

XÂY DỰNG KHUNG QUẢN TRỊ ĐẠO ĐỨC VÀ QUYỀN RIÊNG TƯ AI LẤY CON NGƯỜI LÀM TRUNG TÂM TRONG GIÁO DỤC ĐẠI HỌC

Nguyễn Thị Thu Trang
Khoa CNTT & KTS, Học viện Ngân hàng

Tóm tắt: Sự bùng nổ của Trí tuệ nhân tạo (AI) trong giáo dục đại học mang lại cơ hội cá nhân hóa học tập nhưng cũng đặt ra những thách thức nghiêm trọng về đạo đức và quyền riêng tư. Bài viết này đề xuất một khung quản trị quyền riêng tư lấy con người làm trung tâm (HCP), tích hợp ba trụ cột: Công nghệ, Yếu tố con người và Đạo đức (THE). Thông qua việc áp dụng mô hình PEA-AI (Privacy-Ethics Alignment in AI), bài viết nhấn mạnh việc quản trị AI không phải là một quy tắc tĩnh mà là một quá trình đàm phán năng động giữa các bên liên quan: sinh viên, giảng viên và các chuyên gia công nghệ. Kết quả nghiên cứu hướng tới việc thiết lập các tiêu chuẩn cho sự minh bạch, công bằng thuật toán và bảo vệ quyền tự chủ của người học trong kỷ nguyên số.

Từ khóa: Trí tuệ nhân tạo (AI), Giáo dục đại học, Quyền riêng tư dữ liệu, Đạo đức AI.

ESTABLISHING A HUMAN-CENTERED GOVERNANCE FRAMEWORK FOR AI ETHICS AND PRIVACY IN HIGHER EDUCATION

Abstract: The rapid expansion of Artificial Intelligence (AI) in higher education offers unprecedented opportunities for personalized learning while simultaneously presenting significant ethical and privacy challenges. This paper proposes a Human-Centered Privacy (HCP) governance framework, which integrates three core pillars: Technology, Human Factors, and Ethics (THE). By applying the Privacy-Ethics Alignment in AI (PEA-AI) model, the study highlights that AI governance should not be viewed as a static set of rules but as a dynamic negotiation process among key stakeholders, including students, faculty, and AI professionals. The research aims to establish standards for transparency, algorithmic fairness, and the protection of learner autonomy within the digital era

Keywords: Artificial Intelligence (AI), Higher Education, Data Privacy, AI Ethics.

Nhận bài: 20/04/2026

Phản biện: 20/05/2026

Duyệt đăng: 24/05/2026

I. ĐẶT VẤN ĐỀ

Sự phát triển nhanh chóng và việc áp dụng rộng rãi công nghệ AI trong giáo dục đại học đang làm thay đổi căn bản bối cảnh giảng dạy, học tập và ra quyết định quản lý. Từ các nền tảng học tập thích ứng, hệ thống gia sư thông minh đến các công cụ AI tạo sinh như ChatGPT. AI mang lại những cơ hội chưa từng có để cá nhân hóa trải nghiệm học tập, nâng cao hiệu quả hoạt động của cơ sở giáo dục và thúc đẩy quá trình nghiên cứu khoa học. Tuy nhiên, những tiến bộ này không chỉ dừng lại ở việc tối ưu hóa quy trình mà còn đặt ra những thách thức đạo đức và quyền riêng tư nghiêm trọng, đòi hỏi sự xem xét thấu đáo từ phía các nhà quản lý.

Trong môi trường đại học, AI vận hành dựa trên việc thu thập và phân tích một lượng lớn dữ liệu nhạy cảm, bao gồm học bạ, hồ sơ tài chính, dữ liệu sức khỏe và thậm chí là các dấu hiệu hành vi của sinh viên và giảng viên. Việc thiếu sự minh bạch trong các thuật toán "hộp đen" thường làm dấy lên những lo ngại về quyền sở hữu dữ liệu, khả năng kiểm soát của người dùng và niềm tin vào hệ thống. Hơn nữa, định kiến thuật toán thường k

thừa từ các dữ liệu lịch sử không công bằng, có nguy cơ duy trì hoặc làm trầm trọng thêm các bất bình đẳng xã hội trong tuyển sinh và đánh giá năng lực.

Sự dịch chuyển sang đào tạo trực tuyến sau đại dịch COVID-19 và sự bùng nổ của AI tạo sinh từ năm 2023 đã đẩy các cuộc thảo luận về quyền riêng tư và đạo đức AI lên một tầm cao mới. Thực tế cho thấy, các khung quản trị hiện tại thường chỉ mang tính chất định hướng lý thuyết hoặc tuân thủ pháp lý thuần túy, thiếu đi các hướng dẫn vận hành cụ thể để cân bằng giữa đổi mới công nghệ và bảo vệ giá trị con người. Đặc biệt, quyền riêng tư trong kỷ nguyên AI không nên được xem là một trạng thái tĩnh, mà là kết quả của một quá trình đàm phán năng động giữa các bên liên quan: sinh viên (người có nhu cầu tự chủ số), giảng viên (người giám sát sư phạm) và các chuyên gia công nghệ (người thiết kế hệ thống).

Để giải quyết khoảng cách này, bài viết đề xuất một khung quản trị Quyền riêng tư lấy con người làm trung tâm (Human-Centered Privacy - HCP). Khung lý thuyết này dựa trên tam giác

THE (Technology - Human Factors - Ethics), tích hợp các giải pháp kỹ thuật bảo vệ dữ liệu với việc thấu hiểu mô hình tư duy của người dùng và các nguyên tắc đạo đức cốt lõi như công bằng, trách nhiệm giải trình và sự tôn trọng nhân phẩm. Bằng cách đặt quyền riêng tư vào vị trí trung tâm của thiết kế và quản trị, các cơ sở giáo dục đại học có thể xây dựng một hệ sinh thái AI bền vững, nơi công nghệ đóng vai trò tăng cường năng lực con người thay vì xâm phạm quyền tự quyết và nhân phẩm của người học.

II. NỘI DUNG NGHIÊN CỨU

2.1. Tiếp cận quyền riêng tư lấy con người làm trung tâm (HCP)

Tiếp cận Quyền riêng tư lấy con người làm trung tâm là một sự chuyển dịch quan trọng từ các giải pháp tập trung vào kỹ thuật thuần túy sang một triết lý thiết kế ưu tiên nhu cầu, giá trị và trải nghiệm của con người trong suốt vòng đời của AI. Khung lý thuyết này được xây dựng dựa trên sự tích hợp của ba trụ cột chính, được mô tả thông qua mô hình tam giác THE.

Sự kết hợp giữa ba yếu tố này tạo nên một giải pháp bảo vệ quyền riêng tư vững chắc và đáng tin cậy cho môi trường giáo dục đại học:

Khía cạnh đạo đức (Ethics): Đóng vai trò nền tảng xác định lý do tại sao phải bảo vệ quyền riêng tư. Quyền riêng tư được xem là một quyền cơ bản của con người, là tiền đề để duy trì các giá trị đạo đức khác như sự tự chủ, lòng tin, sự công bằng và nhân phẩm. Trong giáo dục, điều này đảm bảo sinh viên có quyền tự quyết đối với dữ liệu cá nhân của mình mà không bị thao túng bởi các thuật toán dự báo hành vi.

Khía cạnh công nghệ (Technology): Đại diện cho công cụ thực thi “cái gì” để bảo vệ dữ liệu. Thay vì chỉ dựa vào các phương pháp bảo mật truyền thống, HCP nhấn mạnh việc áp dụng các công nghệ tăng cường quyền riêng tư như Học liên kết (Federated Learning), Quyền riêng tư khác biệt (Differential Privacy) và Mã hóa đồng cấu (Homomorphic Encryption). Các kỹ thuật này cho phép các trường đại học khai thác giá trị từ dữ liệu lớn để cải thiện kết quả học tập mà không cần phải tiết lộ thông tin thô nhạy cảm của người học.

Khía cạnh yếu tố con người (Human Factors): Xác định cách thức “như thế nào” để người dùng

trải nghiệm sự an toàn. Khía cạnh này tập trung vào việc thấu hiểu mô hình tư duy của sinh viên và giảng viên về AI, cũng như nâng cao nhận thức và năng lực AI. Lòng tin của người dùng không chỉ được xây dựng từ độ bền kỹ thuật mà còn từ khả năng kiểm soát, giao tiếp minh bạch và sự phù hợp về văn hóa.

Mô hình HCP đề xuất một hệ thống phòng thủ đa tầng bao quanh cốt lõi là quyền riêng tư của con người. Hệ thống này bao gồm ba lớp bảo vệ quan trọng:

Hướng dẫn thiết kế: Đảm bảo các nguyên tắc quyền riêng tư được tích hợp ngay từ giai đoạn đầu của quá trình phát triển hệ thống, giúp ngăn chặn các hành vi xâm phạm trước khi chúng xảy ra.

Quản trị quyền riêng tư: Cung cấp các cấu trúc tổ chức, quy trình và cơ chế trách nhiệm giải trình để giám sát việc thực thi quyền riêng tư trong suốt vòng đời của hệ thống AI tại trường đại học.

Quy định về quyền riêng tư: Lớp bảo vệ ngoài cùng là các tiêu chuẩn pháp lý bắt buộc như GDPR (Châu Âu) hoặc các luật bảo vệ thông tin cá nhân quốc gia, tạo ra khung pháp lý và chuẩn mực tuân thủ.

Khung lý thuyết HCP cũng thừa nhận rằng quyền riêng tư trong AI không phải là một trạng thái tĩnh hay một giải pháp đáp ứng cho tất cả. Thay vào đó, nó là kết quả của một quá trình đàm phán năng động giữa các bên liên quan dựa trên mô hình PEA-AI. Quá trình này yêu cầu sự cân bằng giữa: Quyền tự chủ số và nhu cầu riêng tư cá nhân của sinh viên, yêu cầu về sự minh bạch và hiệu quả sư phạm của giảng viên và các giới hạn công nghệ và mục tiêu hiệu suất hệ thống của chuyên gia AI.

Tóm lại, HCP không chỉ nhìn nhận quyền riêng tư như một rào cản kỹ thuật hay tuân thủ luật pháp, mà là nền tảng thiết yếu để xây dựng một hệ thống Trí tuệ nhân tạo giáo dục có trách nhiệm, nơi công nghệ đóng vai trò tăng cường năng lực con người thay vì thay thế hay xâm phạm.

2.2. Mô hình PEA-AI và quá trình đàm phán đa bên

Sự bùng nổ của các hệ thống AI trong giáo dục đại học, từ các nền tảng học tập thích ứng đến AI tạo sinh, đang mở ra cơ hội cá nhân hóa trải nghiệm học tập nhưng đồng thời đặt ra những

thách thức đạo đức và rủi ro quyền riêng tư chưa từng có. Các hệ thống này thường vận hành dựa trên việc thu thập khối lượng lớn dữ liệu nhạy cảm của sinh viên, bao gồm hồ sơ học tập và hành vi tương tác, nhưng lại hoạt động như những “hộp đen” thiếu minh bạch, gây khó khăn cho việc hiểu cách thức các thuật toán đưa ra quyết định ảnh hưởng đến người học. Trong bối cảnh đó, việc thiếu một khung quản trị toàn diện dẫn đến các rủi ro về định kiến thuật toán, xâm phạm quyền riêng tư và xói mòn quyền tự chủ cá nhân. Thực trạng này đòi hỏi giáo dục đại học phải chuyển dịch từ việc tuân thủ pháp lý thuần túy sang một cách tiếp cận mang tính triết lý sâu sắc hơn, nhằm bảo vệ những giá trị nhân văn cốt lõi trong kỷ nguyên số.

Để giải quyết vấn đề này, bài viết đề xuất áp dụng khung quyền riêng tư lấy con người làm trung tâm (HCP), một mô hình tích hợp được xây dựng trên tam giác THE. Quá trình vận hành quyền riêng tư trong khung HCP không phải là một trạng thái tĩnh mà là kết quả của một quá trình đàm phán năng động giữa các bên liên quan, được mô tả qua mô hình PEA-AI. Mô hình này chỉ ra rằng sinh viên, giảng viên và các chuyên gia phát triển AI thường giữ những kỳ vọng khác nhau về quyền tự chủ dữ liệu, hiệu quả sư phạm và hiệu suất kỹ thuật. Sự lệch pha giữa các nhóm này chính là điểm xung đột mà hệ thống quản trị cần phải dung hòa (ví dụ như nhu cầu về độ chính xác cao hay bảo mật tuyệt đối của người dùng). Việc thiết lập các cơ chế đối thoại và đàm phán đa bên giúp các cơ sở giáo dục xây dựng được ranh giới quyền riêng tư phù hợp với thực tế, thay vì áp đặt các quy định từ trên xuống một cách cứng nhắc.

Để hiện thực hóa các nguyên tắc này, chiến lược thiết kế đạo đức ngay từ đầu cần được triển khai thông qua các hướng dẫn hành động cụ thể trong suốt vòng đời của AI. Các cơ chế đồng ý “đúng lúc” nên thay thế các chính sách dài dòng bằng các thông báo ngắn gọn tại thời điểm thu thập dữ liệu nhạy cảm, từ đó nâng cao tính tự chủ cho người học. Đồng thời, việc ứng dụng các công cụ AI giải thích được (XAI) sẽ giúp minh bạch hóa các thuật toán dự báo rủi ro, cho phép giáo viên hoặc nhà quản lý luôn giữ quyền kiểm soát cuối cùng đối với các quyết định quan trọng.

Ở lớp bảo vệ ngoài cùng, khung quản trị cần được thể chế hóa thông qua các chính sách và cơ chế giám sát chặt chẽ tại mỗi cơ sở đào tạo. Các đại học cần thành lập Hội đồng Đạo đức AI đa ngành để thực hiện kiểm toán thuật toán định kỳ và đánh giá tác động quyền riêng tư trước khi triển khai bất kỳ công cụ mới nào. Việc tuân thủ các chuẩn mực pháp lý phải được kết hợp với việc nâng cao năng lực AI cho toàn bộ cộng đồng trường học, giúp họ có khả năng nhận diện rủi ro và sử dụng công nghệ một cách có trách nhiệm. Kết luận lại, việc xây dựng một hệ sinh thái AI lấy con người làm trung tâm không chỉ là một yêu cầu kỹ thuật mà là một cam kết đạo đức dài hạn, nhằm đảm bảo rằng công nghệ luôn đóng vai trò là công cụ hỗ trợ giáo dục, bảo vệ giá trị tự do học thuật và lòng tin xã hội trong tương lai.

2.3. Thiết kế đạo đức ngay từ đầu

Việc chuyển đổi các nguyên tắc đạo đức trừu tượng thành các thực thể kỹ thuật đòi hỏi một triết lý thiết kế mang tính chủ động, thường được gọi là “Đạo đức ngay từ đầu” hoặc “Riêng tư ngay từ đầu”. Thay vì xem đạo đức là rào cản và là những yếu tố bổ sung sau khi hệ thống đã hoàn thiện, phương pháp tiếp cận này yêu cầu tích hợp các giá trị con người vào ngay lõi kiến trúc của hệ thống AI ngay từ giai đoạn thu thập dữ liệu đến khi triển khai. Trong môi trường đại học, điều này bắt đầu bằng việc áp dụng các hướng dẫn thiết kế hành động cụ thể nhằm hiện thực hóa các nguyên tắc quản trị, thay vì chỉ dừng lại ở các tuyên bố chính sách chung chung. Một hệ thống AI giáo dục có trách nhiệm phải đảm bảo tính minh bạch có khả năng giải thích, cung cấp thông tin dễ hiểu về cách thức và mục đích dữ liệu được sử dụng tại các thời điểm quan trọng thay vì chỉ đưa ra các chính sách bảo mật dài dòng. Việc ứng dụng các kỹ thuật AI giải thích được giúp giáo viên và sinh viên hiểu được logic đằng sau các dự báo về kết quả học tập, từ đó xây dựng lòng tin và giảm bớt tính chất “hộp đen” của thuật toán.

Một thành phần then chốt trong thực thi HCP là cơ chế đồng ý đúng lúc và các thiết lập mặc định bảo mật. Thay vì yêu cầu sự cho phép một lần cho toàn bộ quá trình, hệ thống nên đưa ra các thông báo nhắc nhở tại thời điểm thu thập dữ liệu nhạy cảm nhất, giúp người dùng nhận thức rõ rủi

ro và lợi ích trong ngữ cảnh cụ thể. Đi kèm với đó là quyền Kiểm soát chi tiết, cho phép sinh viên lựa chọn cấp độ chia sẻ dữ liệu khác nhau cho từng mục đích sử dụng (như cá nhân hóa việc học so với nghiên cứu hành vi tổng quát), thay vì chỉ có lựa chọn “đồng ý tất cả” hoặc “từ chối tất cả”.

Từ khía cạnh kỹ thuật, việc đảm bảo công bằng thuật toán phải được xử lý bằng các thuật toán nhận thức công bằng. Các mô hình như Học biểu diễn công bằng có thể được sử dụng để loại bỏ các thuộc tính nhạy cảm như chủng tộc hay giới tính khỏi dữ liệu huấn luyện mà không làm giảm hiệu suất dự báo của AI. Điều này đặc biệt quan trọng trong các quy trình có rủi ro cao như tuyển sinh hoặc đánh giá học bổng, nơi định kiến lịch sử thường bị khuếch đại bởi máy học. Để duy trì tiêu chuẩn này, các trường đại học cần thực hiện kiểm toán thuật toán định kỳ và sử dụng các công cụ giám sát thời gian thực để phát hiện hiện tượng “trôi dạt dữ liệu”, đảm bảo rằng hệ thống AI vẫn hoạt động đúng với các cam kết đạo đức ban đầu sau khi triển khai.

Cuối cùng, thiết kế đạo đức ngay từ đầu không thể hoàn thiện nếu thiếu đi sự tham gia trực tiếp của con người thông qua thiết kế có sự tham gia. Sinh viên và giảng viên không chỉ là đối tượng thụ hưởng mà phải đóng vai trò là những cộng tác viên trong việc xác định các chuẩn mực riêng tư và đồng kiến tạo các tính năng bảo vệ dữ liệu phù hợp với thực tế đời sống học đường. Việc đánh giá hiệu quả của các công nghệ bảo vệ quyền riêng tư cũng cần chuyển dịch từ các thước đo kỹ thuật thuần túy sang đánh giá lấy con người làm trung tâm, tập trung vào mức độ dễ hiểu, sự thoải mái và lòng tin thực tế của người dùng đối với các biện pháp bảo mật đó. Cách tiếp cận tích hợp này đảm bảo rằng các giải pháp đạo đức không chỉ tồn tại trên giấy tờ mà thực sự trở thành một phần hữu cơ của trải nghiệm giáo dục số.

2.4. Chính sách và giám sát tại Đại học

Việc chuyển dịch từ các hướng dẫn thiết kế kỹ thuật sang một hệ thống quản trị bền vững đòi hỏi sự thể chế hóa thông qua các chính sách cụ thể và cơ chế giám sát đa tầng tại mỗi cơ sở đào tạo. Quản trị quyền riêng tư trong kỷ nguyên AI không còn chỉ là vấn đề tuân thủ pháp lý đơn thuần mà phải trở thành một cấu trúc tổ chức toàn diện, đảm

bảo trách nhiệm giải trình và sự minh bạch trong suốt toàn bộ vòng đời của hệ thống AI.

Lớp bảo vệ ngoài cùng của khung quản trị HCP là việc tích hợp chặt chẽ với các chuẩn mực pháp lý quốc tế và quốc gia. Các trường đại học cần xây dựng chính sách nội bộ dựa trên các trụ cột của Quy định Bảo vệ Dữ liệu Chung (GDPR) của Châu Âu, Luật Bảo vệ Thông tin Cá nhân (PIPL) của Trung Quốc hoặc Đạo luật Quyền Giáo dục và Quyền Riêng tư Gia đình (FERPA) của Mỹ. Trong đó, quyền được giải thích đóng vai trò then chốt, đảm bảo rằng sinh viên và giảng viên có quyền nhận được các lý giải minh bạch, dễ hiểu về logic đằng sau các quyết định tự động ảnh hưởng đến lộ trình học tập hoặc đánh giá năng lực của họ.

Một thành phần nòng cốt trong cấu trúc quản trị tại đại học là việc thành lập các Hội đồng Đạo đức AI đa ngành. Các hội đồng này cần có sự tham gia của nhiều bên liên quan, bao gồm lãnh đạo nhà trường, giảng viên, sinh viên, chuyên gia công nghệ và đội ngũ pháp lý để đảm bảo các quyết định không chỉ dựa trên hiệu suất kỹ thuật mà còn tôn trọng các giá trị nhân văn. Vai trò của hội đồng không chỉ dừng lại ở việc phê duyệt triển khai công nghệ mà còn phải thực hiện các đánh giá tác động quyền riêng tư một cách nghiêm ngặt trước khi áp dụng các công cụ có rủi ro cao, chẳng hạn như hệ thống giám sát thi trực tuyến hoặc phân tích dữ liệu sinh trắc học.

Để duy trì tính liên chính và sự tin cậy, các cơ sở giáo dục cần thiết lập cơ chế kiểm toán thuật toán định kỳ và duy trì các sổ đăng ký rủi ro để phát hiện hiện tượng “trôi dạt dữ liệu” hoặc các định kiến phát sinh sau khi triển khai. Hệ thống quản trị này phải mang tính thích ứng, chuyển dịch từ các quy định tĩnh sang các quy trình giám sát liên tục để bắt kịp với sự tiến hóa nhanh chóng và các lỗ hổng mới của AI tạo sinh. Việc công khai các báo cáo kiểm toán và thiết lập các cơ chế giải quyết khiếu nại sẽ giúp sinh viên có thể phản đối hoặc yêu cầu con người can thiệp vào các quyết định sai lệch của thuật toán.

Cuối cùng, sự thành công của công tác thể chế hóa phụ thuộc mật thiết vào việc nâng cao năng lực AI cho toàn bộ cộng đồng trường học. Các chương trình đào tạo chuyên sâu về đạo đức dữ liệu cho giảng viên và nhận thức về quyền tự chủ

số cho sinh viên là điều kiện cần để các chính sách không chỉ tồn tại trên giấy tờ. Khi người dùng được trang bị kiến thức để nhận diện rủi ro và hiểu rõ quyền lợi của mình, họ sẽ trở thành những chủ thể tích cực trong việc đàm phán các ranh giới riêng tư, từ đó xây dựng một hệ sinh thái giáo dục số minh bạch, công bằng và tôn trọng phẩm giá con người.

III. KẾT LUẬN

Việc xây dựng khung quản trị quyền riêng tư lấy con người làm trung tâm (HCP) trong giáo dục đại học cần được nhìn nhận như một cam kết đạo đức dài hạn, vượt ra ngoài yêu cầu kỹ thuật và tuân thủ pháp lý, nhằm bảo vệ phẩm giá và quyền tự chủ của người học trong kỷ nguyên số. Thông qua mô hình tam giác THE, nghiên cứu khẳng định rằng AI giáo dục có trách nhiệm phải được

kiến tạo trên sự tích hợp chặt chẽ giữa công nghệ bảo mật tiên tiến, hiểu biết về hành vi người dùng và các giá trị đạo đức cốt lõi.

Quyền riêng tư, theo mô hình PEA-AI, là một quá trình động, đòi hỏi sự tương tác và thương lượng liên tục giữa các bên liên quan. Việc dung hòa giữa quyền tự trị cá nhân và yêu cầu giám sát sự phạm đóng vai trò then chốt trong việc xây dựng niềm tin.

Để đảm bảo hiệu quả, các cơ sở giáo dục cần thể chế hóa cơ chế giám sát thông qua hội đồng đạo đức AI, kiểm toán định kỳ và nâng cao năng lực AI cho cộng đồng học thuật. Tựu trung, quản trị AI là một bài toán liên ngành phức hợp, chỉ khi duy trì cách tiếp cận lấy con người làm trung tâm, giáo dục đại học mới có thể khai thác tiềm năng của AI một cách an toàn, công bằng và bền vững.

TÀI LIỆU THAM KHẢO

- Y. A. Dadson, Z. Mohammed, B. Hussein, Y. E. Armah and A. Kumi-Yeboah, "Ethics and Privacy in AI Education: Prospects and Challenges in Higher Education," in *Artificial Intelligence and Human Agency in Education: Volume One*, New York, United States, *Springer Nature*, 2025, pp. 215-242.
- P. A. Bourne, "Ethical AI and Higher Education: Navigating Bias, Privacy, Equity, and Governance," *Global Journal on Innovation, Opportunities and Challenges in Applied Artificial Intelligence and Machine Learning*, 2025.
- Y. Xue, V. Chinapah and C. Zhu, "A Comparative Analysis of AI Privacy Concerns in Higher Education: News Coverage in China and Western Countries," *Education sciences*, 2025.
- M. I. I. Alfiras, A. Q. Emran¹ and A. M. Mohame, "Ethics and governance of generative AI in education: a systematic review on responsible adoption," *Discover Education*, 2026.
- A. Barthwal, M. Campbell and A. K. Shrestha, "Privacy Ethics Alignment in AI: A Stakeholder-Centric Framework for Ethical AI," *Journal Systems*, 2025.
- S. Mustofa, Y. R. Emon, S. B. Mamun, S. A. Akhy and M. T. Ahad, "A novel AI-driven model for student dropout risk analysis with explainable AI insights," *Computers and Education: Artificial Intelligence*, 2025.
- L. J. Wiese, I. Patil, D. S. Schiff and A. J. Magana, "AI ethics education: A systematic literature review," *Computers and Education: Artificial Intelligence*, 2025.
- "AI Ethics: Integrating Transparency, Fairness, and Privacy in AI Development," *Applied Artificial Intelligence An International Journal*, vol. 39, 2025.