

TRÁCH NHIỆM PHÁP LÝ VÀ ĐẠO ĐỨC TRONG VIỆC SỬ DỤNG AI ĐỂ TẠO NỘI DUNG GIÁO DỤC TRỰC TUYẾN

● NGUYỄN DIỆU TRINH

Trường Đại học Kinh tế - Luật,
Đại học Quốc gia Thành phố Hồ Chí Minh

DOI: <https://doi.org/10.62831/202526034>

TÓM TẮT:

Bài viết phân tích về trách nhiệm pháp lý và các vấn đề đạo đức trong việc ứng dụng trí tuệ nhân tạo vào giáo dục tại Việt Nam, đặt trong bối cảnh hội nhập với chuẩn mực quốc tế. Nghiên cứu làm rõ nghĩa vụ tuân thủ pháp luật về quyền tác giả, bảo vệ dữ liệu cá nhân, tính chính xác của nội dung và cơ chế bồi thường khi xảy ra thiệt hại. Đồng thời, bài viết cũng nhấn mạnh các nguyên tắc đạo đức như minh bạch, công bằng, tôn trọng sản phẩm con người và bảo vệ giá trị nhân văn, nhằm bảo đảm quyền lợi và sự an toàn của người học trong môi trường giáo dục số.

Từ khóa: trí tuệ nhân tạo, giáo dục trực tuyến, trách nhiệm pháp lý, đạo đức giáo dục, quyền tác giả.

1. Đặt vấn đề

Sự phát triển vượt bậc của trí tuệ nhân tạo (TTNT) trong những năm gần đây, đặc biệt là các mô hình ngôn ngữ lớn (Large Language Models - LLMs) như ChatGPT, Claude hay Gemini,... có khả năng xử lý dữ liệu ở quy mô khổng lồ, tự động tạo ra nhiều loại học liệu số với vài thao tác đơn giản như tạo ra bài giảng, tài liệu học tập, đề kiểm tra hoặc video minh họa, mở ra triển vọng rút ngắn thời gian thiết kế khóa học, giảm chi phí vận hành và cá nhân hóa trải nghiệm học tập. Điều này mang lại lợi ích to lớn không chỉ cho các cơ sở giáo dục mà còn cho người học, bao gồm nhiều đối tượng đang phải chịu rào cản về điều kiện tiếp cận tài nguyên học tập.

Mặc dù mang lại nhiều lợi ích, việc ứng dụng trí tuệ nhân tạo (AI) trong giáo dục trực tuyến cũng đặt ra những thách thức pháp lý và đạo đức cần được kiểm soát chặt chẽ. Nguy cơ sử dụng hoặc tái tạo nội dung có bản quyền mà chưa được phép, xâm phạm quyền riêng tư trong thu thập và xử lý dữ liệu cá nhân hay phát tán thông tin thiếu kiểm chứng đều là những rủi ro hiện hữu. Khả năng hỗ trợ gian lận học thuật có thể làm suy giảm tính liêm chính và minh bạch trong môi trường giáo dục. Trong khi pháp luật của nhiều quốc gia, trong đó có Việt Nam, vẫn đang hoàn thiện để theo kịp tốc độ phát triển công nghệ. Vì vậy, việc xây dựng một khung pháp lý rõ ràng và phù hợp thực tiễn là yêu cầu cấp thiết.

Bài viết này tập trung vào hoạt động tạo nội dung giáo dục trực tuyến bằng công cụ trí tuệ nhân tạo, bao gồm khóa học trực tuyến, e-learning (học qua công nghệ điện tử và internet) và MOOCs (khóa học trực tuyến mở quy mô lớn). Phạm vi phân tích đặt trọng tâm tại Việt Nam, đồng thời tham chiếu kinh nghiệm và quy định của EU cùng khuyến nghị từ UNESCO để rút ra bài học áp dụng. Mục tiêu ở đây là xác định trách nhiệm pháp lý của cá nhân và tổ chức khi sử dụng trí tuệ nhân tạo để sản xuất học liệu; phân tích các yêu cầu về đạo đức và chuẩn mực nghề nghiệp; đề xuất khung pháp lý cùng nguyên tắc đạo đức nhằm bảo đảm ứng dụng công nghệ này một cách an toàn, minh bạch và bền vững.

2. Các vấn đề pháp lý trọng tâm

Theo Luật Sở hữu trí tuệ năm 2005 (được sửa đổi, bổ sung năm 2022), người trực tiếp sáng tạo ra tác phẩm được công nhận là tác giả, trí tuệ nhân tạo do không có ý chí và tư cách pháp nhân, không thể trở thành chủ thể quyền tác giả. Trong trường hợp con người sử dụng trí tuệ nhân tạo như một công cụ hỗ trợ và sản phẩm cuối cùng thể hiện tính nguyên gốc và tính sáng tạo, quyền tác giả có thể sẽ thuộc về người đó. Ngược lại, nếu tác phẩm hoàn toàn do trí tuệ nhân tạo tạo ra mà không có sự can thiệp sáng tạo đáng kể của con người thì không đáp ứng điều kiện bảo hộ theo pháp luật Việt Nam. Điều 28 Luật Sở hữu trí tuệ quy định việc sao chép tác phẩm có bản quyền mà không xin phép là hành vi xâm phạm quyền tác giả. Trí tuệ nhân tạo có thể vô tình vi phạm nếu tái tạo nội dung từ dữ liệu được huấn luyện hoặc trích dẫn không đầy đủ, dẫn đến hành vi đạo văn và xâm phạm quyền nhân thân, quyền tài sản của tác giả gốc. Trên phương diện quốc tế, Đạo luật Trí tuệ nhân tạo của Liên minh châu Âu năm 2024 và Công ước Berne đều khẳng định quyền tác giả chỉ thuộc về con người có đóng góp sáng tạo đáng kể. Tương tự, Tổ chức Giáo dục, Khoa học và Văn hóa Liên Hợp Quốc (UNESCO) trong Khuyến nghị năm 2021 cũng đề cao nguyên tắc

“lấy con người làm trung tâm” nhằm bảo đảm trách nhiệm pháp lý và giá trị đạo đức trong hoạt động sáng tạo có sự tham gia của trí tuệ nhân tạo.

Trong quá trình ứng dụng trí tuệ nhân tạo để cá nhân hóa nội dung học tập, việc xử lý các thông tin nhạy cảm của người học như hồ sơ học tập, kết quả đánh giá hay hành vi tương tác học tập tiềm ẩn nguy cơ rò rỉ dữ liệu nếu không có biện pháp bảo mật phù hợp. Pháp luật Việt Nam, tại Nghị định số 13/2023/NĐ-CP quy định rõ nguyên tắc bảo vệ dữ liệu cá nhân rằng tổ chức, cá nhân thu thập dữ liệu phải thông báo cụ thể về mục đích và phạm vi xử lý thông tin (Điều 11), đồng thời phải có sự đồng ý rõ ràng của chủ thể dữ liệu trước khi thu thập và bảo đảm quyền rút lại sự đồng ý bất kỳ lúc nào. Trường hợp vi phạm, mức xử phạt tối đa có thể lên tới 5% tổng doanh thu của năm tài chính liền trước hoặc tối đa 3 tỷ đồng, cá nhân vi phạm bị phạt tối đa bằng một nửa mức phạt của tổ chức. (theo Luật Bảo vệ dữ liệu cá nhân năm 2025 có hiệu lực thi hành từ ngày 01/01/2026). Về phía quốc tế, Quy định Bảo vệ Dữ liệu Chung của Liên minh Châu Âu (GDPR) đặt ra mức phạt lên đến 20.000.000 EUR hoặc 4% doanh thu toàn cầu, yêu cầu thực hiện đánh giá tác động bảo vệ dữ liệu (DPIA) trước khi đưa hệ thống AI vào sử dụng trong giáo dục. Đạo luật AI của EU bổ sung nghĩa vụ minh bạch về nguồn dữ liệu huấn luyện và quy trình xử lý tự động, trong khi UNESCO (2020 - 2022) khuyến nghị xây dựng cơ chế bảo mật chuyên biệt cho dữ liệu giáo dục trực tuyến, nhằm bảo đảm quyền riêng tư và an toàn thông tin của người học, giáo viên trong môi trường số.

Tại môi trường giáo dục trực tuyến, việc sử dụng trí tuệ nhân tạo có nguy cơ tiềm ẩn khi hệ thống tạo ra thông tin sai sự thật, thiếu căn cứ khoa học hoặc chứa nội dung không phù hợp với mục tiêu giảng dạy. Theo Luật Giáo dục năm 2019, tại Điều 2 và Điều 3, nội dung giáo dục phải bảo đảm tính trung thực, khoa học và tuân thủ chuẩn chương trình, mọi hành vi truyền đạt kiến thức sai lệch đều bị nghiêm cấm. Điều này đặt ra

nghĩa vụ pháp lý cho cơ sở đào tạo và người vận hành AI phải thẩm định, kiểm chứng thông tin trước khi sử dụng trong giảng dạy. Đạo luật Trí tuệ nhân tạo của Liên minh Châu Âu (EU AI Act)¹ xếp các hệ thống AI phục vụ giáo dục vào nhóm “rủi ro cao”, yêu cầu tích hợp cơ chế giám sát của con người, kiểm định chất lượng nội dung và minh bạch về thuật toán, nguồn dữ liệu huấn luyện. Các tiêu chuẩn này nhằm ngăn chặn việc phổ biến thông tin sai và bảo đảm công nghệ AI được vận dụng phù hợp với giá trị và mục tiêu của giáo dục.

Khi nội dung do trí tuệ nhân tạo tạo ra gây thiệt hại, đặc biệt trong các lĩnh vực nhạy cảm như y tế, giáo dục, pháp luật Việt Nam quy định rõ cơ chế trách nhiệm và chế tài áp dụng. Về dân sự, theo Điều 584 đến Điều 588 Bộ luật Dân sự năm 2015, cá nhân hoặc tổ chức có hành vi gây thiệt hại phải bồi thường toàn bộ tổn thất về vật chất và tinh thần, không loại trừ lỗi xuất phát từ việc sử dụng công cụ AI. Ở phương diện hành chính, hành vi xâm phạm quyền sở hữu trí tuệ bị phạt tới 500 triệu đồng, đồng thời, việc cung cấp hoặc phát tán thông tin vi phạm an ninh mạng bị xử lý theo Nghị định số 15/2020/NĐ-CP với mức phạt tiền tương ứng tính chất vi phạm. Nếu hậu quả đặc biệt nghiêm trọng xảy ra, trách nhiệm hình sự sẽ được xem xét theo điều 288 Bộ luật Hình sự quy định tội đưa hoặc sử dụng trái phép thông tin trên mạng máy tính, mạng viễn thông, có khung hình phạt từ 3 tháng đến 7 năm tù giam. Ở cấp độ quốc tế, Đạo luật AI của Liên minh Châu Âu và Dự thảo Chỉ thị về trách nhiệm pháp lý đối với AI (AI Liability Directive)² định hướng thiết lập cơ chế buộc nhà cung cấp hoặc vận hành hệ thống AI chịu trách nhiệm bồi thường thiệt hại nhằm bảo đảm quyền lợi của người bị ảnh hưởng và tăng tính ràng buộc pháp lý đối với công nghệ này.

Tại Việt Nam, các chương trình đào tạo trực tuyến, kể cả khóa học ứng dụng CNTT, phải tuân thủ quy định giáo dục như việc đăng ký khóa học với cơ quan có thẩm quyền và công bố rõ chứng chỉ/văn bằng, báo cáo định kỳ về chất lượng giảng

dạy, tỷ lệ hoàn thành, phản hồi của người học. Bên cạnh đó, Công ước Toàn cầu về Công nhận Bằng cấp của UNESCO năm 2019³ tạo cơ sở pháp lý để các chứng chỉ trực tuyến được công nhận xuyên biên giới. Các hướng dẫn của UNESCO năm 2023⁴ về giáo dục số nhấn mạnh nguyên tắc lấy người học làm trung tâm, bảo đảm mọi người có cơ hội tiếp cận bình đẳng với nguồn lực và công cụ học tập số. Trong khi Tổ chức Hợp tác và Phát triển Kinh tế (OECD)⁵ khuyến nghị các cơ sở đào tạo áp dụng cơ chế báo cáo định kỳ và kiểm định chất lượng như một phần của chiến lược bảo đảm chất lượng giáo dục trực tuyến.

3. Các vấn đề đạo đức trong giáo dục

Trong môi trường giáo dục trực tuyến ứng dụng trí tuệ nhân tạo, minh bạch và trách nhiệm giải trình là nguyên tắc cốt lõi để bảo đảm niềm tin của người học. Các cơ sở đào tạo và giáo viên phải công khai rõ ràng khi nội dung được tạo ra hoặc hỗ trợ bởi AI, đồng thời cung cấp thông tin về nguồn dữ liệu, phương pháp xử lý và các giới hạn của công nghệ. Việc này là yêu cầu đạo đức và là nghĩa vụ pháp lý theo Nghị định số 13/2023/NĐ-CP, đòi hỏi tổ chức xử lý dữ liệu cá nhân phải thông báo rõ mục đích, phạm vi và phương thức sử dụng dữ liệu. Còn Đạo luật AI của Liên minh Châu Âu (2024) yêu cầu minh bạch ở mức cao nhất đối với hệ thống AI rủi ro cao, đặc biệt trong giáo dục và UNESCO (2021) trong Khuyến nghị về đạo đức trí tuệ nhân tạo yêu cầu các tổ chức phát triển và triển khai AI phải công bố đầy đủ quy trình thiết kế, phát triển, vận hành hệ thống, thiết lập cơ chế giải thích và chịu trách nhiệm rõ ràng khi AI tạo ra thông tin sai lệch hoặc gây tác động tiêu cực.

Chính trực học thuật đòi hỏi người học và người dạy phải sử dụng AI như một công cụ hỗ trợ, không thay thế vai trò sáng tạo và nghiên cứu của con người. Việc sử dụng AI làm hộ bài tập, luận văn hay sao chép nội dung mà không trích dẫn đúng quy cách là hành vi vi phạm nghiêm trọng chuẩn mực đạo đức và pháp luật. Luật Giáo dục

Đại học 2018 nghiêm cấm mọi hình thức gian lận học thuật, Luật Sở hữu trí tuệ 2022 bảo vệ quyền tác giả đối với tác phẩm có tính sáng tạo. Quy tắc Ứng xử về Liêm chính trong Nghiên cứu của Châu Âu (2017) và Khuyến nghị của UNESCO (2019) về Tài nguyên Giáo dục Mở (OER) nói về nguyên tắc trung thực, ghi nhận đúng công sức sáng tạo của con người, kể cả trong môi trường số. Điều này bảo đảm rằng AI không làm suy giảm giá trị học thuật và quyền sở hữu trí tuệ của tác giả.

Nội dung giáo dục trực tuyến do AI tạo ra phải phù hợp với chuẩn mực văn hóa, tôn trọng sự đa dạng và không xúc phạm đến tín ngưỡng, giới tính, bản sắc dân tộc. Hiến pháp Việt Nam năm 2013 bảo đảm quyền tự do tín ngưỡng và bảo vệ nhân phẩm con người. Luật An ninh mạng 2018 cấm phát tán thông tin trái với thuần phong mỹ tục hoặc gây chia rẽ cộng đồng. Về phạm vi quốc tế, Tuyên ngôn Quốc tế Nhân quyền 1948 (Universal Declaration of Human Rights) đặt ra nguyên tắc tôn trọng nhân quyền và giá trị văn hóa trong giáo dục. Việc áp dụng cơ chế kiểm duyệt và đánh giá nội dung AI trước khi phát hành là cần thiết để phòng tránh nguy cơ xung đột văn hóa và vi phạm pháp luật.

Bảo vệ dữ liệu cá nhân và quyền lợi người tham gia giáo dục trực tuyến là một yêu cầu đạo đức song hành với nghĩa vụ pháp lý. Khi cá nhân hóa nội dung học tập, hệ thống AI thường xử lý thông tin cá nhân của học viên, giáo viên. Việc thu thập và sử dụng dữ liệu này phải dựa trên sự đồng ý rõ ràng, người tham gia học và dạy được phép rút lại sự đồng ý bất kỳ lúc nào, đồng thời áp dụng biện pháp bảo mật như mã hóa và ẩn danh. Quy định này được cụ thể tại Nghị định số 13/2023/NĐ-CP và Luật Bảo vệ quyền lợi Người tiêu dùng 2010, bảo đảm rằng người học được quyền yêu cầu bồi thường hoặc hỗ trợ.

4. Khuyến nghị và giải pháp

Để bảo đảm trách nhiệm pháp lý và đạo đức trong việc ứng dụng AI tạo nội dung giáo dục trực tuyến, cần triển khai đồng bộ các giải pháp ở

nhiều cấp độ khác nhau. Cần sớm hoàn thiện hành lang pháp lý cho AI trong giáo dục, thông qua việc ban hành mới hoặc sửa đổi các luật, nghị định, thông tư liên quan. Trong đó, bộ quy tắc ứng xử về AI⁶ nên được xây dựng làm cơ sở cho phát triển AI. Trên cơ sở đó, cơ quan quản lý giáo dục (Bộ Giáo dục và Đào tạo) cần nghiên cứu ban hành thông tư hoặc quy định hướng dẫn chuyên ngành về sử dụng AI trong giáo dục, yêu cầu mỗi cơ sở giáo dục xây dựng quy chế và quy tắc ứng xử riêng về AI. Nghiên cứu sửa đổi các luật chuyên ngành như Luật Giáo dục và Luật Giáo dục Đại học bổ sung quy định về ứng dụng công nghệ số và AI trong dạy học (theo tinh thần Nghị quyết số 223/NQ-CP của Chính phủ). Các quy định này hướng tới thúc đẩy cá thể hóa học tập, mở rộng cơ hội tiếp cận giáo dục, đồng thời phải nâng cao hiệu lực quản lý nhà nước và xây dựng môi trường học thuật liêm chính trong kỷ nguyên số. Ngoài ra, pháp luật sở hữu trí tuệ cần được xem xét để làm rõ vấn đề đồng tác giả và bản quyền đối với nội dung do AI tạo ra. Người sử dụng AI tạo nội dung phải được xem là tác giả và chịu hoàn toàn trách nhiệm pháp lý về nội dung đó. Quy định này phù hợp với nguyên tắc quốc tế là tác giả con người có quyền và nghĩa vụ đối với sản phẩm trí tuệ của mình, còn AI chỉ là công cụ hỗ trợ nên không thể chịu trách nhiệm pháp lý như con người. Đồng thời, cơ chế kiểm định nội dung AI cũng cần được thiết lập. Nhà nước có thể giao cho các hội đồng chuyên môn hoặc cơ quan kiểm định chất lượng giáo dục xây dựng bộ tiêu chí đánh giá nội dung học liệu do AI tạo ra, nhằm thẩm định tính chính xác, không sai lệch và phù hợp chương trình trước khi đưa vào sử dụng đại trà. Điều này tương tự với yêu cầu của EU rằng các hệ thống AI có ảnh hưởng lớn tới người học phải qua kiểm tra nghiêm ngặt về an toàn, độ tin cậy trước khi triển khai. Bổ sung quy định bảo vệ dữ liệu cá nhân và quyền riêng tư của người học trong môi trường trực tuyến. Việt Nam đã ban hành Nghị định số 13/2023/NĐ-CP về bảo vệ dữ liệu cá nhân cần đảm bảo các nền tảng giáo dục

AI tuân thủ chặt chẽ Nghị định này tương đương tiêu chuẩn GDPR của EU - vốn trao cho cá nhân quyền kiểm soát dữ liệu của họ và yêu cầu minh bạch trong thu thập, sử dụng thông tin.

Các cơ sở giáo dục giữ vai trò triển khai trực tiếp và cần chủ động xây dựng năng lực quản trị AI nội bộ. Trước hết, mỗi đơn vị phải ban hành quy chế sử dụng AI trong nhà trường phù hợp với đặc thù của mình. Quy chế này nên quy định rõ mục đích và phạm vi áp dụng AI trong giảng dạy, học tập; quy trình phê duyệt khi giảng viên hoặc phòng ban muốn sử dụng công cụ AI để soạn bài giảng, tài liệu, những điều cấm hoặc hạn chế. Việc minh bạch này giúp đảm bảo trung thực học thuật và tạo niềm tin cho người học. Nhà trường cần thiết lập quy trình kiểm duyệt đối với nội dung do AI tạo ra trước khi phân phối đến học viên. Tiếp đó, các cơ sở giáo dục nên đầu tư vào đào tạo và nâng cao nhận thức cho đội ngũ giảng viên, cán bộ về AI. Bộ GD&ĐT định hướng tích hợp nội dung về AI vào chương trình đào tạo giáo viên và cán bộ quản lý giáo dục. Giảng viên cần hiểu rõ giới hạn và rủi ro của AI, AI có thể hỗ trợ chấm bài, tổng hợp tài liệu, nhưng cũng có thể tạo ra thông tin sai lệch hoặc định kiến không lường trước. Do đó, nhà trường phải hướng dẫn giảng viên cách sử dụng AI một cách thận trọng, có phê phán. Bên cạnh đó, năng lực số và AI của người học cũng cần được chú trọng. Nên đưa vào chương trình kỹ năng sống hoặc môn tin học nội dung về học tập cùng AI và dạy học sinh, sinh viên cách khai thác AI để nghiên cứu, gợi ý ý tưởng nhưng không vi phạm quy định, các quy tắc học thuật mới có thể được bổ sung trong nội quy sinh viên. Cơ sở giáo dục cần ứng dụng công nghệ quản lý để giám sát việc sử dụng AI như triển khai hệ thống phát hiện nội dung AI nhằm hỗ trợ phát hiện bài làm hoặc nội dung có khả năng do máy tạo ra, kết hợp các công cụ này trong quy trình chấm bài, kiểm tra luận văn để kịp thời phát hiện trường hợp vi phạm. Đồng thời, cần có cơ chế khuyến khích sáng tạo có trách nhiệm, thay vì

cấm đoán tuyệt đối, nhà trường có thể tổ chức các cuộc thi, dự án để giảng viên và sinh viên thử nghiệm phát triển nội dung e-learning với AI dưới sự hướng dẫn, qua đó rút kinh nghiệm và xây dựng văn hóa sử dụng AI lành mạnh. Mục tiêu sau cùng ở cấp trường là tạo dựng môi trường mà AI trở thành “trợ lý” hữu ích cho người dạy và người học, nhưng luôn nằm trong vòng kiểm soát và tuân thủ các giá trị học thuật cốt lõi.

Đối với cá nhân, giảng viên, giáo viên và người học cần nhận thức rõ trách nhiệm đạo đức khi tương tác với AI. Trước hết, mỗi người dạy phải giữ vững nguyên tắc minh bạch và trung thực học thuật. Nếu có sử dụng AI để soạn bài giảng, tài liệu, họ nên công khai ở mức độ phù hợp. Kế đến, người dạy tuyệt đối không được để AI thay mình chịu trách nhiệm. Dù AI góp phần tạo nội dung, giáo viên vẫn là “tác giả” về mặt học thuật và pháp lý của bài giảng đó, phải kiểm duyệt kỹ lưỡng và chịu trách nhiệm về độ chính xác, tính phù hợp của nội dung. Không thể đổ lỗi cho AI khi xảy ra sai sót, bởi AI không phải là chủ thể pháp lý hay đạo đức để gánh vác trách nhiệm thay con người. Về phía người học, cần đề cao tính trung thực và tự giác trong học tập. Sinh viên, học sinh nên hiểu rằng sao chép hoặc nhờ AI làm bài là hành vi vi phạm nghiêm trọng nguyên tắc học thuật. Khi sử dụng nội dung do AI cung cấp, người học phải biết tôn trọng sở hữu trí tuệ, trích dẫn nguồn nếu AI cung cấp tài liệu có bản quyền, kiểm tra để tránh đạo văn vô ý. Ngoài ra, cả người dạy và người học cần chú trọng bảo mật thông tin khi sử dụng AI. Tuyệt đối không nhập các dữ liệu cá nhân nhạy cảm vào các công cụ AI trực tuyến nếu chưa được phép, đề phòng nguy cơ lộ dữ liệu.

Cùng với giải pháp con người và chính sách, việc ứng dụng các công cụ kỹ thuật có vai trò hỗ trợ đắc lực cho quản lý AI trong giáo dục. Thứ nhất, cần phát triển và tích hợp các công cụ AI Explainability - công cụ giúp con người hiểu được cách AI đưa ra quyết định, bao gồm dữ liệu đầu vào, quá trình xử lý và lý do chọn kết quả, nhằm

đảm bảo minh bạch và tin cậy. Các nhà cung cấp nền tảng e-learning ứng dụng AI nên được khuyến khích (bắt buộc) tích hợp mô-đun giải thích và cảnh báo rủi ro. *Thứ hai*, triển khai các công cụ content moderation AI - hệ thống trí tuệ nhân tạo dùng để phát hiện, lọc và xử lý các nội dung vi phạm, độc hại hoặc không phù hợp (như bạo lực, ngôn từ kích động, thông tin sai lệch) trước khi được công bố hoặc lan truyền. Đây là lớp bảo vệ quan trọng vì AI có thể tạo nội dung mang ý kiến chủ quan hoặc thông tin giả, kết hợp kiểm duyệt AI với giám sát của quản trị viên sẽ giúp môi trường học tập trực tuyến an toàn, lành mạnh. *Thứ ba*, áp dụng giải pháp watermarking (đóng dấu) đối với nội dung do AI sinh ra. Watermarking là kỹ thuật nhúng vào nội dung AI (văn bản, hình ảnh, video) những mẫu nhận biết đặc biệt mà chỉ máy tính có thể phát hiện, giúp phân biệt đâu là sản phẩm của AI. Các mô hình AI hiện đại (như GPT) đang được nghiên cứu tích hợp watermark ở đầu ra, nếu triển khai hiệu quả, trường học hoặc nhà xuất bản có thể dùng công cụ quét watermark để biết tài liệu nào do AI hỗ trợ. Tuy watermark không phải giải pháp hoàn hảo nhưng nó nâng cao rào cản đối với việc giả mạo nội dung con người. Song song đó, EU cũng đã yêu cầu trong một số bối cảnh, người dùng AI phải ghi nhãn nội dung do AI tạo ra để cảnh báo người xem, thông báo cho người tương tác biết họ đang tương tác với hệ thống AI, Việt Nam có thể tham khảo quy định này.

Các công cụ quản lý như hệ thống theo dõi và phân tích dữ liệu nên được ứng dụng để giám sát AI, ví dụ qua dashboard ghi nhận tần suất, mục đích sử dụng của giảng viên và mức độ tương tác của sinh viên. Dữ liệu này giúp phát hiện lạm dụng AI, điều chỉnh chính sách kịp thời. Giải pháp như phân tích học tập cũng hỗ trợ quản lý ra quyết định, phát hiện gian lận hay cảnh báo nguy cơ bỏ học, qua đó vừa tận dụng sức mạnh AI vừa giảm tải cho con người và tăng hiệu quả giám sát.

Trong bối cảnh AI phát triển toàn cầu, Việt

Nam cần tích hợp các chuẩn mực quốc tế về pháp lý và đạo đức vào chính sách trong nước. Đạo luật AI của EU (EU AI Act 2024) là khuôn khổ tiên phong, quản lý theo mức độ rủi ro, xếp AI trong giáo dục vào nhóm rủi ro cao, yêu cầu nghiêm ngặt về an toàn, minh bạch và giám sát của con người, cấm các ứng dụng xâm phạm nhân quyền như phân tích cảm xúc học sinh. Khuyến nghị về Đạo đức AI của UNESCO (2021)⁷ mà Việt Nam đã đồng thuận, nhấn mạnh tôn trọng nhân quyền, nhân phẩm, đa dạng, công bằng và giám sát của con người. Chính sách AI giáo dục cần lấy con người làm trung tâm, tránh bất công cho nhóm yếu thế, bảo đảm minh bạch, giải trình và có cơ chế phản hồi, khiếu nại. Việt Nam nên ban hành quy định bắt buộc các hệ thống AI giáo dục cung cấp kênh để học viên, giáo viên phản hồi hoặc thách thức kết quả do AI tạo ra, tương tự quyền được phản đối kết quả AI chấm và yêu cầu con người xem xét lại mà EU đã quy định. Nguyên tắc AI của OECD⁸ (2019, cập nhật 2024) gồm tăng cường bao trùm, bảo vệ quyền con người, minh bạch, an toàn và trách nhiệm giải trình. Đây là cơ sở để Việt Nam vừa bảo đảm quyền lợi người học vừa hài hòa luật trong nước với tiêu chuẩn quốc tế, tạo môi trường pháp lý rõ ràng và thuận lợi cho triển khai AI trong giáo dục.

5. Kết luận

Trí tuệ nhân tạo với khả năng tạo nội dung nhanh chóng và cá nhân hóa ở quy mô lớn đã đặt ra một loạt vấn đề pháp lý và đạo đức phức tạp, từ quyền tác giả, bảo vệ dữ liệu cá nhân, đến trách nhiệm về tính chính xác của thông tin. Nếu thiếu cơ chế kiểm soát và khung pháp lý phù hợp, các rủi ro như vi phạm sở hữu trí tuệ, xâm phạm quyền riêng tư, thiên kiến thuật toán hay lan truyền thông tin sai lệch có thể gây ra những hậu quả lâu dài, làm suy giảm chất lượng và uy tín của giáo dục trực tuyến. Con đường để AI trở thành động lực thúc đẩy đổi mới giáo dục cần dựa trên sự kết hợp 3 vấn đề chính là pháp lý chặt chẽ, nguyên tắc đạo đức bắt buộc và giải pháp kỹ thuật tiên

tiến. Nhà nước cần đóng vai trò định hình khung pháp lý, các cơ sở giáo dục giữ vai trò giám sát và thẩm định, trong khi nhà phát triển công nghệ và người học phải tuân thủ những chuẩn mực đạo đức và quy định pháp luật. Chỉ khi các chủ thể này cùng chia sẻ trách nhiệm và quyền lợi, nền giáo dục trực tuyến mới có thể phát triển một cách bền vững, an toàn và công bằng.

Giá trị đích thực của việc ứng dụng AI trong giáo dục trực tuyến không nằm ở tốc độ hay số lượng mà ở khả năng duy trì và nâng cao chất lượng tri thức, củng cố niềm tin xã hội vào giáo dục, bảo đảm rằng công nghệ luôn phục vụ con người chứ không thay thế hay làm lu mờ vai trò của họ trong hành trình hình thành và lan tỏa tri thức ■

TÀI LIỆU TRÍCH DẪN:

¹Là khuôn khổ pháp lý toàn diện đầu tiên trên thế giới được Liên minh châu Âu thông qua năm 2024 nhằm quản lý việc phát triển và sử dụng trí tuệ nhân tạo

²Là một đề xuất lập pháp của Ủy ban châu Âu công bố vào năm 2022 nhằm thiết lập khung pháp lý chung cho việc xác định và thực thi trách nhiệm bồi thường thiệt hại liên quan đến hệ thống trí tuệ nhân tạo trong Liên minh châu Âu.

³(Global Convention on the Recognition of Qualifications concerning Higher Education) được thông qua tháng 11/2019 tại Paris trong khuôn khổ Hội nghị toàn thể UNESCO lần thứ 40.

⁴Guidance for Generative AI in Education and Research. UNESCO Digital Learning Week 2023.

⁵Tổ chức quốc tế liên chính phủ, được thành lập năm 1961 với trụ sở đặt tại Paris, Pháp. Mục tiêu chính là thúc đẩy tăng trưởng kinh tế bền vững, nâng cao mức sống, duy trì sự ổn định tài chính, hỗ trợ phát triển kinh tế toàn cầu và mở rộng thương mại quốc tế dựa trên các nguyên tắc kinh tế thị trường và dân chủ.

⁶Do Bộ Khoa học và Công nghệ phối hợp cùng các chuyên gia xây dựng nhằm định hướng sự phát triển và ứng dụng trí tuệ nhân tạo tại Việt Nam theo chuẩn mực đạo đức và phù hợp với pháp luật. Bộ nguyên tắc này bao gồm 7 nội dung trọng tâm: đảm bảo tính bền vững, an toàn và bảo mật của hệ thống AI; bảo vệ dữ liệu cá nhân; duy trì tính minh bạch và khả năng giải thích của công nghệ; bảo đảm sự công bằng, không thiên vị; tôn trọng quyền tự chủ của con người; đề cao trách nhiệm giải trình của các bên liên quan; và thiết lập cơ chế phản hồi, khiếu nại hiệu quả.

⁷Là bộ Khuyến nghị toàn cầu đầu tiên về đạo đức trong phát triển và ứng dụng trí tuệ nhân tạo, được 193 quốc gia thành viên UNESCO (trong đó có Việt Nam) thông qua vào năm 2021.

⁸Được Tổ chức Hợp tác và Phát triển Kinh tế (OECD) thông qua tháng 5/2019 và cập nhật năm 2024, là bộ khuyến nghị quốc tế đầu tiên về quản trị trí tuệ nhân tạo, được hơn 40 quốc gia (trong đó có Việt Nam) ủng hộ.

TÀI LIỆU THAM KHẢO:

Bộ Giáo dục và Đào tạo (2017). Thông tư số 05/2017/TT-BGDĐT ngày 25/01/2017 quy định tiêu chuẩn đánh giá chất lượng giáo dục.

Bộ Giáo dục và Đào tạo (2021). Thông tư số 09/2021/TT-BGDĐT ngày 30/3/2021 quy định quản lý tổ chức dạy học trực tuyến trong cơ sở giáo dục phổ thông và cơ sở giáo dục thường xuyên.

Bộ Khoa học và Công nghệ (2024). Quyết định số 1290/QĐ-BKHCN ngày 11/06/2024 về nguyên tắc nghiên cứu và phát triển hệ thống trí tuệ nhân tạo có trách nhiệm tại Việt Nam.

Chính phủ nước Cộng hòa xã hội chủ nghĩa Việt Nam (2013). Nghị định số 131/2013/NĐ-CP ngày 16/10/2013 quy định xử phạt vi phạm hành chính quyền tác giả, quyền liên quan.

Chính phủ nước Cộng hòa xã hội chủ nghĩa Việt Nam (2020). Nghị định số 15/2020/NĐ-CP ngày 03/2/2020 quy định xử phạt vi phạm hành chính trong lĩnh vực bưu chính, viễn thông, tần số vô tuyến điện, công nghệ thông tin và giao dịch điện tử.

Chính phủ nước Cộng hòa xã hội chủ nghĩa Việt Nam (2020). Nghị định số 140/2020/NĐ-CP ngày 30/11/2020 quy định chi tiết thi hành một số điều của Luật Giáo dục.

Chính phủ nước Cộng hòa xã hội chủ nghĩa Việt Nam (2023). Nghị định số 13/2023/NĐ-CP ngày 17/4/2023 của Chính phủ về bảo vệ dữ liệu cá nhân.

Nguyễn Thị Hạnh. (2024). Bàn về trách nhiệm pháp lý do hậu quả từ sử dụng công nghệ AI của nhà sản xuất. Truy cập tại <https://tapchicongthuong.vn/ban-ve-trach-nhiem-phap-ly-do-hau-qua-tu-su-dung-cong-nghe-ai-cua-nha-san-xuat-123560.htm>.

FPT Cần Thơ (2024). Đạo đức, trách nhiệm của học sinh khi sử dụng Generative AI. Truy cập tại <https://cantho-school.fpt.edu.vn/dao-duc-trach-nhiem-hoc-sinh-khi-su-dung-generative-ai>.

Phan Hải Đăng (ngày 4/6/2025). Các quy định về AI nhằm đảm bảo sự phát triển và sử dụng có trách nhiệm, an toàn, đáng tin cậy. Truy cập tại <https://tuoitre.vn/cac-quy-dinh-ve-ai-nham-dam-bao-su-phat-trien-va-su-dung-co-trach-nhiem-an-toan-dang-tin-cay-20250604105438389.htm>

Quốc hội nước Cộng hòa xã hội chủ nghĩa Việt Nam (2005). Luật Sở hữu trí tuệ năm 2005 (sửa đổi, bổ sung năm 2022).

Quốc hội nước Cộng hòa xã hội chủ nghĩa Việt Nam (2006). Luật Bình đẳng giới năm 2006.

Quốc hội nước Cộng hòa xã hội chủ nghĩa Việt Nam (2010). Luật Bảo vệ quyền lợi người tiêu dùng năm 2010.

Quốc hội nước Cộng hòa xã hội chủ nghĩa Việt Nam (2015). Bộ luật Dân sự năm 2015.

Quốc hội nước Cộng hòa xã hội chủ nghĩa Việt Nam (2017). Bộ luật Hình sự năm 2015 (sửa đổi, bổ sung năm 2017).

Quốc hội nước Cộng hòa xã hội chủ nghĩa Việt Nam (2018). Luật An ninh mạng năm 2018.

Quốc hội nước Cộng hòa xã hội chủ nghĩa Việt Nam (2018). Luật Giáo dục Đại học (sửa đổi, bổ sung năm 2018).

Quốc hội nước Cộng hòa xã hội chủ nghĩa Việt Nam (2019). Luật Giáo dục năm 2019.

Quốc hội nước Cộng hòa xã hội chủ nghĩa Việt Nam (2025). Luật Bảo vệ quyền lợi cá nhân năm 2025.

European Union (2024). Artificial Intelligence Act. Official Journal of the European Union. Available at <https://eur-lex.europa.eu/eli/reg/2024/1689/oj>.

European Union (2019). Directive on Copyright in the Digital Single Market (Directive (EU) 2019/790) . Available at <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32019L0790>.

European Union (2022). Artificial Intelligence Liability Directive. Official Journal of the European Union. Available at <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A52022PC0496>.

European Parliament & Council (2016). General Data Protection Regulation (GDPR), Regulation (EU) 2016/679. Available at <https://gdpr-info.eu/>.

Berne Convention for the Protection of Literary and Artistic Works (Paris Act, 1971). Available at <https://www.wipo.int/treaties/en/ip/berne/>.

OECD Recommendation on Artificial Intelligence. Organisation for Economic Co-operation and Development. Available at <https://oecd.ai/en/ai-principles>.

KPMG (2024). AI and the Laws: Navigating content & IP challenges. Available at <https://kpmg.com/vn/en/home/insights/2024/03/ai-and-the-laws.html>

RMIT (April 2025). Governing AI with ethics and law to guide Vietnam's evolving future. Available at <https://www.rmit.edu.vn/news/all-news/2025/apr/governing-ai-with-ethics-and-law-to-guide-vietnam-evolving-future>.

VietnamPlus (2024). Responsible AI must minimise violations on users' legal rights. Available at <https://en.vietnamplus.vn/responsible-ai-must-minimise-violations-on-users-legal-rights-post288862.vnp>.

UNESCO (2021). Recommendation on the Ethics of Artificial Intelligence. Paris: UNESCO. Available at <https://unesdoc.unesco.org/ark:/48223/pf0000381137>.

UNESCO (2019). Global Convention on the Recognition of Qualifications concerning Higher Education. Paris: UNESCO. Available at <https://unesdoc.unesco.org/ark:/48223/pf0000370215>.

UNESCO. (2019). Recommendation on Open Educational Resources. Paris: UNESCO. Available at <https://unesdoc.unesco.org/ark:/48223/pf0000370936>.

United Nations. (1948). Universal Declaration of Human Rights. Available at <https://www.un.org/en/about-us/universal-declaration-of-human-rights>.

UNESCO. (2023). Guidance for generative AI in education and research. UNESCO Digital Learning Week 2023. Paris: UNESCO. Available at <https://unesdoc.unesco.org/ark:/48223/pf0000386693>.

United Nations. (1948). Universal Declaration of Human Rights. Available at <https://www.un.org/en/about-us/universal-declaration-of-human-rights>.

Ngày nhận bài: 5/6/2025

Ngày phản biện đánh giá và sửa chữa: 20/6/2025

Ngày chấp nhận đăng bài: 7/7/2025

LEGAL AND ETHICAL RESPONSIBILITIES IN APPLYING ARTIFICIAL INTELLIGENCE TO ONLINE EDUCATIONAL CONTENT CREATION

● NGUYEN DIEU TRINH

University of Economics and Law,
Vietnam National University - Ho Chi Minh City

ABSTRACT:

This study examines the legal responsibilities and ethical challenges associated with applying artificial intelligence (AI) in education in Vietnam, within the framework of international standards. The study analyzes key legal obligations related to copyright protection, personal data security, content accuracy, and liability for potential damages arising from AI use. In parallel, it underscores fundamental ethical principles including transparency, fairness, respect for human intellectual contributions, and the preservation of humanistic values. By integrating legal and ethical perspectives, the study aims to propose orientations that ensure the protection of learners' rights and promote a safe, equitable, and responsible digital education environment.

Keywords: artificial intelligence, online education, legal responsibility, educational ethics, copyright.