# RỦI RO HỌC THUẬT VÀ ĐẠO ĐỨC TRONG VIỆC SỬ DỤNG CHATGPT TẠI CÁC CƠ SỞ GIÁO DỤC ĐẠI HỌC: PHÂN TÍCH TOÀN DIỆN VÀ ĐỀ XUẤT GIẢI PHÁP QUẢN LÝ

**CẤN THỊ NGỌC LY**
**Trường Đại học Ngoại thương**

*Nhận bài ngày 15/5/2026. Sửa chữa xong 29/5/2026. Duyệt đăng 15/6/2026.*

***Abstract***

*Since ChatGPT was released in November 2022, higher education institutions worldwide have faced an unprecedented challenge: how to maintain academic integrity when students have access to a tool capable of generating high-quality academic content on demand that traditional plagiarism detection systems may fail to identify. This article provides a comprehensive analysis of the academic and ethical risks associated with the use of ChatGPT in higher education, including AI plagiarism, misinformation, technological dependency, privacy violations, and algorithmic bias. Particular attention is paid to English for Specific Purposes (ESP) training, where these risks have distinctive manifestations and consequences. On that basis, the article proposes an integrated risk management framework for lecturers and educational administrators.*

***Keywords:*** *Academic ethics, academic integrity, academic risks, AI plagiarism, ChatGPT, higher education.*

## 1. Đặt vấn đề

Tốc độ phổ biến của ChatGPT trong môi trường học thuật là chưa từng có trong lịch sử công nghệ giáo dục. Nếu máy tính cá nhân mất nhiều thập kỷ để trở thành công cụ phổ biến trong giảng dạy, và internet mất khoảng một thập kỷ để thay đổi cách sinh viên (SV) tiếp cận thông tin, thì ChatGPT và các công cụ AI tương tự đã thâm nhập vào thói quen học tập của hàng triệu SV chỉ trong vài tháng. Theo nghiên cứu của Welding (2023) trên 1.000 SV đại học Mỹ, 43% đã sử dụng ChatGPT trong ít nhất một bài tập được đánh giá, và trong số đó, hơn một nửa không khai báo việc sử dụng này. Những con số này đặt ra câu hỏi cấp thiết về tính toàn vẹn học thuật, một giá trị nền tảng mà các cơ sở giáo dục đại học đã dành nhiều thập kỷ xây dựng thông qua chính sách, văn hoá và hệ thống đánh giá. Vấn đề không chỉ là gian lận cá nhân mà là một thách thức hệ thống: khi ranh giới giữa 'làm bài tập' và 'nhờ AI làm hộ' trở nên mờ nhạt và khi các công cụ phát hiện đạo văn không theo kịp sự phát triển của AI, hệ thống đảm bảo chất lượng giáo dục đại học đứng trước nguy cơ mất đi công cụ kiểm chứng năng lực quan trọng nhất của mình. Bài viết nhằm phân tích toàn diện các loại rủi ro học thuật từ việc sử dụng ChatGPT, đặt các rủi ro này trong ngữ cảnh đặc thù của đào tạo ESP và đề xuất khung quản lý rủi ro tích hợp ở cả cấp độ giảng dạy và quản trị.

## 2. Nội dung nghiên cứu

### *2.1. Phân loại và phân tích rủi ro học thuật*

*2.1.1. Đạo văn AI*

Đạo văn AI (AI plagiarism) là thuật ngữ chỉ việc SV nộp nội dung do AI tạo ra như thể đó là sản phẩm của chính mình, không có sự khai báo. Điều làm cho loại đạo văn này đặc biệt thách thức so với đạo văn truyền thống là: nội dung do ChatGPT tạo ra là hoàn toàn mới về mặt kỹ thuật (không sao chép từ nguồn nào có thể truy vết), có thể được tùy chỉnh cao độ theo yêu cầu cụ thể, và thường có chất lượng ngôn ngữ cao hơn mức trung bình của SV mục tiêu, điều nghịch lý là có thể khiến nó dễ bị phát hiện bởi giảng viên chú ý nhưng không thể bị phát hiện bởi phần mềm tự động.

*Email: lyctn@ftu.edu.vn*

Các công cụ phát hiện văn bản AI như GPTZero hay Turnitin AI Detector hoạt động dựa trên việc đo lường tính 'bất ngờ' và tính đồng đều của văn bản, vì văn bản người viết có xu hướng bất đều hơn về độ phức tạp cú pháp so với văn bản AI vốn trôi chảy một cách đồng nhất. Tuy nhiên, độ chính xác của các công cụ này vẫn còn hạn chế, đặc biệt khi SV đã học cách chỉnh sửa văn bản AI để tạo ra văn phong tự nhiên hơn. Nghiên cứu của Liang et al. (2023) cho thấy tỷ lệ dương tính giả của các công cụ phát hiện AI lên đến 30%, nghĩa là cứ ba bài bị gắn cờ thì có một bài thực sự do người viết, một hệ quả bất công đáng lo ngại.

*2.1.2. Hallucination và rủi ro thông tin sai lệch*

*Rủi ro thứ hai nghiêm trọng không kém là hiện tượng 'hallucination'*, khi ChatGPT tạo ra thông tin sai lệch với vẻ ngoài tự tin và chuyên nghiệp. Khác với lỗi tìm kiếm truyền thống khi người dùng nhận được không có kết quả hoặc kết quả không liên quan, hallucination của AI đặc biệt nguy hiểm vì nó cung cấp thông tin sai mà trông có vẻ đúng. ChatGPT có thể bịa ra tên tác giả, tiêu đề sách, năm xuất bản, thậm chí DOI và URL của bài báo khoa học không tồn tại và SV thiếu kinh nghiệm kiểm chứng có thể đưa những thông tin này vào bài luận hay báo cáo chuyên ngành mà không nghi ngờ.

Trong bối cảnh ESP, rủi ro này đặc biệt nghiêm trọng trong các ngành đòi hỏi độ chính xác thông tin cao như y tế, pháp lý và kỹ thuật. Bác sĩ tương lai trích dẫn nghiên cứu lâm sàng không tồn tại về tương tác thuốc, hay kỹ sư xây dựng sử dụng tiêu chuẩn kỹ thuật bịa đặt trong thiết kế công trình, là những kịch bản có thể gây hậu quả nghiêm trọng ngoài đời thực. Đây không phải lo ngại lý thuyết vì tháng 6 năm 2023, một luật sư tại New York đã bị toà phán xử vì nộp hồ sơ với 6 án lệ được ChatGPT tạo ra, không có án lệ nào tồn tại trong thực tế.

*2.1.3. Phụ thuộc công nghệ và suy giảm năng lực nhận thức*

Rủi ro thứ ba có tính chất dài hạn và ít nhìn thấy hơn nhưng có lẽ là đáng lo ngại nhất về mặt giáo dục: sự phụ thuộc vào ChatGPT có thể dẫn đến suy giảm năng lực nhận thức quan trọng. Khi SV quen với việc nhận phản hồi tức thì và câu trả lời sẵn có từ AI, họ có ít cơ hội thực hành những quá trình nhận thức bậc cao như tổng hợp thông tin từ nhiều nguồn, lập luận logic có cấu trúc, và sáng tạo ra giải pháp nguyên bản cho vấn đề mới. Nghiên cứu của Luo và Xie (2023) trên 240 SV đại học cho thấy nhóm sử dụng ChatGPT thường xuyên trong một học kỳ (>5 giờ/tuần) có điểm trung bình trong bài kiểm tra tư duy phê phán thấp hơn đáng kể so với nhóm sử dụng ít hơn (<1 giờ/tuần), ngay cả khi kiểm soát về trình độ đầu vào. Mặc dù nghiên cứu này cần được nhân rộng và kiểm chứng, nó gợi ra một cảnh báo quan trọng: điểm số học phần cao có thể che giấu sự sụt giảm thực sự về năng lực nhận thức của SV.

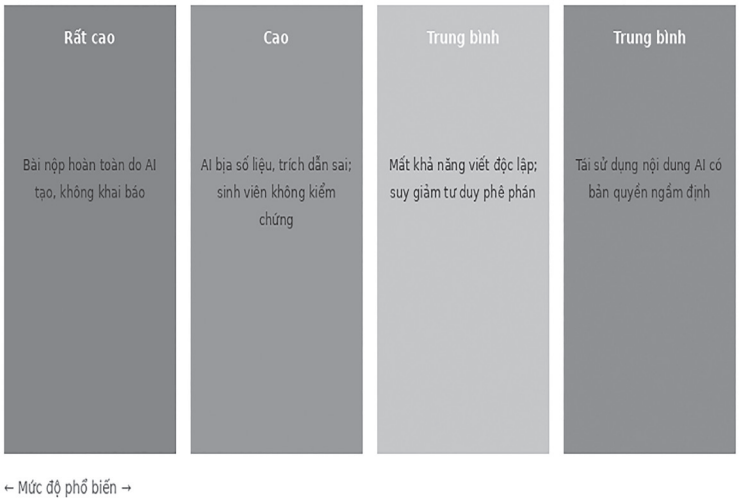
Bảng 1 tổng hợp hệ thống các loại rủi ro học thuật từ việc sử dụng ChatGPT, biểu hiện cụ thể trong đào tạo ESP và hệ quả tiềm ẩn đối với chất lượng giáo dục.

*Bảng 1: Phân loại rủi ro học thuật khi sử dụng ChatGPT trong đào tạo đại học*

| Loại rủi ro | Biểu hiện cụ thể trong ESP | Hệ quả đối với chất lượng đào tạo |
|---|---|---|
| Đạo văn AI (AI Plagiarism) | Bài luận ESP hoàn toàn do ChatGPT viết, không khai báo | Mất giá trị đánh giá; bất bình đẳng giữa SV |
| Hallucination thông tin | Số liệu, tên tác giả, năm xuất bản bịa đặt nhưng trông đáng tin | Sai sót chuyên môn nghiêm trọng; mất thói quen kiểm chứng |
| Phụ thuộc công nghệ | Mất khả năng viết độc lập; giảm tư duy phê phán | Năng lực thực sự thấp hơn kết quả học tập trên hồ sơ |
| Vi phạm quyền riêng tư | Chia sẻ dữ liệu cá nhân, luận văn, đề tài nghiên cứu với OpenAI | Rủi ro bảo mật thông tin cá nhân và sở hữu trí tuệ |
| Thiên kiến thuật toán | Nội dung AI phản ánh quan điểm văn hoá phương Tây | Thiếu đa dạng quan điểm; đồng nhất hoá văn hoá học thuật |

*Nguồn: Tổng hợp từ Perkins (2023); Liang et al. (2023); Luo & Xie (2023)*

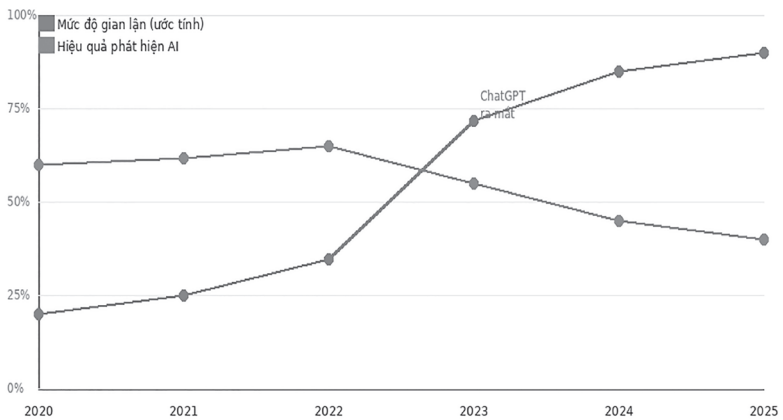
*Hình 1: Phân loại và mức độ nghiêm trọng của rủi ro học thuật từ việc sử dụng ChatGPT*




*2.1.4. Chiều kích thời gian: Sự gia tăng rủi ro song hành với sự phát triển AI*

Điều quan trọng là các rủi ro này không đứng yên mà biến động theo sự phát triển của AI. Mỗi thế hệ mô hình AI mới (GPT-3.5 → GPT-4 → GPT-4o) tạo ra nội dung chất lượng cao hơn, khó phát hiện hơn và có khả năng thực hiện các nhiệm vụ học thuật phức tạp hơn. Hình 2 dưới đây minh hoạ xu hướng đối nghịch giữa mức độ gian lận tiềm năng và hiệu quả phát hiện của các công cụ hiện có theo thời gian, một 'khoảng cách' ngày càng mở rộng đặt ra thách thức hệ thống cho toàn ngành Giáo dục.

*Hình 2: Xu hướng gia tăng rủi ro gian lận AI và sự sụt giảm hiệu quả phát hiện (2020–2025)*




***2.2. Rủi ro đặc thù trong đào tạo ESP***

*2.2.1. Sự vô hình hóa của quá trình thụ đắc ngôn ngữ*

Trong đào tạo ESP, một rủi ro đặc thù cần được nhấn mạnh là việc sử dụng ChatGPT để 'làm hộ' các nhiệm vụ ngôn ngữ không chỉ là vấn đề đạo đức học thuật mà còn trực tiếp triệt tiêu quá trình thụ đắc ngôn ngữ . Theo lý thuyết của Krashen (1985), ngôn ngữ được thụ đắc thông qua tiếp xúc với đầu vào có thể hiểu được kết hợp với nỗ lực xử lý ngôn ngữ của người học. Khi SV nhờ ChatGPT viết báo cáo kỹ thuật bằng tiếng Anh thay vì tự viết và vật lộn với ngôn ngữ, họ bỏ lỡ chính xác quá trình xử lý ngôn ngữ là điều kiện cần thiết để phát triển năng lực ESP thực sự.

Nghịch lý của tình huống này là ở chỗ: một SV kỹ thuật có thể hoàn thành toàn bộ học phần ESP với điểm A bằng cách sử dụng ChatGPT, nhưng khi bước vào môi trường làm việc quốc tế, năng lực ESP thực tế của họ không khác biệt so với người không học học phần đó. Đây là một hình thức thất bại của giáo dục mà bằng chứng không xuất hiện trong thống kê điểm số mà chỉ lộ ra trong môi trường nghề nghiệp thực.

*2.2.2. Rủi ro thiên kiến văn hóa trong nội dung ESP*

Một rủi ro ít được thảo luận hơn nhưng quan trọng không kém là thiên kiến văn hoá trong nội dung do ChatGPT tạo ra. Các mô hình LLMs được huấn luyện chủ yếu trên dữ liệu ngôn ngữ tiếng Anh từ các nguồn phương Tây, dẫn đến xu hướng tạo ra văn bản phản ánh các chuẩn mực giao tiếp và tư duy của các nền văn hóa Anglo-Saxon. Khi SV Việt Nam sử dụng nội dung từ ChatGPT mà không có sự phê phán, họ có thể vô tình nội hóa một khung tham chiếu văn hóa không phù hợp với bối cảnh nghề nghiệp và xã hội của mình.

Trong ngữ cảnh ESP, điều này có thể biểu hiện qua những tình huống như: SV sử dụng phong cách giao tiếp trực tiếp của ChatGPT trong bối cảnh kinh doanh châu Á vốn coi trọng sự gián tiếp và thể diện; hay áp dụng quy ước viết học thuật Anglo-Saxon cho những bối cảnh đòi hỏi phong cách học thuật khác. Giảng viên ESP cần trang bị cho SV khả năng nhận ra và đánh giá thiên kiến văn hoá trong nội dung AI.

***2.3. Khung quản lý rủi ro tích hợp***

*2.3.1. Cấp độ thiết kế học phần*

Giải pháp căn bản nhất ở cấp độ giảng dạy là thiết kế lại hệ thống đánh giá theo hướng khó có thể thực hiện bằng AI mà không có sự tham gia thực sự của người học. Điều này không có nghĩa là 'AI-proof' theo nghĩa cứng nhắc – thực tế không có bài tập nào là hoàn toàn không thể làm bằng AI nếu SV đủ sáng tạo. Thay vào đó, triết lý thiết kế nên hướng đến các nhiệm vụ yêu cầu trải nghiệm cá nhân không thể tổng quát hoá, phán đoán theo ngữ cảnh cụ thể, hoặc thực hiện trong điều kiện không thể sử dụng AI.

Bảng 2 dưới đây trình bày các chiến lược cụ thể đã được kiểm nghiệm trong thực tiễn giảng dạy ESP, bao gồm mô tả, phương thức áp dụng cụ thể và hiệu quả ghi nhận, cung cấp tài nguyên thực tiễn cho giảng viên.

*Bảng 2: Chiến lược quản lý rủi ro ChatGPT trong thiết kế học phần ESP*

| Chiến lược | Mô tả | Áp dụng trong ESP | Hiệu quả ghi nhận |
|---|---|---|---|
| Thiết kế lại đánh giá | Bài tập yêu cầu tư duy bậc cao, trải nghiệm cá nhân | Phân tích diễn ngôn tình huống thực; phỏng vấn miệng | Cao – khó thay thế bằng AI |
| Yêu cầu khai báo AI | SV ghi chú cách sử dụng AI trong bài nộp | Mục 'AI Statement' trong rubric đánh giá | Trung bình – phụ thuộc ý thức người học |
| Dạy kỹ năng kiểm chứng | Huấn luyện đánh giá và kiểm chứng thông tin từ AI | So sánh phản hồi AI với nguồn học thuật Scopus/PubMed | Cao khi kết hợp thực hành liên tục |
| Học tập có người chứng kiến | Bài tập thực hiện trước mặt giảng viên hoặc nhóm | Thuyết trình không chuẩn bị; viết tay trong lớp | Rất cao – xác thực năng lực thực |
| Giáo dục đạo đức AI | Thảo luận định kỳ về đạo đức sử dụng AI trong học thuật | Case study về tình huống đạo đức AI trong ESP | Trung bình đến cao theo thời gian |

*Nguồn: Tổng hợp từ Perkins (2023); Cheung (2023); Lodge et al. (2023)*

*2.3.2. Cấp độ cơ sở đào tạo*

*Ở cấp độ tổ chức, các cơ sở giáo dục cần xây dựng chính sách AI rõ ràng, minh bạch và khả thi*, không phải chính sách cấm đoán không thể thực thi mà là chính sách hướng dẫn có nguyên tắc. Một chính sách AI hiệu quả cần trả lời được ba câu hỏi cơ bản: (1) Những hành vi sử dụng AI nào được phép, không được phép và được phép có điều kiện? (2) SV cần khai báo việc sử dụng AI như thế nào? (3) Hệ quả của vi phạm chính sách là gì và làm thế nào để phát hiện?

Ngoài chính sách, cần đầu tư vào giáo dục đạo đức AI như một phần bắt buộc của chương trình học. Khác với giáo dục đạo đức học thuật truyền thống (vốn tập trung vào việc không sao chép), giáo dục đạo

đức AI cần phát triển khả năng phán đoán đạo đức trong các tình huống mơ hồ và không có tiền lệ rõ ràng, một thách thức sư phạm phức tạp hơn nhiều đòi hỏi tiếp cận dựa trên thảo luận và tình huống thực tế.

*2.3.3. Cấp độ sinh viên: Phát triển tư duy phê phán AI*

Ở cấp độ cá nhân người học, giải pháp dài hạn và bền vững nhất là phát triển tư duy phê phán về AI, khả năng đánh giá có hệ thống nội dung nhận được từ AI, nhận diện giới hạn và thiên kiến của nó, và đưa ra quyết định có ý thức về khi nào và cách nào sử dụng AI một cách có trách nhiệm. Kỹ năng này không tự nhiên phát triển mà cần được dạy tường minh, với ví dụ cụ thể và cơ hội thực hành đủ nhiều.

Trong ngữ cảnh ESP, điều này có thể được tích hợp vào hoạt động lớp học thông qua các bài tập yêu cầu SV sử dụng ChatGPT để tạo ra nội dung chuyên ngành, sau đó hệ thống kiểm chứng thông tin với nguồn học thuật, xác định những gì ChatGPT đúng, sai và thiếu, và viết phản tư về những gì họ học được từ quá trình kiểm chứng đó. Đây là mẫu hoạt động vừa phát triển năng lực ESP vừa hình thành thói quen tư duy phê phán AI.

**3. Kết luận**

ChatGPT là một công cụ mạnh mẽ và trung tính về mặt đạo đức, tác động của nó đối với tính toàn vẹn học thuật và chất lượng giáo dục phụ thuộc hoàn toàn vào cách nó được sử dụng và trong khuôn khổ nào. Các rủi ro được phân tích trong bài viết đạo văn AI, hallucination, phụ thuộc công nghệ, vi phạm quyền riêng tư và thiên kiến văn hoá là thực sự báo động và nghiêm trọng, đòi hỏi phản ứng có hệ thống từ các cơ sở giáo dục. Tuy nhiên, phản ứng này không nên là sự hoảng loạn hay cấm đoán vội vàng mà nên là sự điều chỉnh có suy xét: điều chỉnh hệ thống đánh giá để đo lường được những gì AI không thể thay thế, điều chỉnh văn hóa học thuật để đạo đức AI trở thành một phần của chuẩn mực chung và điều chỉnh mục tiêu giáo dục để phát triển những năng lực nhận thức bậc cao mà AI làm nổi bật thêm sự quan trọng của chúng, không phải làm giảm đi.

**Tài liệu tham khảo**


Cheung, C. K. F. (2023). The use of generative AI in higher education and concerns about academic integrity. *Computers and Education: Artificial Intelligence, 5*, 100190.

Ji, Z., Lee, N., Frieske, R., Yu, T., Su, D., Xu, Y., … Fung, P. (2023). Survey of hallucination in natural language generation. *ACM Computing Surveys, 55(12*), 1–38.

Krashen, S. D. (1985). *The input hypothesis: Issues and implications.* Longman.

Liang, W., Yuksekgonul, M., Mao, Y., Wu, E., & Zou, J. (2023). GPT detectors are biased against non-native English writers. *Patterns, 4(7),* 100779.

Lodge, J. M., Thompson, K., & Corrin, L. (2023). Mapping out a research agenda for generative artificial intelligence in highly regulated learning contexts. *Australasian Journal of Educational Technology, 39(1),* 1–8.

Luo, T., & Xie, K. (2023). AI tools in higher education: Student dependency, self-regulation and academic performance. *Journal of Educational Technology Research and Development, 71(4),* 1–22.

Perkins, M. (2023). Academic integrity considerations of AI large language models in the post-pandemic era. *Journal of University Teaching & Learning Practice, 20(2),* 7.

Welding, L. (2023). *Half of college students say using AI on schoolwork is cheating or plagiarism*. BestColleges.com Survey Report.