TÀI LIỆU THAM KHẢO DO AI HỖ TRỢ: VẤN ĐỀ XÁC THỰC VÀ ĐẢM BẢO LIÊM CHÍNH TRONG NGHIÊN CỨU KHOA HỌC TẠI VIỆT NAM

Trần Triệu Hải[1*], Lê Thị Ngọc Trâm[1], Lê Thị Thu Vân[1], Lê Hữu Nam[1]
[*]Tác giả liên hệ, Email:haitt@hou.edu.vn

Ngày tòa soạn nhận được bài báo: 02/06/2025
Ngày phản biện đánh giá: 15/08/2025
Ngày bài báo được duyệt đăng: 29/08/2025

DOI: 10.59266/houjs.2025.665

Tóm tắt: *Sự bùng nổ của trí tuệ nhân tạo (AI) đang tạo nên làn sóng thay đổi sâu rộng trong xã hội loài người. Một số mô hình AI được công bố, đặc biệt là Chat GPT đã có thể tạo ra các văn bản giống con người thực sự và giúp con người viết thư, soạn thảo các bài luận và vô số công việc (Rane et al., n.d.). Chỉ khoảng 2 năm gần đây, ta có thể thấy từ khoá "AI", "ChatGPT" đã trở thành từ khoá quen thuộc với mọi người Việt Nam. Tuy nhiên, bên cạnh việc ứng dụng các hệ thống mới này là một số thách thức và hạn chế, nhất là với lĩnh vực đạo đức (Kooli, 2023). Một trong những vấn đề đáng lo ngại nhất hiện nay chính là khả năng AI tạo các tài liệu tham khảo hoặc tạo các dữ liệu sai lệch nhưng rất khó bị các công cụ kiểm tra đạo văn truyền thống phát hiện. Bài viết này sẽ nói về hiện tượng sử dụng AI để tạo các tài liệu tham khảo sẽ gây ra những thách thức với việc duy trì liêm chính học thuật và đề xuất các giải pháp để giải quyết vấn đề này. Các giải pháp bao gồm việc xây dựng chính sách rõ ràng với các trường hợp sử dụng trí tuệ nhân tạo, tăng cường đào tạo nhận thức về đạo đức AI, cải thiện các biện pháp đánh giá, công nghệ nhận dạng AI.*

Từ khoá: *trí tuệ nhân tạo (AI), chatGPT, liêm chính học thuật, ngụy tạo tài liệu tham khảo, đạo văn, gian lận học thuật, đạo đức AI, thách thức giáo dục*

I. Giới thiệu

Trong những năm gần đây, Trí tuệ nhân tạo (AI) đã có những bước phát triển nhảy vọt, đặc biệt là những mô hình ngôn ngữ lớn (LLMs) như ChatGPT (Dempere et al., 2023) (Rahman & Watanobe, 2023). AI hứa hẹn một tiềm năng phát triển mạnh khi bước đầu đã có thể hỗ trợ cải thiện năng suất, hỗ trợ sáng tạo, tối ưu hoá quá trình giảng dạy và học tập (Rane et al., n.d.). Việc sử dụng các công cụ AI như ChatGPT, Gemini vào việc dịch tài liệu học thuật, tìm kiếm thông tin, phân tích tài liệu, tạo bài giảng chiếm tỷ lệ rất cao, lên tới hơn 70% (*Khảo sát về việc sử dụng AI tại ĐHQGHN tháng 2/2025*). Có thể thấy,

[1] Khoa Đào tạo Cơ bản, Trường Đại học Mở Hà Nội

sự phát triển mạnh của AI đã trở thành một cuộc cách mạng trong việc tiếp cận trí thức, công nghệ nhưng cũng đặt ra những thách thức to lớn cho liêm chính khoa học. AI càng phổ biến, càng dễ dùng thì hiện tượng sử dụng các tài liệu tham khảo được AI hỗ trợ trong nghiên cứu khoa học cũng tăng theo. Bài tham luận này sẽ phân tích hiện trạng, nguyên nhân và đề xuất các giải pháp nhằm bảo vệ liêm chính khoa học trước thách thức mà AI mang lại.

II. Phương pháp nghiên cứu

Để hoàn thành bài nghiên cứu này, chúng tôi sử dụng phương pháp tổng quan tài liệu kết hợp với phân tích định tính. Đầu tiên, chúng tôi tiến hành rà soát, tổng hợp và phân tích các nguồn tài liệu đa dạng trong và ngoài nước. Các nguồn này bao gồm các bài báo khoa học đã được bình duyệt từ những cơ sở dữ liệu uy tín như Scopus, Web of Science, và Google Scholar, các quy định về liêm chính học thuật do các trường đại học, tổ chức khoa học ban hành, cùng với các bài viết, báo cáo từ những tạp chí khoa học hàng đầu như Nature và Science, các nguồn dữ liệu trong nước như CSDL của Bộ GD&ĐT, Quỹ Nafosted được sử dụng để truy xuất những công trình liên quan. Các nghiên cứu tập trung vào những vấn đề: hiện tượng "AI hallucination" và "fabricated references" trong sản phẩm học thuật, các khảo sát về gian lận học thuật và liêm chính nghiên cứu cả trong nước và quốc tế, chính sách và quy định về sử dụng AI trong nghiên cứu ở bối cảnh quốc tế và Việt Nam.

Quá trình này giúp xác định các khái niệm cốt lõi, thực trạng và những thách thức chính liên quan đến việc sử dụng AI trong nghiên cứu khoa học. Chúng tôi lựa chọn các trường hợp điển hình đã được công bố công khai, trong đó bài báo khoa học hoặc báo cáo nghiên cứu bị thu hồi do phát hiện tài liệu tham khảo hoặc dữ liệu ngụy tạo bởi AI. Các trường hợp này được phân tích nhằm rút ra đặc điểm chung về cách AI tạo ra trích dẫn sai lệch hoặc tài liệu không tồn tại, lỗ hổng trong quá trình phản biện và kiểm duyệt, bài học kinh nghiệm trong việc phát hiện và xử lý.

III. Thực trạng về các tài liệu do AI hỗ trợ

Sự ra đời của các công cụ trí tuệ nhân tạo làm gia tăng những mối lo ngại về sự không trung thực trong học thuật, trong nghiên cứu khi các tác giả sử dụng các bài viết do AI tạo ra hoặc dựa trên các tài liệu tham khảo do AI hỗ trợ. Hiện nay, số lượng các bài nghiên cứu bị thu hồi vì sử dụng dữ liệu ảo, dữ liệu do AI hỗ trợ đang tăng vọt trên thế giới. Một nghiên cứu tại Hà Lan cho thấy 8% các nhà khoa học tại đây thừa nhận có gian lận nghiên cứu (Gopalakrishna et al., 2022) tỷ lệ này đã tăng gấp đôi so với năm trước. Thậm chí, trong một nghiên cứu, họ đã tạo 288 bài viết học thuật được AI tạo ra có đầy đủ các thành tố học thuật tiêu chuẩn, mô tả toàn diện về dữ liệu, phương pháp, thảo luận chi tiết về kết quả, kết luận để chứng minh về khả năng của AI (Novy-Marx & Velikov, 2025). Tuy đây chỉ là một ví dụ nhưng chúng ta cũng có thể thấy được nguy cơ hiện hữu việc sử dụng AI trong nghiên cứu rất dễ tạo ra các kết quả giả mạo, sửa đổi dữ liệu thử nghiệm hoặc các mục đích xấu khác.

Sự phát triển nhanh chóng của AI đã mang lại cả cơ hội và thách thức cho

liêm chính học thuật tại Việt Nam. Các công cụ AI được ứng dụng để hỗ trợ nhiều khía cạnh trong nghiên cứu khoa học, từ việc đưa ra ý tưởng, thu thập và phân tích dữ liệu, đến xác minh kết quả và dự đoán phản hồi của người xem xét. ChatGPT-4, với khả năng viết, tạo đồ họa, phân tích dữ liệu và truy cập internet, đóng vai trò quan trọng trong việc tăng cường quy trình nghiên cứu hàng ngày. Việc lạm dụng AI tại Việt Nam đã đặt ra những lo ngại nghiêm trọng về đạo văn, các tài liệu tham khảo giả mạo hoặc bị sai lệch do AI tạo ra. Hiện nay, một số trường Đại học, Viện Nghiên cứu cũng đã bắt đầu đưa vấn đề đạo đức, tính xác thực của dữ liệu do AI tạo ra vào thảo luận cũng như ban hành một số quy định công khai tỷ lệ AI sử dụng trong bài viết (Vu & Thành Long, 2021). Quỹ Nafosted đã ban hành quy định về liêm chính nghiên cứu, nhấn mạnh trách nhiệm của các bên trong đảm bảo tính trung thực và minh bạch - mặc dù chưa đề cập sâu đến AI (NAFOSTED, 2022). Hay Đại học Khoa học xã hội và nhân văn - ĐH Quốc gia tpHCM cũng có ban hành quy định về liêm chính học thuật bắt đầu có hiệu lực từ 15/01/2025 trong đó có Yêu cầu minh bạch mức độ sử dụng AI, công khai phần trăm nội dung do AI phụ trợ, đảm bảo người dùng chịu trách nhiệm về dữ liệu và trích dẫn từ AI (Trường ĐH KHXH&NV, 2025). Về phía Nhà nước, Luật Khoa học, Công nghệ và Đổi mới sáng tạo cũng chưa bổ sung thêm các quy định hay nội dung về trí tuệ nhân tạo và đang ở mức dự thảo. Có thể thấy các giải pháp này chưa có tính đồng bộ, chỉ mang tính nội bộ các đơn vị, thiếu các hành lang pháp lý rõ ràng, tổng thể, chưa có các tiêu chí đánh giá và kiểm định cụ thể.

IV. Thách thức từ tài liệu tham khảo do AI hỗ trợ

Tài liệu tham khảo do AI tạo ra (AI-generated references) là các trích dẫn do trí tuệ nhân tạo tự sinh ra dựa trên cú pháp và ngữ cảnh giống thật. Các tài liệu này có thể chính xác, cũng có thể không tồn tại hoặc bị gán sai nguồn dẫn tới nhầm lẫn, sai lệch hay còn gọi là nguỵ tạo tài liệu tham khảo cho các lập luận học thuật.

Một trong những vấn đề lớn nhất mà các công cụ AI tác động tới liêm chính khoa học chính là hiện tượng "Ảo giác AI - AI hallucinations" (York College of Pennsylvania, 2025) (Princeton University, 2025). Đây là hiện tượng dùng AI tạo ra văn bản mạch lạc và trôi chảy nhưng khi không có đủ tham chiếu hoặc giới hạn rõ ràng, AI có thể tạo ra các kết nối không tồn tại hoặc thông tin sai lệch để duy trì tính nhất quán của văn bản. Hiện tượng này thường do các mô hình ngôn ngữ lớn LLMs dựa trên mô hình Transformer để dự đoán từ tiếp theo trong một chuỗi (next-token prediction). Do đó, các trích dẫn được tạo ra xem qua thì rất hợp lý nhưng lại không có thật do bản chất nó không phải là kết quả của tìm kiếm tuyệt đối mà là do AI "điền vào chỗ trống" những thông tin có xác suất xuất hiện cao nhất. Ví dụ: ***"Nguyen & Le (2022), Journal of AI Ethics"***, viết đúng chuẩn nhưng khi tra cứu thì thấy bài nghiên cứu này không tồn tại. Thậm chí, trên một nghiên cứu mới đây về lập trình máy tính, nhóm nghiên cứu đã phát hiện ra 52% câu trả lời của AI với các câu hỏi về lập trình có chứa lỗi (Kabir et al., 2024). Có thể thấy, AI không thực sự hiểu thông tin mà chỉ tái tạo lại dựa vào các mẫu đã học được trong quá trình huấn luyện (training). Việc đào tạo chatbot bằng dữ liệu khổng lồ từ internet không còn đủ

để cải thiện chất lượng. OpenAI, Google và nhiều hãng khác giờ chuyển sang mô hình huấn luyện bằng reinforcement learning, tức để AI "thử sai" rồi học từ phản hồi. Phương pháp này cải thiện rõ ở các bài toán logic, nhưng lại không giúp ích mấy trong việc kiểm soát tính xác thực thông tin. Ngoài ra, ảo giác còn có thể xuất phát từ những yếu tố như Mô hình AI quá tự tin, các chương trình xử lý ngôn ngữ tự nhiên NLP quá phức tạp, quá trình giải mã/mã hoá của AI có lỗi, ….Những kết quả như vậy kết hợp với sự chủ quan và phụ thuộc quá mức vào AI dẫn tới việc lan truyền những thông tin sai lệch, hiểu nhầm các kết quả của các nghiên cứu và đưa ra các kết luận không chính xác. Bên cạnh đó, các mô hình AI như ChatGPT còn có thể tạo ra các tài liệu tham khảo chứa thông tin sai lệch, lỗi thời được gọi là "confabulated references" hoặc «bịa đặt tham chiếu/thông tin thư mục» - fabrication and errors in the bibliographic citations làm bài viết mất đi tính chính xác (Stokel-Walker, 2023). Theo một thí nghiệm, nhóm tác giả đã tạo các bài đánh giá tài liệu ngắn về 42 chủ đề đa ngành, biên soạn dữ liệu về 636 trích dẫn tài liệu tham khảo có trong 84 bài báo bằng ChatGPT 3.5 và ChatGPT4. Kết quả là với 55% trích dẫn ChatGPT 3.5 là bịa đặt, tỷ lệ này giảm xuống còn 18% ở Chat GPT 4 nhưng có 24% trích dẫn bằng Chat GPT 4 có lỗi trích dẫn quan trọng(Walters & Wilder, 2023). Hai nghiên cứu năm 2023 với ChatGPT 3.5 chỉ ra rằng 87% các trích dẫn là các bài báo có thật nhưng lại chứa ít nhất 1 trong 7 lỗi như sai số PubMedID, tên tác giả, tiêu đề bài viết, ngày xuất bản, tên tạp chí, số tập hoặc số trang (Bhattacharyya et al., 2023) và (Day, 2023) cũng cho kết quả tương tự. Mặc dù hiện nay các công cụ AI liên tục nâng cấp, nhưng những nguy cơ này vẫn có thể xảy ra. Điều này có thể do lượng dữ liệu văn bản rất lớn từ nhiều nguồn tham khảo khác nhau và sự không nhất quán hoặc có sai lệch trong dữ liệu gốc đã khiến ChatGPT gặp khó khăn trong việc phân biệt các thông tin của các tài liệu tham khảo "có thật" như ngày đăng bản nháp, ngày đăng bản hoàn chỉnh và ngày xuất bản chính thức (Sanchez-Ramos et al., 2023).

Bên cạnh đó, các công cụ phát hiện AI và phát hiện đạo văn cũng gặp thách thức rất lớn. Với các công cụ kiểm tra đạo văn truyền thống như Plagiarism Detector (trang web miễn phí)và iThenticate(công cụ chuyên nghiệp trả phí) khi đánh giá 50 bản tóm tắt từ các tạp chí lớn và 50 bản tóm tắt do ChatGPT tạo ra thì kết quả nhận được là các tóm tắt do AI tạo ra có điểm đạo văn thấp hơn đáng kể. Với Plagiarism Detector, tỷ lệ phần trăm văn bản bị trùng hoặc "bị nghi đạo văn" là 62,5% với các bản gốc, 0% so với bản do AI tạo. Với iThenticate, các bản gốc có chỉ số là 100 so với các bản tóm tắt do AI tạo nhận được điểm là 27. Cũng với dữ liệu test này, để các đánh giá viên thực hiện "đánh giá mù", tỷ lệ nhận ra văn bản AI tạo là 68% và tỷ lệ nhận ra văn bản gốc là 86% nhưng các đánh giá viên lại xác định sai tới 32% các văn bản do AI tạo ra là văn bản thật và nhầm 14% các bản gốc là do AI tạo (Gao et al., 2023). Một số nghiên cứu khác về các công cụ khác như Turnitin cũng cho thấy kết quả không chính xác, có sai lệch (Spennemann et al., 2023).

V. Các giải pháp đề xuất

Để giải quyết vấn đề sử dụng các tài liệu tham khảo tạo bởi AI nói riêng và các hành vi vi phạm liêm chính khoa học trong thời đại trí tuệ nhân tạo nói chung, cần một

cách tiếp cận kết hợp đồng bộ giữa công nghệ, chính sách và giáo dục. Đầu tiên, quan trọng nhất đó là hoàn thiện chính sách và quy định. Trên bình diện quốc tế, nhiều tổ chức đã ban hành các chính sách khai báo bắt buộc khi sử dụng AI. Ví dụ như Science Journals và Springer Nature, đã yêu cầu tác giả phải khai báo vai trò của AI trong phần phương pháp nghiên cứu hoặc cảm ơn, nhưng không được coi AI là đồng tác giả (Thorp, 2023)(Chris Stokel-Walker, 2023)(Springer Nature, 2023). Tại Việt Nam, Bộ Khoa học Công nghệ cần nhanh chóng hoàn thiện và công bố Luật Khoa học, Công nghệ và Đổi mới sáng tạo, trong đó bổ sung thêm những chính sách, quy định, cơ chế xử lý nghiêm minh đối với các trường hợp sử dụng AI. Đây chính là xương sống, là bộ quy tắc chung cho các cơ sở nghiên cứu khoa học, các trường đại học, các nhà nghiên cứu có thể lấy làm căn cứ. Bên cạnh đó, các nhà xuất bản, các cơ sở khoa học, các trường học cần hợp tác để phát triển các hướng dẫn nhất quán, chi tiết hơn cho việc sử dụng và khai thác và công bố về mức độ áp dụng AI trong bài nghiên cứu. Tiếp theo, vấn đề nâng cấp hạ tầng công nghệ cũng cần được quan tâm. Việc cải thiện hạ tầng tính toán, nâng cấp phần mềm như hỗ trợ các tài khoản AI trả phí và tích hợp AI vào các hệ thống quản lý học tập LMS, hệ thống cơ sở dữ liệu về nghiên cứu khoa học sẽ giúp việc phân tích dữ liệu học tập, nghiên cứu, kiểm tra đạo văn được nâng cao, kết quả kiểm tra đạo văn sẽ chính xác hơn.

Thứ ba, xác minh nghiêm ngặt các kết quả do AI tạo ra. Các nhà nghiên cứu phải áp dụng tư duy phê phán, phải coi toàn bộ các sản phẩm đầu ra của AI là bản nháp sơ bộ, vẫn cần được con người xác minh nghiêm ngặt về tính chính xác, đặt biệt là tài liệu tham khảo và các trích dẫn. Để đảm bảo tính chính xác, chúng ta cần kiểm tra chéo tỉ mỉ mọi trích dẫn do AI tạo ra với các cơ sở dữ liệu học thuật đáng tin cậy như Scopus, Web of Science, Google Scholar,… Bên cạnh đó, giáo dục toàn diện nhằm trang bị kiến thức cho mọi người về các lợi ích, rủi ro và các vấn đề đạo đức của việc sử dụng AI trong nghiên cứu học thuật. Việc giáo dục này phải bao gồm đào tạo cách trích dẫn bằng AI đúng cách, hiểu biết về cách các mô hình AI hoạt động để nhận biết các hạn chế của AI để luôn cẩn trọng trước các nội dung do AI tạo ra, cung cấp hướng dẫn rõ ràng, dễ tiếp cận và đào tạo toàn diện cho các tác giả, người đánh giá đồng cấp và biên tập viên về việc sử dụng AI có đạo đức, các hạn chế của công cụ phát hiện và tầm quan trọng tối cao của sự giám sát của con người.

VI. Kết luận

Trí tuệ nhân tạo mang lại cả cơ hội và thách thức to lớn đối với liêm chính khoa học. Hiện tượng nguỵ tạo tài liệu tham khảo là một minh chứng rõ ràng cho những rủi ro tiềm ẩn khi lạm dụng công nghệ. Tuy nhiên, tại Việt Nam, bằng cách sử dụng các nguyên tắc đạo đức, xây dựng chính sách rõ ràng, tăng cường việc giáo dục và liên tục đổi mới phương pháp giảng dạy cũng như đánh giá, chúng ta có thể tận dụng những lợi ích của AI để thúc đẩy nghiên cứu khoa học mà vẫn đảm bảo sự liêm chính, tin cậy của tri thức. AI không phải là "kẻ thù" của khoa học mà nó có thể trở thành "cánh tay đắc lực" cho mọi người nếu được sử dụng minh bạch, có trách nhiệm và tuân thủ các quy tắc liêm chính. Ở đất nước nào, mục tiêu giáo dục cũng là hình thành nên những công dân có khả năng tư duy độc lập, trung thực và sáng tạo. Trong kỷ nguyên AI, mục tiêu đó vẫn đúng, chỉ có cách thực hiện là cần phải thích ứng và đổi mới.

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*. https://doi.org/10.7759/cureus.39238

[2]. Chris Stokel-Walker. (2023). ChatGPT listed as author on research papers: many scientists disapprove. *Nature Briefing Newsletter*. https://doi.org/10.1038/d41586-023-00107-z

[3]. Day, T. (2023). A Preliminary Investigation of Fake Peer-Reviewed Citations and References Generated by ChatGPT. *The Professional Geographer*, *75*(6), 1024-1027. https://doi.org/10.1080/00330124.2023.2190373

[4]. Dempere, J., Modugu, K., Hesham, A., & Ramasamy, L. K. (2023). The impact of ChatGPT on higher education. In *Frontiers in Education* (Vol. 8). Frontiers Media SA. https://doi.org/10.3389/feduc.2023.1206936

[5]. Gao, C. A., Howard, F. M., Markov, N. S., Dyer, E. C., Ramesh, S., Luo, Y., & Pearson, A. T. (2023). Comparing scientific abstracts generated by ChatGPT to real abstracts with detectors and blinded human reviewers. *Npj Digital Medicine*, *6*(1). https://doi.org/10.1038/s41746-023-00819-6

[6]. Gopalakrishna, G., ter Riet, G., Vink, G., Stoop, I., Wicherts, J. M., & Bouter, L. M. (2022). Prevalence of questionable research practices, research misconduct and their potential explanatory factors: A survey among academic researchers in The Netherlands. *PLOS ONE*, *17*(2), e0263023. https://doi.org/10.1371/journal.pone.0263023

[7]. Kabir, S., Udo-Imeh, D. N., Kou, B., & Zhang, T. (2024, May 11). Is Stack Overflow Obsolete? An Empirical Study of the Characteristics of ChatGPT Answers to Stack Overflow Questions. *Conference on Human Factors in Computing Systems - Proceedings*. https://doi.org/10.1145/3613904.3642596

[8]. Kooli, C. (2023). Chatbots in Education and Research: A Critical Examination of Ethical Implications and Solutions. *Sustainability (Switzerland)*, *15*(7). https://doi.org/10.3390/su15075614

[9]. NAFOSTED. (2022). *NAFOSTED-Quy-dinh-ve-liem-chinh-nghien-cuu*. https://nafosted.gov.vn/wp-content/uploads/2022/02/NAFOSTED-Quy-dinh-ve-liem-chinh-nghien-cuu.pdf

[10]. Novy-Marx, R., & Velikov, M. (2025). *AI-Powered (Finance) Scholarship **.

[11]. Princeton University. (2025, January 13). *Disclosing the Use of AI*. Jan 13, 2025 5:41 PM. https://scholarlyintegrity.princeton.edu/students/disclosing-generative-ai-use

[12]. Rahman, M. M., & Watanobe, Y. (2023). ChatGPT for Education and Research: Opportunities, Threats, and Strategies. Applied Sciences (Switzerland), 13(9). https://doi.org/10.3390/app13095783

[13]. Rane, N. L., Paramesha, M., & Desai, P. V. (n.d.). Artificial intelligence, ChatGPT, and the new cheating dilemma: Strategies for academic integrity. Artificial Intelligence and Industry in Society 5.0 (Deep Science Publishing, 2024). https://doi.org/10.70593/978-81-981271-1-2_1

[14]. Sanchez-Ramos, L., Lin, L., & Romero, R. (2023). Beware of references when using ChatGPT as a source of information to write scientific articles. American Journal of Obstetrics and Gynecology, 229(3), 356-357. https://doi.org/https://doi.org/10.1016/j.ajog.2023.04.004

[15]. Spennemann, D., Biles, J., Lachlan Brown, Ireland, M., Laura Longmore, Clare Singh, Wallis, A., & Catherine Ward. (2023). ChatGPT giving advice on how to cheat in university assignments—how workable are its suggestions? https://doi.org/10.21203/rs.3.rs-3365084/v1

[16]. Springer Nature. (2023). Tools such as ChatGPT threaten transparent science;

here are our ground rules for their use. Nature, 613(7945), 612-612. https://doi.org/10.1038/d41586-023-00191-1

[17]. Stokel-Walker, C. (2023). ChatGPT listed as author on research papers: many scientists disapprove. Nature, 613. https://doi.org/10.1038/d41586-023-00107-z

[18]. Thorp, H. H. (2023). ChatGPT is fun, but not an author. Science, 379(6630), 313-313. https://doi.org/10.1126/science.adg7879

[19]. Trường ĐH KHXH&NV. (2025, January 15). Quy định về liêm chính học thuật tại Trường ĐH KHXH&NV, ĐHQG-HCM. 15/01/2025.

[20]. Vu, Đ., & Thành Long, N. (2021). Đánh giá liêm chính học thuật của sinh viên qua nhận thức của sinh viên về môi trường học thuật và hành vi không trung thực học thuật. KINH TẾ VÀ QUẢN TRỊ KINH DOANH, 16, 46-63. https://doi.org/10.46223/HCMCOUJS.econ.vi.16.1.1389.2021

[21]. Walters, W. H., & Wilder, E. I. (2023). Fabrication and errors in the bibliographic citations generated by ChatGPT. Scientific Reports, 13(1). https://doi.org/10.1038/s41598-023-41032-5

[22]. York College of Pennsylvani. (2025, June 16). Hallucinations - Artificial Intelligence in Research. Jun 16, 2025 10:58 AM. https://library.ycp.edu/c.php?g=1387233&p=10262462.

AI-GENERATED REFERENCES: THE CHALLENGE OF VERIFICATION AND ENSURING INTEGRITY IN SCIENTIFIC RESEARCH IN VIETNAM

Tran Trieu Hai[2], Le Thi Ngoc Tram[1], Le Thi Thu Van[1], Le Huu Nam[1]

Abstract: *The rapid rise of artificial intelligence (AI) is creating a profound wave of transformation across human society. Several AI models have been released—most notably ChatGPT—which are now capable of generating human-like texts and assisting people in writing letters, composing essays, and performing countless other tasks (Rane et al., n.d.). In just the past two years, keywords such as "AI" and "ChatGPT" have become familiar terms to nearly every Vietnamese person. However, alongside the adoption of these new systems are a number of challenges and limitations, especially in the domain of ethics (Kooli, 2023). One of the most concerning issues today is AI's ability to generate fabricated references or inaccurate data that traditional plagiarism detection tools struggle to identify. This article will explore the phenomenon of using AI to generate references and how it presents serious challenges to maintaining academic integrity. It will also propose a set of solutions to address this issue. These include establishing clear policies for the use of artificial intelligence, strengthening awareness and training on AI ethics, improving assessment practices, and adopting advanced AI detection technologies.*

Keywords: *artificial Intelligence (AI), ChatGPT, academic integrity, fabricated references, plagiarism, academic misconduct, AI ethics, educational challenges*

---

[2] Faculty of General Education, Hanoi Open University