Quyền riêng tư của người học khi ứng dụng trí tuệ nhân tạo vào cá nhân hóa học tập

Trần Bình Hậu[1,*], Nguyễn Kim Quốc[2]

[1]Phòng Quản trị Thông tin, Trường Đại học Nguyễn Tất Thành, Thành phố Hồ Chí Minh
[2]Khoa Công nghệ Thông tin, Trường Đại học Nguyễn Tất Thành, Thành phố Hồ Chí Minh
[*]tbhau@ntt.edu.vn

Tóm tắt

Các hệ thống học tập cá nhân hóa dựa vào trí tuệ nhân tạo để mô hình hóa người học, dự đoán kết quả học tập và hỗ trợ quyết định đào tạo cho từng đối tượng. Nghiên cứu này đánh giá có hệ thống các phương pháp bảo vệ quyền riêng tư, gồm bảo mật vi sai, học tập liên kết, tính toán bảo mật đa phương, mã hóa đồng hình và môi trường thực thi đáng tin cậy. Đồng thời, nghiên cứu xem xét việc áp dụng các kỹ thuật này trong phân tích học tập, đề xuất nội dung, hướng dẫn thông minh và đánh giá tự động. Bên cạnh đó, nghiên cứu xác định các mối đe dọa về quyền riêng tư trong giáo dục như lập hồ sơ người học, tấn công suy luận hành vi học tập và rò rỉ dữ liệu cá nhân trong các kịch bản học tập cộng tác. Kết quả là một hệ thống phân loại liên kết mục tiêu cá nhân hóa, các rủi ro về quyền riêng tư và kỹ thuật trí tuệ nhân tạo bảo vệ dữ liệu. Trên cơ sở đó, nghiên cứu đề xuất chương trình nghiên cứu nhằm làm rõ những thách thức còn tồn tại và định hướng phát triển các hệ thống học tập cá nhân hóa dựa trên trí tuệ nhân tạo đáng tin cậy và có ý thức bảo vệ quyền riêng tư.

Nhận 30/01/2026
Được duyệt 08/02/2026
Công bố 28/02/2026

Từ khóa

Học tập cá nhân hóa, phân tích học tập, bảo mật dữ liệu giáo dục, trí tuệ nhân tạo có ý thức về quyền riêng tư.

1 Đặt vấn đề

Ngày nay, trí tuệ nhân tạo (AI) đã trở thành một thành phần cốt lõi của các hệ thống học tập, đặc biệt là ở các hệ thống cá nhân hóa, định hình lại một cách cơ bản các hoạt động giáo dục trên khắp các trường học, và nền tảng học tập trực tuyến [1].

Sự gia tăng đáng kể lượng dữ liệu đào tạo làm gia tăng mối lo ngại liên quan đến việc lạm dụng dữ liệu, lập hồ sơ trái phép và suy luận ngoài ý muốn về các thuộc tính nhạy cảm của người học. Đáng chú ý, ngay cả khi các định danh rõ ràng bị loại bỏ, các mô hình AI vẫn có thể suy luận thông tin riêng tư từ các hành vi học tập tổng hợp hoặc từ chính đầu ra của mô hình, có khả năng khiến người học bị vi phạm quyền riêng tư [2].

Để giải quyết những lo ngại này, một lượng lớn nghiên cứu đã tập trung vào các kỹ thuật AI và học máy bảo vệ quyền riêng tư. Các nghiên cứu hiện có khám phá nhiều phương pháp nhằm giảm thiểu việc lộ dữ liệu, bảo vệ các thuộc tính nhạy cảm và giảm thiểu rủi ro về quyền riêng tư trong các hệ thống AI [4-6, 15]. Tuy nhiên, theo nhận định của nhóm nghiên cứu, phần lớn các tài liệu này được phát triển theo cách không phụ thuộc vào lĩnh vực cụ thể hoặc tập trung vào các lĩnh vực ứng dụng như chăm sóc sức khỏe, tài chính và internet vạn vật. Những đặc điểm riêng biệt của dữ liệu giáo dục, bao gồm tính chất theo thời gian, các ràng buộc về mặt sư phạm và các cân nhắc đạo đức lấy người học làm trung tâm vẫn chưa được giải quyết đầy đủ trong các công trình nghiên cứu này.

https://doi.org/10.55401/jmcefe42

Nghiên cứu này dựa trên quan sát thực tiễn, đưa ra một số nhận định và gợi ý để phát triển sự hiểu biết mạch lạc về cách AI cá nhân hóa, và quyền riêng tư giao thoa trong các hệ thống học tập hiện đại, đồng thời định hướng cả nghiên cứu trong tương lai và thực tiễn giáo dục, có trách nhiệm trong lĩnh vực đang phát triển nhanh chóng này.

2 Tổng quan

Khảo sát này được thực hiện dựa trên phương pháp quan sát thực tiễn, tổng hợp các nghiên cứu về cá nhân hóa học tập và ứng dụng AI. Dựa trên tiêu chí ứng dụng AI vào học tập, và rủi ro có thể xảy ra, từ đó đưa ra các nhận định cho quyền riêng tư của người học khi ứng dụng AI vào cá nhân hóa học tập. Phạm vi thực hiện tại hệ thống LMS và Chatbot của Trường Đại học Nguyễn Tất Thành (NTTU). Khảo sát này tập trung vào ba mô hình dựa trên AI được áp dụng rộng rãi nhất trong các hệ thống học tập cá nhân hóa và có liên quan xét ở góc độ bảo mật: phân tích học tập, hệ thống đề xuất và hệ thống hướng dẫn học tập thông minh (Intelligent Tutoring Systems - ITS).

2.1 Phân tích học tập (Learning Analytics)

Phân tích học tập tập trung vào việc đo lường, thu thập, phân tích và báo cáo dữ liệu về người học, bối cảnh học tập của họ với mục tiêu hiểu và tối ưu hóa các quá trình học tập. Hệ thống phân tích học tập dựa trên AI tận dụng mô hình thống kê, học máy và kỹ thuật khai thác dữ liệu để xác định các mẫu trong dữ liệu giáo dục, chẳng hạn như xu hướng hiệu suất, mức độ tham gia và quỹ đạo hành vi. Những hiểu biết này thường được sử dụng để hỗ trợ các nhiệm vụ bao gồm: dự đoán hiệu suất, cảnh báo sớm cho người học có nguy cơ và đánh giá hiệu quả giảng dạy [3].

Trong môi trường học tập cá nhân hóa, phân tích học tập đóng vai trò là xương sống phân tích để thích ứng bằng cách liên tục cập nhật các mô hình người học dựa trên dữ liệu quan sát được. Tuy nhiên, từ góc độ quyền riêng tư, việc dựa vào dữ liệu học tập chi tiết và theo chiều dọc mang lại những rủi ro đáng kể. Theo quan điểm của bài nghiên cứu, các mô hình phân tích có thể vô tình tiết lộ thông tin nhạy cảm về khả năng, thói quen hoặc điểm yếu của người học theo thời gian, ngay cả khi các thuộc tính đó không được thu thập một cách rõ ràng [2]. Vai trò kép này vừa là yếu tố cho phép cá nhân hóa, vừa là rủi ro tiềm tàng gây lộ thông tin riêng tư.

2.2 Hệ thống đề xuất/khuyến nghị (Recommendation Systems)

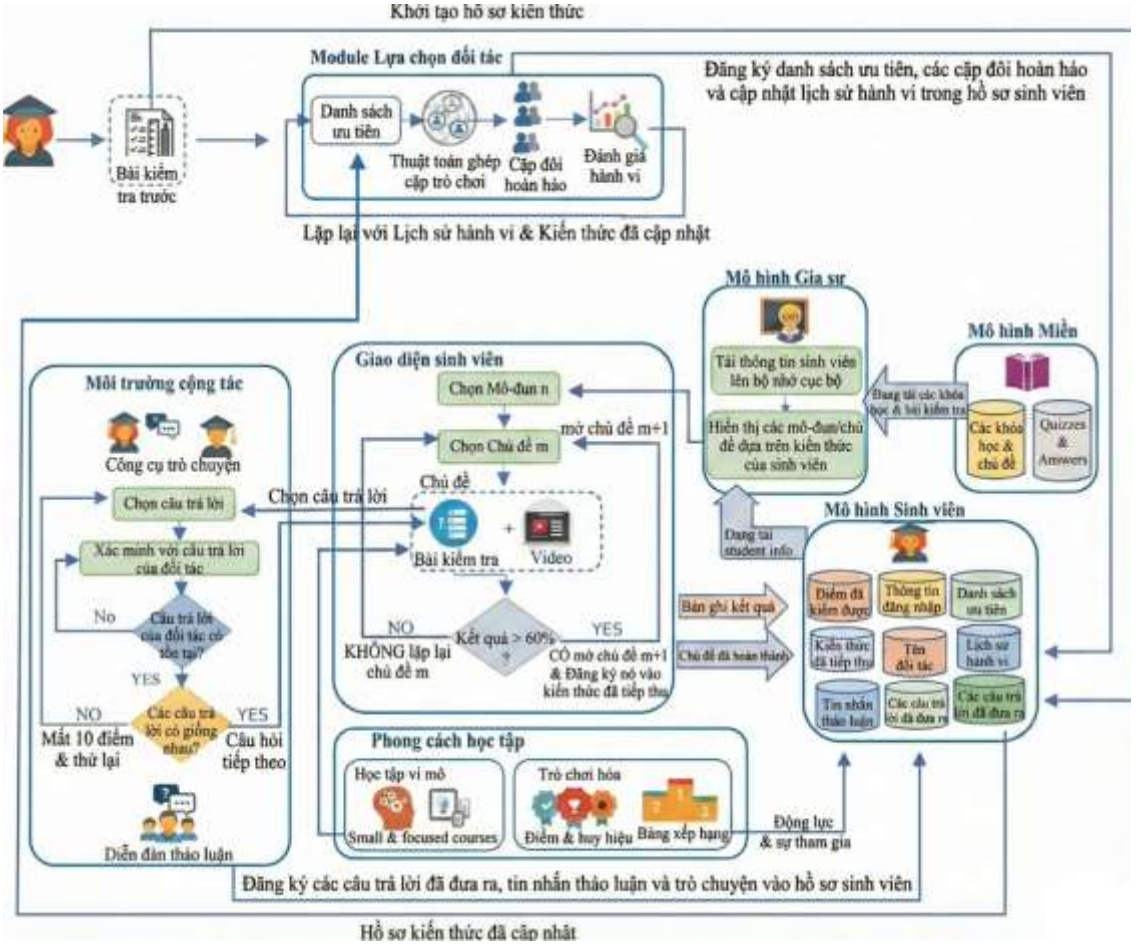
Hệ thống đề xuất được sử dụng rộng rãi trong học tập cá nhân hóa để đề xuất các nguồn tài nguyên học tập, các hoạt động hoặc lộ trình học tập phù hợp với sở thích và nhu cầu của từng người học. Lấy cảm hứng từ các hệ thống đề xuất trong thương mại điện tử và nền tảng truyền thông, các hệ thống đề xuất giáo dục thường sử dụng phương pháp lọc cộng tác, lọc dựa trên nội dung và các phương pháp kết hợp để cá nhân hóa trải nghiệm học tập [4].

Trong bối cảnh giáo dục, hệ thống đề xuất có thể dựa trên nhiều nguồn dữ liệu khác nhau, bao gồm hồ sơ người học, hồ sơ thành tích trước đây, nhật ký tương tác và hành vi của bạn bè. Trên thực tế, sự đa dạng dữ liệu này hỗ trợ các đề xuất chính xác hơn và phù hợp với ngữ cảnh hơn. Đồng thời, nó cũng tiềm ẩn rủi ro về quyền riêng tư, đặc biệt khi các đề xuất được đưa ra dựa trên các mẫu ở cấp độ nhóm hoặc dữ liệu người học được chia sẻ.

2.3 Hệ thống hướng dẫn học tập thông minh (ITS)

ITS là hệ thống giáo dục do AI điều khiển được thiết kế để cung cấp hướng dẫn và phản hồi cá nhân hóa bằng cách mô phỏng các khía cạnh của việc dạy kèm trực tiếp giữa người với người. ITS thường tích hợp mô hình người học, biểu diễn kiến thức miền và các chiến lược sư phạm để điều chỉnh hướng dẫn trong thời gian thực. Bằng cách theo dõi hành động và phản hồi của người học, các hệ thống này có thể chẩn đoán những quan niệm sai lầm, điều chỉnh độ khó của nhiệm vụ và cung cấp phản hồi phù hợp [4].

Việc thu thập dữ liệu chuyên sâu này hỗ trợ cá nhân hóa ở mức độ cao nhưng cũng làm gia tăng các thách thức về quyền riêng tư. Cụ thể, các suy luận nhạy cảm về trạng thái kiến thức, chiến lược học tập hoặc đặc điểm nhận thức của người học có thể được mã hóa ngầm trong các mô hình và quy trình ra quyết định của hệ thống. Do đó, ITS đại diện cho một trong những hình thức cá nhân hóa dựa trên AI nhạy cảm nhất về quyền riêng tư trong giáo dục, bất chấp những lợi ích sư phạm mà chúng mang lại. Hình 1 mô tả kiến trúc tổng thể, bao gồm các thành phần có thể tham gia vào hệ thống để tăng cường tri thức, cá nhân hóa học tập.



Hình 1 Kiến trúc tổng thể hệ thống học tập cá nhân hóa có bảo vệ quyền riêng tư

3 Kết quả khảo sát

Phần này tập trung vào hai khía cạnh đặc biệt liên quan đến việc học tập cá nhân hóa có tính đến quyền riêng tư: đặc điểm nội tại của dữ liệu người học và bối cảnh pháp lý chi phối việc sử dụng. Thay vì cung cấp một phân tích pháp lý toàn diện, phần thảo luận này nhấn mạnh cách các yếu tố này định hình rủi ro về quyền riêng tư trong môi trường giáo dục được hỗ trợ bởi AI.

3.1 Các đặc tính dữ liệu người học

Dữ liệu giáo dục có đặc điểm là độ chi tiết cao, phụ thuộc mạnh mẽ vào ngữ cảnh và cấu trúc theo thời gian rõ rệt. Các nền tảng học tập hỗ trợ bởi AI thường xuyên thu thập dấu vết học tập chi tiết, bao gồm nhật ký tương tác, lịch sử đánh giá, hồ sơ thời gian làm bài và các mô hình hành vi. Theo thời gian, những dữ liệu này được tổng hợp để xây dựng hồ sơ người học đang phát triển, phản ánh không chỉ thành tích học tập mà còn cả chiến lược học tập, mức độ tham gia và xu hướng hành vi [3].

Dữ liệu học tập được tích lũy qua các khóa học, học kỳ, và trong một số trường hợp, qua các cơ sở giáo dục, cho phép mô hình hóa dài hạn lộ trình học tập của người học. Khía cạnh theo chiều dọc này đại diện cho một trong những thách thức về quyền riêng tư quan trọng nhất trong học tập cá nhân hóa. Hơn nữa, dữ liệu giáo dục thường bao gồm các khía cạnh xã hội và hợp tác, trong đó thông tin của người học cá nhân có thể bị lộ gián tiếp thông qua phân tích cấp nhóm hoặc tương tác giữa các đồng nghiệp [5].

Tại NTTU, hệ thống LMS phục vụ học tập trực tuyến (các môn học thông thường, ngoại ngữ) dùng nền tảng moodle, Smartclass và tự phát triển tính năng quản lý. Đã phát triển AI để phân tích các dữ liệu này, phân tích và tổng hợp thông tin khóa học. Phần mềm Chatbot khi triển khai, có hiện tượng tri thức từ lĩnh vực này gây ảnh hưởng tới lĩnh vực khác, cần phải phân tách những đặc trưng dữ liệu và phạm vi ứng dụng để không xảy ra hiện tượng thông tin lĩnh vực này thấy lĩnh vực khác.

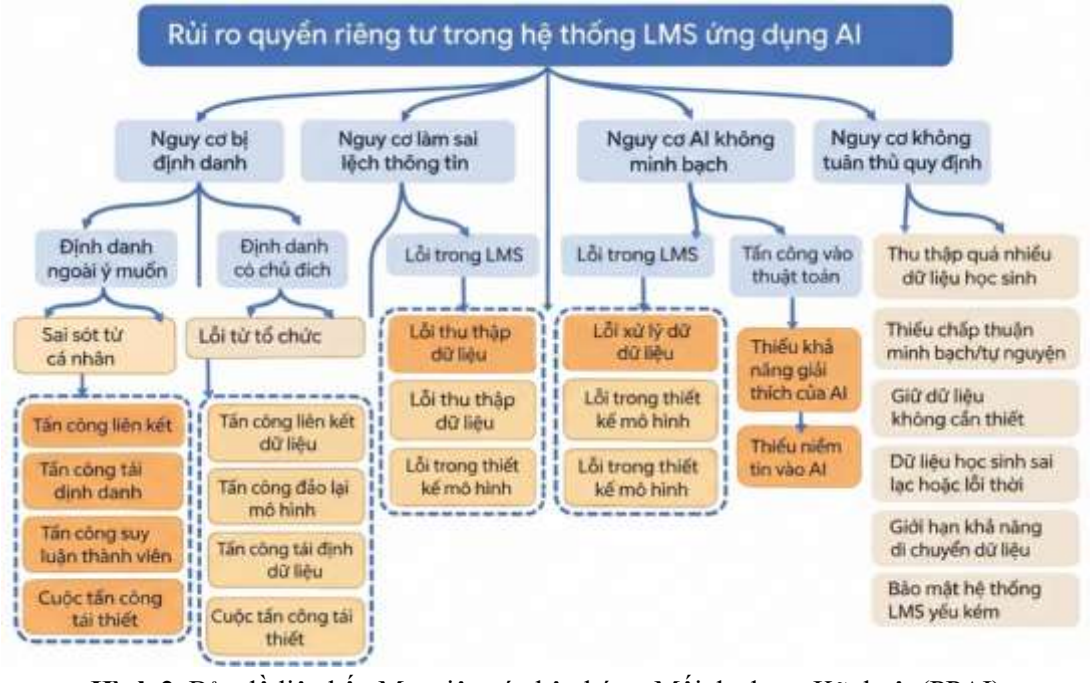
Bảng 1 Bảng so sánh các nhóm dữ liệu người học

Tiêu chí	Dữ liệu nhân thân & hành chính	Dữ liệu kết quả học tập	Dữ liệu tương tác và hành vi	Dữ liệu đa phương thức và cảm xúc
Bản chất	Tĩnh, mang tính định danh và quản lý.	Kết quả đầu ra của quá trình đào tạo.	Động, phản ánh quá trình và thói quen học tập.	Phức tạp, phản ánh trạng thái tâm sinh lý ẩn.
Ví dụ cụ thể	Tên, địa chỉ, giới tính, dân tộc, thông tin tài chính (vay vốn, hỗ trợ).	Điểm thi, kết quả bài tập, bài luận, chứng chỉ, hồ sơ kỷ luật.	Lượt nhấp chuột (clicks), mẫu điều hướng, thời gian làm bài, lịch sử thảo luận.	Theo dõi ánh mắt (eye tracking), cảm biến sinh lý, nhận diện khuôn mặt, trạng thái cảm xúc.
Ứng dụng chính	Dự báo khả năng trúng tuyển, phân loại đối tượng để hỗ trợ tài chính.	Đánh giá năng lực, cấp chứng chỉ, dự báo thành công học thuật.	Cá nhân hóa lộ trình, cảnh báo sớm nguy cơ bỏ học (early warning).	Tối ưu hóa giao diện ITS, hiểu sâu về mức độ cam kết và căng thẳng của người học.
Rủi ro đạo đức/Bảo mật	Lộ diện thông tin riêng tư nhạy cảm, định kiến dựa trên sắc tộc hoặc vùng miền.	Đánh giá sai lệch nếu thuật toán chấm điểm tự động có định kiến.	Xâm phạm quyền riêng tư khi theo dõi liên tục, tạo cảm giác bị giám sát.	Rủi ro cao về xâm hại quyền tự trị và giám sát vượt mức (surveillance).

3.2 Bối cảnh pháp lý

Việc thu thập và sử dụng dữ liệu giáo dục phải tuân theo các khuôn khổ pháp lý và quy định nhằm bảo vệ quyền riêng tư của người học và quyền dữ liệu. Trong bối cảnh các nước châu Âu, Quy định chung về Bảo vệ dữ liệu (GDPR) thiết lập các yêu cầu liên quan đến việc giảm thiểu dữ liệu, giới hạn mục đích và tính minh bạch, đặc biệt chú trọng đến các hệ thống ra quyết định tự động. Tại Hoa Kỳ, Đạo luật về Quyền riêng tư và Giáo dục gia đình (FERPA) điều chỉnh việc truy cập và tiết lộ hồ sơ giáo dục của học sinh, nhấn mạnh sự đồng ý và trách nhiệm của tổ chức [6]. Hiện nay, ở Việt Nam chưa có khung pháp lý cho vấn đề này.

Mặc dù, các quy định này cung cấp các biện pháp bảo vệ thiết yếu, phần lớn không phụ thuộc vào công nghệ và không giải quyết rõ ràng những phức tạp do các hệ thống học tập cá nhân hóa dựa trên AI gây ra. Đáng chú ý, các khung pháp lý thường tập trung vào việc truy cập và lưu trữ dữ liệu, trong khi nhiều rủi ro về quyền riêng tư trong học tập cá nhân hóa lại xuất phát từ hành vi của mô hình và các quy trình suy luận.



Hình 2 Bản đồ liên kết: Mục tiêu cá nhân hóa – Mối đe dọa – Kỹ thuật (PPAI)

Bảng 2 Phân loại dữ liệu giáo dục trong học tập cá nhân hóa và rủi ro quyền riêng tư

Loại dữ liệu	Ví dụ	Mục đích cá nhân hóa	Rủi ro quyền riêng tư
Dữ liệu hành vi	Clickstream, thời gian làm bài	Dự đoán mức độ tham gia	Suy luận thói quen, phong cách học
Dữ liệu đánh giá	Điểm số, phản hồi chi tiết	Đánh giá thích ứng	Hồ sơ năng lực dài hạn
Dữ liệu tương tác xã hội	Thảo luận, làm việc nhóm	Cá nhân hóa theo nhóm	Rò rỉ gián tiếp thông tin cá nhân
Dữ liệu theo thời gian	Lịch sử học tập nhiều học kỳ	Mô hình quỹ đạo học tập	Tái nhận dạng người học

3.3 Phân tích hồ sơ người học và suy luận hành vi

Suy luận hành vi có thể dẫn đến hồ sơ người học tồn tại lâu dài, ảnh hưởng đến các quyết định giáo dục trên các khóa học, nền tảng hoặc tổ chức. Từ góc độ quyền riêng tư, mối quan ngại không chỉ nằm ở việc thu thập dữ liệu người học mà còn ở khả năng giải thích, tính bền vững và việc sử dụng tiếp theo các thuộc tính người học được suy luận [7].

3.4 Tấn công suy luận thành viên và các thuộc tính

Các cuộc tấn công dựa trên suy luận thuộc tính và tư cách thành viên đại diện cho một loại mối đe dọa quyền riêng tư, trong đó kẻ tấn công cố gắng suy luận xem dữ liệu của người học có được sử dụng trong quá trình huấn luyện mô hình hay không, hoặc suy ra các thuộc tính nhạy cảm từ kết quả đầu ra của mô hình. Trong các hệ thống học tập cá nhân hóa, những mối đe dọa này có thể xuất hiện khi các mô hình dự đoán, công cụ đề xuất hoặc bảng điều khiển phân tích học tập tiết lộ các dự đoán được cá nhân hóa hoặc tổng hợp [7].

Các lỗ hổng về quyền riêng tư có thể phát sinh thông qua tương tác hệ thống thông thường chứ không phải thông qua các vi phạm dữ liệu rõ ràng, thách thức các giả định truyền thống về kiểm soát truy cập dữ liệu [8].

3.5 Rủi ro về quyền riêng tư trong học tập cộng tác và học tập xã hội

Môi trường học tập hợp tác và tương tác xã hội đặt ra những thách thức bổ sung về quyền riêng tư bằng cách kết hợp dữ liệu của từng người học với phân tích cấp độ nhóm. Mặc dù, các cơ chế trí tuệ tập thể như vậy có thể làm phong phú thêm trải nghiệm học tập, nhưng cũng có thể vô tình làm lộ thông tin cá nhân của người học thông qua số liệu thống kê tổng hợp hoặc suy luận về hành vi nhóm.

Việc quản lý quyền riêng tư trong học tập cộng tác đòi hỏi phải chú ý đến cả việc sử dụng dữ liệu ở cấp độ cá nhân và tập thể, một sự phân biệt thường bị xem nhẹ trong các cuộc thảo luận về quyền riêng tư hiện có [7].

Hình 2 đã tổng hợp các nguy cơ trong hệ thống LMS có ứng dụng AI để thực hiện cá nhân hóa học tập. Lưu ý những sai sót từ cá nhân có thể được lập trình ngăn chặn bỏ qua thu thập dữ liệu này. Nếu định danh có chủ đích, cần xem xét mức độ vi phạm quyền riêng tư. Nguy cơ từ thuật toán, có thể được khắc phục từ việc liên tục cập nhật thuật toán mới.

3.6 Dữ liệu học tập theo thời gian dài và tái nhận dạng

3.3 Phân tích hồ sơ người học và suy luận hành vi

Suy luận hành vi có thể dẫn đến hồ sơ người học tồn tại lâu dài, ảnh hưởng đến các quyết định giáo dục trên các khóa học, nền tảng hoặc tổ chức. Từ góc độ quyền riêng tư, mối quan ngại không chỉ nằm ở việc thu thập dữ liệu người học mà còn ở khả năng giải thích, tính bền vững và việc sử dụng tiếp theo các thuộc tính người học được suy luận [7].

3.4 Tấn công suy luận thành viên và các thuộc tính

Các cuộc tấn công dựa trên suy luận thuộc tính và tư cách thành viên đại diện cho một loại mối đe dọa quyền riêng tư, trong đó kẻ tấn công cố gắng suy luận xem dữ liệu của người học có được sử dụng trong quá trình huấn luyện mô hình hay không, hoặc suy ra các thuộc tính nhạy cảm từ kết quả đầu ra của mô hình. Trong các hệ thống học tập cá nhân hóa, những mối đe dọa này có thể xuất hiện khi các mô hình dự đoán, công cụ đề xuất hoặc bảng điều khiển phân tích học tập tiết lộ các dự đoán được cá nhân hóa hoặc tổng hợp [7].

Các lỗ hổng về quyền riêng tư có thể phát sinh thông qua tương tác hệ thống thông thường chứ không phải thông qua các vi phạm dữ liệu rõ ràng, thách thức các giả định truyền thống về kiểm soát truy cập dữ liệu [8].

3.5 Rủi ro về quyền riêng tư trong học tập cộng tác và học tập xã hội

Môi trường học tập hợp tác và tương tác xã hội đặt ra những thách thức bổ sung về quyền riêng tư bằng cách kết hợp dữ liệu của từng người học với phân tích cấp độ nhóm. Mặc dù các cơ chế trí tuệ tập thể như vậy có thể làm phong phú thêm trải nghiệm học tập, nhưng chúng cũng có thể vô tình làm lộ thông tin cá nhân của người học thông qua số liệu thống kê tổng hợp hoặc suy luận về hành vi nhóm.

Việc quản lý quyền riêng tư trong học tập cộng tác đòi hỏi phải chú ý đến cả việc sử dụng dữ liệu ở cấp độ cá nhân và tập thể, một sự phân biệt thường bị xem nhẹ trong các cuộc thảo luận về quyền riêng tư hiện có [7].

Hình 2 đã tổng hợp các nguy cơ trong hệ thống LMS có ứng dụng AI để thực hiện cá nhân hóa học tập. Lưu ý những sai sót từ cá nhân có thể được lập trình ngăn chặn bỏ qua thu thập dữ liệu này. Nếu định danh có chủ đích, cần xem xét mức độ vi phạm quyền riêng tư. Nguy cơ từ thuật toán, có thể được khắc phục từ việc liên tục cập nhật thuật toán mới.

3.6 Dữ liệu học tập theo thời gian dài và tái nhận dạng

Một đặc điểm nổi bật của các hệ thống học tập cá nhân hóa là việc tích lũy dữ liệu người học trong thời gian dài, qua nhiều khóa học và nhiều nền tảng khác nhau. Dữ liệu học tập theo thời gian hỗ trợ việc thích ứng liên tục và mô hình hóa người học sâu sắc hơn, nhưng chúng cũng làm tăng đáng kể nguy cơ bị nhận dạng lại. Ngay cả khi các tập dữ liệu được ẩn danh, sự kết hợp của các mô hình theo thời gian, dấu hiệu hành vi và thông tin ngữ cảnh vẫn có thể cho phép người học bị nhận dạng lại [6].

Những thách thức này nhấn mạnh tầm quan trọng của việc xem xét các rủi ro về quyền riêng tư trong toàn bộ vòng đời của dữ liệu người học, thay vì coi quyền riêng tư là một thuộc tính tĩnh của các tập dữ liệu riêng lẻ [10].

Bảng 3 Các mối đe dọa quyền riêng tư trong học tập cá nhân hóa

Mối đe dọa	Mô tả	Nguồn phát sinh	Tác động
Phân tích hồ sơ người học (Learner profiling)	Lập hồ sơ người học dài hạn	Phân tích học tập	Kỳ thị, thiên lệch
Suy luận tư cách thành viên (Membership inference)	Suy luận dữ liệu huấn luyện	Đầu ra mô hình	Lộ dữ liệu cá nhân
Suy luận thuộc tính (Attribute inference)	Suy luận thuộc tính nhạy cảm	Hệ khuyến nghị (Recommendation / ITS)	Vi phạm riêng tư
Nhận dạng lại (Re-identification)	Nhận dạng lại qua dữ liệu thời gian	Dữ liệu theo chiều dọc (Longitudinal data)	Mất ẩn danh

3.7 Bảo mật vi sai trong phân tích dữ liệu học tập (Differential Privacy in Learning Analytics)

Bảo mật vi sai (Differential Privacy) đã nổi lên như một trong những khuôn khổ được áp dụng rộng rãi nhất để bảo vệ các đóng góp dữ liệu cá nhân trong phân tích thống kê và học máy. Trong phân tích học tập, bảo mật vi sai thường được áp dụng cho thống kê tổng hợp, mô hình dự đoán và bảng phân tích nhằm hạn chế rủi ro nhận dạng hoặc suy luận thông tin về từng người học. Bằng cách đưa tính ngẫu nhiên có kiểm soát vào phân tích dữ liệu hoặc đầu ra của mô hình, bảo mật vi sai cung cấp các đảm bảo chính thức rằng việc bao gồm hoặc loại trừ dữ liệu của một người học duy nhất sẽ có tác động giới hạn đến kết quả phân tích [4].

Sự đánh đổi giữa bảo vệ quyền riêng tư và hiệu suất cá nhân hóa này đặc biệt nổi bật trong các ứng dụng giáo dục, nơi đầu ra của mô hình có thể ảnh hưởng trực tiếp đến các quyết định giảng dạy và hỗ trợ người học [9].

Công thức 1: định nghĩa cốt lõi của quyền riêng tư vi sai [17].

$$\Pr[\mathcal{M}(D_1) \in S] \leq e^{\varepsilon} \cdot \Pr[\mathcal{M}(D_2) \in S]$$

M: cơ chế

D_1, D_2: hai bộ dữ liệu lân cận

S: tập các kết quả đầu ra

Pr[M(D)∈S]: xác suất đầu ra

ε(epsilon): tham số riêng tư

e^{ε}: hệ số giới hạn rủi ro

Công thức này giúp phát triển hệ thống đưa ra xác suất về quyền riêng tư vi sai, dùng để tinh chỉnh với những yếu tố khác.

3.8 Học tập liên kết và dữ liệu giáo dục phân tán

Học tập liên kết cho phép huấn luyện mô hình hợp tác giữa nhiều bên nắm giữ dữ liệu mà không cần thu thập dữ liệu tập trung. Trong môi trường giáo dục, phương pháp này đặc biệt phù hợp khi dữ liệu người học được phân tán trên nhiều tổ chức, nền tảng hoặc thiết bị của người học.

Trong một thiết lập học tập liên kết điển hình, mô hình toàn cục được tạo ra bằng cách tổng hợp các mô hình được huấn luyện cục bộ như sau:

Công thức 2: học tập liên kết (Federated Learning aggregation [16])

$$w^{(t-1)} = \sum_{k=1}^{K} \frac{n_k}{n} w_k^{(t)}$$

$w^{(t+1)}$: mô hình toàn cục sau vòng huấn luyện thứ t, do server tạo ra để gửi lại cho các client ở vòng huấn luyện kế tiếp.

$w_k^{(t)}$: mô hình cục bộ của client k sau vòng huấn luyện t, mỗi client cập nhật mô hình toàn cục $w^{(t)}$ bằng dữ liệu riêng của mình.

K: số client/thiết bị tham gia ở vòng huấn luyện này.

n: tổng dữ liệu của tất cả client.

$\frac{n_k}{n}$: trọng số đóng góp của client k.

Bảng 4 Minh họa mô hình của client A ảnh hưởng gấp 4 lần client B

Client	Dữ liệu	Trọng số
A	2000	0,8
B	500	0,2

Đối với các hệ thống học tập cá nhân hóa, học tập liên kết mang lại một cơ chế đầy hứa hẹn để tận dụng dữ liệu đào tạo đa dạng, đồng thời giải quyết các hạn chế về quy định và tổ chức. Trên thực tế, những vấn đề này trở nên rõ rệt hơn trong các kịch bản học tập quy mô lớn hoặc liên tổ chức, đã hạn chế việc áp dụng trực tiếp các phương pháp liên kết [10].

3.9 Tính toán bảo mật đa phương trong học tập cộng tác (Secure Multi-Party Computation in Collaborative Learning)

Tính toán bảo mật đa phương (Secure multi-party computation - SMPC) cung cấp các giao thức mật mã cho phép nhiều bên cùng nhau tính toán các hàm trên dữ liệu của họ mà không tiết lộ các đầu vào cơ bản. Trong bối cảnh học tập cộng tác, SMPC có thể hỗ trợ phân tích cấp nhóm, so sánh ngang hàng hoặc cá nhân hóa trong tập thể khi ngăn chặn truy cập trực tiếp vào dữ liệu của từng người học [11].

Mặc dù SMPC cung cấp các đảm bảo mạnh mẽ về quyền riêng tư, nhưng chi phí tính toán có thể hạn chế việc sử dụng trong các hệ thống học tập cá nhân hóa theo thời gian thực. Do đó, SMPC thường được xem xét cho các nhiệm vụ phân tích cụ thể, có rủi ro cao hơn là một giải pháp đa năng cho tất cả các thành phần của quy trình học tập cá nhân hóa.

3.10 Mã hóa đồng hình để huấn luyện mô hình học tập an toàn (Homomorphic Encryption for Secure Model Training)

Mã hóa đồng hình cho phép thực hiện các phép tính trực tiếp trên dữ liệu đã mã hóa, cho phép huấn luyện hoặc đánh giá các mô hình máy học mà không cần giải mã các dữ liệu đầu vào nhạy cảm.

Mặc dù sở hữu các đặc tính bảo mật mạnh mẽ, mã hóa đồng hình vẫn đòi hỏi chi phí tính toán cao. Hạn chế này đặt ra thách thức cho các ứng dụng học tập cá nhân hóa quy mô lớn hoặc cần xử lý nhanh chóng, nơi khả năng phản hồi và khả năng mở rộng là rất quan trọng.

3.11 Môi trường thực thi đáng tin cậy và các phương pháp lai (Hybrid Approaches)

Môi trường thực thi đáng tin cậy (Trusted execution environments - TEE) cung cấp các cơ chế cách ly dựa trên phần cứng, cho phép xử lý dữ liệu an toàn trong các vùng bảo vệ. Trong các hệ thống giáo dục, TEE có thể được sử dụng để bảo vệ dữ liệu nhạy cảm của người học và các mô hình AI trong quá trình thực thi, ngay cả khi hoạt động trong môi trường điện toán không đáng tin cậy.

Nhiều hệ thống học tập cá nhân hóa áp dụng các phương pháp kết hợp nhiều kỹ thuật bảo vệ quyền riêng tư, để cân bằng giữa bảo vệ quyền riêng tư, hiệu quả hệ thống và độ phức tạp triển khai. Ví dụ, học tập liên kết có thể được kết hợp với bảo mật vi sai (differential privacy) hoặc tổng hợp an toàn (secure aggregation), trong khi TEE có thể được sử dụng cùng với xử lý dữ liệu được mã hóa. Các thiết kế kết hợp như vậy, phản ánh nhu cầu thực tế là điều chỉnh các giải pháp AI bảo vệ quyền riêng tư cho các yêu cầu, và hạn chế cụ thể của bối cảnh giáo dục thay vì chỉ dựa vào một kỹ thuật duy nhất [12].

Bảng 5 So sánh các kỹ thuật AI bảo vệ quyền riêng tư trong giáo dục

Kỹ thuật	Mức bảo vệ	Ảnh hưởng hiệu suất	Phù hợp giáo dục
Bảo mật khác biệt (Differential Privacy)	Trung bình – Cao	Giảm độ chính xác	Phân tích học tập
Học tập liên kết (Federated Learning)	Cao	Chi phí giao tiếp	Liên tổ chức
SMPC	Rất cao	Chi phí tính toán	Phân tích nhóm
Mã hóa đồng hình (Homomorphic Encryption)	Rất cao	Rất chậm	Huấn luyện an toàn
TEE	Cao	Phụ thuộc phần cứng	Triển khai thực tế

4 Bàn luận và đề xuất

4.1 Nội dung cá nhân hóa

Hệ thống đề xuất nội dung cá nhân hóa nhằm mục đích gợi ý các tài liệu học tập, các hoạt động hoặc lộ trình học tập phù hợp với nhu cầu, sở thích và tiến độ của từng người học. Các hệ thống đề xuất (Recommendation System) dựa trên AI thường tận dụng hồ sơ người học, lịch sử tương tác và dữ liệu hiệu suất để mang lại trải nghiệm học tập thích ứng.

Trong bối cảnh chú trọng đến quyền riêng tư, hệ thống đề xuất phải cân bằng giữa hiệu quả cá nhân hóa với việc bảo vệ thông tin nhạy cảm của người học.

Các phương pháp đề xuất chú trọng đến quyền riêng tư thường ưu tiên hạn chế việc tiết lộ dữ liệu, trong khi vẫn duy trì tính phù hợp theo ngữ cảnh của các tài nguyên học tập được đề xuất [14].

Ví dụ: có thể dựa vào hồ sơ người học (khoa/khóa/chuyên ngành), mức độ tương tác nhiều hay ít (các khóa học/tỷ lệ hoàn thành bài tập/đặt câu hỏi/điểm số) và kết quả học tập để đề xuất nội dung phù hợp cho người học.

4.2 Đánh giá thích ứng

Hệ thống đánh giá thích ứng (Adaptive Assessment Systems) tự động điều chỉnh độ khó, định dạng, hoặc thứ tự các câu hỏi đánh giá dựa trên phản hồi của người học, và hiệu suất quan sát được. Từ góc độ bảo mật, điều này làm dấy lên những lo ngại liên quan đến việc lập hồ sơ người học, suy luận về đặc điểm nhận thức và việc lưu giữ hồ sơ đánh giá trong thời gian dài.

Từ góc độ bảo mật, điều này làm dấy lên những lo ngại liên quan đến việc lập hồ sơ người học, suy luận về đặc điểm nhận thức và việc lưu giữ hồ sơ đánh giá trong thời gian dài. Các phương pháp trí tuệ nhân tạo bảo vệ quyền riêng tư trong bối cảnh này, nhằm mục đích giảm thiểu những rủi ro trong khi vẫn duy trì tính hợp lệ, độ tin cậy và tính hữu ích về mặt sư phạm của các quy trình đánh giá thích ứng.

Ví dụ: dựa trên mức độ hoàn thành bài tập/câu trả lời tương tác trong LMS để hệ thống tự động điều chỉnh mức độ cho bài học tiếp theo, giúp người học có cảm hứng hiệu quả khi học tập.

4.3 Hệ thống cảnh báo sớm (dự đoán sự bỏ học)

Hệ thống cảnh báo sớm (Early Warning Systems) sử dụng phân tích dữ liệu học tập và mô hình dự đoán, để xác định những người học có nguy cơ mất hứng thú hoặc bỏ học.

Các hệ thống như vậy cũng liên quan đến những dự đoán nhạy cảm về quỹ đạo học tập trong tương lai của người học. Đặc biệt, các chỉ số rủi ro do các mô hình dự đoán tạo ra có thể ảnh hưởng đến nhận thức và quyết định theo những cách mà người học khó có thể phản bác.

Việc triển khai chú trọng đến quyền riêng tư nhấn mạnh việc sử dụng dữ liệu một cách có trách nhiệm, hạn chế tiết lộ thông tin rủi ro và các biện pháp bảo vệ chống lại sự kỳ thị hoặc lạm dụng những hiểu biết dự đoán [14].

Ví dụ: dựa trên thời lượng vào hệ thống, điểm số, tỷ lệ hoàn thành bài tập, dự đoán khả năng bỏ học giữa chừng, hoặc không tham gia khóa tiếp theo.

4.4 Cá nhân hóa dựa trên nhóm

Cá nhân hóa theo nhóm mở rộng khả năng thích ứng ở cấp độ cá nhân, bằng cách kết hợp các động lực học tập cộng tác và xã hội. Xét từ góc độ quyền riêng tư, việc cá nhân hóa dựa trên nhóm lại phát sinh sự phức tạp khác. Dữ liệu của từng người học không chỉ đóng góp vào các đề xuất cá nhân, mà còn đóng góp vào các mô hình và sự điều chỉnh ở cấp độ nhóm. Sự phụ thuộc lẫn nhau này làm tăng nguy cơ lộ dữ liệu gián tiếp, trong đó thông tin nhạy cảm về người học có thể được suy ra từ hành vi nhóm.

Các hệ thống dựa trên AI có thể cá nhân hóa trải nghiệm học tập dựa trên hiệu suất nhóm, tương tác giữa các bạn cùng lớp hoặc các mô hình học tập tập thể.

Ví dụ: dựa trên các cá nhân tương đồng về cấp độ, đã đạt được các kết quả hoặc tỷ lệ tương tác vào những khóa

học, những câu hỏi nào, từ đó cá nhân hóa cho những thành viên thuộc nhóm.

4.5 Mô hình đánh giá mối đe dọa quyền riêng tư trong giáo dục

Việc phát triển các mô hình về mối nguy cơ có nhận thức trong giáo dục (education-aware threat models) sẽ cung cấp một nền tảng phù hợp hơn để đánh giá các kỹ thuật AI bảo vệ quyền riêng tư trong môi trường học tập, và để hiểu cách các rủi ro về quyền riêng tư phát triển theo các lộ trình học tập kéo dài [15].

Ví dụ: ứng dụng AI với các tham số đầu vào là các rủi ro trong việc cá nhân hóa, từ đó dự đoán tỷ lệ rủi ro có thể xảy ra.

4.6 Cân bằng giữa việc bảo vệ quyền riêng tư và hiệu quả cá nhân hóa

Các nghiên cứu trong tương lai nên xem xét cách các phương pháp bảo vệ quyền riêng tư khác nhau ảnh hưởng đến kết quả học tập, chất lượng cá nhân hóa và sự tham gia của người học trong các bối cảnh giáo dục đa dạng.

$$\max U = \alpha \cdot Accuracy - \beta \cdot Privacy\_Loss$$

Sự đánh đổi cơ bản này có thể được diễn đạt một cách trừu tượng như một bài toán tối ưu hóa nhằm cân bằng giữa tiện ích cá nhân hóa và sự mất mát quyền riêng tư. Công thức 3: hàm đánh đổi giữa cá nhân hóa và quyền riêng tư [10].

Khi áp dụng công thức này vào hệ thống, sẽ thấy được không thể triệt tiêu hoàn toàn các rủi ro để đạt được mức độ cá nhân hóa.

4.7 Quyền riêng tư trong trí tuệ học tập cộng tác và học tập tập thể

Cần có nghiên cứu để khám phá cách các rủi ro về quyền riêng tư biểu hiện trong các kịch bản học tập cộng tác, và cách thiết kế các cơ chế bảo vệ mà không làm suy yếu lợi ích sư phạm của tương tác giữa các bạn đồng học và sự thích ứng tập thể. Quyền riêng tư trong học tập tập thể vẫn là hướng nghiên cứu chưa được khám phá đầy đủ nhưng ngày càng quan trọng khi cá nhân hóa dựa trên xã hội và nhóm trở nên phổ biến hơn [4].

Ví dụ: quy định các loại thông tin cá nhân nào hệ thống sẽ bỏ qua để không lộ quá nhiều yếu tố cá nhân.

4.8 Khung đánh giá và tiêu chuẩn bảo mật thông tin trong giáo dục

Các nghiên cứu trong tương lai nên hướng đến việc phát triển các phương pháp đánh giá đồng thời về bảo vệ quyền riêng tư, hiệu quả cá nhân hóa, tính công bằng và giá trị giáo dục.

Ví dụ: có thể dùng khung đánh giá của châu Âu hoặc Mỹ để Việt hóa và áp dụng cho các khóa học ở Việt Nam trong thời gian chờ Bộ Giáo dục và Đào tạo ban hành bộ khung riêng.

4.9 Thiết kế hệ thống AI có nhận thức về chính sách và bối cảnh

Hình 3 Đánh đổi giữa Quyền riêng tư – Hiệu suất – Khả năng giải thích

Việc tích hợp nhận thức về chính sách vào thiết kế hệ thống AI có thể giúp thu hẹp khoảng cách giữa các giải pháp bảo mật và thực tiễn giáo dục.

Hình 3 diễn giải mối tương quan giữa quyền riêng tư, hiệu suất, khả năng giải thích của hệ thống, cho thấy có sự đánh đổi giữa 3 yếu tố này khi yếu tố kia thay đổi.

Ví dụ: hệ thống cần xây dựng tính năng trực quan, mô hình hóa từ tỷ lệ đánh đổi giữa 3 yếu tố trụ cột, nhằm dự đoán được mức độ hiệu quả của hệ thống.

5 Kết luận

Khảo sát này đã xem xét sự giao thoa giữa AI, học tập cá nhân hóa và quyền riêng tư của người học bằng cách xem xét một cách có hệ thống các khái niệm chính, rủi ro về quyền riêng tư, và các phương pháp bảo vệ quyền riêng tư trong bối cảnh giáo dục.

Bắt đầu bằng việc xem xét các kỹ thuật AI cốt lõi làm nền tảng cho việc học tập cá nhân hóa, bao gồm phân tích học tập (learning analytics), hệ thống đề xuất (recommendation systems) và hệ thống hướng dẫn thông minh (intelligent tutoring systems). Những

phương pháp này tạo nên nền tảng của công nghệ giáo dục thích ứng (adaptive educational technologies) bằng cách mô hình hóa hành vi người học và điều chỉnh quy trình giảng dạy.

Những đặc điểm riêng biệt của dữ liệu giáo dục và các khung pháp lý điều chỉnh việc sử dụng chúng. Dữ liệu người học rất chi tiết và mang tính dài hạn, thường trải rộng trên nhiều bối cảnh học tập và khoảng thời gian dài. Mặc dù các công cụ pháp lý như GDPR và FERPA cung cấp các biện pháp bảo vệ quan trọng, nhưng chúng không giải quyết triệt để các thách thức về quyền riêng tư do cá nhân hóa dựa trên AI tiên tiến gây ra.

Dựa trên nền tảng này, nghiên cứu đã tóm tắt các mối đe dọa chính về quyền riêng tư trong các hệ thống học tập cá nhân hóa. Những mối đe dọa này bao gồm việc lập hồ sơ người học và suy luận hành vi, suy luận tư cách thành viên và thuộc tính thông qua đầu ra của mô hình, rủi ro về quyền riêng tư phát sinh trong môi trường học tập cộng tác và xã hội, rủi ro nhận dạng lại liên quan đến dữ liệu học tập theo chiều dọc.

Cuối cùng, các loại kỹ thuật AI bảo vệ quyền riêng tư chính và ứng dụng của chúng trong các kịch bản học tập cá nhân hóa. Các kỹ thuật như bảo mật vi sai, học tập liên kết, tính toán bảo mật đa phương, mã hóa đồng hình và môi trường thực thi đáng tin cậy cung cấp các cơ chế bổ sung để giảm thiểu việc lộ dữ liệu trong khi vẫn hỗ trợ cá nhân hóa.

Tài liệu tham khảo

1. Holmes, W. (2020). Artificial intelligence in education. In *Encyclopedia of education and information technologies* (pp. 88-103). Cham: Springer International Publishing.

2. Richter, O. Z., Juarros, V. I. M., Bond, M., & Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education: where are the educators? *International Journal of Educational Technology in Higher Education*, (16), 6.

3. Dawson, S., Gašević, D., Siemens, G., & Joksimovic, S. (2014, March). Current state and future trends: A citation network analysis of the learning analytics field. In *Proceedings of the fourth international conference on learning analytics and knowledge* (pp. 231-240).

4. Amo-Filva, D., et al. (2023). Learning analytics in higher education: A systematic review of privacy-aware infrastructures. *Computers & Education, 194*, Article 104678.

5. Mutimukwe, C., Viberg, O., Oberg, L. M., & Cerratto-Pargman, T. (2022). Students' privacy concerns in learning analytics: Model development. *British Journal of Educational Technology*, *53*(4), 932-951.

6. Viberg, O., Khalil, A., & Ebner, M. (2022). Learning analytics, privacy, and ethics: A systematic literature review. *Journal of Learning Analytics, 9*(2), 1-20.

7. Niu, J., Liu, P., Zhu, X., Shen, K., Wang, Y., Chi, H., ... & Zhang, Y. (2024). A survey on membership inference attacks and defenses in machine learning. *Journal of Information and Intelligence*, *2*(5), 404-454.

8. Kabir, J., et al. (2023). *Balancing differential privacy and utility in educational data mining. In Proceedings of the Educational Data Mining Conference (EDM).* Paris, France.

9. Liu, B., Lv, N., Guo, Y., & Li, Y. (2024). *Recent advances on federated learning: A systematic survey*. Neurocomputing, 597, 128019.

10. Li, Q., Wen, Z., Wu, Z., Hu, S., Wang, N., Li, Y., ... & He, B. (2021). A survey on federated learning systems: Vision, hype and reality for data privacy and protection. *IEEE Transactions on Knowledge and Data Engineering*, *35*(4), 3347-3366.

11. Fang, H., & Qian, Q. (2021). Privacy preserving machine learning with homomorphic encryption and federated learning. *Future Internet*, *13*(4), 94.
12. Zhang, J., et al. (2022). Hybrid privacy-preserving machine learning with TEEs and federated learning. *IEEE Internet of Things Journal, 9*(12), 10234-10246.
13. Chen, S., et al. (2022). Privacy-aware recommendation systems in e-learning. *IEEE Access, 10*, 55678-55690.
14. Pedro, F., Subosa, M., Rivas, A., & Valverde, P. (2019). Artificial intelligence in education: Challenges and opportunities for sustainable development.
15. Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. *Nature machine intelligence*, *1*(9), 389-399.
16. McMahan, H. B., Moore, E., Ramage, D., Hampson, S., & Arcas, B. A. y. (2017). Communication-efficient learning of deep networks from decentralized data. In Proceedings of the 20th International Conference on Artificial Intelligence and Statistics (AISTATS) (pp. 1273-1282). *Fort Lauderdale, FL, United States.*
17. Dwork, C., McSherry, F., Nissim, K., & Smith, A. (2006). Calibrating noise to sensitivity in private data analysis. *Theory of Cryptography Conference*, 265-284.

Learner privacy when applying artificial intelligence to personalized learning

Hau Tran Binh[1], Quoc Nguyen Kim[2]

[1]Nguyen Tat Thanh University, Department of Information Administration, Ho Chi Minh City, Viet Nam
[2]Nguyen Tat Thanh University, Faculty of Information Technology, Ho Chi Minh City, Viet Nam
tbhau@ntt.edu.vn

Abstract Personalized learning systems rely on artificial intelligence to model learners, predict learning outcomes, and support instructional decisions for different individuals. This study systematically reviews privacy-preserving methods, including differential privacy, federated learning, secure multi-party computation, homomorphic encryption, and trusted execution environments. It also examines how these techniques are applied to personalized learning tasks such as learning analytics, content recommendation, intelligent tutoring, and automated assessment. In addition, the paper identifies key privacy threats in education, including learner profiling, inference attacks on learning behaviors, and personal data leakage in collaborative learning scenarios. These results provide a taxonomy linking personalization goals, privacy risks, and privacy-preserving AI techniques. Based on this framework, the study outlines a research agenda to highlight remaining challenges and future directions for developing trustworthy, privacy-aware AI-driven personalized learning systems.

Keywords Personalized learning, learning analytics, educational data privacy, privacy-aware artificial intelligence