# 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 bối cảnh Giáo dục 4.0, các hệ thống quản lý học tập (LMS) truyền thống thường thiếu khả năng cá nhân hóa, áp dụng một lộ trình cố định cho mọi đối tượng người học. Để giải quyết hạn chế này trong giáo dục STEM, bài báo đề xuất khung giải pháp học tập thích ứng sử dụng thuật toán Q-learning, được tích hợp vào nền tảng Moodle thông qua chuẩn LTI 1.3. Nghiên cứu mô hình hóa quá trình học tập dưới dạng Quy trình Quyết định Markov (MDP), kết hợp với kỹ thuật phân cụm hành vi để xây dựng không gian trạng thái đa chiều. Đặc biệt, để khắc phục vấn đề "khởi động lạnh" (cold-start) và khan hiếm dữ liệu, chúng tôi đề xuất quy trình mô phỏng hướng dữ liệu (Data-driven Simulation), tái tạo chính xác các đặc trưng phân phối của người học thực tế. Kết quả thực nghiệm mô phỏng trên 500 vòng lặp cho thấy chiến lược của tác nhân AI vượt trội so với phương pháp truyền thống, giúp tăng 22.5% điểm số trung bình và giảm 51.0% số lượng kỹ năng yếu, chứng minh tiềm năng hiện thực hóa đào tạo cá nhân hóa quy mô lớn***

***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 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 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.

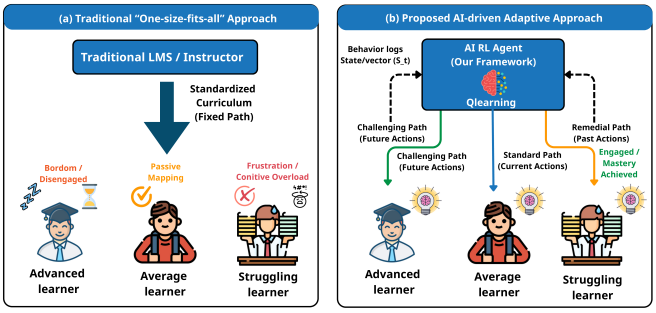


Figure 1. So sánh các phương pháp tiếp cận học tập. (a) Phương pháp truyền thống áp dụng một chương trình cố định cho mọi đối tượng người học, dẫn đến mức độ tương tác không tối ưu (gây nhàm chán hoặc quá tải). (b) Khung giải pháp đề xuất sử dụng Tác nhân Học tăng cường (AI Agent) để phân tích trạng thái người học từ dữ liệu log, từ đó đưa ra các gợi ý hành động cá nhân hóa (ôn tập, tiêu chuẩn, hoặc nâng cao), nhằm tối đa hóa sự tương tác và mức độ thành thạo kiến thức.

## 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).

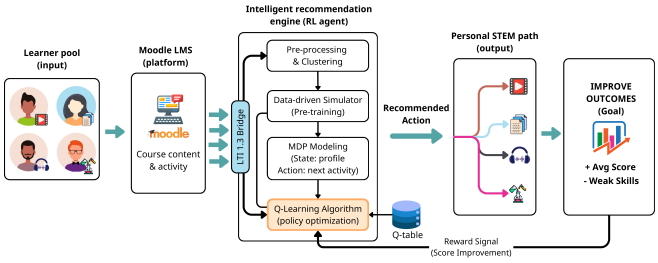


Figure 2. Tổng quan phương pháp đề xuất: Dữ liệu hành vi từ Moodle được chuyển đổi thành các trạng thái (State), qua đó Agent lựa chọn hành động (Action) tối ưu để nhận phần thưởng (Reward).

### 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 $< S, A, R >$ 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:

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

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 ($A = \{a_0, a_1, ..., a_{m-1}\}$). Các hành động này được phân loại theo trục 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 nhiễu 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 ba chỉ số đánh giá: phương pháp Elbow (Inertia), chỉ số Silhouette (tách biệt cụm), và chỉ số Davies-Bouldin (độ nén cụm). Kết quả từ ba chỉ số này được hợp nhất bằng cơ chế Bỏ phiếu đa số (Majority Voting) để xác định số cụm tối ưu. 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 $\alpha$ là tốc độ học và $\gamma$ là hệ số chiết khấu. Chiến lược $\epsilon$-greedy được áp dụng để cân bằng giữa khám phá (Exploration) và khai thác (Exploitation).

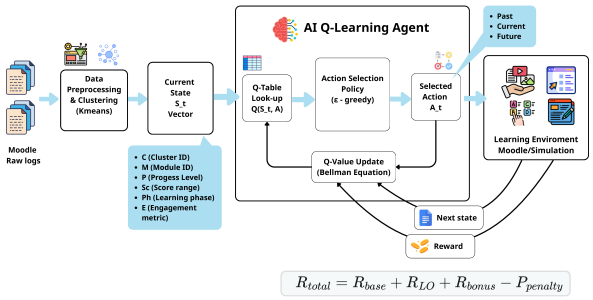


Figure 3. Sơ đồ chi tiết luồng hoạt động của thuật toán Q-learning trong việc đưa ra gợi ý sư phạm.

### D. Khung giải thích mô hình (Explainability Framework)

Để giải quyết tính "hộp đen" của bảng Q-table, nghiên cứu tích hợp phương pháp **SHAP (SHapley Additive exPlanations)** [6].

*1) Cơ sở toán học:* Giá trị SHAP $\phi_i(s)$ đo lường đóng góp biên của đặc trưng $i$ vào giá trị Q dự đoán:

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

Tính chất cộng tính đảm bảo: $f(s) = \phi_0 + \sum_{i=1}^{6} \phi_i(s)$. Để đánh giá độ quan trọng toàn cục, chúng tôi tính trung bình trị tuyệt đối SHAP trên một tập mẫu kiểm tra gồm $N$ trạng thái ngẫu nhiên:

$$I_i = \frac{1}{N} \sum_{j=1}^{N} |\phi_i(s_j)| \quad (5)$$

*2) Triển khai kỹ thuật với KernelExplainer:* Quy trình xấp xỉ giá trị Shapley được thực hiện qua 3 bước:

1) **Hàm dự đoán:** Xây dựng hàm wrapper $f(s) = \max_a Q(s,a)$, ánh xạ trạng thái $s$ sang giá trị lợi ích tối đa kỳ vọng.
2) **Lấy mẫu nền (Background Sampling):** Sử dụng tập dữ liệu nền $D_{bg}$ ($N_{bg} = 100$) để đại diện cho kỳ vọng cơ sở $E[f(x)]$.
3) **Tính toán (Computation):** Với mỗi trạng thái trong tập kiểm tra, KernelExplainer thực hiện tính toán trên toàn bộ $2^6 = 64$ liên minh đặc trưng, đảm bảo độ chính xác cao trong thời gian $O(N \cdot 2^D)$.

### *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 trước khi triển khai thực tế, nghiên cứu đề xuất quy trình **Khai phá tham số (Parameter Mining)**. Quy trình này chuyển đổi dữ liệu log thô thành các tham số xác suất để vận hành "Bản sao số" (Digital Twin) của lớp học thực tế.

*1) Chiến lược Ánh xạ và Ngữ cảnh hóa:* Hệ thống chuẩn hóa các sự kiện kỹ thuật của Moodle về không gian hành động $A$ gồm 15 hành động chuẩn. Để tăng độ mịn cho không gian trạng thái, mỗi hành động được gắn nhãn ngữ cảnh thời gian dựa trên tiến độ học tập $P_t$:

$$Context(a_t) = \begin{cases} \text{Past}, & \text{if } P_t < 25\% \\ \text{Current}, & \text{if } 25\% \leq P_t < 85\% \\ \text{Future}, & \text{if } P_t \geq 85\% \end{cases} \quad (6)$$

*2) Mô hình hóa Xác suất chuyển đổi (Transition Dynamics):* Dựa trên dữ liệu lịch sử $\mathcal{D}_{history}$, hệ thống xây dựng Ma trận xác suất chuyển đổi (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:

$$P(s'|s,a) = \frac{count(s,a,s')}{\sum_{s^*} count(s,a,s^*)} \quad (7)$$

Ma trận $\mathcal{P}$ đóng vai trò là quy luật vận hành của môi trường, đảm bảo phản ứng của hệ thống giả lập sát với thực tế.

*3) Mô hình hóa Chính sách Đối chứng (Param Policy Baseline):* Để đánh giá hiệu quả, chúng tôi xây dựng một **Param Policy** ($\pi_{param}$) làm baseline so sánh. Đây là chính sách tham số hóa (parametric policy) dựa trên phân phối xác suất hành động từ dữ liệu lịch sử, mô phỏng lại thói quen tự nhiên của sinh viên trong quá khứ:

$$\begin{aligned} \pi_{param}(a|s) &= P(a \mid \text{Phase}(s), \text{Cluster}(s)) \\ &= \frac{\text{count}\big(\text{Phase}(s), \text{Cluster}(s), a\big)}{\sum_{a'} \text{count}\big(\text{Phase}(s), \text{Cluster}(s), a'\big)} \end{aligned} \quad (8)$$

Với mỗi trạng thái $s$, hành động được gợi ý dựa trên tần suất xuất hiện trong tập huấn luyện, phân tầng theo giai đoạn học tập (Phase) và nhóm người học (Cluster). Đây là baseline để so sánh với chiến lược tối ưu hóa phần thưởng ($\pi^*$) của Q-learning.

*4) Quy trình Huấn luyện và Đánh giá:* Quy trình mô phỏng được thực hiện theo chu trình khép kín:

1) **Khởi tạo (Initialization):** Thiết lập các Tác nhân Người học (Student Agents) với đặc trưng hành vi được tham số hóa từ các cụm dữ liệu (Clusters).
2) **Huấn luyện (Training):** Agent tương tác với Student Agents trong môi trường $\mathcal{P}$ để tối ưu bảng Q-table thông qua cơ chế thử-sai.
3) **Đối sánh (Comparison):** So sánh hiệu suất (Reward, Score, LO Mastery) giữa chiến lược tối ưu $\pi^*$ của Q-learning Agent và chính sách đối chứng $\pi_{param}$ (Param Policy baseline).

## 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 lộ 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 diễn ra qua 500 episodes. Trong mỗi episode, hệ thống khởi tạo một quần thể gồm 100 tác nhân người học ảo (Student Agents). Như vậy, Agent được học từ tổng cộng $50,000$ lượt tương tác mô phỏng, đảm bảo độ bao phủ không gian trạng thái. 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 ($\sigma$):** 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:

$$S_{real} = \text{clip}(S_{base} + \mathcal{U}(-\sigma_c, \sigma_c), 0, 1) \quad (9)$$

  Trong đó $\sigma_c$ 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 |
| $Seed$ | Hạt giống ngẫu nhiên | 42 |
| $N_{modules}$ | Số module trong khóa học | 6 |
| *Tham số theo phân cụm:* | | |
| $P_{success}$ (Weak) | Xác suất thành công cơ sở | 0.72 |
| $P_{success}$ (Medium) | Xác suất thành công cơ sở | 0.78 |
| $P_{success}$ (Strong) | Xác suất thành công cơ sở | 0.90 |
| $\alpha_{learn}$ (Weak) | Tốc độ học tập | 0.22 |
| $\alpha_{learn}$ (Medium) | Tốc độ học tập | 0.32 |
| $\alpha_{learn}$ (Strong) | Tốc độ học tập | 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) Quy trình Tiền xử lý và Phân cụm (Data Preprocessing & Clustering):* Hệ thống sử dụng thuật toán K-Means nhờ ưu thế về tốc độ xử lý dữ liệu số. Để đảm bảo tính chính xác và tránh chọn tham số chủ quan, nghiên cứu áp dụng quy trình lọc hai lớp kết hợp với cơ chế "Bỏ phiếu đa số" (Majority Voting) để xác định cấu trúc phân nhóm.

**1) Chọn lọc đặc trưng (Feature Selection):** Dữ liệu đầu vào được chuẩn hóa (Z-score normalization), sau đó đi qua bộ lọc hai lớp để tinh gọn không gian chiều:

- **Lọc phương sai (Variance Filtering):** Loại bỏ các đặc trưng có phương sai dưới ngưỡng 0.01 (ít biến động, không mang giá trị phân loại).
- **Lọc tương quan (Correlation Filtering):** Với ngưỡng tương quan Pearson $> 0.95$, thuật toán đã loại bỏ **78 đặc trưng** bị trùng lặp thông tin, giải quyết triệt để hiện tượng đa cộng tuyến.

**2) Tối ưu hóa số cụm ($K$):** Số lượng cụm $K$ được quyết định dựa trên sự đồng thuận của 3 chỉ số đánh giá độ gom cụm (Hình 4):

- **Elbow Method:** Biểu đồ độ lỗi (Inertia) xuất hiện điểm uốn (knee point) rõ nét tại $K = 6$.
- **Silhouette Score:** Đạt giá trị cực đại tại $K = 2$ (0.42). Tuy nhiên, tại $K = 6$, chỉ số này vẫn duy trì mức ổn định chấp nhận được ($\approx 0.35$).
- **Davies-Bouldin Index:** Chỉ số đo độ nén (càng thấp càng tốt) cho thấy điểm trũng cục bộ khả quan tại $K = 6$.

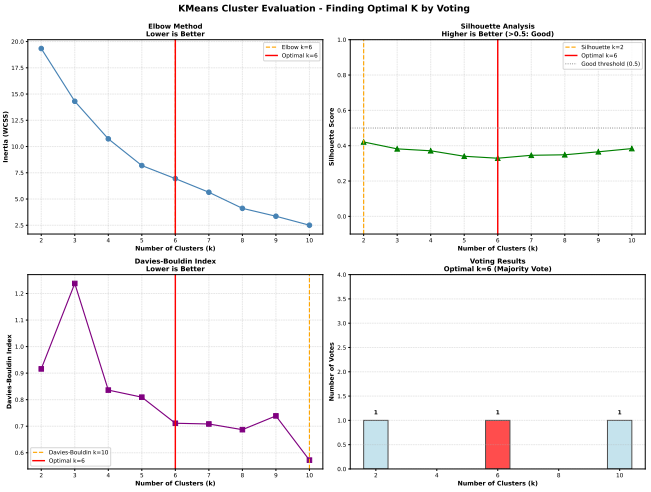


Figure 4. Các chỉ số đánh giá phân cụm. Elbow và Davies-Bouldin ủng hộ $K = 6$, trong khi Silhouette chấp nhận được ở ngưỡng này.

**3) Biện luận lựa chọn $K = 6$:** Mặc dù $K = 2$ cho chỉ số tách biệt tốt nhất, nhưng việc phân chia sinh viên chỉ thành 2 nhóm (Giỏi/Yếu) là quá thô sơ, không đủ độ mịn để Agent tối ưu hóa chiến lược sư phạm. Ngược lại, $K \geq 10$ quá phức tạp và gây bùng nổ không gian trạng thái (State Explosion). Do đó, $K = 6$ là điểm cân bằng tối ưu, đại diện cho 6 hồ sơ năng lực điển hình của người học.

Trong quá trình phân tích cụm, một cụm được phát hiện là đại diện cho hành vi của giảng viên (tần suất cao các sự kiện chấm điểm và hoạt động diễn đàn). Cụm này được loại bỏ khỏi tập dữ liệu huấn luyện của Agent, chỉ giữ lại 5 cụm hành vi sinh viên. Các cụm sinh viên này sau đó được ánh xạ về ba nhóm sư phạm (Weak, Medium, Strong) để cấu hình hàm phần thưởng thích ứng.

*3) Mô hình hóa Quy luật Chuyển đổi và Chính sách Đối chứng (Transition Dynamics & Baseline):* Tính hợp lý của môi trường mô phỏng (Bản sao số) được xác định bởi Ma trận xác suất chuyển trạng thái. Kết quả khai phá dữ liệu (Parameter Mining) từ tập huấn luyện được định lượng chi tiết thông qua các bảng phân phối xác suất, đóng vai trò là tham số đầu vào cho chính sách đối chứng Param Policy ($\pi_{param}$).

**1) Phân phối hành động theo Giai đoạn học tập (Learning Phase):** Dữ liệu tại Bảng II cho thấy sự chuyển dịch hành vi rõ rệt giữa các giai đoạn, từ thụ động sang kiến tạo và cuối cùng là phản tư.

**2) Phân phối theo Mức độ Tương tác (Engagement Level):** Bảng III thể hiện sự phân hóa trong chiến lược chọn hành động giữa các nhóm sinh viên. Nhóm tương tác cao (High) có xu hướng duy trì cân bằng giữa xem nội dung và

Table II
XÁC SUẤT CHỌN HÀNH ĐỘNG THEO GIAI ĐOẠN (LEARNING PHASE)

| **Giai đoạn** | **Hành động ($a_t$)** | **Xác suất ($P$)** |
|---|---|---|
| Phase 0 (Pre-learning) | view_assignment<br>view_content | 56.4%<br>43.6% |
| Phase 1 (Active) | **attempt_quiz**<br>submit_assignment | **74.4%**<br>25.6% |
| Phase 2 (Reflective) | post_forum | 100.0% |

làm bài tập, trong khi nhóm thấp (Low) tập trung chủ yếu vào việc xem yêu cầu bài tập (*view_assignment*).

Table III
PHÂN PHỐI XÁC SUẤT HÀNH ĐỘNG THEO MỨC ĐỘ TƯƠNG TÁC

| **Mức độ** | **View Assign** | **View Content** | **Quiz** | **Submit** | **Forum** |
|---|---|---|---|---|---|
| **High** | 52.2% | 40.5% | 5.8% | 1.5% | 0.02% |
| **Medium** | 51.7% | 41.9% | 4.3% | 2.2% | - |
| **Low** | **57.6%** | 34.0% | 4.8% | 3.6% | - |

Các tham số này được xuất ra dưới dạng cấu hình JSON để điều khiển hành vi ngẫu nhiên có hướng của các Tác nhân người học (Student Agents) trong quá trình mô phỏng.

*4) 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:
$$S_t = (C, M, P, Sc, Ph, E) \tag{10}$$
Trong đó: $C$ là chỉ số cụm hành vi sinh viên được suy ra từ mô hình K-means ban đầu với $K = 6$; sau khi loại bỏ một cụm ứng với hành vi giảng viên, còn lại 5 cụm sinh viên ($C \in \{0, 1, 2, 3, 4\}$) được mã hóa vào không gian trạng thái. $M$ 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 V.
- **Tham số Q-learning:** Tốc độ học $\alpha = 0.1$, hệ số chiết khấu $\gamma = 0.95$, và chiến lược $\epsilon$-greedy giảm dần.

### *C. Mô hình hóa Không gian trạng thái (State Space Definition)*

Bài toán được mô hình hóa dưới dạng Quy trình Quyết định Markov (MDP). Trạng thái $S_t$ tại thời điểm $t$ là một vector 6 chiều, kết hợp giữa đặc điểm phân cụm cố định và các biến hành vi động:

$$S_t = (C, M, P, Sc, Ph, E) \tag{11}$$

Để đảm bảo khả năng hội tụ của bảng Q-table trong giới hạn tính toán, các biến trạng thái liên tục được rời rạc hóa (discretized) theo các ngưỡng sư phạm cụ thể, được chi tiết hóa tại Bảng IV.

Table IV
ĐỊNH NGHĨA CHI TIẾT VÀ MIỀN GIÁ TRỊ CỦA KHÔNG GIAN TRẠNG THÁI

| **Ký hiệu** | **Thành phần** | **Định nghĩa / Rời rạc hóa (Bins)** | **Kích thước** |
|---|---|---|---|
| $C$ | Cluster | Nhóm hành vi sinh viên sau khi loại bỏ cụm giảng viên từ mô hình K-means ban đầu ($ID \in \{0, 1, 2, 3, 4\}$) | 5 |
| $M$ | Module | Chỉ số bài học hiện tại ($ID \in \{0, \ldots, 5\}$) | 6 |
| $P$ | Progress | Mức 0.25: Mới bắt đầu ($< 25\%$)<br>Mức 0.50: Đang học ($25\% - 50\%$)<br>Mức 0.75: Gần xong ($50\% - 99\%$)<br>Mức 1.00: Hoàn thành ($100\%$) | 4 |
| $Sc$ | Score | Mức 0.25 (Yếu): $< 2.5$<br>Mức 0.50 (TB): $2.5 \leq s < 5.0$<br>Mức 0.75 (Khá): $5.0 \leq s < 7.5$<br>Mức 1.00 (Giỏi): $\geq 7.5$ | 4 |
| $Ph$ | Phase | 0: Pre-learning (Tiếp thu thụ động)<br>1: Active-learning (Tương tác/Làm bài)<br>2: Reflective-learning (Ôn tập/Thảo luận) | 3 |
| $E$ | Engagement | 0: Thấp ($S_{total} < 8$)<br>1: Trung bình ($8 \leq S_{total} < 16$)<br>2: Cao ($S_{total} \geq 16$) | 3 |

*1) Cơ chế tính toán mức độ tương tác (Engagement Calculation):* Khác với các nghiên cứu trước chỉ đếm số lượng hành động, chúng tôi đề xuất công thức tính điểm tương tác tổng hợp $S_{total}$ dựa trên khung lý thuyết ICAP, kết hợp ba yếu tố: trọng số hành động, thời gian và tính nhất quán.

$$S_{total} = S_{weighted} + S_{time} + S_{consistency} \tag{12}$$

Trong đó:

- $S_{weighted}$ **(Chất lượng hành động):** Tổng trọng số ICAP của các hành động trong cửa sổ quan sát. Các hành động kiến tạo (như *submit_assignment*) có trọng số cao ($w = 5$), trong khi hành động thụ động (như *view*) có trọng số thấp ($w = 1$):
$$S_{weighted} = \sum_{i=1}^{n} w(action_i)$$
- $S_{time}$ **(Hiệu quả thời gian):** So sánh thời gian thực tế $T_{real}$ với thời gian kỳ vọng $T_{exp}$:
$$S_{time} = \begin{cases} 2 & \text{if } T_{real} \geq 50\% T_{exp} \\ 1 & \text{if } 30\% \leq T_{real} < 50\% T_{exp} \\ 0 & \text{else} \end{cases}$$
- $S_{consistency}$ **(Tính đều đặn):** Dựa trên khoảng cách trung bình $\Delta t$ giữa các lần tương tác:
$$S_{consistency} = \begin{cases} 2 & \text{if } 1\text{min} \leq \Delta t \leq 60\text{min} \\ 1 & \text{if } 30\text{s} \leq \Delta t \leq 2\text{h} \\ 0 & \text{else} \end{cases}$$

Trên cơ sở các định nghĩa trên, không gian trạng thái được xác định hoàn toàn bởi công thức kích thước đã nêu ở phần "Tham số mô hình" (Section 3.2.2), đảm bảo tính khả thi để thuật toán Q-learning hội tụ.

Table V
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 |

*2) Cấu hình Hàm phần thưởng (Reward Configuration):* Để giải quyết bài toán phân hóa đối tượng, hàm phần thưởng được thiết kế thích ứng (Adaptive Reward Shaping). Thay vì sử dụng các giá trị cố định, hệ thống điều chỉnh trọng số phần thưởng dựa trên 3 cụm người học (Weak, Medium, Strong), cụ thể hóa các giá trị $R_{comp}$ và $R_{LO}$ như sau:

$$R(s,a) = R_{comp}(C) + R_{score}(\Delta S) + R_{LO}(\Delta\Omega, C) - P_{fail}(C) \quad (13)$$

Trong đó, chiến lược phân phối phần thưởng được thiết kế dựa trên triết lý sư phạm riêng biệt cho từng nhóm (Bảng VI):

- **Nhóm Yếu (Weak):** Ưu tiên động lực ngắn hạn (Small wins) với phần thưởng hoàn thành cao nhất ($R_{comp} = 10.0$) và điểm thưởng lớn cho việc cải thiện kỹ năng còn yếu ($R_{LO} = 15.0$)[cite: 740].
- **Nhóm Trung bình (Medium):** Tập trung vào sự phát triển cân bằng (Balanced Growth). Phần thưởng được thiết lập ở mức trung gian ($R_{comp} = 7.0$) để khuyến khích sinh viên duy trì tiến độ và cải thiện điểm số ổn định[cite: 748].
- **Nhóm Giỏi (Strong):** Hướng tới sự thành thạo (Mastery-oriented). Phần thưởng cơ bản thấp ($R_{comp} = 5.0$) buộc sinh viên phải tìm kiếm các hành động mang lại hiệu quả cao về thời gian (Time Efficiency) hoặc giải quyết các bài tập khó để tối đa hóa điểm số[cite: 756].

### D. 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).

Table VI
THAM SỐ CHI TIẾT HÀM PHẦN THƯỞNG THEO PHÂN CỤM

| Thành phần | Ký hiệu | Giá trị thiết lập (Config Value) | | |
|---|---|---|---|---|
| | | **Weak** | **Medium** | **Strong** |
| Hoàn thành module | $R_{comp}$ | **10.0** | 7.0 | 5.0 |
| Cải thiện LO Mastery | $R_{LO}$ | **15.0** | 10.0 | 7.0 |
| Tăng điểm số | $R_{score}$ | $\Delta Score \times 10.0$ | | |
| Hiệu quả thời gian | $R_{time}$ | - | - | +1.5 |
| Phạt thất bại/Kẹt | $P_{fail}$ | -3.0 | -2.0 | -1.0 |

- **Số kỹ năng yếu (Weak Skills Count):** Số lượng Chuẩn đầu ra (LO) có độ thông thạo $< 0.5$.

### E. 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 (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.

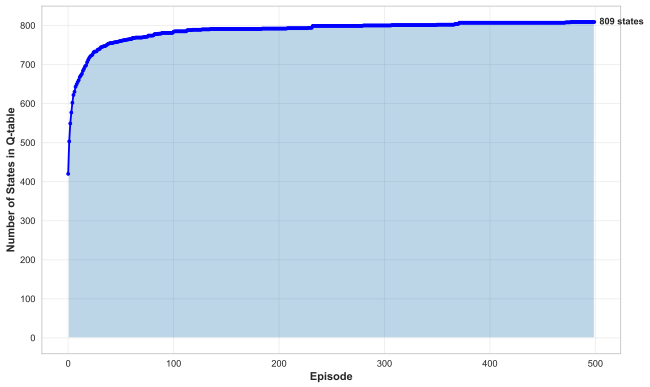


Figure 5. Biểu đồ hội tụ phần thưởng tích lũy (Cumulative Reward) qua 500 episodes. Sự ổn định bắt đầu xuất hiện rõ rệt sau episode 350.

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 $\epsilon$ 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.

### F. 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

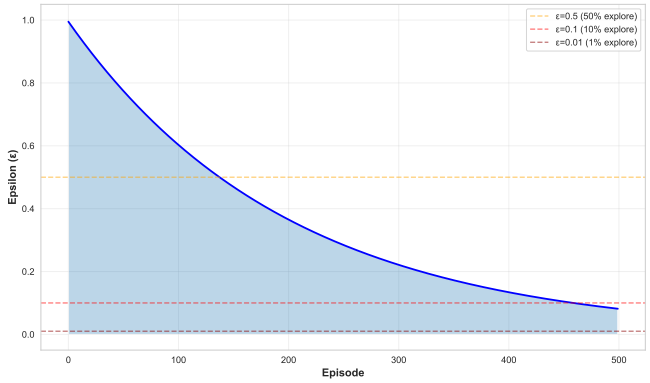


Figure 6. Chiến lược giảm dần tham số Epsilon ($\epsilon$) qua 3 giai đoạn: Khám phá, Chuyển đổi và Khai thác.

VII và chi tiết từng phân cụm tại Hình 7 cho thấy sự vượt trội rõ rệt của thuật toán Q-learning.

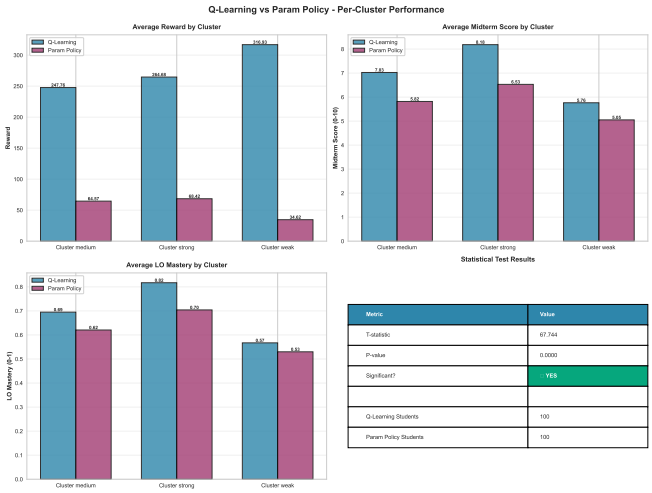


Figure 7. So sánh hiệu suất giữa Q-Learning và Param Policy trên các chỉ số: Phần thưởng (Reward), Điểm số (Score), và Độ thành thạo (LO Mastery).

Table VII
KẾT QUẢ SO SÁNH HIỆU SUẤT TRUNG BÌNH

| Chỉ số | Param Policy | Q-learning | Cải thiện |
|---|---|---|---|
| Phần thưởng TB | $59.95 \pm 12.38$ | $264.26 \pm 27.33$ | **+340.8%** |
| Điểm TB (thang 10) | $5.82 \pm 0.48$ | $7.14 \pm 0.82$ | **+22.5%** |
| Độ thành thạo LO | $0.621 \pm 0.056$ | $0.707 \pm 0.085$ | **+13.9%** |
| Số kỹ năng yếu | 3.02 | 1.48 | **-51.0%** |

### *G. Phân tích độ quan trọng đặc trưng*

Từ các giá trị SHAP thu được trên 802 trạng thái cốt lõi, chúng tôi tính toán hai chỉ số chính cho từng đặc trưng:

- **Mean Absolute SHAP**: $I_i = \frac{1}{N}\sum_{j=1}^{N}|\phi_i(s_j)|$ - đo mức độ ảnh hưởng trung bình của đặc trưng $i$ lên Q-value.
- **SHAP Variance**: $V_i = \text{Var}(\phi_i)$ - đo tính ổn định hay phụ thuộc ngữ cảnh của ảnh hưởng (variance cao = phụ thuộc ngữ cảnh nhiều).

Kết quả định lượng được tổng hợp tại Bảng VIII và trực quan hóa tại Hình 8. Phân tích cho thấy *Module ID* (mean $|\text{SHAP}| = 28.32$) 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 8) 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 khoảng $-50$ đến $+50$, chứng tỏ tác động của yếu tố này mang tính ngữ cảnh cao (context-dependent): mức tương tác cao có thể làm tăng Q-value mạnh (điểm màu đỏ, phía dương) khi sinh viên đang ở module và cụm phù hợp, nhưng cũng có thể làm giảm Q-value (điểm màu xanh, phía âm) nếu tương tác không đúng hướng (ví dụ xem nội dung thụ động quá nhiều trước kỳ đánh giá).

Ngược lại, *Progress Level* có mean $|\text{SHAP}|$ 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 tối ưu. 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.

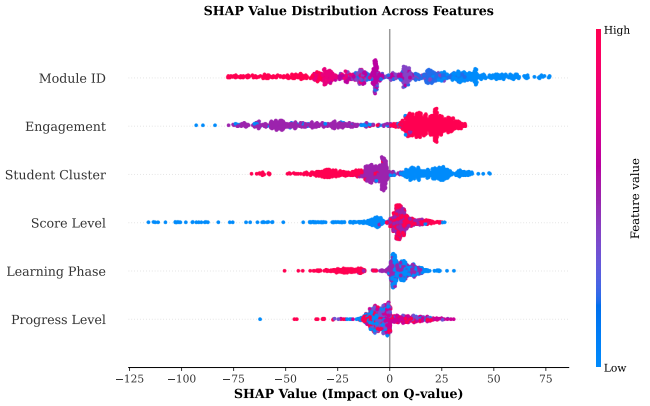


Figure 8. Phân phối SHAP values trên 802 trạng thái từ Q-table. Mỗi điểm biểu diễn SHAP value của một trạng thái; màu sắc thể hiện giá trị đặc trưng (đỏ = cao, xanh = thấp). Trục hoành cho thấy mức độ ảnh hưởng lên Q-value (dương = tăng, âm = giảm). Engagement và Module ID có khoảng tác động (impact range) rộng nhất, trong khi Progress Level tập trung quanh 0.

### *H. 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 gợi 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].

Table VIII
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 MỨC ĐỘ BIẾN ĐỘNG (CAO = PHỤ THUỘC NGỮ CẢNH NHIỀU, 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 |

- **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.

## LỜI CAM KẾT ĐẠO ĐỨC

Nghiên cứu sử dụng dữ liệu khóa học công khai từ Kaggle (đã anonymized). Nếu triển khai Live trên lớp học thực, sẽ xin phép từ Hội đồng Đạo đức [IRB].

## 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.