

ỨNG DỤNG TRÍ TUỆ NHÂN TẠO TRONG GIÁO DỤC ĐẠI HỌC, NGHIÊN CỨU TRƯỜNG HỢP GIẢNG DẠY MÔN GIÁO DỤC QUỐC PHÒNG VÀ AN NINH CHO SINH VIÊN VIỆT NAM

*TS Đào Anh Tuấn**

Tóm tắt: Bài viết hệ thống hóa cơ sở lý luận và tổng hợp bằng chứng thực tiễn về ứng dụng trí tuệ nhân tạo trong giáo dục đại học, đồng thời xem xét khả năng vận dụng vào dạy học môn Giáo dục quốc phòng và an ninh cho sinh viên Việt Nam. Bằng phương pháp tổng quan tích hợp kết hợp phân tích nội dung văn bản, kết quả cho thấy trí tuệ nhân tạo, đặc biệt là trí tuệ nhân tạo tạo sinh, có tính ứng dụng cao và hiệu quả hơn khi hỗ trợ phát triển học liệu, luyện tập, phản hồi, truy xuất nguồn và xây dựng tình huống học tập; trong khi bằng chứng về tác động bền vững đối với tư duy bậc cao, năng lực ra quyết định và kỹ năng thực hành còn phân tán và phụ thuộc mạnh vào thiết kế nhiệm vụ. Trên cơ sở đó, bài viết đề xuất khung năm trụ cột và ma trận ứng dụng trí tuệ nhân tạo tạo sinh theo bốn học phần, làm cơ sở triển khai có trách nhiệm trong dạy học Giáo dục quốc phòng và an ninh.

Từ khóa: giáo dục quốc phòng và an ninh; ma trận ứng dụng GenAI; sinh viên; trí tuệ nhân tạo.

1. ĐẶT VẤN ĐỀ

Trí tuệ nhân tạo (AI) đang chuyển từ một công nghệ chuyên biệt sang lớp hạ tầng hỗ trợ, ngày càng phổ biến trong giáo dục đại học. Các ứng dụng hiện diện ở nhiều khâu, từ phân tích dữ liệu học tập, dự báo nguy cơ bỏ học, cá nhân hóa nội dung, hỗ trợ đánh giá, tạo học liệu và phản hồi đến quản trị cơ sở đào tạo. Tổng quan của Zawacki-Richter và cộng sự (2019) chỉ ra rằng nghiên cứu AI trong giáo dục đại học phát triển nhanh nhưng còn thiếu sự tham gia của các nhà giáo dục và chưa gắn chặt với lý thuyết sư phạm. Tổng quan bậc ba của Bond và cộng sự (2024) tiếp tục ghi nhận những hạn chế về đạo đức, tính đa dạng bối cảnh và độ chặt chẽ phương pháp.

Sự phổ cập của trí tuệ nhân tạo tạo sinh (Generative AI - GenAI) từ cuối năm 2022 làm thay đổi đáng kể điều kiện tiếp cận công nghệ. Giảng viên và sinh viên có thể tạo văn bản, hình ảnh, câu hỏi, phản hồi hoặc bản nháp học thuật bằng tương tác ngôn ngữ tự nhiên. Khả năng tạo nội dung nhanh làm tăng hiệu suất, nhưng cũng che khuất một đặc điểm nền tảng: đầu ra của mô hình mang tính xác suất và không tự bảo đảm tính đúng đắn, nguồn gốc hay sự phù hợp với mục tiêu dạy học. Một câu trả lời trôi chảy có thể chứa nguồn không tồn tại, văn bản đã hết hiệu lực, diễn giải thiếu bối cảnh hoặc thiên lệch được kế thừa từ dữ liệu và mô hình. Vì vậy, mức độ dễ dùng, sự hài lòng của người dùng và hiệu quả giáo dục phải được xem là ba nhóm kết quả khác nhau.

Đối với môn Giáo dục quốc phòng và an ninh (GDQP&AN), yêu cầu kiểm soát nghiêm ngặt hơn so với nhiều bối cảnh đào tạo thông thường. Môn học đồng thời bao gồm tri thức chính trị - pháp lý, kiến thức quốc phòng - an ninh, giáo dục trách nhiệm

* Khoa Kinh tế - Quản trị kinh doanh, Trường Đại học Mở - Địa chất.

công dân, tổ chức kỷ luật và các nội dung thực hành. Chương trình hiện hành gồm bốn học phần với tổng thời lượng 165 tiết (Bộ Giáo dục và Đào tạo, 2020). Sai lệch về khái niệm, nguồn hoặc hiệu lực văn bản có thể dẫn đến nhận thức không chính xác; còn trong thực hành, hệ thống tự động không thể thay thế đầy đủ việc quan sát trực tiếp, giám sát an toàn và phán đoán chuyên môn của giảng viên.

Nghiên cứu trực tiếp về AI trong dạy học GDQP&AN ở Việt Nam hiện còn khá ít. Cao Trần Trí và Nguyễn Linh Phong (2025) đã nhận diện các cơ hội và thách thức, nhưng chủ yếu ở mức độ khái quát. Tổng quan thư mục của Nguyễn Linh Phong, Lê Đức Sơn và Nguyễn Thị Hà My (2025) cho thấy nền tảng nghiên cứu GDQP&AN trong nước đang được mở rộng, song số nghiên cứu thực nghiệm về AI vẫn còn hạn chế, đặc biệt là các nghiên cứu có nhóm so sánh, chỉ báo kết quả rõ ràng và đánh giá rủi ro. Khoảng trống cốt lõi vì thế không còn là câu hỏi AI có lợi ích và rủi ro hay không, mà là ứng dụng nào phù hợp với nhiệm vụ nào; dữ liệu nào được phép xử lý; ai có quyền kiểm soát đầu ra; và ngưỡng bằng chứng nào đủ để chuyển từ thử nghiệm sang triển khai ở mức độ rộng hơn.

Bài viết tập trung giải quyết bốn mục tiêu chính: (1) làm rõ khái niệm và các tiếp cận lý luận về AI và GenAI trong giáo dục đại học; (2) phân tích các yêu cầu pháp lý và chính sách liên quan tại Việt Nam; (3) tổng hợp bằng chứng thực tiễn quốc tế và trong nước có khả năng chuyển giao sang GDQP&AN; (4) đề xuất khung ứng dụng GenAI có trách nhiệm cùng ma trận vận dụng theo bốn học phần. Đóng góp chính của nghiên cứu là đặt vấn đề sư phạm, quản trị - pháp lý và đặc thù nội dung GDQP&AN trong cùng một khung phân tích thay vì xem chúng như ba nội dung riêng biệt.

2. NỘI DUNG

2.1. Cơ sở lý luận và khung phân tích

2.1.1. Khái niệm và phạm vi trí tuệ nhân tạo trong giáo dục

Nghiên cứu sử dụng định nghĩa cập nhật của OECD, theo đó hệ thống AI là hệ thống dựa trên máy, suy luận từ dữ liệu đầu vào để tạo ra dự báo, nội dung, khuyến nghị hoặc quyết định có khả năng ảnh hưởng đến môi trường vật lý hoặc số; các hệ thống khác nhau về mức độ tự chủ và khả năng thích ứng sau triển khai (OECD, 2024). Định nghĩa này giúp phân biệt AI với số hóa thông thường. Hệ thống quản lý học tập, bài giảng điện tử, mô hình ba chiều hoặc môi trường thực tế ảo không tự động trở thành AI nếu không có chức năng nhận dạng, dự báo, sinh nội dung, thích ứng hoặc khuyến nghị dựa trên dữ liệu.

Trong giáo dục đại học, có thể phân biệt ba phạm vi: học về AI, học với AI và sử dụng AI để quản trị giáo dục. Trí tuệ nhân tạo tạo sinh thuộc nhóm công nghệ có khả năng tạo văn bản, hình ảnh, âm thanh, video hoặc mã máy tính trên cơ sở các mẫu thống kê trong dữ liệu huấn luyện. Giá trị nổi bật của GenAI là khả năng tương tác linh hoạt và tạo nhiều biến thể nội dung trong thời gian ngắn; giới hạn căn bản là đầu ra của nó không kèm theo cơ chế kiểm chứng tính đúng đắn. Do đó, đầu ra của GenAI phải được xử lý như một đề xuất cần kiểm tra, không phải như một nguồn học thuật hoặc căn cứ pháp lý độc lập.

2.1.2. Các tiếp cận lý thuyết

Tiếp cận lấy con người làm trung tâm đặt quyền tự chủ, năng lực hành động và lợi ích của người học và nhà giáo lên trên mục tiêu tự động hóa. Hướng dẫn của UNESCO yêu cầu việc sử dụng AI trong giáo dục phải được thẩm định cả về đạo đức lẫn sự phạm, duy trì sự giám sát có ý nghĩa của con người và bảo vệ quyền riêng tư (Miao & Holmes, 2023). Khung năng lực AI dành cho giáo viên của UNESCO gồm năm phương diện: tư duy lấy con người làm trung tâm, đạo đức AI, nền tảng và ứng dụng AI, sự phạm AI, và phát triển nghề nghiệp với AI (Miao & Cukurova, 2024). Với GDQP&AN, “giám sát có ý nghĩa” không chỉ là sự hiện diện hình thức của giảng viên, mà đòi hỏi người giám sát có đủ thông tin, năng lực và thẩm quyền để phát hiện, bác bỏ hoặc dừng sử dụng đầu ra không phù hợp.

Khung TPACK nhấn mạnh sự tương tác giữa kiến thức công nghệ, sự phạm và nội dung (Mishra & Koehler, 2006). Theo logic này, lựa chọn công cụ phải xuất phát từ chuẩn đầu ra và đặc điểm nội dung, không phải từ tính mới của công nghệ. Một ứng dụng chỉ có ý nghĩa khi tạo ra giá trị bổ sung cho một nhiệm vụ cụ thể và không làm suy yếu hoạt động nhận thức mà sinh viên cần tự thực hiện. Trong GDQP&AN, tri thức nội dung và năng lực kiểm chứng nguồn giữ vai trò đặc biệt vì mô hình có thể diễn giải sai thuật ngữ chuyên ngành hoặc sử dụng văn bản không còn hiệu lực.

Tiếp cận hệ thống kỹ thuật - xã hội coi kết quả triển khai là sản phẩm của tương tác giữa công nghệ, con người, nhiệm vụ, quy trình và cấu trúc tổ chức (Bostrom & Heinen, 1977a, 1977b). Một công cụ đạt độ chính xác cao trong thử nghiệm vẫn có thể thất bại trong lớp học nếu dữ liệu không đại diện, người dùng không hiểu giới hạn, quyền quyết định không rõ hoặc không có quy trình xử lý sự cố. Vì vậy, đơn vị phân tích phù hợp không phải là “AI” nói chung, mà là một hệ thống AI trong một nhiệm vụ giáo dục xác định, với dữ liệu đầu vào, đầu ra, người sử dụng, người kiểm soát và hậu quả của sai sót được mô tả rõ.

Các mô hình chấp nhận công nghệ hữu ích khi giải thích ý định sử dụng, nhưng không đủ để kết luận về chất lượng sự phạm. Tần suất dùng, tính dễ dùng hoặc mức độ hài lòng không chứng minh rằng người học hiểu sâu hơn, ghi nhớ bền hơn hoặc thực hiện kỹ năng chính xác hơn. Bởi vậy, đánh giá AI trong giáo dục cần tách các chỉ báo chấp nhận công nghệ khỏi các chỉ báo về kết quả học tập, công bằng, an toàn và chi phí cơ hội.

2.1.3. Khái niệm ứng dụng trí tuệ nhân tạo có trách nhiệm

Từ các tiếp cận trên có thể hiểu ứng dụng AI có trách nhiệm trong giáo dục đại học là quá trình lựa chọn, thiết kế, triển khai, sử dụng và giám sát hệ thống AI để hỗ trợ một mục tiêu giáo dục xác định; trong đó quyền kiểm soát và trách nhiệm của con người được duy trì, dữ liệu được xử lý hợp pháp và an toàn, đầu ra có thể kiểm tra, việc sử dụng được thông báo minh bạch, và tác động được đánh giá trong toàn bộ vòng đời. Định nghĩa này loại trừ hai cách hiểu cực đoan: coi AI là công cụ trung tính chỉ cần “dùng đúng”, hoặc coi mọi ứng dụng AI là rủi ro cao như nhau. Mức kiểm soát phải tương xứng với nhiệm vụ, dữ liệu và hậu quả có thể xảy ra.

2.2. Phương pháp nghiên cứu

Nghiên cứu sử dụng tổng quan tích hợp có định hướng kết hợp phân tích nội dung văn bản. Tổng quan tích hợp phù hợp khi câu hỏi nghiên cứu đòi hỏi kết nối nghiên cứu lý thuyết, nghiên cứu thực nghiệm, báo cáo thực tiễn và văn bản chính sách (Whittemore & Knafl, 2005). Phân tích tài liệu được dùng để nhận diện phạm vi điều chỉnh, nguyên tắc, trách nhiệm và giới hạn pháp lý (Bowen, 2009).

Nguồn tài liệu học thuật được tổng hợp từ Scopus, Web of Science, ERIC, Google Scholar và trang của nhà xuất bản; nguồn trong nước được đối chiếu với Tạp chí Giáo dục, Tạp chí Quốc phòng toàn dân và cổng thông tin của cơ quan ban hành. Google Scholar chỉ được sử dụng để phát hiện tài liệu; thông tin thư mục và nội dung trọng yếu được kiểm tra lại trên nguồn xuất bản hoặc văn bản chính thức.

Dữ liệu được mã hóa theo các nhóm: loại công nghệ; nhiệm vụ giáo dục; chủ thể sử dụng; lợi ích; rủi ro; thiết kế nghiên cứu; yêu cầu pháp lý; mức độ phù hợp với GDQP&AN; và cơ chế kiểm soát. Mỗi phát biểu được phân biệt thành kết quả thực nghiệm, yêu cầu pháp luật, nhận định của nguồn hoặc suy luận của nghiên cứu. Độ tin cậy được xem xét theo thiết kế nghiên cứu, quy mô và tính đại diện của mẫu, mức độ trực tiếp của kết quả, khả năng kiểm chứng và tính nhất quán giữa các nguồn. Không thực hiện phân tích gộp do khác biệt lớn về công nghệ, bối cảnh, đối tượng và biến kết quả.

Bảng 1. Cấu trúc phân tích và tiêu chí đánh giá bằng chứng

Thành phần	Nội dung trích xuất	Câu hỏi kiểm tra	Cách sử dụng kết quả
Công nghệ	Loại mô hình, dữ liệu, đầu ra	Hệ thống thực hiện chức năng AI nào?	Phân biệt AI với số hóa thông thường
Sự phạm	Mục tiêu, nhiệm vụ, đánh giá	AI tạo giá trị bổ sung cho chuẩn đầu ra nào?	Đánh giá mức phù hợp theo nhiệm vụ
Con người	Năng lực, quyền kiểm soát, trách nhiệm	Ai kiểm tra và ai quyết định cuối cùng?	Xác định giám sát có ý nghĩa
Dữ liệu - pháp lý	Nguồn, căn cứ xử lý, độ nhạy cảm	Dữ liệu có cần thiết, hợp pháp và an toàn không?	Xác định nghĩa vụ và biện pháp kiểm soát
Bằng chứng	Thiết kế, mẫu, chỉ báo, giới hạn	Kết luận có vượt quá dữ liệu không?	Gán mức tin cậy và giới hạn suy rộng

Nguồn: tác giả tổng hợp.

2.3. Kết quả nghiên cứu và thảo luận

2.3.1. Bằng chứng thực tiễn quốc tế và Việt Nam

Các tổng quan gần đây cho thấy AI trong giáo dục đại học được sử dụng nhiều nhất cho đánh giá - phản hồi, cá nhân hóa, khuyến nghị và dạy học thông minh. Luo và cộng sự (2025) tổng hợp 63 nghiên cứu thực nghiệm công bố trong giai đoạn 2014-2024. Kết quả cho thấy công cụ AI nhìn chung có tác động tích cực đến tiếp thu kiến thức và một số kết quả cảm xúc, nhưng tác động đối với quá trình nhận thức bậc cao và kỹ năng lại biến động đáng kể. Đáng lưu ý, 40 trong 63 nghiên cứu sử dụng thiết kế bán thực nghiệm không ngẫu nhiên; vì vậy, bằng chứng tích cực không nên được diễn giải như quan hệ nhân quả ổn định trong mọi bối cảnh.

Tổng quan của Qian (2025) cho thấy ứng dụng GenAI tập trung vào viết, ngôn ngữ, phản hồi, tự học, sáng tạo và năng lực sử dụng câu lệnh. Nghiên cứu này đồng thời chỉ ra một căng thẳng quan trọng: công nghệ có thể nâng hiệu suất nhưng cũng tạo điều kiện chuyển giao các hoạt động nhận thức và siêu nhận thức cho công cụ. Điều đó giải thích vì sao AI không tự động phát triển tư duy phản biện. Giá trị sư phạm chỉ xuất hiện khi nhiệm vụ buộc sinh viên phải kiểm tra, so sánh, bác bỏ, sửa chữa hoặc giải thích đầu ra; ngược lại, bài tập chỉ yêu cầu sản phẩm cuối cùng có thể làm tăng tốc độ hoàn thành nhưng giảm phần tư duy do người học trực tiếp thực hiện.

Trong đánh giá, Xia và cộng sự (2024) cho rằng sự xuất hiện của AI tạo sinh đòi hỏi chuyển trọng tâm từ kiểm soát sản phẩm sang thiết kế đánh giá theo quá trình, tự điều chỉnh, trách nhiệm học tập và năng lực AI. Thang đánh giá mức sử dụng AI của Perkins và cộng sự (2024) cung cấp một cách tiếp cận thực hành: quy định rõ từng nhiệm vụ cho phép AI ở mức nào, phần nào phải do người học tự thực hiện và bằng chứng quá trình nào cần nộp. Cách làm này đáng tin cậy hơn việc chỉ dựa vào công cụ phát hiện văn bản do AI tạo, vốn không đủ để làm chứng cứ duy nhất về vi phạm.

Bằng chứng tại Việt Nam phản ánh việc người học đã sử dụng AI trước khi nhiều cơ sở đào tạo có chính sách thống nhất. Maheshwari (2024), trên mẫu 108 sinh viên đại học và sau đại học, cho thấy cảm nhận về tính dễ sử dụng có ảnh hưởng đến ý định chấp nhận ChatGPT; tuy nhiên, đây là bằng chứng về hành vi công nghệ, không phải bằng chứng về hiệu quả học tập. Khảo sát 694 sinh viên sư phạm của Nguyễn Nam Phương và cộng sự (2025) cho thấy ChatGPT được dùng nhiều cho tóm tắt và phân tích tài liệu; dữ liệu chủ yếu dựa trên tự báo cáo và chỉ phản ánh hai cơ sở đào tạo.

Nghiên cứu hỗn hợp của Tran, Le và Pham (2025) cho thấy 85,6% người tham gia có hiểu biết tốt về đạo văn và 61,3% biết hướng dẫn trích dẫn, nhưng cách công khai đóng góp của ChatGPT vẫn dao động từ thừa nhận đến che giấu hoàn toàn. Phỏng vấn 15 sinh viên cho thấy áp lực thời gian, cách hiểu khác nhau về đạo văn và thái độ của giảng viên ảnh hưởng đến hành vi. Kết quả này hàm ý rằng lệnh cấm chung, nếu không đi kèm thiết kế lại đánh giá và hướng dẫn minh bạch, có thể đẩy việc sử dụng AI sang trạng thái khó quan sát hơn.

Đối với GDQP&AN, độ trực tiếp của bằng chứng còn thấp. Cao Trần Trí và Nguyễn Linh Phong (2025) chủ yếu phân tích cơ hội và thách thức ở bình diện khái

niệm. Báo cáo thực tiễn của Phan Xuân Dũng (2025) cho thấy chuyển đổi số tại các trung tâm chưa đồng đều; nhiều nơi mới dừng ở quản lý học tập, điểm danh, lưu trữ hồ sơ và trình chiếu điện tử, trong khi học liệu số, LMS và mô phỏng còn hạn chế. Báo cáo này có giá trị mô tả bối cảnh nhưng không phải đánh giá tác động có đối chứng. Vì vậy, mức tin cậy hiện tại có thể xem là trung bình đối với các ứng dụng hỗ trợ học liệu, phản hồi và luyện tập; thấp đối với tuyên bố về cải thiện bền vững tư duy bậc cao hoặc kỹ năng thực hành GDQP&AN.

2.3.2. Cơ sở pháp lý và nguyên tắc quản trị

Khung pháp lý liên quan được cấu thành từ nhiều lớp: pháp luật trực tiếp về AI; pháp luật giáo dục đại học; pháp luật về dữ liệu cá nhân và bí mật nhà nước; pháp luật chuyên ngành GDQP&AN; cùng các chiến lược, đề án chuyển đổi số. Các lớp văn bản này có chức năng khác nhau. Chính sách khuyến khích đổi mới không thay thế căn cứ pháp lý cho việc thu thập dữ liệu, tự động hóa đánh giá hoặc sử dụng nội dung nhạy cảm.

Luật Trí tuệ nhân tạo số 134/2025/QH15, có hiệu lực từ ngày 01/3/2026, áp dụng cách tiếp cận phân loại theo rủi ro. Điều 9 xác định mức rủi ro dựa trên tác động đến quyền con người, an toàn, an ninh, lĩnh vực sử dụng, lợi ích công cộng, phạm vi người dùng và quy mô ảnh hưởng; Điều 11 quy định trách nhiệm minh bạch; Điều 14 yêu cầu đối với hệ thống rủi ro cao, trong đó có quản trị dữ liệu, lưu hồ sơ, khả năng giám sát và can thiệp của con người, cùng trách nhiệm giải trình. Từ cấu trúc này, không thể suy ra rằng mọi hệ thống AI dùng trong giáo dục đều mặc nhiên là rủi ro cao. Việc phân loại phải dựa trên chức năng thực tế và hậu quả: công cụ gợi ý chỉnh sửa câu chữ khác căn bản với hệ thống chấm điểm, phân loại người học, nhận dạng hành vi hoặc tác động trực tiếp đến quyền lợi học tập.

Luật Giáo dục đại học số 125/2025/QH15 đặt hoạt động ứng dụng công nghệ trong khuôn khổ trách nhiệm của cơ sở giáo dục về chất lượng, minh bạch và giải trình. Quyền tự chủ không chuyển trách nhiệm sang nhà cung cấp công nghệ. Nếu một hệ thống tạo nội dung sai, đánh giá thiên lệch hoặc làm lộ dữ liệu, cơ sở đào tạo vẫn phải chịu trách nhiệm về quyết định triển khai và quy trình kiểm soát.

Luật Bảo vệ dữ liệu cá nhân số 91/2025/QH15 yêu cầu việc xử lý dữ liệu phải có mục đích và căn cứ phù hợp, tôn trọng quyền của chủ thể dữ liệu và bảo đảm an toàn. Trong giáo dục, dữ liệu học tập, thông tin định danh, hồ sơ kỷ luật, hình ảnh hoặc dữ liệu sinh trắc học có thể tạo ra rủi ro đáng kể nếu được chuyển lên nền tảng công cộng. Việc bắt buộc sinh viên sử dụng một dịch vụ khi chưa thẩm định các điều khoản về lưu trữ, tái sử dụng, chuyển dữ liệu ra nước ngoài và quyền xóa dữ liệu là một khiếm khuyết quản trị.

Khung đạo đức trí tuệ nhân tạo quốc gia nhấn mạnh an toàn, độ tin cậy, công bằng, minh bạch, kiểm soát của con người và trách nhiệm giải trình (Bộ Khoa học và Công nghệ, 2026). Khung năng lực số cho người học bổ sung yêu cầu phát triển năng lực sử dụng công nghệ có trách nhiệm (Bộ Giáo dục và Đào tạo, 2025). Đây là các căn cứ quan trọng để thiết kế quy tắc nội bộ, nhưng không thay thế nghĩa vụ pháp lý về dữ liệu, an ninh và chất lượng giáo dục.

Đối với GDQP&AN, Văn bản hợp nhất số 48/VBHN-VPQH xác lập vị trí, mục tiêu, trách nhiệm quản lý và tổ chức thực hiện môn học; Thông tư số 05/2020/TT-BGDĐT quy định chương trình; Thông tư số 39/2026/TT-BGDĐT quy định quy trình biên soạn, thẩm định và ban hành giáo trình. Trong môi trường có nội dung nhạy cảm, Luật Bảo vệ bí mật nhà nước số 117/2025/QH15 tạo thêm một ranh giới bắt buộc: tài liệu mật, tài liệu nội bộ chưa được phép công bố hoặc dữ liệu có khả năng suy diễn thông tin nhạy cảm không được đưa vào dịch vụ AI công cộng. Giáo trình và văn bản chính thức phải giữ vị trí nguồn kiểm chứng; đầu ra AI chỉ là tài liệu hỗ trợ.

Bảng 2. Căn cứ pháp lý và hàm ý đối với ứng dụng AI trong dạy học GDQP&AN

Văn bản/nhóm quy định	Yêu cầu trọng tâm	Hàm ý đối với cơ sở đào tạo	Mức kiểm soát ưu tiên
Luật Trí tuệ nhân tạo 2025	Phân loại theo rủi ro; minh bạch; giám sát và can thiệp của con người; giải trình	Phân loại từng ứng dụng theo chức năng, dữ liệu và hậu quả; lưu hồ sơ kiểm thử	Cao với đánh giá, phân loại, nhận dạng
Luật Giáo dục đại học 2025	Tự chủ gắn với chất lượng, minh bạch và trách nhiệm	Đưa AI vào hệ thống bảo đảm chất lượng, không chỉ quy trình mua sắm	Đánh giá trước, trong và sau triển khai
Luật Bảo vệ dữ liệu cá nhân 2025	Mục đích, căn cứ xử lý, quyền chủ thể dữ liệu, an toàn	Thẩm định nhà cung cấp, thời hạn lưu, quyền truy cập/xóa, chuyển dữ liệu	Cao với dữ liệu học tập và sinh trắc học
Luật Bảo vệ bí mật nhà nước 2025	Bảo vệ thông tin thuộc danh mục bí mật nhà nước	Loại trừ tài liệu mật, nội bộ hoặc chưa được phép công bố khỏi nền tảng công cộng	Rất cao với nội dung nhạy cảm
Khung đạo đức AI 2026 và Khung năng lực số 2025	An toàn, công bằng, minh bạch, năng lực sử dụng có trách nhiệm	Thông báo cho người học; bồi dưỡng năng lực; có đầu mối khiếu nại và phương án dừng	Áp dụng xuyên suốt vòng đời
Luật, chương trình và quy định giáo trình GDQP&AN	Nguồn chính thống; chuẩn đầu ra; lý thuyết gắn thực hành	AI không thay thế giáo trình, thực hành, giám sát an toàn và quyết định chuyên môn	Cao với nội dung chuyên ngành và thực hành

Nguồn: tác giả tổng hợp.

2.3.3. Khung năm trụ cột ứng dụng trí tuệ nhân tạo tạo sinh có trách nhiệm trong giáo dục đại học

Trên cơ sở tổng hợp các phân tích ở trên, tác giả đề xuất khung năm trụ cột cho ứng dụng GenAI có trách nhiệm trong giáo dục đại học

Trụ cột thứ nhất là mục tiêu và thiết kế sư phạm. Mỗi ứng dụng phải được gắn với một chuẩn đầu ra, nhiệm vụ học tập và cơ chế tạo giá trị bổ sung xác định. Việc “có thể dùng AI” không phải là lý do đủ để triển khai; cần chứng minh rằng công nghệ cải thiện chất lượng, khả năng tiếp cận hoặc hiệu suất mà không làm mất đi hoạt động nhận thức cốt lõi.

Trụ cột thứ hai là năng lực con người và giám sát có ý nghĩa. Cán bộ quản lý, giảng viên và sinh viên cần có năng lực tương ứng với vai trò. Người giám sát phải hiểu giới hạn của hệ thống, có đủ thời gian kiểm tra và có quyền bác bỏ đầu ra. Đối với đánh giá hoặc thực hành, quyết định cuối cùng phải thuộc về người có thẩm quyền chuyên môn.

Trụ cột thứ ba là nội dung, dữ liệu và an toàn thông tin. Nguồn tri thức phải được kiểm duyệt, dữ liệu được phân loại theo mức nhạy cảm, nguyên tắc tối thiểu hóa dữ liệu được áp dụng ngay từ thiết kế. Hệ thống không được tiếp nhận tài liệu mật, dữ liệu cá nhân không cần thiết hoặc nội dung chưa được phép công bố.

Trụ cột thứ tư là quản trị, minh bạch và trách nhiệm giải trình. Cơ sở đào tạo phải quy định phạm vi sử dụng, chủ thể chịu trách nhiệm, yêu cầu khai báo, lưu vết, cơ chế khiếu nại và quy trình xử lý sự cố. Sinh viên phải biết khi nào họ đang tương tác với GenAI, dữ liệu nào được thu thập và GenAI được phép tham gia đến mức nào trong từng nhiệm vụ.

Trụ cột thứ năm là hạ tầng, đánh giá và cải tiến liên tục. Quyết định mở rộng phải dựa trên kết quả học tập, chất lượng đầu ra, công bằng, an toàn, chi phí và khả năng duy trì. Khung này tương thích với tiếp cận chính sách ba chiều sư phạm - quản trị - vận hành của Chan (2023) và nguyên tắc phân mức sử dụng AI trong đánh giá của Perkins và cộng sự (2024), nhưng được điều chỉnh cho yêu cầu nội dung và an toàn của GDQP&AN.

2.3.4. Ma trận ứng dụng trí tuệ nhân tạo tạo sinh theo các học phần

Mức độ phù hợp của GenAI được xác định theo năm tiêu chí: bản chất chuẩn đầu ra; khả năng kiểm chứng; hậu quả của sai sót; độ nhạy cảm của dữ liệu; và khả năng giám sát của giảng viên. GenAI phù hợp hơn với các nhiệm vụ nhận biết, giải thích, so sánh, hệ thống hóa và phản biện thông tin. Mức phù hợp giảm khi chuẩn đầu ra đòi hỏi kỹ năng vận động, phối hợp tập thể, chấp hành quy trình hoặc xử lý trong môi trường thực tế.

Ở học phần Đường lối quốc phòng và an ninh, GenAI có thể hỗ trợ sơ đồ hóa khái niệm, tạo câu hỏi tự luyện, truy xuất văn bản và tổ chức hoạt động phản biện. Rủi ro chính là nguồn giả, văn bản hết hiệu lực và diễn giải thiếu bối cảnh. Ở học phần Công

tác quốc phòng và an ninh, công nghệ có thể hỗ trợ xây dựng tình huống giả định và phản hồi bài tập, nhưng phải phân biệt rõ giả định với sự kiện thực tế và không được tự động đưa ra quyết định ảnh hưởng đến người học.

Ở học phần Quân sự chung và Kỹ thuật chiến đấu bộ binh và chiến thuật, GenAI chỉ nên giữ vai trò hỗ trợ trực quan hóa, chuẩn bị trước thực hành, mô phỏng giáo dục và phản tư sau hoạt động. Thị giác máy tính có thể được nghiên cứu để hỗ trợ nhận diện một số biểu hiện hình thức, nhưng không được dùng làm căn cứ duy nhất để đánh giá kỹ năng. AI không thay thế tổ chức lớp, giám sát an toàn, thực hành trực tiếp hoặc quyết định đạt - không đạt.

Bảng 3. Ma trận ứng dụng GenAI theo bốn học phần GDQP&AN

Học phần	Mức phù hợp	Ứng dụng có thể xem xét	Rủi ro chính	Giới hạn và kiểm soát
I. Đường lối QP&AN	Cao, có điều kiện	Sơ đồ khái niệm; câu hỏi tự luyện; đối chiếu nguồn; phản biện đầu ra	Nguồn giả; văn bản hết hiệu lực; diễn giải sai	Kho nguồn chính thức; giảng viên thẩm định; AI không phải nguồn cuối cùng
II. Công tác QP&AN	Trung bình - cao	Tình huống giả định; chatbot dựa trên nguồn; phản hồi bài tập	Nhằm giả định với thực tế; tư vấn thiếu căn cứ; rò rỉ dữ liệu	Gắn nhãn; dẫn nguồn; ẩn danh dữ liệu; con người quyết định
III. Quân sự chung	Hạn chế, có điều kiện	Học liệu trực quan; mô phỏng; câu hỏi chuẩn bị; nhận dạng trong thử nghiệm	Nhận dạng sai; đồng nhất biểu hiện với năng lực	Chỉ dùng hỗ trợ luyện tập; kiểm định độ chính xác; đánh giá trực tiếp
IV. Kỹ thuật chiến đấu bộ binh và chiến thuật	Hạn chế	Trực quan hóa; mô phỏng giáo dục; phản tư trước - sau thực hành	Dữ liệu nhạy cảm; mô phỏng không phản ánh đầy đủ thực tế; rủi ro an toàn	Không thay thế thực hành, giám sát an toàn và quyết định đạt - không đạt

Nguồn: tác giả đề xuất.

2.3.5. Điều kiện và lộ trình triển khai

Cơ sở đào tạo cần ban hành quy định nội bộ trước khi dùng GenAI cho nhiệm vụ có dữ liệu cá nhân, nội dung nhạy cảm hoặc ảnh hưởng đến đánh giá. Quy định tối thiểu phải xác định mục đích; danh mục ứng dụng được phép, hạn chế và không được phép;

dữ liệu có thể sử dụng; trách nhiệm của các bên; yêu cầu khai báo; quy trình phê duyệt; cơ chế khiếu nại; và phương án dừng. Nhóm phụ trách cần có đại diện chuyên môn GDQP&AN, công nghệ, pháp chế hoặc quản trị dữ liệu, khảo thí và bảo đảm chất lượng.

Bồi dưỡng đội ngũ phải vượt ra ngoài kỹ năng viết câu lệnh. Giảng viên cần biết kiểm tra nguồn, phát hiện sai lệch, thiết kế nhiệm vụ có GenAI, bảo vệ dữ liệu và đánh giá phần đóng góp thực tế của sinh viên. Sinh viên cần hiểu rằng đầu ra GenAI không phải bằng chứng; phải khai báo phạm vi sử dụng; và vẫn chịu trách nhiệm đối với sản phẩm nộp. Khung năng lực số cho người học và khung năng lực AI cho giáo viên có thể được dùng để xây dựng chuẩn bồi dưỡng (Bộ Giáo dục và Đào tạo, 2025; Miao & Cukurova, 2024).

Lộ trình phù hợp gồm ba giai đoạn. Giai đoạn chuẩn bị tập trung khảo sát việc sử dụng tự phát, rà soát dữ liệu, ban hành hướng dẫn và lựa chọn nhiệm vụ rủi ro thấp. Giai đoạn thử nghiệm có kiểm soát phải xác định nhóm người học, thời gian, chỉ báo, phương án so sánh và tiêu chí dừng. Giai đoạn chuẩn hóa - mở rộng chỉ được thực hiện khi lợi ích có ý nghĩa giáo dục, rủi ro nằm trong khả năng kiểm soát và đơn vị có đủ năng lực vận hành. Hứng thú của người học hoặc số lượng tài khoản sử dụng không phải căn cứ đủ để mở rộng.

Bảng 4. Lộ trình và chỉ báo đánh giá triển khai

Giai đoạn	Hoạt động chính	Sản phẩm/điểm quyết định	Chỉ báo và tiêu chí dừng
1. Chuẩn bị	Khảo sát thực trạng; phân loại dữ liệu; bồi dưỡng; chọn nhiệm vụ rủi ro thấp	Quy định tạm thời; danh mục ứng dụng; kế hoạch thử nghiệm	Mức sẵn sàng; tỷ lệ giảng viên được bồi dưỡng; rủi ro chưa có biện pháp kiểm soát
2. Thử nghiệm	Triển khai quy mô nhỏ; có phương án so sánh; theo dõi lỗi và phản hồi	Báo cáo tác động; quyết định điều chỉnh, tiếp tục hoặc dừng	Kết quả học tập; độ chính xác; thời gian; sự cố dữ liệu; chênh lệch tiếp cận; lỗi nghiêm trọng
3. Chuẩn hóa - mở rộng	Rà soát pháp lý; chuẩn hóa quy trình; tính chi phí; đánh giá lại định kỳ	Quy trình vận hành; cơ chế giám sát; kế hoạch mở rộng	Chi phí - hiệu quả; công bằng; khả năng giải trình; tính bền vững; xu hướng rủi ro tăng

Nguồn: tác giả đề xuất.

3. KẾT LUẬN VÀ HÀM Ý

3.1. Kết luận

AI có tiềm năng hỗ trợ đổi mới giáo dục đại học và dạy học GDQP&AN, nhưng công nghệ không tự tạo ra chất lượng. Tác động phụ thuộc vào mục tiêu, thiết kế sư phạm, năng lực con người, nguồn dữ liệu, cơ chế quản trị và cách đánh giá. Bằng chứng hiện có tương đối nhất quán đối với các chức năng hỗ trợ học liệu, luyện tập, phản hồi và truy xuất; nhưng yếu hơn và kém đồng đều đối với tư duy bậc cao, năng lực ra quyết định và kỹ năng thực hành.

Khung pháp lý mới của Việt Nam chuyển trọng tâm sang quản lý theo rủi ro, minh bạch, bảo vệ dữ liệu, khả năng can thiệp của con người và trách nhiệm giải trình. AI dùng trong giáo dục không mặc nhiên là rủi ro cao; mức kiểm soát phải được xác định theo chức năng, dữ liệu, quy mô và hậu quả. Với GDQP&AN, giáo trình và văn bản chính thức giữ vị trí nguồn kiểm chứng; AI không thay thế thực hành, giám sát an toàn hoặc quyết định chuyên môn của giảng viên.

Ứng dụng GenAI trong dạy học GDQP&AN chỉ có giá trị khi công nghệ được đặt dưới mục tiêu sư phạm, nguồn nội dung chính thống, quyền kiểm soát của con người, cơ chế bảo vệ dữ liệu và quy trình đánh giá dựa trên bằng chứng. Cách tiếp cận phù hợp không phải là cấm hoàn toàn hoặc mở rộng vô điều kiện, mà là triển khai theo nhiệm vụ, phân loại rủi ro, thử nghiệm có kiểm soát và cải tiến liên tục.

3.2. Hàm ý quản lý và thực tiễn

Đối với cơ quan quản lý, cần xây dựng hướng dẫn chuyên biệt về AI trong giáo dục đại học và GDQP&AN, bao gồm tiêu chí phân loại ứng dụng, yêu cầu tối thiểu đối với nhà cung cấp, biểu mẫu đánh giá trước triển khai và cơ chế báo cáo sự cố. Hướng dẫn đạo đức chung là cần thiết nhưng chưa đủ để vận hành.

Đối với cơ sở đào tạo, ưu tiên không phải là mua nhiều công cụ mà là xây dựng năng lực lựa chọn, kiểm soát và đánh giá. Một đơn vị sử dụng ít công cụ nhưng có kho nguồn được kiểm duyệt, trách nhiệm rõ và bằng chứng tác động đáng tin cậy có mức trưởng thành cao hơn đơn vị triển khai nhiều công cụ theo cách tự phát.

Đối với giảng viên, GenAI nên được dùng để thiết kế nhiệm vụ yêu cầu sinh viên kiểm chứng, giải thích và phản biện, thay vì chỉ tạo sản phẩm nhanh. Mỗi nhiệm vụ cần công bố rõ mức sử dụng GenAI được phép. Không dùng công cụ phát hiện GenAI làm bằng chứng duy nhất cho vi phạm; cần xem xét quá trình làm bài, lịch sử chỉnh sửa, vấn đáp và các nguồn dữ liệu phù hợp khác.

Đối với sinh viên, năng lực AI phải gắn với trách nhiệm học thuật và an toàn thông tin. Sinh viên cần biết kiểm tra nguồn, nhận diện sai lệch, khai báo sự hỗ trợ của công cụ và không đưa dữ liệu cá nhân, tài liệu nội bộ hoặc thông tin chưa được phép công bố lên nền tảng công cộng.

TÀI LIỆU THAM KHẢO

1. Bộ Giáo dục và Đào tạo, (2020), *Thông tư số 05/2020/TT-BGDĐT ngày 18/3/2020 ban hành Chương trình Giáo dục quốc phòng và an ninh trong trường trung cấp sư phạm, cao đẳng sư phạm và cơ sở giáo dục đại học*, Hà Nội.
2. Bộ Giáo dục và Đào tạo, (2025), *Thông tư số 02/2025/TT-BGDĐT ngày 24/01/2025 quy định Khung năng lực số cho người học*, Hà Nội.
3. Bộ Giáo dục và Đào tạo, (2026), *Thông tư số 39/2026/TT-BGDĐT ngày 12/5/2026 quy định quy trình biên soạn, thẩm định và ban hành Giáo trình Giáo dục quốc phòng và an ninh*, Hà Nội.
4. Bộ Khoa học và Công nghệ, (2026), *Thông tư số 05/2026/TT-BKHCN ngày 10/3/2026 ban hành Khung đạo đức trí tuệ nhân tạo quốc gia*, Hà Nội.
5. Quốc hội, (2025a), *Luật Bảo vệ dữ liệu cá nhân*, Luật số 91/2025/QH15 ngày 26/6/2025.
6. Quốc hội, (2025b), *Luật Bảo vệ bí mật nhà nước*, Luật số 117/2025/QH15 ngày 10/12/2025.
7. Quốc hội, (2025c), *Luật Giáo dục đại học*, Luật số 125/2025/QH15 ngày 10/12/2025.
- Quốc hội, (2025d), *Luật Trí tuệ nhân tạo*, Luật số 134/2025/QH15 ngày 10/12/2025.
8. Thủ tướng Chính phủ, (2021), *Quyết định số 127/QĐ-TTg ngày 26/01/2021 ban hành Chiến lược quốc gia về nghiên cứu, phát triển và ứng dụng trí tuệ nhân tạo đến năm 2030*, Hà Nội.
9. Thủ tướng Chính phủ, (2022), *Quyết định số 131/QĐ-TTg ngày 25/01/2022 phê duyệt Đề án tăng cường ứng dụng công nghệ thông tin và chuyển đổi số trong giáo dục và đào tạo giai đoạn 2022-2025, định hướng đến năm 2030*, Hà Nội.
10. Văn phòng Quốc hội, (2026), *Văn bản hợp nhất số 48/VBHN-VPQH ngày 17/3/2026: Luật Giáo dục quốc phòng và an ninh*.
11. Phan Xuân Dũng, (2025, December 3), *Đẩy mạnh chuyển đổi số trong tổ chức đào tạo ở các trung tâm Giáo dục quốc phòng và an ninh*, Tạp chí Quốc phòng toàn dân.
12. Nguyễn Linh Phong, Lê Đức Sơn & Nguyễn Thị Hà My, (2025), *Tổng quan một số công trình nghiên cứu về Giáo dục quốc phòng và an ninh trên các tạp chí khoa học tại Việt Nam từ năm 2000 đến năm 2025 theo mô hình phân tích thư mục định lượng*, Tạp chí Giáo dục, 25(số đặc biệt 9), 1-6.
13. Nguyễn Nam Phương, Nguyễn Hoài Thu, Lại Thùy Linh & Nguyễn Thị Thanh Hiền, (2025), *Thực trạng sử dụng ChatGPT của sinh viên sư phạm trong tìm kiếm thông tin phục vụ học tập: Nghiên cứu tại Trường Đại học Hùng Vương và Trường Đại học Sư phạm Hà Nội*, Tạp chí Giáo dục, 25 (số đặc biệt 4), 351-356.