Bảo vệ liêm chính học thuật trong hoạt động biên tập và xuất bản trước thách thức của trí tuệ nhân tạo

PGS, TS. NGUYỄN THỊ TRƯỜNG GIANG
Học viện Báo chí và Tuyên truyền; Email: truonggiangbmdt@yahoo.com.vn

Nhận ngày 15 tháng 12 năm 2025; chấp nhận đăng tháng 01 năm 2026.

Tóm tắt: *Làn sóng AI tạo sinh đang làm thay đổi căn bản cách tri thức được hình thành và công bố, đồng thời kéo theo những rủi ro mới đối với liêm chính học thuật trong biên tập, xuất bản. Bài viết chỉ ra rằng gian lận không còn dừng ở sao chép dễ nhận biết, mà chuyển sang các thủ thuật ngụy trang tinh vi, từ tái diễn đạt nhằm né phần mềm phát hiện, đến tạo dữ liệu và trích dẫn "có vẻ đúng" nhưng sai hoặc không tồn tại. Đáng lo hơn, mô hình "sản xuất công nghiệp" bài báo và thao túng phản biện khiến quy trình thẩm định truyền thống trở nên quá tải. Trong bối cảnh Việt Nam còn độ trễ về quy định và năng lực số, tác giả đề xuất một gói giải pháp đồng bộ giúp hệ thống xuất bản vừa thích ứng công nghệ vừa giữ vững chuẩn mực khoa học.*
Từ khóa: *Trí tuệ nhân tạo; AI; AI tạo sinh; liêm chính học thuật; liêm chính học thuật trong xuất bản.*

Abstract: *The wave of generative artificial intelligence (AI) is fundamentally transforming how knowledge is produced and disseminated, while simultaneously introducing new risks to academic integrity in editorial and publishing practices. This paper argues that misconduct no longer takes the form of easily detectable plagiarism, but has shifted toward increasingly sophisticated forms of concealment, ranging from paraphrasing designed to avoid detection software to the fabrication of data and citations that appear plausible but are incorrect or entirely nonexistent. More alarmingly, the emergence of an "industrial-scale" model of article production and the manipulation of peer review processes have placed traditional evaluation mechanisms under severe strain. In a context where Vietnam still faces regulatory and digital capacity gaps, the author proposes a comprehensive set of solutions to help the publishing system both adapt to technological change and uphold scientific standards.*
Keywords: *Artificial intelligence; AI; generative AI; academic integrity; academic integrity in publishin.*

1. Mở đầu

Sự bùng nổ của cuộc Cách mạng công nghiệp lần thứ tư, với tâm điểm là các mô hình ngôn ngữ lớn và trí tuệ nhân tạo tạo sinh, đang tạo ra những bước ngoặt căn bản trong phương thức lao động trí óc. Trong hoạt động nghiên cứu khoa học, công nghệ này không chỉ đóng vai trò là công cụ hỗ trợ xử lý dữ liệu mà đang dần trở thành một tác nhân tham gia trực tiếp vào quá trình sáng tạo tri thức. Tuy nhiên, bên cạnh những giá trị tích cực về hiệu suất, sự thâm nhập sâu rộng của trí tuệ nhân tạo cũng đặt nền tảng học thuật trước những thách thức phi truyền thống và chưa từng có tiền lệ.

Thực tiễn cho thấy, các hành vi vi phạm liêm chính học thuật đang có sự chuyển dịch phức tạp: từ những hình thức sao chép thủ công dễ dàng nhận diện sang các quy trình gian lận tinh vi dựa trên thuật toán. Sự xuất hiện của các công cụ "làm sạch văn bản", hiện tượng ngụy tạo dữ liệu hay sự hình thành các mạng lưới gian lận có tổ chức đang làm xói mòn nghiêm trọng các giá trị cốt lõi của khoa học. Đáng quan ngại hơn, tại Việt Nam, công tác quản trị hoạt động xuất bản vẫn còn tồn tại độ trễ nhất định về cơ chế so với tốc độ phát triển của công nghệ, tạo ra

những khoảng trống rủi ro trong kiểm soát chất lượng công bố.

Trong bối cảnh đó, bài viết tập trung phân tích và nhận diện các nguy cơ mới dựa trên dữ liệu thực chứng quốc tế giai đoạn 2024 - 2025. Trên cơ sở đó, bài viết đề xuất hệ thống giải pháp đồng bộ từ hoàn thiện thể chế, đổi mới quy trình kỹ thuật đến nâng cao năng lực con người, nhằm góp phần kiến tạo một môi trường học thuật minh bạch và trách nhiệm.

2. Nội dung

2. 1. Những thách thức phi truyền thống đối với liêm chính học thuật dưới tác động của trí tuệ nhân tạo

Sự biến đổi của hành vi đạo văn đến công nghệ "rửa văn bản"

Thách thức đầu tiên và tinh vi nhất chính là sự thay đổi về chất của hành vi gian lận: sự chuyển dịch từ đạo văn sao chép đơn thuần sang quy trình "rửa văn bản" công nghệ cao. Hiện nay, giới học thuật đang chứng kiến sự trỗi dậy của một hệ sinh thái các công cụ được thiết kế chuyên biệt để "nhân hóa" văn bản máy (như Walter Writes, BypassGPT hay HIX Bypass). Các công cụ này công khai quảng cáo khả năng viết lại nội dung do trí tuệ nhân tạo tạo ra nhằm vô hiệu hóa các hệ thống kiểm duyệt uy tín như Turnitin hay GPTZero.

Hiệu quả của các công cụ này đã được chứng minh qua nhiều nghiên cứu thực nghiệm. Các kỹ thuật diễn giải lại (paraphrasing) dựa trên mô hình ngôn ngữ lớn đang đặt ra thách thức nghiêm trọng đối với khả năng phát hiện của máy móc. Cụ thể, Krishna và cộng sự (2023) ghi nhận tỷ lệ phát hiện chính xác của DetectGPT giảm mạnh từ 70,3% xuống còn 4,6% sau khi văn bản được xử lý lại. Đáng chú ý hơn, kỹ thuật "diễn giải lại đối kháng" (adversarial paraphrasing) thậm chí có thể làm giảm hiệu suất nhận diện trên hệ thống Fast-DetectGPT tới 98,96% (Cheng và cộng sự, 2025).

Hệ quả của xu hướng này được phản ánh rõ rệt qua sự gia tăng số lượng các công bố khoa học bị rút lại. Phân tích dữ liệu từ PubMed cho thấy các bài báo liên quan đến trí tuệ nhân tạo bị rút bỏ đã đạt đỉnh điểm vào năm 2023 với 667 trường hợp, trong đó có 238 trường hợp được xác định cụ thể là do vi phạm đạo đức trong sử dụng AI (Kocyigit và cộng sự, 2025). Những số liệu này cho thấy các phương thức kiểm soát truyền thống đang trở nên mong manh trước sự phát triển của công nghệ gian lận thế hệ mới.

Nguy cơ sai lệch chuẩn mực khoa học từ hiện tượng "ảo giác" thông tin và ngụy tạo dữ liệu

Không chỉ dừng lại ở vấn đề câu chữ, thách thức nguy hiểm hơn nằm ở tính chính xác của dữ liệu. Do cơ chế hoạt động dựa trên xác suất thống kê thay vì tư duy kiểm chứng, trí tuệ nhân tạo thường mắc phải lỗi "ảo giác", tức là tự động bịa đặt ra các thông tin sai lệch nhưng trình bày dưới hình thức trông có vẻ hoàn hảo nhưng thực chất lại chứa đầy các bẫy thông tin. Thực nghiệm cho thấy các mô hình ngôn ngữ lớn có xu hướng trộn lẫn các nguồn tài liệu có thật với các thông tin ngụy tạo, khiến gần một nửa số lượng tham chiếu trong bài viết (khoảng 47%) chứa các lỗi nghiêm trọng như sai lệch ngữ cảnh, sai mã định danh (DOI), hoặc thậm chí dẫn nguồn đến những công trình hoàn toàn không tồn tại (Májovský và cộng sự, 2023)".

Các nghiên cứu thực chứng đã chỉ ra một con số đáng báo động. Nghiên cứu của Bhattacharyya và cộng sự (2023) trên các bài báo y tế cho thấy ChatGPT tạo ra tỷ lệ trích dẫn giả lên đến 47%, và chỉ có 7% trích dẫn là vừa chính xác vừa tồn tại thực tế.

Sự nguy hại của vấn đề này nằm ở chỗ các trích dẫn giả mạo thường được ngụy tạo rất tinh vi với đầy đủ tên tác giả thật và chỉ số định danh nhìn như chuẩn xác, ví dụ như trường hợp trí tuệ nhân tạo tự bịa ra bài báo của nhóm tác giả có thật. Nguy hiểm hơn, các công cụ tìm kiếm học thuật hiện đại như Elicit hay Perplexity trong nhiều trường hợp đã vô tình tham chiếu cả những bài báo đã bị gỡ bỏ, tạo ra một vòng lặp thông tin sai lệch.

Khi những trích dẫn và dữ liệu "giả" này lọt qua khâu biên tập và được xuất bản, chúng sẽ được các nhà nghiên cứu khác kế thừa, tạo ra một hiệu ứng domino sai lệch. Thực tế năm 2023 đã ghi nhận hơn 13.000 bài báo bị rút - con số cao nhất trong lịch sử, trong đó một phần không nhỏ liên quan đến việc lạm dụng AI để giả mạo dữ liệu thực nghiệm và hình ảnh y sinh (Stauffer, 2025).

Sự gia tăng các hình thức gian lận có tổ chức và thách thức đối với quy trình thẩm định chuyên môn

Sự tiếp sức của AI đã biến các "lò sản xuất bài báo" thành những cỗ máy gian lận quy mô lớn. Thay

vì viết thủ công, các tổ chức này tận dụng AI để sản xuất hàng loạt nghiên cứu có vẻ ngoài hoàn chỉnh nhưng hoàn toàn rỗng về giá trị khoa học. Để qua mặt các tạp chí, họ thậm chí còn thao túng cả quy trình đánh giá đồng cấp.

Dữ liệu phân tích từ Retraction Watch của Martínez-Rojas và Zahn-Muñoz (2025) cho thấy một sự thay đổi rõ rệt trong các lý do rút bài. Nếu như giai đoạn 2015 - 2019, lỗi phổ biến nhất là đạo văn truyền thống (chiếm 21,6%), thì đến giai đoạn 2020-2024, các hình thức gian lận công nghệ cao đã chiếm ưu thế. Cụ thể, tỷ lệ bài bị rút do xuất phát từ "lò sản xuất bài báo" (30,1%) và "nội dung tạo ngẫu nhiên bởi máy" (23,3%) đã tăng vọt, trở thành những nguyên nhân hàng đầu.

Sự thiếu hụt các quy chuẩn kiểm soát và độ trễ trong cơ chế quản trị hoạt động biên tập tại Việt Nam

Lý do rút bài	Giai đoạn 2015-2019	Giai đoạn 2020-2024	Xu hướng
Đạo văn truyền thống	21.6%	6.8%	Giảm mạnh
Lò sản xuất bài báo	-	30.1%	Mới nổi & Tăng nhanh
Nội dung tạo bởi AI/Máy	-	23.3%	Mới nổi & Tăng nhanh

Bảng 1. Sự chuyển dịch cơ cấu các nguyên nhân dẫn đến việc rút bài báo khoa học: So sánh giai đoạn 2015-2019 và 2020-2024(Nguồn: Tổng hợp từ dữ liệu phân tích của Martínez-Rojas và Zahn-Muñoz, 2025)

Thách thức bao trùm lên toàn bộ hệ thống xuất bản trong nước hiện nay là sự bất cập của các quy chế hoạt động trước tốc độ phát triển vũ bão của công nghệ. Trên bình diện quốc tế, mặc dù chưa có một bộ luật thống nhất, nhưng cộng đồng xuất bản học thuật đã nhanh chóng hình thành các quy chuẩn đạo đức chung. Phân tích của Perkins và Roe (2024) trên các nhà xuất bản hàng đầu thế giới cho thấy đã xuất hiện sự đồng thuận ở một số nguyên tắc cốt lõi: vai trò tác giả của con người là tối thượng (AI không được đứng tên tác giả), và việc sử dụng công cụ AI tạo sinh dù được chấp nhận cho mục đích hỗ trợ, bắt buộc phải đảm bảo tính minh bạch tuyệt đối. Nghiên cứu cũng nhấn mạnh rằng các chính sách cần duy trì sự linh hoạt và phải luôn có sự giám sát của con người trước những hạn chế cố hữu của công nghệ này.

Đối sánh với các tiêu chuẩn quốc tế nêu trên, công tác quản trị tại Việt Nam đang bộc lộ độ trễ nhất định về mặt chính sách. Hiện nay, các hướng dẫn chi tiết về liêm chính học thuật liên quan đến trí tuệ nhân tạo vẫn chưa được luật hóa cụ thể hoặc cập nhật đồng bộ trong quy chế của nhiều tạp chí khoa học trong nước. Sự thiếu hụt các quy định kỹ thuật về việc trích dẫn AI hay giới hạn can thiệp của thuật toán đang tạo ra những "khoảng xám" về đạo đức nghiên cứu. Điều này đặt ra thách thức lớn cho các hội đồng biên tập trong việc phân định ranh giới giữa việc "sử dụng công cụ hỗ trợ hợp lệ" và hành vi "lệ thuộc công nghệ", gây rủi ro cho uy tín và khả năng hội nhập của các công bố khoa học trong nước.

Giới hạn của các giải pháp kỹ thuật và hạn chế về năng lực số của đội ngũ biên tập

Song hành với những khoảng trống về pháp lý là sự bất cân xứng nghiêm trọng giữa năng lực "tấn công" của công nghệ tạo sinh và năng lực "phòng thủ" của các công cụ kiểm soát hiện hữu. Thực tế cho thấy, giới xuất bản đang bị cuốn vào một cuộc đua không cân sức khi tốc độ hoàn thiện của các mô hình ngôn ngữ lớn luôn đi trước một bước so với các giải pháp phát hiện gian lận. Các phần mềm rà soát phổ biến hiện nay thường xuyên rơi vào trạng thái "báo động giả" hoặc bỏ lọt vi phạm khi đối mặt với các văn bản đã qua xử lý. Nghiên cứu của Liang và cộng sự (2023) đăng trên tạp chí *Patterns* đã chỉ ra một thực tế đáng quan ngại: các công cụ phát hiện GPT có tỷ lệ sai sót rất cao khi phân tích văn bản của những tác giả không sử dụng tiếng Anh như ngôn ngữ mẹ đẻ (non-native English writers). Cụ thể, hơn một nửa số bài luận của nhóm tác giả này bị phân loại nhầm là do AI viết, gây ra những rủi ro oan sai nghiêm trọng và làm gia tăng sự bất bình đẳng trong môi trường học thuật.

Bên cạnh sự giới hạn của công cụ kỹ thuật, yếu tố con người - chốt chặn cuối cùng của quy trình xuất bản cũng đang bộc lộ những hạn chế về năng lực số. Trước làn sóng ồ ạt các bài báo được sản xuất công nghiệp bởi AI, đội ngũ biên tập viên và chuyên gia phản biện đang rơi vào tình trạng quá tải cục bộ. Phần lớn các chuyên gia hiện nay được đào tạo trong môi trường học thuật truyền thống, do đó còn thiếu sự nhạy bén cần thiết để nhận diện những dấu hiệu bất thường tinh vi như các "cụm từ bị tra tấn" hay các lỗi

logic đặc trưng của máy móc. Sự thiếu hụt các kỹ năng kiểm chứng dữ liệu số khiến cho quy trình thẩm định chủ yếu vẫn dựa vào niềm tin và kinh nghiệm cảm tính, tạo ra những kẽ hở lớn để các ấn phẩm kém chất lượng lọt lưới và xuất bản.

2.2. Một số giải pháp đảm bảo liêm chính học thuật trong hoạt động biên tập và xuất bản hiện nay

2.2.1. Hoàn thiện cơ chế quản lý biên tập và thiết lập chuẩn mực về tính minh bạch và trách nhiệm giải trình

Để giải quyết các thách thức nêu trên, giải pháp căn cơ đầu tiên nằm ở việc các tạp chí phải lấp đầy khoảng trống trong quy chế hoạt động. Dựa trên các khuyến nghị quốc tế, quy trình quản trị mới cần được thiết lập dựa trên ba trụ cột chính sau đây:

Thứ nhất, kiên quyết xác lập nguyên tắc "chỉ có con người mới đáp ứng đủ điều kiện để đứng tên tác giả". Theo các hướng dẫn từ Ủy ban Quốc tế các Tổng biên tập Tạp chí Y học (ICMJE, cập nhật tháng 1 năm 2024) và Hiệp hội Biên tập viên Y khoa Thế giới (WAME, cập nhật tháng 5 năm 2023), các tạp chí cần kiên quyết áp dụng nguyên tắc "trí tuệ nhân tạo không có tư cách tác giả". Cơ sở của nguyên tắc này rất rõ ràng: máy móc không thể chịu trách nhiệm pháp lý về nội dung công bố, không thể xác nhận các xung đột lợi ích và không thể ký phê duyệt bản thảo cuối cùng. Do đó, mọi hành vi liệt kê AI là tác giả hoặc đồng tác giả đều bị coi là vi phạm quy chế xuất bản.

Thứ hai, thể lệ hóa yêu cầu về minh bạch thông tin. Thay vì cấm đoán cực đoan, tạp chí cần yêu cầu tác giả công khai việc sử dụng công cụ hỗ trợ. Cụ thể, các tạp chí cần bổ sung vào thể lệ gửi bài quy định bắt buộc tác giả phải có tuyên bố trong phần *"Phương pháp nghiên cứu"* hoặc *"Lời cảm ơn"*. Nội dung khai báo cần chi tiết, bao gồm: tên công cụ (ví dụ: ChatGPT-4), phiên bản, mục đích sử dụng (như xử lý số liệu, chỉnh sửa ngữ pháp hay tóm tắt văn bản) và các câu lệnh quan trọng đã dùng. Điều này giúp minh bạch hóa quá trình nghiên cứu, giúp biên tập viên và độc giả phân định rõ đâu là đóng góp của con người, đâu là sản phẩm của thuật toán.

Thứ ba, ràng buộc chặt chẽ trách nhiệm giải trình của con người. Theo các chỉ dẫn từ Viện Y tế Quốc gia Hoa Kỳ (NIH) năm 2025. Các tạp chí cần yêu cầu tác giả ký cam kết chịu trách nhiệm hoàn toàn về độ chính xác của công trình, kể cả những phần có sự hỗ trợ của máy móc. Điều này có nghĩa là tác giả phải trực tiếp rà soát từng nguồn trích dẫn để xóa bỏ các tham chiếu giả và đảm bảo dữ liệu không bị sai lệch. Việc chuẩn hóa mẫu cam kết này ngay từ khâu nộp bài sẽ tạo ra rào cản cần thiết, giúp ngăn chặn tình trạng lạm dụng công nghệ trong công bố khoa học.

Cuối cùng, thiết lập khung chế tài xử lý vi phạm phân tầng. Một quy định chỉ có hiệu lực khi đi kèm chế tài răn đe. Tạp chí cần phân biệt rõ giữa "lỗi kỹ thuật" (như quên khai báo công cụ sửa lỗi chính tả) và "gian lận học thuật" (như dùng AI để ngụy tạo dữ liệu hoặc viết toàn bộ bài báo). Đối với các vi phạm nghiêm trọng liên quan đến tính nguyên bản, cần áp dụng các biện pháp xử lý mạnh như: gỡ bỏ bài báo, công bố danh tính người vi phạm trên hệ thống cảnh báo và từ chối nhận bài có thời hạn.

2.2.2. Đổi mới quy trình thẩm định thông qua cơ chế rà soát kép và tiêu chuẩn dữ liệu mở

Trước sự gia tăng của các hành vi gian lận sử dụng công nghệ cao, quy trình biên tập truyền thống vốn dựa nhiều vào niềm tin mặc định đang bộc lộ những lỗ hổng nghiêm trọng. Để khắc phục tình trạng này, các tạp chí cần chuyển đổi sang mô hình kiểm soát đa tầng, lấy cơ chế rà soát kép và minh bạch hóa dữ liệu làm trọng tâm.

Thứ nhất, thiết lập quy trình "kiểm soát kép" trong khâu tiền kiểm. Quy trình này đòi hỏi sự phối hợp chặt chẽ giữa công cụ kỹ thuật và thẩm định chuyên môn, cụ thể:

(1) Về rà soát kỹ thuật: Sử dụng các công cụ chuyên dụng (như *Originality.ai* hoặc các tính năng tích hợp của *Turnitin*) để sàng lọc bước đầu. Tuy nhiên, Hội đồng biên tập cần quán triệt nguyên tắc: kết quả từ phần mềm chỉ mang tính chất cảnh báo, không phải là kết luận cuối cùng. Việc này nhằm tránh các sai sót "dương tính giả", gây oan sai cho các tác giả, đặc biệt là những người không sử dụng tiếng Anh như ngôn ngữ mẹ đẻ.

(2) Về thẩm định chuyên môn: Đây là chốt chặn quan trọng để phát hiện hiện tượng "ảo giác" của trí tuệ nhân tạo. Biên tập viên cần thực hiện quy trình "thẩm định xác suất" đối với danh mục tài liệu tham khảo. Cụ thể, cần tra cứu ngẫu nhiên từ 10 - 20% các trích dẫn trên các cơ sở dữ liệu định danh uy tín (như

Scopus, PubMed) để phát hiện các tài liệu giả mạo hoặc sai lệch chỉ số DOI - những dấu hiệu đặc trưng của văn bản do máy tạo lập mà mắt thường dễ bỏ qua.

Thứ hai, áp dụng tiêu chuẩn "dữ liệu mở" để ngăn chặn hành vi ngụy tạo. Đây được xem là giải pháp căn cơ để bảo vệ tính chân thực của khoa học. Tham khảo mô hình quản trị của các hệ thống xuất bản hàng đầu như *Nature Portfolio* hay *Science*, các tạp chí khoa học trong nước cần xây dựng lộ trình bắt buộc tác giả cung cấp dữ liệu gốc đi kèm bài báo.

(1) Đối với các nghiên cứu định lượng, tác giả phải nộp kèm các tệp dữ liệu thô (bảng tính, mã nguồn phân tích, hình ảnh gốc chưa qua xử lý) hoặc cung cấp đường dẫn lưu trữ trên các kho dữ liệu mở uy tín (như *Zenodo, Figshare*).

(2) Khi dữ liệu nguồn được công khai, khả năng tác giả sử dụng công nghệ để làm sai lệch hoặc ngụy tạo kết quả nghiên cứu sẽ bị triệt tiêu đáng kể. Đồng thời, cơ chế này tạo điều kiện thuận lợi cho cộng đồng khoa học thực hiện việc tái lập và kiểm chứng kết quả sau công bố.

Thứ ba, kích hoạt cơ chế "hậu kiểm" và giám sát cộng đồng. Nhận thức rõ tính giới hạn của các biện pháp kỹ thuật, quy trình đảm bảo liêm chính không nên chấm dứt tại thời điểm bài báo được xuất bản. Các tạp chí cần thiết lập kênh tiếp nhận phản hồi để cộng đồng khoa học và độc giả có thể báo cáo các nghi vấn về dữ liệu hoặc văn phong bất thường. Đối với các bài báo đã xuất bản nhưng sau đó bị phát hiện có dấu hiệu vi phạm do sử dụng trí tuệ nhân tạo, tạp chí cần thực hiện quy trình tái thẩm định độc lập. Trong trường hợp xác định có vi phạm, cần kiên quyết thực hiện việc thu hồi bài báo. Hành động này cần được nhìn nhận là biểu hiện của tinh thần trách nhiệm và sự cam kết về chất lượng khoa học, thay vì là sự yếu kém trong công tác quản lý.

2.2.3. Nâng cao năng lực số cho đội ngũ biên tập viên và xây dựng văn hóa liêm chính trong cộng đồng học thuật

Công nghệ và thể chế chỉ có thể phát huy hiệu quả tối đa khi yếu tố con người - chủ thể của hoạt động nghiên cứu có đủ năng lực và nhận thức đúng đắn. Do đó, giải pháp về nhân lực cần được xem là nhiệm vụ chiến lược, tiến hành đồng bộ trên ba phương diện: đào tạo kỹ năng, định hình văn hóa và liên kết hệ thống.

Thứ nhất, chú trọng công tác đào tạo, bồi dưỡng "năng lực số" cho đội ngũ biên tập và phản biện. Trước sự biến đổi không ngừng của công nghệ, các tòa soạn cần triển khai các chương trình tập huấn định kỳ về kỹ năng kiểm chứng dữ liệu trong môi trường số. Nội dung đào tạo cần đi sâu vào các kỹ thuật nhận diện những dấu hiệu bất thường trong văn phong máy móc (như cấu trúc câu lặp lại, thiếu tính biện chứng, sử dụng thuật ngữ sai ngữ cảnh) và kỹ năng tra cứu chéo để phát hiện thông tin giả mạo. Mục tiêu là trang bị cho đội ngũ "gác gôn" tri thức tư duy phản biện sắc bén và khả năng sử dụng thành thạo các công cụ hỗ trợ để không bị động trước các hành vi gian lận tinh vi.

Thứ hai, kiến tạo văn hóa liêm chính dựa trên sự minh bạch và trung thực trí tuệ. Cần chuyển dịch tư duy quản lý từ "cấm đoán và trừng phạt" sang "quản trị dựa trên sự minh bạch". Theo tổng hợp của Zhou và Soulière (2025) về Diễn đàn COPE, các tạp chí cần truyền thông điệp rõ ràng rằng việc sử dụng công nghệ để khắc phục rào cản ngôn ngữ hoặc xử lý dữ liệu là chấp nhận được, miễn là được công khai minh bạch. Tuy nhiên, quyền lợi phải đi đôi với trách nhiệm. Cần xây dựng bầu không khí học thuật nơi sự trung thực được tôn vinh, đồng thời áp dụng các chế tài xử lý nghiêm khắc đối với các hành vi cố tình che giấu hoặc gian lận, nhằm tạo tính răn đe và bảo vệ những nhà nghiên cứu chân chính.

Thứ ba, thiết lập mạng lưới hợp tác và chia sẻ thông tin vi phạm. Trong bối cảnh các hình thức gian lận có tổ chức (như các "nhà máy bài báo") hoạt động ngày càng tinh vi, sự đơn độc của từng tòa soạn sẽ tạo ra kẽ hở cho các hành vi phi đạo đức. Do đó, cần thúc đẩy cơ chế liên kết, chia sẻ dữ liệu giữa các tạp chí khoa học trong và ngoài hệ thống. Việc xây dựng một cơ sở dữ liệu chung hoặc hệ thống cảnh báo sớm về các tác giả, nhóm tác giả có lịch sử vi phạm liêm chính sẽ giúp các tòa soạn chủ động phòng ngừa rủi ro, đồng thời tạo ra một mặt bằng tiêu chuẩn đạo đức đồng nhất cho nền khoa học nước nhà.

3. Kết luận

Sự hiện diện của trí tuệ nhân tạo trong đời sống khoa học là xu thế khách quan và không thể đảo ngược. Đối với hoạt động biên tập và xuất bản, công nghệ này mang tính hai chiều: vừa là đòn bẩy nâng

cao năng suất, vừa là nguy cơ đe dọa sự toàn vẹn của tri thức nếu thiếu đi các cơ chế kiểm soát hữu hiệu. Những phân tích trong nghiên cứu này đã chỉ ra rằng, các phương thức "gác gôn" truyền thống hay sự phụ thuộc đơn thuần vào công cụ kỹ thuật đã không còn đủ dư địa để đối phó với làn sóng gian lận công nghệ cao ngày càng tinh vi.

Để giải quyết thấu đáo bài toán này, nền khoa học Việt Nam cần một tư duy quản trị mới, tiếp cận vấn đề theo hướng tổng thể và đa chiều. Đó là sự kết hợp chặt chẽ giữa "hàng rào cứng", bao gồm các quy chế quản lý minh bạch, chế tài nghiêm khắc; và "hàng rào kỹ thuật" với trọng tâm là quy trình rà soát kép và tiêu chuẩn dữ liệu mở. Song, yếu tố quyết định và bền vững nhất vẫn nằm ở con người. Việc bồi dưỡng năng lực số cho đội ngũ biên tập viên và xây dựng văn hóa liêm chính, nơi lòng tự trọng nghề nghiệp được đặt cao hơn áp lực thành tích chính là chốt chặn cuối cùng để bảo vệ sự trong sạch của khoa học.

Thay vì né tránh hay áp đặt các lệnh cấm cực đoan, chúng ta cần chủ động thích ứng và thiết lập các chuẩn mực đạo đức mới. Chỉ khi xác lập được vị thế chủ động của con người trong việc kiểm soát và sử dụng công nghệ, chúng ta mới có thể tận dụng tối đa sức mạnh của trí tuệ nhân tạo, đồng thời giữ vững được vị thế và uy tín của nền khoa học quốc gia trong tiến trình hội nhập quốc tế./.

TÀI LIỆU THAM KHẢO

1. Bhattacharyya, M., Miller, V. M., Bhattacharyya, D., & Miller, L. E. (2023). High rates of fabricated and inaccurate references in ChatGPT-generated medical content. Cureus, 15(5), e39238. https://doi.org/10.7759/cureus.39238.

2. Breugelmans, R. (2025, June 30). How to disclose the use of AI in your research paper. Japan Society for Medical Education (JASMEE). https://jasmee.jp/how-to-disclose-the-use-.of-ai-in-your-research-paper/.

3. Cheng, Y., Sadasivan, V. S., Saberi, M., Saha, S., & Feizi, S. (2025). Adversarial paraphrasing: A universal attack for humanizing AI-generated text. arXiv. .https://doi.org/10.48550/arXiv.2506.07001.

4. Kocyigit, B. F., Okyay, R. A., Seiil, B., Qumar, A. B., & Sumbul, H. E. (2025). Analysis of retracted publications on artificial intelligence: Trends, ethical concerns, and scientific integrity. Journal of Korean Medical Science, 40(44), e280. https://doi.org/10.3346/jkms.2025.40.e280.

5. Krishna, K., Song, Y., Karpinska, M., Wieting, J., & Iyyer, M. (2023). Paraphrasing evades detectors of AI-generated text, but retrieval is an effective defense. arXiv. https://doi.org/10.48550/arXiv.2303.13408.

6. Liang, W., Yuksekgonul, M., Mao, Y., Wu, E., & Zou, J. (2023). GPT detectors are biased against non-native English writers. Patterns, 4(7), 100779. https://doi.org/10.1016/j.patter.2023.100779.

7. Májovský, M., Černý, M., Kasal, M., Komarc, M., & Netuka, D. (2023). Artificial intelligence can generate fraudulent but authentic-looking scientific medical articles: Pandora's box has been opened. Journal of Medical Internet Research, 25, e46924. https://doi.org/10.2196/46924.

8. Martínez-Rojas, E., & Zahn-Muñoz, C. (2025). New forms of fraud in science: Deceptive practices such as article mills, fraudulent peer review, and automatic content generation. Data and Metadata, 4, Article 655. https://doi.org/10.56294/dm2025655.

9. National Institutes of Health. (2025, July 17). Supporting fairness and originality in NIH research applications (Notice No. NOT-OD-25-132). U.S. Department of Health and Human Services. https://grants.nih.gov/grants/guide/notice-files/NOT-OD-25-132.html.

10. Perkins, M., & Roe, J. (2024). Academic publisher guidelines on AI usage: A ChatGPT supported thematic analysis. F1000Research, 12, 1398. https://doi.org/10.12688/f1000research.142411.2.

11. Stauffer, R. (2025, February). Retractions increase 10-fold in 20 years – and now AI is involved. AAPS Newsmagazine. Truy cập ngày 10/12/2025, từ https://www.aapsnewsmagazine.org/aapsnewsmagazine/articles/new-page3/feb25/meetings-feb25.

12. Zhou, H., & Soulière, M. (2025, August 25). From detection to disclosure - Key takeaways on AI ethics from COPE's forum. The Scholarly Kitchen. Truy cập ngày 1/12/2025 từ https://scholarlykitchen.sspnet.org/2025/08/25/from-detection-to-disclosure-key-takeaways-on-ai-ethics-from-copes-forum/.

13. Zielinski, C., Winker, M. A., Aggarwal, R., Ferris, L. E., Heinemann, M., Lapeña, J. F., Jr., Pai, S. A., Ing, E., Citrome, L., Alam, M., Voight, M., & Habibzadeh, F. (2023, May 31). Chatbots, generative AI, and scholarly manuscripts: WAME recommendations on chatbots and generative artificial intelligence in relation to scholarly publications. World Association of Medical Editors. https://wame.org/page3.php?id=106.