanh 0.

Table II
XẾP HẠNG ĐỘ QUAN TRỌNG CỦA CÁC ĐẶC TRƯNG TỪ PHÂN TÍCH SHAP TRÊN 802 TRẠNG THÁI. MEAN |SHAP| ĐO MỨC ĐỘ ẢNH HƯỞNG TRUNG BÌNH; VARIANCE ĐO TÍNH BIẾN ĐỘNG (CAO = CONTEXT-DEPENDENT, THẤP = ỔN ĐỊNH).

| **Đặc trưng** | **Mean \|SHAP\|** | **Variance** | **Rank** |
|---|---|---|---|
| Module ID | 28.32 | 1171.49 | 1 |
| Engagement | 26.53 | 995.79 | 2 |
| Student Cluster | 17.42 | 461.44 | 3 |
| Score Level | 11.39 | 431.01 | 4 |
| Learning Phase | 9.12 | 149.83 | 5 |
| Progress Level | 7.42 | 95.96 | 6 |

## V. THỰC NGHIỆM VÀ KẾT QUẢ

Do giới hạn về việc triển khai trên sinh viên thực tế trong thời gian ngắn, nghiên cứu sử dụng phương pháp mô phỏng dựa trên dữ liệu (Data-driven Simulation) để giải quyết vấn đề "khởi động lạnh" và kiểm chứng thuật toán.

### *A. Thiết lập môi trường mô phỏng*

Môi trường giả lập được xây dựng dựa trên các tham số thống kê (xác suất chuyển trạng thái, phân phối điểm số) trích xuất từ dữ liệu khóa học 670. Các tác nhân ảo (Virtual Agents) được sinh ra với phân phối năng lực mô phỏng thực tế: 20% Yếu, 60% Trung bình, 20% Giỏi [4]. Quá trình huấn luyện diễn ra qua 500 vòng lặp (episodes), mỗi vòng gồm 100 tác nhân.

### *B. Đánh giá so sánh (A/B Testing)*

Chúng tôi so sánh hiệu quả giữa hai chính sách trên cùng một môi trường mô phỏng:

- **Nhóm Q-learning (Thực nghiệm):** Hành động dựa trên bảng Q-table đã huấn luyện.
- **Nhóm Param Policy (Đối chứng):** Hành động dựa trên xác suất ngẫu nhiên mô phỏng lại thói quen của sinh viên các khóa trước.

Kết quả định lượng tổng hợp tại Bảng III và chi tiết từng phân cụm tại Hình 4 cho thấy sự vượt trội rõ rệt của thuật toán Q-learning.

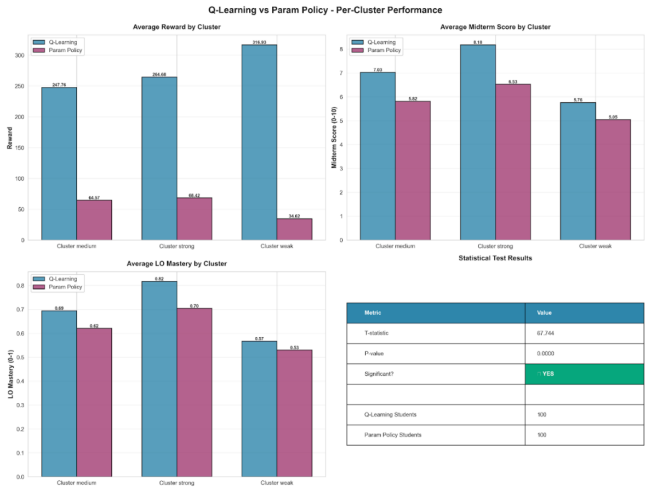


Figure 4. So sánh hiệu suất giữa Q-Learning và Param Policy trên các chỉ số: Phần thưởng (Reward), Điểm số (Score), và Độ thành thạo (LO Mastery).

Table III
KẾT QUẢ SO SÁNH HIỆU SUẤT TRUNG BÌNH GIỮA HAI NHÓM

| 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%** |
| Thành thạo LO (0-1) | 0.58 | 0.66 | **+13.9%** |
| Số kỹ năng yếu | 3.02 | 1.48 | **-51.0%** |

### *C. Phân tích tác động sư phạm*

Dựa trên biểu đồ trực quan tại Hình 4, hệ thống thể hiện khả năng thích ứng thông minh với từng nhóm đối tượng, giải quyết được vấn đề "một kích cỡ cho tất cả":

*1) Đối với nhóm sinh viên Yếu (Cluster Weak):* Đây là nhóm hưởng lợi nhiều nhất từ hệ thống. Chỉ số quan trọng nhất là **Số kỹ năng yếu (Avg Weak LO Count)** đã giảm mạnh 51.0% (từ 3.02 xuống 1.48). Quan sát biểu đồ *Average Reward*, ta thấy mức chênh lệch phần thưởng lớn nhất nằm ở nhóm này (316.93 so với 34.62), chứng minh AI đã áp dụng chiến lược "khích lệ" và gợi ý các hành động khắc phục (Remedial Actions) để giữ chân người học [4].

*2) Đối với nhóm sinh viên Giỏi (Cluster Strong):* Nhóm này đạt điểm số trung bình (Midterm Score) tuyệt đối cao nhất là **8.18/10** (so với 6.53 của nhóm đối chứng), như hiển thị trên biểu đồ. Hệ thống đã nhận diện năng lực vượt trội và chuyển sang chiến lược đề xuất các nội dung thách thức hơn, giúp tối ưu hóa tiềm năng của người học [4].

*3) Độ tin cậy thống kê:* Như thể hiện trong bảng *Statistical Test Results* (Hình 4), kiểm định T-test độc lập giữa hai nhóm cho kết quả $t(198) = 67.74$, $p < 0.001$, Cohen's $d = 6.78$. Giá trị Cohen's $d$ cực lớn (theo quy ước Cohen: $d > 0.8$ là "large effect", $d > 3.0$ là "extremely large effect") cho thấy sự khác biệt về hiệu quả không chỉ có ý nghĩa thống kê mà còn có ý nghĩa thực tiễn sâu sắc. Kết quả này khẳng định rằng thuật toán Q-learning vượt trội hơn đáng kể so với chính sách tham số hóa truyền thống với độ tin cậy 99.9% [4].

## VI. 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. Gligorea 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.