



Working Paper 2026.1.5.1

- Vol. 1, No. 5

**NGOẠI LỆ SỬ DỤNG QUYỀN SAO CHÉP TRONG HUẤN LUYỆN TRÍ TUỆ
NHÂN TẠO: KINH NGHIỆM QUỐC TẾ VÀ KHUYẾN NGHỊ HOÀN THIỆN
PHÁP LUẬT VIỆT NAM**

Nguyễn Hương Giang¹

Sinh viên K61 Kinh tế đối ngoại - Viện Kinh tế và Kinh doanh quốc tế

Trường Đại học Ngoại thương, Hà Nội, Việt Nam

Lê Thị Ngọc Ánh, Vũ Uyên Nhi, Vũ Thu Hiền

Sinh viên K62 Luật thương mại quốc tế - Khoa Luật

Trường Đại học Ngoại thương, Hà Nội, Việt Nam

Phùng Thị Yến

Viện Nghiên cứu Sáng tạo

Trường Đại học Ngoại thương, Hà Nội, Việt Nam

Tóm tắt

Sự phát triển nhanh chóng của trí tuệ nhân tạo (AI) làm gia tăng nhu cầu thu thập và sao chép lượng lớn dữ liệu để huấn luyện mô hình, trong đó nhiều dữ liệu được bảo hộ quyền tác giả. Điều này đặt ra câu hỏi liệu

¹ Tác giả liên hệ, email: k61.2211110098@ftu.edu.vn

việc sao chép phục vụ huấn luyện AI có cấu thành xâm phạm quyền sao chép hay thuộc ngoại lệ bản quyền. Một số hệ thống pháp luật như EU và Hoa Kỳ đã thiết lập cơ chế ngoại lệ (TDM, fair use) nhằm cân bằng giữa bảo hộ quyền tác giả và thúc đẩy đổi mới. Trên cơ sở so sánh kinh nghiệm quốc tế, bài viết chỉ ra khoảng trống pháp lý trong hệ thống pháp luật Việt Nam và đề xuất hoàn thiện quy định về ngoại lệ quyền sao chép trong bối cảnh AI.

Từ khóa: trí tuệ nhân tạo, quyền sao chép, ngoại lệ bản quyền, khai thác văn bản và dữ liệu (TDM), pháp luật Việt Nam

Mã bài viết: LFW_2026-04

REPRODUCTION RIGHT EXCEPTIONS IN ARTIFICIAL INTELLIGENCE TRAINING: INTERNATIONAL EXPERIENCE AND RECOMMENDATIONS FOR IMPROVING VIETNAM'S LEGAL FRAMEWORK

Abstract

The rapid development of artificial intelligence (AI) increases the need to collect and reproduce large datasets for model training, many of which are protected by copyright. This raises the question of whether such reproduction constitutes infringement or falls under copyright exceptions. Jurisdictions such as the EU and the United States have developed exception mechanisms (TDM, fair use) to balance copyright protection and innovation. Based on comparative analysis, this paper identifies legal gaps in Vietnam's legal framework and proposes improvements to regulations on reproduction right exceptions in the AI context.

Keywords: artificial intelligence, reproduction right, copyright exception, text and data mining (TDM), Vietnamese law

Article code: LFW_2026-04

Lời mở đầu

Sự phát triển nhanh chóng của trí tuệ nhân tạo (AI) đã làm gia tăng nhu cầu thu thập, xử lý và sao chép khối lượng lớn dữ liệu để huấn luyện mô hình, trong đó nhiều dữ liệu thuộc phạm vi bảo hộ của quyền tác giả. Điều này đặt ra câu hỏi pháp lý quan trọng: liệu hành vi sao chép tác phẩm trong quá trình huấn luyện AI có cấu thành xâm phạm quyền sao chép hay có thể được coi là ngoại lệ bản quyền. Trên bình diện quốc tế, tranh luận học thuật tập trung vào ba hướng chính: học thuyết sử dụng hợp lý (fair use), các ngoại lệ chuyên biệt như khai thác văn bản và dữ liệu (text and data mining - TDM), và cách tiếp cận kinh tế học pháp luật. Tuy nhiên, các cách tiếp cận này chủ yếu được phát triển trong bối cảnh các nền kinh tế tiên tiến

và chưa giải quyết đầy đủ bài toán thể chế tại các quốc gia đang phát triển như Việt Nam. Trên cơ sở đó, thông qua phương pháp phân tích pháp lý, so sánh và kinh tế học pháp luật về cách tiếp cận của Hoa Kỳ và EU, bài viết nhằm làm rõ giới hạn của quyền sao chép trong huấn luyện AI và đề xuất khung hoàn thiện pháp luật Việt Nam theo hướng cân bằng giữa đổi mới và bảo hộ quyền tác giả.

1. Cơ sở lý luận về quyền sao chép và các ngoại lệ đối với quyền sao chép trong hoạt động huấn luyện trí tuệ nhân tạo

Theo Khoản 10 Điều 4 Luật Sở hữu trí tuệ 2025: “Sao chép là việc tạo ra bản sao của tác phẩm hoặc bản ghi âm, ghi hình bằng bất kỳ phương tiện hay hình thức nào”, qua đó cho thấy việc lưu trữ tác phẩm dưới dạng dữ liệu số trong hệ thống máy tính cũng có thể được coi là vi sao chép. Cách tiếp cận này phù hợp với pháp luật quốc tế khi Công ước Berne và các văn kiện sau này như tuyên bố được thông qua tại Hội nghị ngoại giao của WIPO khẳng định rằng việc lưu trữ tác phẩm trong phương tiện điện tử hoặc bộ nhớ máy tính về nguyên tắc cũng được xem là hành vi sao chép.

Về bản chất pháp lý, quyền sao chép là một quyền tài sản cơ bản của chủ sở hữu quyền tác giả gắn liền với khả năng khai thác giá trị kinh tế của tác phẩm bởi việc sao chép là tiền đề cho các hoạt động phân phối, truyền đạt và thương mại hóa. Từ góc độ này, việc sử dụng tác phẩm trong quá trình huấn luyện trí tuệ nhân tạo có thể làm phát sinh vấn đề liên quan đến quyền sao chép. Trong thực tiễn, quá trình xây dựng và huấn luyện mô hình trí tuệ nhân tạo thường bao gồm các giai đoạn thu thập dữ liệu, tiền huấn luyện và tinh chỉnh, trong đó dữ liệu phải được tải xuống, lưu trữ và xử lý với khối lượng lớn trong hệ thống máy tính. Mặc dù mục đích của hoạt động này chủ yếu nhằm phân tích dữ liệu và huấn luyện thuật toán chứ không phải phổ biến tác phẩm đến công chúng nhưng về phương diện kỹ thuật và pháp lý, việc sử dụng dữ liệu trong quá trình huấn luyện trí tuệ nhân tạo vẫn có làm phát sinh các hành vi sao chép tác phẩm (Trình, 2025).

Tuy nhiên, mặc dù được thừa nhận là độc quyền của chủ thể quyền, quyền sao chép không phải là quyền tuyệt đối. Pháp luật luôn thiết lập các giới hạn và ngoại lệ nhằm bảo đảm sự cân bằng giữa lợi ích của chủ sở hữu và lợi ích chung của xã hội, chẳng hạn các trường hợp sao chép cho mục đích nghiên cứu, giảng dạy hoặc sử dụng cá nhân trong phạm vi nhất định. Vì vậy, sao chép không chỉ là công cụ khai thác thương mại của chủ thể quyền mà còn là cơ chế cho phép công chúng tiếp cận tác phẩm trong những điều kiện hợp lý, miễn là không gây phương hại bất hợp lý đến quyền và lợi ích hợp pháp của tác giả hoặc chủ sở hữu. Về bản chất, các ngoại lệ này cho phép những chủ thể không phải là chủ sở hữu quyền tác giả được thực hiện hành vi sao chép tác phẩm trong những trường hợp nhất định mà không cần xin phép và không phải trả tiền nhuận bút, thù lao cho chủ thể quyền. Cơ chế này phản ánh sự cân bằng giữa yêu cầu bảo hộ

quyền tác giả và nhu cầu tiếp cận, sử dụng tri thức của xã hội, đồng thời tạo điều kiện cho các hoạt động nghiên cứu, giáo dục và phát triển khoa học công nghệ.

2. Kinh nghiệm quốc tế điều chỉnh ngoại lệ quyền sao chép trong huấn luyện AI

2.1. Nguyên tắc sử dụng quyền sao chép trong huấn luyện trí tuệ nhân tạo của Hoa Kỳ

Kinh nghiệm của Hoa Kỳ về quyền sao chép tác phẩm và giới hạn quyền tác giả trong hoạt động huấn luyện trí tuệ nhân tạo cho thấy một bức tranh pháp lý phức tạp. Dựa trên các tiêu chuẩn nghiêm ngặt tại Điều 106 Đạo luật Bản quyền, việc thu thập và lưu trữ dữ liệu đầu vào về nguyên tắc đã cấu thành hành vi xâm phạm độc quyền sao chép (Medina, 2025; Sidebari, 2023). Tuy nhiên, khác với cách tiếp cận thiết lập ngoại lệ cụ thể của Liên minh châu Âu, Hoa Kỳ giải quyết bài toán bản quyền thông qua cơ chế linh hoạt của học thuyết sử dụng hợp lý được quy định tại Điều 107. Điều luật này đưa ra một khung phân tích bốn yếu tố, trong đó trọng tâm định đoạt tính hợp pháp đối với trí tuệ nhân tạo nằm ở yếu tố tính chất chuyển đổi. Dựa trên lý thuyết sử dụng phi biểu đạt, các học giả lập luận rằng việc sao chép tác phẩm thuần túy để phân tích các quy luật thống kê hoàn toàn không xâm phạm đến cốt lõi bảo hộ của bản quyền (Neill, Lee và Thomas, 2025).

Từ góc độ thực tiễn tư pháp, án lệ Authors Guild kiện Google năm 2015 đã tạo tiền lệ nền tảng khi xác nhận hoạt động số hóa quy mô lớn là hành vi sử dụng hợp lý vì mang tính chuyển đổi cao (Authors Guild v. Google Inc., 2015). Dưới lăng kính kinh tế học pháp luật, cơ chế này vận hành như một công cụ tối thiểu hóa chi phí giao dịch, giải quyết triệt để vấn đề phân mảnh quyền sở hữu. Nếu buộc các công ty đàm phán với từng chủ thể quyền, chi phí giao dịch khổng lồ sẽ làm tê liệt hoạt động đổi mới sáng tạo. Do đó, cơ chế sử dụng hợp lý đóng vai trò như một van an toàn, cho phép trích xuất dữ liệu với chi phí đàm phán bằng không, tạo điều kiện cho các mô hình lớn phát triển (Torrance và Tomlinson, 2023).

Dù vậy, sự phụ thuộc quá mức vào cơ chế này cũng bộc lộ những hạn chế nghiêm trọng. Sự linh hoạt của học thuyết sử dụng hợp lý luôn song hành với những bất định pháp lý và chi phí rủi ro tố tụng đáng kể. Việc thiếu vắng các tiêu chuẩn định lượng cụ thể buộc tòa án phải đánh giá theo từng vụ việc, biến gánh nặng pháp lý thành một rào cản gia nhập thị trường. Thực tiễn từ các vụ kiện như The New York Times kiện OpenAI cho thấy các doanh nghiệp khởi nghiệp nhỏ lẻ thường e ngại rủi ro, vô tình củng cố vị thế độc quyền dữ liệu cho các tập đoàn công nghệ khổng lồ có đủ tiềm lực tài chính để theo đuổi kiện tụng (Singh, 2025; Lucchi, 2023).

Trước áp lực từ sự bất định pháp lý, thị trường đang dịch chuyển từ mô hình thu thập tự do sang các thỏa thuận cấp phép thương mại song phương đất đỏ, kiến tạo nên các khu vườn dữ liệu có tường bao khóa

chặt nguồn tài nguyên trí thức. Từ thực trạng này, bài học rút ra từ Hoa Kỳ là sự linh hoạt của giới hạn quyền tác giả bắt buộc phải có sự can thiệp chủ động từ cơ quan quản lý. Việc kết hợp các quy tắc minh bạch dữ liệu từ Văn phòng Bản quyền và cơ chế giám sát chống độc quyền của Ủy ban Thương mại Liên bang là yếu tố then chốt để bảo vệ môi trường cạnh tranh lành mạnh trong kỷ nguyên số (Samuelson, 2023; Medina, 2025).

2.2. Nguyên tắc sử dụng quyền sao chép trong huấn luyện trí tuệ nhân tạo của Liên minh Châu Âu

Trong hệ thống pháp luật của Liên minh châu Âu, việc điều chỉnh hành vi sao chép trong môi trường số được thiết kế theo hướng thừa nhận tính tất yếu kỹ thuật của quá trình sao chép nhưng đặt chúng trong khuôn khổ giới hạn có điều kiện. Cơ chế này được thể hiện tại Chi thị 2001/29/EC, theo đó: “Các hành vi sao chép tạm thời được đề cập tại Điều 2, mang tính nhất thời hoặc ngẫu nhiên, là một phần không thể tách rời và thiết yếu của một quy trình công nghệ và có mục đích duy nhất nhằm cho phép: (a) việc truyền tải trong mạng giữa các bên thứ ba thông qua một trung gian; hoặc (b) việc sử dụng hợp pháp một tác phẩm hoặc đối tượng được bảo hộ khác và không có ý nghĩa kinh tế độc lập thì được miễn trừ khỏi quyền sao chép quy định tại Điều 2”. Quy định này nhận định rằng khi các bản sao này chỉ mang tính kỹ thuật trung gian và không có giá trị kinh tế độc lập, chúng không bị coi là xâm phạm quyền sao chép.

Trên nền tảng đó, Chi thị (EU) 2019/790 đã xác lập cơ chế chuyên biệt cho hoạt động khai thác văn bản và dữ liệu (TDM). Chi thị này định nghĩa TDM là “bất kỳ kỹ thuật phân tích tự động nào nhằm phân tích văn bản và dữ liệu dưới dạng kỹ thuật số để tạo ra thông tin, bao gồm nhưng không giới hạn ở các mẫu, xu hướng và mối tương quan”. Cụ thể, chi thị đã thiết lập ngoại lệ bắt buộc cho mục đích nghiên cứu khoa học khi có quyền truy cập hợp pháp và mở rộng cho các chủ thể khác, nhưng kèm theo cơ chế bảo lưu quyền, góp phần duy trì quyền kiểm soát của chủ thể quyền đối với hành vi sao chép phục vụ huấn luyện AI nhằm bảo đảm lợi ích của chủ thể quyền. Ngoài ra, Đạo luật Trí tuệ nhân tạo châu Âu năm 2024 tiếp tục củng cố cấu trúc khi yêu cầu nhà cung cấp mô hình trí tuệ nhân tạo đa năng phải tuân thủ pháp luật bản quyền của EU và cơ chế opt-out theo Chi thị DSM, đồng thời công bố bản tóm tắt đủ chi tiết về nội dung được sử dụng để huấn luyện mô hình, qua đó thiết lập cơ chế giám sát gắn với nghĩa vụ minh bạch dữ liệu.

Về mặt thực tiễn áp dụng, trong vụ Like Company v. Google Ireland (C-250/25), Tòa án Công lý Liên minh châu Âu (CJEU) đã được yêu cầu làm rõ liệu việc huấn luyện mô hình ngôn ngữ lớn trên tác phẩm được bảo hộ có cấu thành hành vi sao chép theo Điều 2 Chi thị 2001/29 và Điều 15 Chi thị DSM, đồng thời xác định việc hoạt động này có thuộc phạm vi ngoại lệ TDM tại Điều 4 DSM hay không. Phán quyết dự kiến sẽ có ý nghĩa quan trọng trong việc xác định ranh giới giữa khai thác hợp pháp và xâm phạm quyền trong bối cảnh kinh tế số.

2.3. Bài học kinh nghiệm cho Việt Nam

Thứ nhất, cần thiết lập ngoại lệ pháp lý rõ ràng, có cấu trúc và khả năng dự đoán cao đối với hành vi khai thác dữ liệu phục vụ huấn luyện AI, thay vì chỉ dựa vào các nguyên tắc chung mang tính khái quát. Kinh nghiệm từ Hoa Kỳ cho thấy việc phụ thuộc chủ yếu vào học thuyết “fair use” tuy tạo dư địa linh hoạt cho đôi mắt sáng tạo, nhưng đồng thời làm gia tăng đáng kể mức độ bất định pháp lý và chi phí tuân thủ do phụ thuộc vào diễn giải tư pháp (Lemley & Casey, 2021; Samuelson, 2023). Ngược lại, cách tiếp cận của Liên minh châu Âu với các ngoại lệ TDM được luật hóa cụ thể, kèm theo điều kiện tiếp cận hợp pháp và phạm vi áp dụng minh bạch, đã góp phần định hình một khung pháp lý ổn định hơn, cân bằng lợi ích giữa chủ thể quyền và các chủ thể phát triển công nghệ. Trong bối cảnh đó, Việt Nam cần xây dựng một ngoại lệ chuyên biệt cho hoạt động huấn luyện AI với các tiêu chí rõ ràng về mục đích sử dụng, phạm vi sao chép và điều kiện tiếp cận dữ liệu, qua đó nâng cao tính minh bạch và khả năng dự báo của hệ thống pháp luật.

Thứ hai, cần thiết kế cơ chế bảo lưu quyền (opt-out) hiệu quả nhằm bảo đảm quyền kiểm soát thực chất của chủ thể quyền đối với việc khai thác tác phẩm trong môi trường số. So với EU, nơi cơ chế opt-out đã được ghi nhận tương đối rõ ràng và có xu hướng được chuẩn hóa thông qua các công cụ kỹ thuật như metadata hoặc điều khoản sử dụng máy đọc được thì Việt Nam hiện chưa có quy định cụ thể về cách thức thể hiện và thực thi quyền này (European Parliament and Council, 2019). Bài học đặt ra là không chỉ ghi nhận về mặt nguyên tắc, mà còn phải quy định chi tiết về hình thức, hiệu lực và cơ chế thực thi của opt-out, bảo đảm tính khả thi trong thực tiễn và tránh nguy cơ hình thức hóa quyền của tác giả trong bối cảnh khai thác dữ liệu quy mô lớn.

Thứ ba, pháp luật bản quyền cần được gắn kết chặt chẽ với khung quản trị AI theo hướng tăng cường minh bạch và trách nhiệm giải trình. Trong khi EU đã bước đầu tích hợp các yêu cầu về minh bạch dữ liệu huấn luyện trong các quy định về AI, còn Hoa Kỳ chủ yếu xử lý vấn đề này thông qua thực tiễn xét xử, thì Việt Nam cần chủ động xây dựng một cơ chế phối hợp đồng bộ giữa pháp luật sở hữu trí tuệ và pháp luật điều chỉnh AI (WIPO, 2023). Cụ thể, cần xem xét thiết lập các nghĩa vụ như công bố có chọn lọc về nguồn dữ liệu huấn luyện, cơ chế kiểm toán độc lập và trách nhiệm giải trình của nhà phát triển hệ thống AI. Cách tiếp cận này không chỉ góp phần giảm thiểu rủi ro pháp lý và xung đột quyền, mà còn thúc đẩy sự phát triển bền vững và có trách nhiệm của hệ sinh thái AI trong nền kinh tế số.

Tổng thể, bài học kinh nghiệm cho Việt Nam không nằm ở việc sao chép nguyên mẫu bất kỳ mô hình nào, mà ở việc lựa chọn và điều chỉnh linh hoạt giữa tính mở của Hoa Kỳ và tính chuẩn hóa của EU, từ đó

xây dựng một khung pháp lý vừa bảo đảm thúc đẩy đổi mới sáng tạo, vừa duy trì hiệu quả bảo hộ quyền sở hữu trí tuệ trong kỷ nguyên trí tuệ nhân tạo.

3. Thực trạng và một số đề xuất hoàn thiện pháp luật về giới hạn quyền tác giả trong huấn luyện AI tại Việt Nam

3.1. Thực trạng quy định tại Việt Nam

Theo Khoản 1 Điều 20 Luật Sở hữu trí tuệ sửa đổi năm 2025, chủ sở hữu quyền tác giả có quyền độc quyền tạo ra bản sao của toàn bộ hoặc một phần tác phẩm dưới bất kỳ hình thức hay phương tiện nào. Quy định này phù hợp với cách tiếp cận phổ biến của pháp luật bản quyền quốc tế khi xem quyền sao chép là nền tảng cho việc khai thác kinh tế của tác phẩm (WIPO, 2023). Trong bối cảnh đó, việc sử dụng dữ liệu để huấn luyện các mô hình trí tuệ nhân tạo có thể làm phát sinh hành vi sao chép tác phẩm dưới dạng số, từ đó đặt ra những vấn đề pháp lý mới liên quan đến phạm vi và giới hạn của quyền này.

Một điểm đáng chú ý trong quá trình hoàn thiện pháp luật sở hữu trí tuệ là việc bổ sung khoản 5 Điều 7 Luật Sở hữu trí tuệ sửa đổi năm 2025, theo đó nhấn mạnh yêu cầu bảo đảm sự cân bằng giữa bảo hộ quyền sở hữu trí tuệ và lợi ích của xã hội, đặc biệt trong hoạt động khoa học, công nghệ và đổi mới sáng tạo. Quy định này cho thấy sự chuyển dịch trong tư duy lập pháp khi thừa nhận nhu cầu khai thác dữ liệu trong kỷ nguyên số và tạo cơ sở chính sách cho việc điều chỉnh các hoạt động sử dụng dữ liệu phục vụ phát triển trí tuệ nhân tạo (Senftleben, 2020).

Pháp luật Việt Nam hiện chủ yếu điều chỉnh quan hệ giữa chủ sở hữu quyền tác giả và người sử dụng tác phẩm theo mô hình truyền thống. Trong khi đó, hoạt động phát triển trí tuệ nhân tạo thường liên quan đến nhiều chủ thể như doanh nghiệp phát triển mô hình AI, tổ chức thu thập và xử lý dữ liệu hoặc các đơn vị khai thác thương mại kết quả của hệ thống AI. Tuy nhiên, pháp luật hiện hành chưa xác định rõ vị trí pháp lý cũng như trách nhiệm của các chủ thể trung gian công nghệ này trong trường hợp việc thu thập hoặc xử lý dữ liệu có liên quan đến các tác phẩm được bảo hộ quyền tác giả (OECD, 2021).

Trên thực tế, pháp luật Việt Nam vẫn tồn tại một số khoảng trống đáng kể trong việc điều chỉnh hoạt động này. Thứ nhất, pháp luật chưa đưa ra định nghĩa pháp lý rõ ràng về hoạt động khai thác văn bản và dữ liệu (TDM), dẫn đến khó khăn trong việc phân định giữa hành vi phân tích dữ liệu phục vụ nghiên cứu công nghệ và hành vi sao chép tác phẩm thuộc phạm vi độc quyền của chủ thể quyền (Ginsburg, 2021). Thứ hai, pháp luật chưa có quy định cụ thể về tính hợp pháp của các bản sao tạm thời phát sinh trong quá trình xử lý dữ liệu bằng thuật toán, mặc dù các bản sao này thường chỉ mang tính kỹ thuật trung gian và không có ý nghĩa kinh tế độc lập (Margoni và Kretschmer, 2022). Ngoài ra, pháp luật Việt Nam cũng chưa thiết lập cơ chế bảo lưu quyền (opt-out) cho phép tác giả từ chối việc sử dụng tác phẩm của mình trong các tập dữ liệu

huấn luyện trí tuệ nhân tạo, đồng thời chưa có nghĩa vụ minh bạch rõ ràng về nguồn dữ liệu huấn luyện của các nhà phát triển hệ thống AI (OECD, 2021).

3.2. Một số đề xuất hoàn thiện pháp luật Việt Nam

3.2.1 Khuyến nghị về khung pháp lý

Quán triệt định hướng chiến lược của Đảng tại Nghị quyết số 52 của Bộ Chính trị, việc hoàn thiện thể chế sở hữu trí tuệ phải đóng vai trò bệ phóng cho công cuộc chuyển đổi số quốc gia. Dưới lăng kính kinh tế học pháp luật, hệ thống bản quyền truyền thống đang đối diện rủi ro thất bại thị trường do chi phí giao dịch khổng lồ khi phải đàm phán cấp phép với hàng triệu chủ thể quyền, tạo ra bi kịch chống tài sản chung làm triệt tiêu thặng dư xã hội (Barker, 2019). Mặc dù khoản 5 Điều 7 Luật Sở hữu trí tuệ sửa đổi năm 2025 đã thiết lập ngoại lệ khai thác dữ liệu cho trí tuệ nhân tạo, nhưng quy phạm khung này cần được pháp điển hóa chi tiết thông qua một Nghị định hướng dẫn thi hành chuyên biệt của Chính phủ. Trước hết, Nghị định này cần bổ sung quy định nội luật hóa cơ chế miễn trừ tuyệt đối đối với các bản sao tạm thời mang tính kỹ thuật. Cụ thể, cần quy định rõ các bản sao nhất thời phát sinh trong bộ nhớ đệm trong quá trình thu thập dữ liệu và tiền huấn luyện thuật toán có giá trị biên bằng không và không có ý nghĩa kinh tế độc lập, do đó tự động được loại trừ khỏi phạm vi độc quyền sao chép quy định tại Điều 20 Luật Sở hữu trí tuệ. Việc giải phóng nhà phát triển khỏi nghĩa vụ xin phép cho các hành vi kỹ thuật thuần túy này sẽ ngăn chặn tổn thất vô ích cho toàn xã hội.

Bên cạnh đó, để giải quyết mâu thuẫn giữa nhu cầu khai thác dữ liệu và rủi ro tước đoạt tài sản sáng tạo, Nghị định hướng dẫn thi hành Luật Sở hữu trí tuệ cần bổ sung một điều khoản quy định rõ về cơ chế bảo lưu quyền. Theo định lý Coase, cơ chế từ chối quyền chính là công cụ phục hồi quy tắc tài sản trong không gian số. Quy định mới cần trao quyền pháp lý cho chủ sở hữu từ chối việc bị các hệ thống máy học khai thác dữ liệu thông qua các biện pháp công nghệ như siêu dữ liệu máy có thể đọc được hoặc các giao thức kỹ thuật được nhúng trực tiếp vào tệp tin (Maidanik, 2020). Yêu cầu lập pháp này sẽ buộc các doanh nghiệp công nghệ phải rà soát tín hiệu từ chối bằng thuật toán tự động, từ đó chuyển dịch từ mô hình chiếm dụng dữ liệu miễn phí sang việc đàm phán các thỏa thuận cấp phép thương mại tự nguyện nhằm duy trì động lực kinh tế cốt lõi cho hoạt động sáng tạo.

Về thực thi pháp luật, sự tồn tại của các ngoại lệ quy định tại Điều 7 Luật Sở hữu trí tuệ phải chịu sự chi phối tối cao của nguyên tắc kiểm tra ba bước theo Công ước Berne. Để tránh sự tùy tiện trong áp dụng, Tòa án Nhân dân Tối cao cần ban hành Nghị quyết hướng dẫn xét xử, lượng hóa tiêu chí định tính không gây ảnh hưởng bất hợp lý thành tiêu chuẩn định lượng về ngoại ứng tiêu cực dựa trên học thuyết thay thế thị trường. Cụ thể, cơ quan tài phán cần xác định rõ nếu mô hình trí tuệ nhân tạo tạo sinh được huấn luyện

để tạo ra sản phẩm cạnh tranh trực tiếp và giành giật thị phần của chính tác giả gốc thì hành vi đó làm xói mòn khả năng thu hồi chi phí chìm và bị loại trừ khỏi ngoại lệ hợp pháp (WIPO, 2023). Đồng thời, để giải quyết sự bất cân xứng về chi phí tố tụng trong các vụ án dữ liệu lớn, Quốc hội cần xem xét sửa đổi Điều 205 Luật Sở hữu trí tuệ về căn cứ xác định mức bồi thường thiệt hại, bổ sung trực tiếp chế định bồi thường theo luật định. Cơ chế này trao quyền cho hội đồng xét xử ấn định một khoản tiền bồi thường trừng phạt cố định trên mỗi vi phạm mà không yêu cầu nguyên đơn chứng minh thiệt hại vật chất cụ thể, qua đó triệt tiêu động cơ vi phạm có tính toán và cung cấp công cụ thực thi hiệu quả cho các tác giả nhỏ lẻ.

3.2.2. Khuyến nghị về thực tiễn thực thi

Để chuyển hóa quy phạm pháp luật vào thực tiễn, việc kiện toàn thiết chế thực thi đòi hỏi một hệ sinh thái quản trị tối thiểu hóa chi phí giám sát. Về mặt kiến tạo hạ tầng, Nhà nước cần đầu tư xây dựng các trung tâm dữ liệu lớn và cơ sở hạ tầng siêu máy tính cấp quốc gia bám sát định hướng tại Quyết định số 127 của Thủ tướng Chính phủ. Dưới góc độ kinh tế học công cộng, các công nghệ bảo vệ bản quyền tiên tiến như thủy vân kỹ thuật số và công nghệ chuỗi khối cần được xem là hàng hóa công cộng thiết yếu giúp toàn bộ nền kinh tế dễ dàng truy vết hành vi sao chép trái phép mà không phải gánh chịu rào cản tài chính.

Về phương diện năng lực quản lý, cơ quan chức năng cần thiết lập cơ chế giám sát chặt chẽ tính minh bạch của chuỗi cung ứng dữ liệu. Tiếp thu chuẩn mực lập pháp tiên tiến, Việt Nam cần ban hành quy định bắt buộc các doanh nghiệp cung cấp mô hình trí tuệ nhân tạo phải thiết lập hồ sơ lưu vết và công khai minh bạch nguồn dữ liệu đầu vào. Nghĩa vụ báo cáo này không chỉ cung cấp chứng cứ kỹ thuật cho cơ quan thanh tra mà còn giúp các tác giả khắc phục sự bất cân xứng thông tin khi chứng minh thiệt hại trước cơ quan tài phán.

Cuối cùng, nhằm giải quyết triệt để bài toán chi phí giao dịch khổng lồ trong đàm phán cấp phép đơn lẻ, cơ quan quản lý cần phát huy vai trò trung gian và trao thêm quyền năng pháp lý cho các tổ chức đại diện tập thể quyền tác giả thông qua cơ chế cấp phép tập thể mở rộng. Thay vì để hàng ngàn doanh nghiệp tự đàm phán dẫn đến sự bế tắc của thị trường, các tổ chức đại diện sẽ đứng ra thiết lập các hợp đồng cấp phép trọn gói cho khối lượng dữ liệu huấn luyện khổng lồ. Sự phối hợp đồng bộ này dỡ bỏ rào cản chi phí giao dịch cho các doanh nghiệp khởi nghiệp, đồng thời bảo đảm cơ chế phân phối thù lao minh bạch và nội hóa lợi ích kinh tế về cho cộng đồng sáng tạo.

Kết luận

Hoạt động huấn luyện trí tuệ nhân tạo đặt ra những thách thức mới đối với phạm vi và giới hạn của quyền sao chép trong pháp luật bản quyền. Kinh nghiệm của Hoa Kỳ và Liên minh châu Âu cho thấy việc thiết lập các cơ chế ngoại lệ phù hợp có ý nghĩa quan trọng trong việc cân bằng giữa việc bảo hộ quyền tác

giả và việc thúc đẩy đổi mới sáng tạo. Đối với Việt Nam, việc hoàn thiện pháp luật cần hướng tới xây dựng quy định rõ ràng về khai thác dữ liệu, bản sao tạm thời và cơ chế bảo lưu quyền, đồng thời tăng cường minh bạch nguồn dữ liệu huấn luyện. Những cải cách này sẽ góp phần tạo dựng khung pháp lý ổn định, hỗ trợ sự phát triển của trí tuệ nhân tạo và hệ sinh thái sáng tạo trong nền kinh tế số.

TÀI LIỆU THAM KHẢO

Authors Guild v. Google, Inc., 804 F.3d 202 (2d Cir. 2015).

Barker, G. (2019). The law and economics of copyright law and copyright exceptions, limitations, and immunities. *ERPN: Copyright & Trademark (Sub-Topic)*. Available at: <https://doi.org/10.2139/ssrn.3315619>.

Berne Convention for the Protection of Literary and Artistic Works (1971 revision).

Bộ Chính trị. (2019). *Nghị quyết số 52-NQ/TW về một số chủ trương, chính sách chủ động tham gia cuộc Cách mạng công nghiệp lần thứ tư*.

European Parliament and Council of the European Union. (2001). *Directive 2001/29/EC on the harmonisation of certain aspects of copyright and related rights in the information society*, 22 May 2001.

European Parliament and Council of the European Union. (2019). *Directive (EU) 2019/790 on copyright and related rights in the Digital Single Market and amending Directives 96/9/EC and 2001/29/EC*, 17 April 2019.

European Parliament and Council of the European Union. (2024). *Regulation (EU) 2024/1689 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act)*, 13 June 2024.

Ginsburg, J.C. (2021). Copyright and control over new uses of works: AI, text and data mining, and copyright law. New York: Columbia University School of Law.

Ginsburg, J.C. (2021). Fair use in the digital age. *Columbia Journal of Law & the Arts*, Vol 44 no 3, pp. 315–340.

Jain, V. & Verma, R. (2025). AI and IP: Rethinking ownership, creativity and rights. *International Journal for Multidisciplinary Research*. Available at: <https://doi.org/10.36948/ijfmr.2025.v07i02.42466>

Lucchi, N. (2023). ChatGPT: A case study on copyright challenges for generative artificial intelligence systems. *European Journal of Risk Regulation*, 15, pp. 602–624. <https://doi.org/10.1017/err.2023.59>.

Marcel, M.T.J.J. (2025). AI and copyright: The training of general purpose AI. European Parliamentary Research Service. Available at: [https://www.europarl.europa.eu/thinktank/en/document/EPRS_ATA\(2025\)769585](https://www.europarl.europa.eu/thinktank/en/document/EPRS_ATA(2025)769585).

Margoni, T. & Kretschmer, M. (2022). Text and data mining: Copyright and data access for artificial intelligence. Oxford: Oxford University Press.

Margoni, T. & Kretschmer, M. (2022). The text and data mining exception in the EU: Legal and economic perspectives. *Journal of Intellectual Property Law & Practice*, 17(2), pp. 85–95.

Medina, A. (2025). Applying copyright law to artificial intelligence. *Journal of Science Policy & Governance*, 26(1). <https://doi.org/10.38126/jspg260105>.

Neill, A., Lee, E. & Thomas, J. (2025). A framework for applying copyright law to the training of textual generative artificial intelligence. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.5362231>.

OECD. (2021). Artificial intelligence, data and intellectual property. Paris: OECD Publishing.

OECD. (2021). *Artificial intelligence, intellectual property and innovation*. Paris: OECD Publishing.

Quốc hội Việt Nam. (2025). *Luật Sở hữu trí tuệ sửa đổi*.

Samuelson, P. (2023). Generative AI meets copyright. *Science*, 381(6654), pp. 158–161. <https://doi.org/10.1126/science.adi0656>.

Senftleben, M. (2020). *Copyright, data mining and artificial intelligence: Emerging legal challenges*. Amsterdam: University of Amsterdam.

Senftleben, M. (2020). Copyright, limitations and the three-step test. The Hague: Kluwer Law International.

Sidebari, L. (2023). Generative artificial intelligence and copyright law.

Singh, A. (2025). Artificial intelligence and copyright: Navigating limitations and unlocking challenges. *International Journal for Multidisciplinary Research*, 7(2). <https://doi.org/10.36948/ijfmr.2025.v07i02.39671>.

Škiljić, A. (2021). When art meets technology or vice versa: Key challenges at the crossroads of AI-generated artworks and copyright law. *IIC – International Review of Intellectual Property and Competition Law*, 52, pp. 1338–1369. <https://doi.org/10.1007/s40319-021-01119-w>.

Thủ tướng Chính phủ. (2021). *Quyết định số 127/QĐ-TTg ngày 26 tháng 1 năm 2021 ban hành Chiến lược quốc gia về nghiên cứu, phát triển và ứng dụng trí tuệ nhân tạo đến năm 2030*.

Torrance, A. & Tomlinson, B. (2023). Training is everything: Artificial intelligence, copyright, and fair training. *ArXiv*. <https://doi.org/10.48550/arxiv.2305.03720>.

Trịnh, Q. (2025). Ngoại lệ quyền tác giả đối với huấn luyện trí tuệ nhân tạo sinh – kinh nghiệm một số quốc gia và gợi mở cho Việt Nam. <https://tapchicongthuong.vn/ngoai-le-quyen-tac-gia-doi-voi-huan-luyen-tri-tue-nhan-tao-sinh-kinh-nghiem-mot-so-quoc-gia-va-goi-mo-cho-viet-nam-219760.htm>.

United States Copyright Act

WIPO. (2023). Generative artificial intelligence and copyright. Geneva: World Intellectual Property Organization.

WIPO. (2023). *Intellectual property and artificial intelligence*. Geneva: World Intellectual Property Organization.