Báo cáo Nghiên cứu Chuyên sâu: Khung Đạo đức, Chính sách và Ứng dụng Trí tuệ Nhân tạo trong Nhiệm vụ Tổng hợp Tài liệu Học thuật

Sự bùng nổ của trí tuệ nhân tạo tạo sinh (Generative AI) đã tạo ra một cuộc cách mạng sâu sắc trong phương thức giáo dục và nghiên cứu tại các cơ sở giáo dục đại học.[1] Các mô hình ngôn ngữ lớn (LLM) như ChatGPT, Gemini hay Claude không chỉ là những công cụ truy xuất thông tin mà đã trở thành những "đối tác nhận thức" có khả năng tham gia vào quá trình tư duy, phân tích và sản xuất văn bản phức tạp.[2] Mặc dù mang lại tiềm năng to lớn trong việc tối ưu hóa hiệu suất học tập, việc ứng dụng AI đồng thời đặt ra những thách thức chưa từng có đối với hệ thống giá trị cốt lõi của môi trường hàn lâm, đặc biệt là liêm chính học thuật, đạo đức nghiên cứu và quyền sở hữu trí tuệ.[4] Dựa trên cấu trúc yêu cầu của một nhiệm vụ học tập điển hình tại Trường Đại học Công nghệ – Đại học Quốc gia Hà Nội (UET - VNU), báo cáo này được thiết kế để phân tích một cách toàn diện đa chiều về hệ sinh thái học thuật số. Thông qua việc đánh giá các chính sách hiện hành, thực nghiệm một nhiệm vụ tổng hợp tài liệu khoa học cụ thể (chủ đề An ninh Internet Vạn vật - IoT Security), phân tích chuyên sâu các khía cạnh đạo đức, và đề xuất các khung nguyên tắc cùng giải pháp trực quan hóa, nghiên cứu này thiết lập một nền tảng lý luận vững chắc cho việc sử dụng AI một cách minh bạch, có trách nhiệm và tuân thủ các chuẩn mực quốc tế.

1. Phân tích Chính sách Quản trị Trí tuệ Nhân tạo và Liêm chính Học thuật tại Đại học Quốc gia Hà Nội (VNU)

Trong bối cảnh hội nhập và chuyển đổi số toàn diện, hệ thống giáo dục đại học Việt Nam, với sự dẫn dắt của các tổ chức trọng điểm như Đại học Quốc gia Hà Nội (VNU), đang từng bước hoàn thiện hành lang pháp lý và các quy chuẩn học thuật để thích ứng với làn sóng công nghệ mới.[7] Các chính sách này không mang tính cấm đoán cực đoan mà hướng tới việc tích hợp công nghệ một cách có kiểm soát, tôn trọng quyền sở hữu trí tuệ và bảo vệ giá trị thực chất của tri thức.[9]

1.1. Khung Định hướng Chiến lược Quốc gia và ĐHQGHN

Ở cấp độ vĩ mô, Bộ Giáo dục và Đào tạo Việt Nam đã liên tục thúc đẩy các dự thảo và thông tư quy định về ứng dụng AI trong giáo dục đại học.[9] Các văn bản này nhấn mạnh yêu cầu phải chuẩn hóa hạ tầng dữ liệu, đảm bảo khả năng liên thông, kết nối an toàn, đồng thời bắt buộc phát triển năng lực số (digital competencies) cho cả người học, giảng viên và cán bộ quản lý.[9]

Việc phát triển này được kỳ vọng sẽ tạo ra cơ sở pháp lý vững chắc cho việc tiếp nhận công nghệ có trách nhiệm, đảm bảo chất lượng và bảo mật dữ liệu.[9]

Đối với Đại học Quốc gia Hà Nội, chiến lược chuyển đổi số đã được ban hành chính thức từ năm 2022 (Quyết định số 1323/QĐ-ĐHQGHN), bao trùm toàn diện các khía cạnh từ quản trị, đào tạo, nghiên cứu đến quản lý tài nguyên dữ liệu.[8] Tầm nhìn chiến lược này được tái khẳng định qua sự tham gia của lãnh đạo VNU tại các diễn đàn đối thoại chính sách cấp cao như EducAItion 2026, nơi các vấn đề về quản trị AI có trách nhiệm (responsible AI governance) và bảo vệ liêm chính học thuật được đặt lên hàng đầu nhằm thu hẹp khoảng cách kỹ thuật số.[12] Đáng chú ý, để hiện thực hóa chiến lược này thành hành động, VNU đã định hướng triển khai học phần mới mang mã số VNU1001 ("Nhập môn Công nghệ số và ứng dụng Trí tuệ nhân tạo") thay thế cho học phần Tin học cơ sở truyền thống đối với sinh viên tuyển sinh từ năm 2025.[13] Học phần này không chỉ cung cấp kỹ năng vận hành công cụ mà còn tập trung giáo dục ý thức đạo đức, an toàn không gian mạng và liêm chính học thuật, đảm bảo người học làm chủ công nghệ thay vì lệ thuộc thụ động.[13]

1.2. Chế tài Xử lý Vi phạm và Tiêu chuẩn Liêm chính tại UET và Các Trường Thành viên

Khái niệm "Liêm chính học thuật" tại VNU được định nghĩa là hệ thống các giá trị đòi hỏi sự trung thực, ngay thẳng, công bằng, tôn trọng và trách nhiệm trong mọi hoạt động học tập, giảng dạy và nghiên cứu khoa học.[14] Các hành vi vi phạm đạo đức học thuật, đặc biệt là đạo văn (plagiarism) – việc sử dụng câu chữ, dữ liệu, hoặc ý tưởng của người khác (bao gồm cả nội dung do máy móc tạo ra) mà không trích dẫn minh bạch – bị nghiêm cấm tuyệt đối.[14]

Tại các trường thành viên như Trường Quản trị và Kinh doanh (HSB) hay Trường Đại học Công nghệ (UET), các quy định chi tiết về xử lý vi phạm đã được luật hóa thành văn bản.[14] Các hình thức kỷ luật được phân cấp rõ ràng dựa trên mức độ vi phạm, dao động từ việc trừ 50% số điểm, hủy kết quả bài kiểm tra (0 điểm), yêu cầu làm lại sản phẩm học thuật, cho đến đình chỉ học tập có ghi chú vào hồ sơ.[14] Các hội đồng thẩm định tính liêm chính trong nghiên cứu khoa học được thiết lập để đánh giá các trường hợp có tỷ lệ trùng lặp cao. Ví dụ, một đề tài nghiên cứu sẽ bị từ chối nghiệm thu nếu mức độ trùng lặp vượt quá 25% đối với ấn phẩm công bố lần đầu, hoặc trên 70% đối với các sản phẩm tái bản.[17]

Sự xuất hiện của công cụ tạo sinh văn bản đã buộc hệ thống kiểm soát phải nâng cấp. Các đơn vị thuộc VNU đã áp dụng các hệ thống quét đạo văn chuyên sâu, trong đó giảng viên được khuyến khích sử dụng chức năng phát hiện AI của Turnitin để giám sát tiểu luận và báo cáo.[14] Các văn bản có tỷ lệ nghi ngờ được tạo bởi AI (AI suspicion level) vượt mức 90% sẽ tự động bị chuyển lên Hội đồng Liêm chính Học thuật của trường để điều tra.[14] Tương tự như chính sách tại các cơ sở tiên tiến khác ở Việt Nam (ví dụ: VinUni), giảng viên có toàn quyền sử dụng các công cụ phát hiện AI; khi có sự nghi ngờ, sinh viên phải có nghĩa vụ cung cấp các bằng chứng chứng minh quá trình làm việc độc lập của mình như bản nháp, ghi chú nghiên cứu, hoặc thậm chí phải thực hiện lại một phần bài làm dưới sự giám sát trực tiếp.[18] Sinh viên phải hoàn toàn

chịu trách nhiệm về nội dung do AI tạo ra, bởi lẽ hệ thống có thể sinh ra các thông tin sai lệch (hallucinations), định kiến hoặc vi phạm bản quyền.[19]

1.3. Vai trò của Hệ thống DoIT tại UET

Một trong những công cụ nội bộ đóng vai trò then chốt trong việc duy trì liêm chính học thuật tại Trường Đại học Công nghệ (UET) là hệ thống DoIT (Hệ thống hỗ trợ nâng cao chất lượng tài liệu).[15] DoIT không chỉ đơn thuần là một công cụ kiểm tra chính tả tiếng Việt mà còn là một nền tảng đối chiếu tài liệu trực tuyến mạnh mẽ, giúp sinh viên và giảng viên phát hiện tỷ lệ trùng lặp ý tưởng trước khi chính thức nộp bài.[15]

Từ nền tảng ban đầu dành riêng cho UET (http://doit.uet.vnu.edu.vn), hệ thống này đã được chiến lược hóa để mở rộng phục vụ toàn bộ Đại học Quốc gia Hà Nội thông qua hệ thống tập trung của Trung tâm Thư viện và Tri thức Số (http://doit.lic.vnu.edu.vn), yêu cầu xác thực bảo mật thông qua tài khoản email chính danh @vnu.edu.vn.[20] Việc tích hợp DoIT vào quy trình nộp bài tập chuẩn hóa giúp người học hình thành thói quen kiểm tra nguồn gốc, tạo màng lọc ban đầu chống lại sự lạm dụng việc sao chép từ mạng Internet và hệ thống AI.[15]

2. Quy trình Thực nghiệm Nhiệm vụ Học tập: Tổng hợp Tài liệu với Sự hỗ trợ của Trí tuệ Nhân tạo

Nhằm minh họa cụ thể cho năng lực phân tích của AI cũng như cách áp dụng các nguyên tắc đạo đức vào thực tiễn, một nhiệm vụ "Tổng hợp tài liệu học thuật" (Document Synthesis) đã được tiến hành. Chủ đề được lựa chọn là "Các thách thức và giải pháp An ninh mạng trong hệ sinh thái Internet Vạn vật (IoT Security) giai đoạn 2024-2025". Quá trình này đòi hỏi sự tương tác liên tục giữa tư duy phân tích của con người và năng lực xử lý ngôn ngữ của hệ thống mô hình ngôn ngữ lớn (LLM), kết hợp với việc áp dụng nghiêm ngặt các quy chuẩn trích dẫn quốc tế.

2.1. Quá trình Tương tác và Khởi tạo Nội dung (Prompts & AI Outputs)

Quá trình tổng hợp tài liệu được thực hiện thông qua hai vòng lặp truy vấn (iterative prompting) nhằm thu thập, phân loại và hệ thống hóa khối lượng lớn các nghiên cứu chuyên sâu về an ninh IoT.[21] Sự can thiệp của AI được thiết kế theo hướng "tương tác phản biện", tức là AI đóng vai trò như một bộ lọc dữ liệu thô, trong khi việc hình thành khung lý thuyết do con người định hướng.

Vòng Truy vấn	Câu lệnh đầu vào (Prompt) được thiết kế	Tóm tắt Đầu ra của Hệ thống AI (AI Output)
Vòng 1: Phân tích Lỗ hổng	"Hãy đóng vai trò là một chuyên gia an ninh mạng	AI cung cấp một phân tích cấu trúc.[24] Lớp Cảm nhận

	độc lập. Dựa trên các tài liệu từ IEEE và MDPI (2024-2025), hãy tổng hợp các lỗ hổng bảo mật cốt lõi trong kiến trúc IoT. Phân loại theo mô hình ba lớp (Cảm nhận, Mạng, Ứng dụng) và giải thích nguyên nhân gốc rễ."	gặp rủi ro thao túng vật lý do thiết bị bị hạn chế tài nguyên (resource-constrained) không thể chạy mã hóa nặng.[25] Lớp Mạng đối diện với sự cố hủy xác thực (deauthentication) trên Wi-Fi và tấn công DDoS.[24] Lớp Ứng dụng bị ảnh hưởng bởi lỗ hổng Zero-day.[21]
Vòng 2: Phân tích Giải pháp AI	"Phân tích cách các công nghệ AI tiên tiến (như Federated Learning, XAI, GANs) đang được sử dụng để xây dựng hệ thống phòng thủ cho IoT. Cung cấp cơ chế hoạt động ngắn gọn và lợi ích của từng công nghệ."	AI hệ thống hóa các giải pháp phòng thủ hiện đại.[21] Federated Learning (Học liên kết) bảo vệ quyền riêng tư bằng cách huấn luyện mô hình cục bộ không cần chia sẻ dữ liệu gốc.[21] Explainable AI (XAI) minh bạch hóa hệ thống phát hiện xâm nhập (IDS).[21] GANs giúp mô phỏng tấn công để tạo dữ liệu huấn luyện.[21]

2.2. Đánh giá, Chỉnh sửa và Tích hợp Kết quả (Human Oversight & Integration)

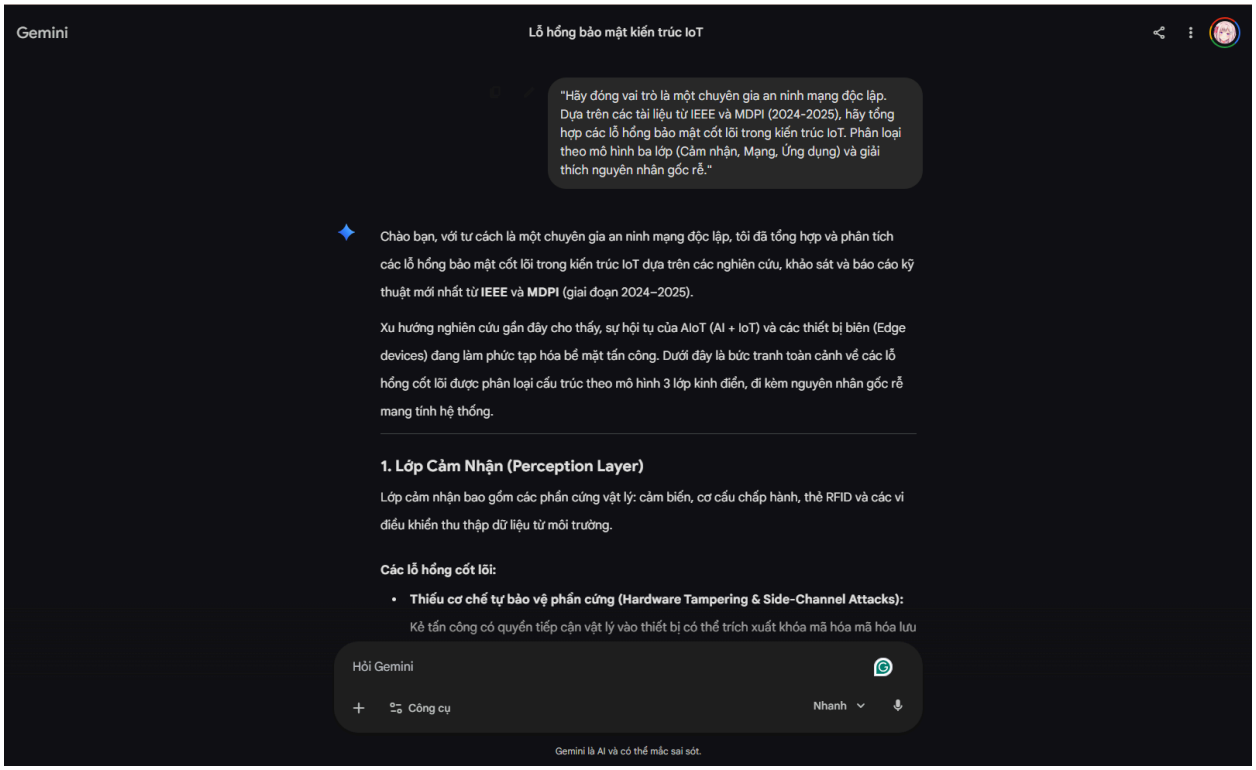
Việc tiếp nhận toàn bộ nội dung do AI sinh ra mang tính rủi ro rất cao do hiện tượng "ảo giác" (hallucinations), nơi máy học tạo ra các tài liệu tham khảo giả mạo hoặc kết luận sai lệch.[2] Do đó, quá trình giám sát của con người (human oversight) được thiết lập thông qua các bước kiểm chứng (fact-checking) chéo với hệ thống cơ sở dữ liệu học thuật.

Đầu tiên, đối với lỗ hổng mạng không dây do AI đề xuất, thông tin đã được đối chiếu chéo với báo cáo thực tế của Nozomi Networks năm 2025, xác nhận chính xác số liệu 94% mạng Wi-Fi công nghiệp thiếu bảo vệ chống tấn công hủy xác thực.[26] Thứ hai, đối với các giải pháp chuẩn hóa, AI bỏ sót bối cảnh quản trị rủi ro tổng thể. Nhà nghiên cứu đã chủ động tích hợp thêm thông tin từ tài liệu tiêu chuẩn ISO/IEC 30141 (tái bản 2024) và NIST 8259 để bổ sung khung lý thuyết về quản lý tính toàn vẹn và xác thực danh tính.[22] Cuối cùng, văn phong của AI (thường mang tính liệt kê cứng nhắc) đã được viết lại hoàn toàn (paraphrased) thành dạng văn xuôi

tường thuật học thuật mượt mà, đảm bảo sự tích hợp hữu cơ giữa lý thuyết an ninh mạng và các bằng chứng thực tiễn, đồng thời loại bỏ những khái niệm mà AI diễn giải chưa đủ độ sâu. Điều này bảo vệ tiếng nói cá nhân (authorial voice) của người nghiên cứu và xóa bỏ dấu vết của lối hành văn công nghiệp.[4]

Kết quả của quá trình tích hợp trên đã hình thành nên một đoạn tổng hợp tài liệu chuyên sâu như sau: Kiến trúc phi tập trung và tính chất không đồng nhất của IoT đang tạo ra một "bề mặt tấn công" khổng lồ.[24] Sự ràng buộc nghiêm ngặt về năng lượng và bộ nhớ khiến các thiết bị phần cứng không thể tiếp nhận các giao thức mã hóa thông thường, buộc phải dùng mật khẩu tĩnh, từ đó dễ dàng bị xâm nhập.[24] Để đối phó, các công nghệ phòng thủ hướng dữ liệu đã ra đời, tiêu biểu là Học liên kết (Federated Learning) giúp bảo vệ quyền riêng tư trong chăm sóc sức khỏe thông minh, và Trí tuệ Bầy đàn (Swarm Intelligence) giúp tăng cường khả năng thích ứng của mạng không dây trước sự biến đổi liên tục của mã độc.[21]

2.3. Trích dẫn việc sử dụng AI



Gemini

Lỗ hổng bảo mật kiến trúc IoT

1. Lớp Cảm Nhận (Perception Layer)

Lớp cảm nhận bao gồm các phần cứng vật lý: cảm biến, cơ cấu chấp hành, thẻ RFID và các vi điều khiển thu thập dữ liệu từ môi trường.

Các lỗ hổng cốt lõi:

- **Thiếu cơ chế tự bảo vệ phần cứng (Hardware Tampering & Side-Channel Attacks):** Kẻ tấn công có quyền tiếp cận vật lý vào thiết bị có thể trích xuất khóa mã hóa mã hóa lưu trong bộ nhớ Flash hoặc EEPROM thông qua phân tích điện năng tiêu thụ (DPA) hoặc phân tích điện từ (EMA).
- **Chèn mã độc thông qua cập nhật Firmware (Firmware Injection):** Nhiều thiết bị IoT giá rẻ không có cơ chế **Secure Boot** (Khởi động an toàn). Kẻ tấn công có thể nạp một bản firmware đã chỉnh sửa để biến thiết bị thành mã độc nằm vùng.
- **Giả mạo nút/thiết bị (Node Cloning / Spoofing):** Do thiếu định danh duy nhất không thể làm giả (như PUF - Physical Unclonable Function), kẻ tấn công dễ dàng nhân bản một thiết bị hợp lệ để đưa dữ liệu rác vào hệ thống.

Nguyên nhân gốc rễ:

- **Ràng buộc ngặt nghèo về tài nguyên (Resource Constraints):** Các chip vi điều khiển (MCU) ở lớp này thường chạy trên kiến trúc 8-bit hoặc 32-bit công suất thấp (Low-power). Chúng không đủ tài nguyên tính toán (CPU, RAM, PIN) để thực thi các thuật toán mã hóa hết đối xứng phức tạp (như RSA hoặc ECC

Hỏi Gemini

Công cụ

Nhanh

Gemini là AI và có thể mắc sai sót.

Gemini

AI Phòng Thủ IoT Tiên Tiến

"Phân tích cách các công nghệ AI tiên tiến (như Federated Learning, XAI, GANs) đang được sử dụng để xây dựng hệ thống phòng thủ cho IoT. Cung cấp cơ chế hoạt động ngắn gọn và lợi ích của từng công nghệ."

Sự bùng nổ của các thiết bị IoT mang lại sự tiện lợi vô song, nhưng cũng biến chúng thành miếng mồi ngon cho tội phạm mạng do cấu hình phần cứng hạn chế và tính bảo mật lỏng lẻo. Để đối phó với các cuộc tấn công ngày càng tinh vi, các công nghệ AI tiên tiến như **Federated Learning (Học máy liên kết)**, **XAI (AI có thể giải thích)**, và **GANs (Mạng đối nghịch tạo ra)** đang trở thành những trụ cột mới trong việc xây dựng hệ thống phòng thủ vững chắc cho IoT.

Dưới đây là phân tích chi tiết về cơ chế hoạt động và lợi ích của từng công nghệ:

---

1. Federated Learning (Học máy liên kết) – Bảo mật dữ liệu tại nguồn

Các hệ thống IoT truyền thống thường phải gom toàn bộ dữ liệu về một máy chủ trung tâm để huấn luyện AI phát hiện mã độc. Điều này vừa gây nghẽn băng thông vừa làm rò rỉ thông tin riêng tư. Federated Learning giải quyết triệt để bài toán này.

Cơ chế hoạt động

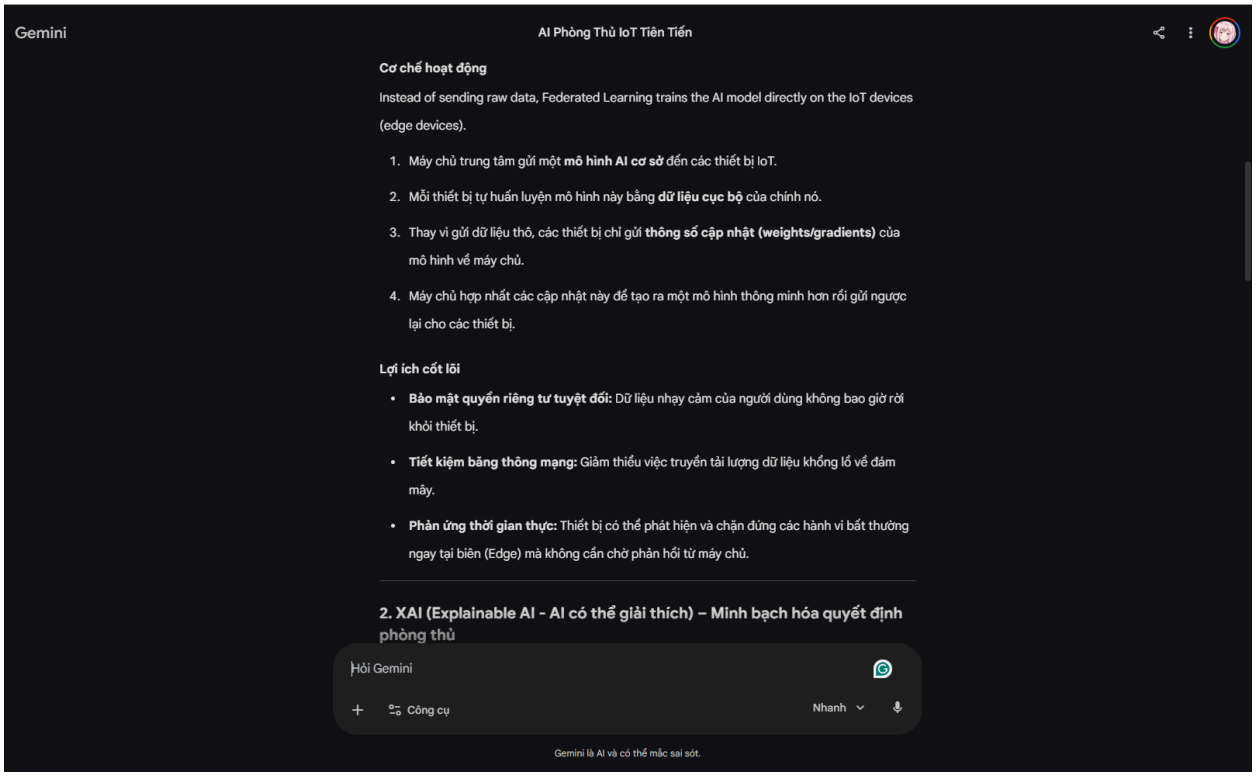
Instead of sending raw data, Federated Learning trains the AI model directly on the IoT devices (edge devices).

Hỏi Gemini

Công cụ

Nhanh

Gemini là AI và có thể mắc sai sót.



3. Phân tích Chuyên sâu Các Vấn đề Đạo đức Phát sinh từ Ứng dụng AI Học thuật

Việc sử dụng các công cụ trí tuệ nhân tạo tạo sinh trong quá trình kiến tạo tri thức học thuật không đơn thuần là một thao tác công nghệ, mà đang tạo ra những rạn nứt cấu trúc trong các định chế đạo đức học thuật truyền thống. Sự giao thoa giữa trí tuệ con người và thuật toán dấy lên những tranh luận nảy lửa về định nghĩa thực sự của sự sáng tạo, giới hạn của sự trung thực và tính toàn vẹn của một sản phẩm nghiên cứu khoa học.[1]

3.1. Ranh giới Giữa Hỗ trợ Hợp lý và Gian lận Học thuật (Academic Cheating)

Xác định ranh giới giữa việc dùng công cụ để hỗ trợ quá trình nhận thức (cognitive scaffolding) và hành vi gian lận (academic cheating) là thử thách phức tạp nhất đối với các nhà hoạch định giáo dục hiện nay.[41] Hỗ trợ hợp lý (reasonable support) được xác định khi công cụ công nghệ đóng vai trò xúc tác, giúp làm sáng tỏ cấu trúc logic, hiệu đính lỗi cú pháp (đặc biệt đối với rào cản ngôn ngữ của người học quốc tế), hoặc kích thích các ý tưởng sơ khởi.[29] Trong trường hợp này, AI hoạt động tương tự như một máy tính điện tử trong môn toán học: nó xử lý phần việc cơ học, nhưng con người vẫn kiểm soát định hướng giải quyết vấn đề và là người đưa ra kết luận cuối cùng.[43]

Ngược lại, gian lận học thuật diễn ra khi sự hỗ trợ vượt ngưỡng thành sự thay thế toàn diện, tạo ra hiện tượng "làm mướn hợp đồng thông qua AI" (AI-based contract cheating).[41] Khi sinh

viên sử dụng câu lệnh yêu cầu AI tạo ra toàn bộ cấu trúc, luận điểm và văn bản của một bài luận, sau đó chỉ nộp lại dưới danh nghĩa cá nhân, hành vi này phá vỡ hợp đồng cốt lõi của giáo dục – đo lường năng lực thực chất của người học.[39] Hệ lụy của việc này là sự phai mòn giá trị của hệ thống đánh giá. Khảo sát cho thấy phần lớn sinh viên nhận thức được rằng việc dùng AI để viết bài thay là hành vi phi đạo đức, gây phương hại đến sự công bằng đối với những sinh viên lao động trí óc chân chính.[44] Ranh giới này càng trở nên mờ nhạt do sự thiếu vắng các tài liệu hướng dẫn thể chế chi tiết ở một số khu vực, khiến sự nghi ngờ về năng lực sáng tạo của người học ngày càng gia tăng, buộc các trường đại học phải triển khai diện rộng các nền tảng quét AI nhằm duy trì sự công bằng.[4]

3.2. Quyền Sở hữu Trí tuệ (Intellectual Property) và Lỗ hổng Trích dẫn

Một rào cản triết học và pháp lý khổng lồ của AI tạo sinh liên quan đến vấn đề quyền tác giả (authorship) và bản quyền dữ liệu.[6] Trong mọi văn bản quy phạm về xuất bản học thuật của các tổ chức quốc tế, có một sự đồng thuận mang tính phổ quát: AI không có quyền nhân thân, không có năng lực hành vi dân sự, không chịu trách nhiệm giải trình và không thể đồng thuận với các thỏa thuận cấp phép bản quyền.[45] Do đó, AI hoàn toàn bị tước bỏ tư cách tác giả (author) hoặc đồng tác giả (co-author) trong mọi công bố khoa học.[6]

Hơn thế nữa, tính chất thiết kế của các mô hình như GPT dựa trên nền tảng "Hộp đen" (Black-box), nơi chúng được huấn luyện bằng cách "cào" (scraping) hàng triệu văn bản, sách và nghiên cứu của con người từ Internet mà không cần xin phép hay bồi thường cho tác giả gốc.[6] Khi AI tổng hợp các chuỗi từ vựng này thành một đoạn văn bản mới, nó tạo ra một cấu trúc ngôn ngữ chưa từng xuất hiện, đánh lừa các công cụ phát hiện đạo văn truyền thống chỉ dựa trên đối sánh chuỗi ký tự (string-matching).[41] Tuy nhiên, về mặt ý tưởng và lý luận, hành vi này thực chất là một sự đạo văn ẩn danh quy mô lớn (plagiarism of uncredited human ideas).[6] Các nhà xuất bản đều khẳng định rằng việc đệ trình các phân đoạn do AI viết mà không minh bạch nguồn gốc cấu thành tội danh đạo văn, phá vỡ nguyên tắc minh bạch xuất xứ (provenance) của khoa học.[45] Điều này đòi hỏi các nhà khoa học phải cực kỳ thận trọng: khi đưa ý tưởng của AI vào bài, họ phải có khả năng truy xuất ngược lý luận đó về các tài liệu nghiên cứu của con người đã được thẩm định (peer-reviewed), nếu không sẽ dẫn đến việc lồng ghép các kết quả thực nghiệm giả tưởng vào chuỗi tri thức nhân loại.[29]

3.3. Tác động Nhận thức đến Quá trình Học tập và Phát triển Kỹ năng

Đằng sau những rủi ro về thủ tục hành chính, tác động sâu xa nhất của AI nằm ở việc định hình lại năng lực nhận thức (cognitive development) và thói quen giải quyết vấn đề của sinh viên.[48] Việc sử dụng AI đòi hỏi người dùng chia sẻ "tải trọng nhận thức" (cognitive load) với máy tính.[49] Tùy thuộc vào chiến lược sử dụng, sự chia sẻ này có thể mang tính hủy diệt hoặc nâng tầm.

Các nghiên cứu gần đây trong lĩnh vực tương tác Người-Máy (HAI) phân tích thói quen sử dụng AI qua lăng kính các hệ thống tư duy. Khi sinh viên hình thành một vòng lặp thói quen "Hệ thống 1" (System-1 habit) – tức là tiếp cận một vấn đề khó (problem) và lập tức chuyển giao (offload) toàn bộ thao tác tư duy cho AI để nhận kết quả tức thì – họ đã tước bỏ cơ hội của não

bộ trong việc rèn luyện sự bền bỉ nhận thức.[48] Sự ủy thác cơ học (routine usage) này bào mòn động lực tìm kiếm sự thật, làm thui chột nhu cầu thấu hiểu (need for understanding), kỹ năng phản tư (reflection) và năng lực đánh giá phản biện.[48] Hậu quả lâu dài là sự suy giảm khả năng sáng tạo gốc và tạo ra một lực lượng lao động tri thức phụ thuộc hoàn toàn vào công cụ.[4]

Tuy nhiên, nếu được định hướng đúng đắn, sự can thiệp của AI có thể đóng vai trò "sử dụng chuyển đổi" (transformative AI use).[49] Khi sinh viên áp dụng các chiến lược đối thoại phản biện (critical dialogic strategies), AI trở thành một tấm gương phản chiếu, giúp họ mở rộng năng lực tâm lý (mind expansion), liên kết các lĩnh vực liên ngành và khám phá các giả thuyết mới (frame generation).[49] Thống kê chỉ ra rằng nhóm sinh viên sử dụng AI với động lực học hỏi sâu thường có mức độ tự tin trong nghiên cứu (research self-efficacy) và mức độ tham gia vào bài học (engagement) cao hơn hẳn, do AI giúp giải phóng họ khỏi những bế tắc ngôn từ ban đầu, tạo không gian cho những suy nghĩ bậc cao (higher-order thinking).[49]

Loại hình Tương tác với AI	Đặc điểm Nhận thức (Cognitive Characteristics)	Hệ quả đối với Năng lực Học thuật
Giao phó Cơ học (Routine Offloading / System 1)	Tự động hóa giải pháp, né tránh tư duy phức tạp, chấp nhận mọi kết quả đầu ra vô điều kiện.[48]	Bào mòn tư duy phản biện, giảm sự sáng tạo, gia tăng nguy cơ gian lận hợp đồng (contract cheating).[4]
Sử dụng Chuyển đổi (Transformative / Dialogic)	Phản biện liên tục với đầu ra, tái cấu trúc logic, tìm kiếm sự thấu hiểu sâu sắc (Need for understanding).[49]	Nâng cao kỹ năng lập luận, mở rộng năng lực giải quyết vấn đề, củng cố sự tự tin nghiên cứu (self-efficacy).[49]

4. Xây dựng Bộ Nguyên tắc Cá nhân về Sử dụng Trí tuệ Nhân tạo Có trách nhiệm

Để định vị bản thân một cách an toàn và có đạo đức giữa dòng chảy công nghệ, mỗi cá nhân trong môi trường học thuật cần tự thiết lập một hệ trục giá trị nội tâm. Dựa trên việc tổng hợp các chính sách của các viện hàn lâm hàng đầu, bộ 7 nguyên tắc dưới đây được đề xuất như một kim chỉ nam hành động, đảm bảo việc khai thác AI không đi ngược lại mục đích tối thượng của giáo dục:

Số Thứ Tự	Tên Nguyên tắc	Giải nghĩa và Cơ sở Lý luận Đạo đức

01	**Bảo toàn Tư duy Độc lập (Cognitive Sovereignty)**	Chỉ sử dụng AI như một bệ phóng ý tưởng hoặc công cụ kiểm tra thứ cấp. Tuyệt đối không nhượng quyền tư duy phân tích, đánh giá và định hướng nghiên cứu cốt lõi cho thuật toán. Sự thấu hiểu cá nhân luôn phải đi trước sự hỗ trợ của máy móc.[43]
02	**Trách nhiệm Giải trình Vô điều kiện (Absolute Accountability)**	Người đứng tên sản phẩm học thuật phải chịu trách nhiệm 100% đối với mọi lỗi sai, dữ liệu giả mạo (hallucinations), định kiến hoặc vi phạm bản quyền xuất hiện trong kết quả cuối cùng. AI không phải là tác giả và không chịu bất kỳ hệ lụy đạo đức nào.[19]
03	**Minh bạch và Ghi công Tuyệt đối (Transparency & Attribution)**	Mọi hình thức can thiệp của AI (viết nháp, chỉnh sửa ảnh, lập trình mã code) phải được công bố rõ ràng theo quy chuẩn học thuật của tổ chức sở tại (IEEE, APA). Sự minh bạch loại bỏ hoàn toàn ranh giới mờ nhạt của hành vi đạo văn tinh vi.[32]
04	**Kiểm chứng Kép qua Dữ liệu Thực tế (Rigorous Fact-Checking)**	Bất kỳ số liệu, trích dẫn, nguyên lý hay kết luận nào do AI sinh ra đều phải bị coi là "thông tin chưa được xác thực" cho đến khi được đối chiếu chéo (cross-reference) với các nguồn tài liệu thứ cấp đáng tin cậy đã qua bình duyệt (peer-reviewed).[6]

05	**Bảo mật và Tôn trọng Quyền riêng tư (Data Confidentiality)**	Không đưa các dữ liệu chưa công bố, bản thảo học thuật của cá nhân/tổ chức, mã nguồn độc quyền, hoặc thông tin định danh cá nhân vào các nền tảng AI mã nguồn mở để tránh rủi ro rò rỉ và trở thành bộ dữ liệu huấn luyện ngoài ý muốn.[19]
06	**Chống lại Sự Ảo tưởng Năng lực (Prevent Competence Illusion)**	Liên tục phản tư (reflect) để nhận thức rõ giới hạn của bản thân. Điểm số hay bài viết xuất sắc nhờ AI không đồng nghĩa với việc cá nhân đã hấp thụ tri thức đó. Phải đảm bảo kỹ năng nghề nghiệp thực tế phát triển song song với năng lực sử dụng công cụ.[48]
07	**Tuân thủ Hệ sinh thái Đạo đức Thể chế (Institutional Compliance)**	Hiểu rõ và chấp hành vô điều kiện các ngưỡng phần trăm cho phép, hướng dẫn kiểm duyệt và chế tài xử lý của trường đại học sở tại (như VNU, UET). Công nghệ thay đổi nhưng quy tắc danh dự của học thuật mang tính trường tồn.[19]

5. Thiết kế Trực quan: Phát triển Infographic "Sử dụng AI Có trách nhiệm trong Học thuật"



Nguồn trích dẫn

1. EJ1435770 - The More Things Change: The Ethical Impacts of Artificial Intelligence in Higher Education, Research Issues in Contemporary Education, 2024 - ERIC, truy cập vào tháng 5 16, 2026, https://eric.ed.gov/?id=EJ1435770
2. Artificial Intelligence: Ethical Considerations In Academia - MDPI Blog, truy cập vào tháng 5 16, 2026, https://blog.mdpi.com/2024/02/01/ethical-considerations-artificial-intelligence/
3. How to cite ChatGPT - APA Style, truy cập vào tháng 5 16, 2026, https://apastyle.apa.org/blog/how-to-cite-chatgpt
4. Ethical Problems in the Use of Artificial Intelligence by University Educators - MDPI, truy cập vào tháng 5 16, 2026, https://www.mdpi.com/2227-7102/15/10/1322
5. Ethics of the Use of Artificial Intelligence in Academia and Research: The Most Relevant Approaches, Challenges and Topics - MDPI, truy cập vào tháng 5 16, 2026, https://www.mdpi.com/2227-9709/12/4/111
6. Research ethics and issues regarding the use of ChatGPT-like artificial intelligence platforms by authors and reviewers: a narrative review - Science Editing, truy cập vào tháng 5 16, 2026, https://www.escienceediting.org/journal/view.php?number=344
7. Applying AI in higher education – An inevitable trend for the future | Nhan Dan Online, truy cập vào tháng 5 16, 2026, https://en.nhandan.vn/applying-ai-in-higher-education-an-inevitable-trend-for-the-future-post155207.html
8. Vietnam University Development of VNU-Hanoi, VNU-HCM, and UD (P166656) -

Documents & Reports - World Bank, truy cập vào tháng 5 16, 2026, https://documents1.worldbank.org/curated/en/099152502062335188/pdf/P1666560462d7d0e509f3304ca131a97289.pdf
9. Vietnam government's draft circular to promote responsible use of tech, AI in higher education - Asia News Network, truy cập vào tháng 5 16, 2026, https://asianews.network/minister-urges-graduates-to-shape-cambodias-ai-future-with-integrity-ethics/
10. Ministry of Education and Training proposes regulations on AI in higher education, truy cập vào tháng 5 16, 2026, https://news.laodong.vn/giao-duc/bo-giao-duc-va-dao-tao-de-xuat-quy-dinh-ai-trong-giao-duc-dai-hoc-1697101.ldo
11. OVERVIEW OF AI APPLICATIONS IN THE FIELD EDUCATION IN VIETNAM, truy cập vào tháng 5 16, 2026, https://tdmujournal.vn/uploads/paper/files/31-Ngo-Thi-Ngoc-Diu-S.pdf
12. VNU's participation in EducAItion 2026 in Israel affirms the promotion of education through the application of artificial intelligence - Vietnam National University, Hanoi, truy cập vào tháng 5 16, 2026, https://en.vnu.edu.vn/-post36763.html
13. VNU first-year students will study AI and digital technologies, truy cập vào tháng 5 16, 2026, https://en.vnu.edu.vn/-post36590.html
14. Academic Integrity | HSB - VNU, truy cập vào tháng 5 16, 2026, https://www.hsb.edu.vn/recruitment-policy/academic-integrity
15. Bài giảng Công nghệ thông tin: Học tập và làm việc trong môi trường số - Chương 3, truy cập vào tháng 5 16, 2026, https://tailieu.vn/doc/bai-giang-cong-nghe-thong-tin-chuong-3-hoc-tap-va-lam-viec-trong-moi-truong-so-2975178.html
16. HR Policies | HSB - VNU, truy cập vào tháng 5 16, 2026, https://www.hsb.edu.vn/audience/hr-policies
17. 1786-QĐ-ĐHKTQD-12-10-2021-Quyết định quy định về Liêm chính học thuật.pdf - PHÒNG KHẢO THÍ VÀ ĐẢM BẢO CHẤT LƯỢNG GIÁO DỤC, truy cập vào tháng 5 16, 2026, https://khaothi-dbclgd.neu.edu.vn/Resources/Docs/SubDomain/khaothi-dbclgd/Khao%20thi/1786-Q%C4%90-%C4%90HKTQD-12-10-2021-Quy%E1%BA%BFt%20%C4%91%E1%BB%8Bnh%20quy%20%C4%91%E1%BB%8Bnh%20v%E1%BB%81%20Li%C3%AAm%20ch%C3%ADnh%20h%E1%BB%8Dc%20thu%E1%BA%ADt.pdf
18. Guidelines on Student Use of Generative Artificial Intelligence - VinUni Policy - VinUniversity, truy cập vào tháng 5 16, 2026, https://policy.vinuni.edu.vn/all-policies/guidelines-on-student-use-of-generative-artificial-intelligence/
19. AI Guidelines | Villanova University, truy cập vào tháng 5 16, 2026, https://www.villanova.edu/university/ai/governance/guidelines.html
20. Đăng ký | DoIT - Hệ thống hỗ trợ nâng cao chất lượng tài liệu - VNU, truy cập vào tháng 5 16, 2026, http://doit.uet.vnu.edu.vn/register
21. Securing the future: AI-driven cybersecurity in the age of autonomous IoT - Frontiers, truy cập vào tháng 5 16, 2026,

https://www.frontiersin.org/journals/the-internet-of-things/articles/10.3389/friot.2025.1658273/full
22. A Literature Review on Security in the Internet of Things: Identifying and Analysing Critical Categories - MDPI, truy cập vào tháng 5 16, 2026, https://www.mdpi.com/2073-431X/14/2/61
23. IoT, Volume 6, Issue 1 (March 2025) – 19 articles, truy cập vào tháng 5 16, 2026, https://www.mdpi.com/2624-831X/6/1
24. A Systematic Review on IoT Security (Threats, Mitigations) Strategies and Future Directions, truy cập vào tháng 5 16, 2026, https://www.researchgate.net/publication/394113036_A_Systematic_Review_on_IoT_Security_Threats_Mitigations_Strategies_and_Future_Directions
25. Harnessing Generative AI and Large Language Models for Revolutionizing Cybersecurity in the Internet of Things: Ethical and Privacy Implications - Beadle Scholar, truy cập vào tháng 5 16, 2026, https://scholar.dsu.edu/cgi/viewcontent.cgi?article=1465&context=bispapers
26. OT/IoT Cybersecurity Trends and Insights - GitHub, truy cập vào tháng 5 16, 2026, https://raw.githubusercontent.com/jacobdjwilson/awesome-annual-security-reports/main/Annual%20Security%20Reports/2025/Nozomi-Networks-OT-IoT-Security-Report-2025.pdf
27. The Future of AI-Driven Cybersecurity for Advanced IoT | Request PDF - ResearchGate, truy cập vào tháng 5 16, 2026, https://www.researchgate.net/publication/396211797_The_Future_of_AI-Driven_Cybersecurity_for_Advanced_IoT
28. IoT Security - MDPI, truy cập vào tháng 5 16, 2026, https://mdpi-res.com/bookfiles/book/11642/IoT_Security.pdf?v=1775437892
29. The promise and perils of using AI for research and writing, truy cập vào tháng 5 16, 2026, https://www.apa.org/topics/artificial-intelligence-machine-learning/ai-research-writing
30. Security Challenges in IoT: A Systematic Literature Review - Everant Journals, truy cập vào tháng 5 16, 2026, https://everant.org/index.php/etj/article/view/2062
31. Securing Internet of Things (IoT) ecosystems: Addressing scalability, authentication, and privacy challenges, truy cập vào tháng 5 16, 2026, https://wjarr.com/sites/default/files/fulltext_pdf/WJARR-2025-0999.pdf
32. Citation Guide for Use of AI in Research and Academic Writing - Augusta University, truy cập vào tháng 5 16, 2026, https://www.augusta.edu/ai/documents/ai_citation_guide_v2.pdf
33. Generative AI (eg., ChatGPT) - Concordia Library, truy cập vào tháng 5 16, 2026, https://library.concordia.ca/help/citing/ieee.php?guid=chatgpt
34. IEEE Referencing: A Quick Guide with Clear Examples - FinalScanPro, truy cập vào tháng 5 16, 2026, https://finalscanpro.com/site/blog/ieee-referencing-style-with-examples/
35. Guidelines for Generative AI Usage - IEEE Robotics and Automation Society Website, truy cập vào tháng 5 16, 2026,

https://www.ieee-ras.org/publications/guidelines-for-generative-ai-usage/
36. How do I cite generative AI in MLA style? (Updated and Revised), truy cập vào tháng 5 16, 2026, https://style.mla.org/citing-generative-ai-updated-revised/
37. Citing generative AI in APA Style: Part 1—Reference formats, truy cập vào tháng 5 16, 2026, https://apastyle.apa.org/blog/cite-generative-ai-references
38. AI references - APA Style - American Psychological Association, truy cập vào tháng 5 16, 2026, https://apastyle.apa.org/style-grammar-guidelines/references/examples/ai-references
39. ChatGPT Citations | Formats & Examples - Scribbr, truy cập vào tháng 5 16, 2026, https://www.scribbr.com/ai-tools/chatgpt-citations/
40. When and how to cite ChatGPT and AI in MLA / APA formats - Turnitin, truy cập vào tháng 5 16, 2026, https://www.turnitin.com/blog/when-and-how-to-cite-chatgpt-and-ai-in-mla-apa-formats
41. Generative AI and Academic Integrity in Higher Education: A Systematic Review and Research Agenda - MDPI, truy cập vào tháng 5 16, 2026, https://www.mdpi.com/2078-2489/16/4/296
42. Chatting and cheating: Ensuring academic integrity in the era of ChatGPT - Taylor & Francis, truy cập vào tháng 5 16, 2026, https://www.tandfonline.com/doi/full/10.1080/14703297.2023.2190148
43. Unlocking Your Creative Potential with AI: An Ethical Guide for College Students, truy cập vào tháng 5 16, 2026, https://www.gonzaga.edu/news-events/stories/2024/3/26/unlocking-your-creative-potential-with-ai
44. AI and Academic Integrity: Exploring Student Perceptions and Implications for Higher Education - the College of Information - University of North Texas, truy cập vào tháng 5 16, 2026, https://ci.unt.edu/computational-humanities-information-literacy-lab/aiandai.pdf
45. Clarifying Ethical Dilemmas of Using Artificial Intelligence in Research Writing: A Rapid Review - ScholarWorks, truy cập vào tháng 5 16, 2026, https://scholarworks.waldenu.edu/cgi/viewcontent.cgi?article=1549&context=hlrc
46. Artificial Intelligence in Peer Review: Ethical Risks and Practical Limits - PMC - NIH, truy cập vào tháng 5 16, 2026, https://pmc.ncbi.nlm.nih.gov/articles/PMC12481007/
47. The ethics of using artificial intelligence in scientific research: new guidance needed for a new tool - PMC, truy cập vào tháng 5 16, 2026, https://pmc.ncbi.nlm.nih.gov/articles/PMC12057767/
48. Thinking Less, Trusting More: GenAI's Impacts on Students' Cognitive Habits - arXiv, truy cập vào tháng 5 16, 2026, https://arxiv.org/html/2601.22430v2
49. The Impact of Generative AI Use on Graduate Students' Research Competence: The Mediating Role of Critical Thinking and the Moderating Role of Research Self-Efficacy - PMC, truy cập vào tháng 5 16, 2026, https://pmc.ncbi.nlm.nih.gov/articles/PMC12937814/
50. (PDF) The Impact of Generative AI on University Students' Learning Experience:

A Study on Cognitive and Affective Outcomes - ResearchGate, truy cập vào tháng 5 16, 2026, https://www.researchgate.net/publication/398460632_The_Impact_of_Generative_AI_on_University_Students'_Learning_Experience_A_Study_on_Cognitive_and_Affective_Outcomes
51. The Impact of Artificial Intelligence (AI) on Students' Academic Development - MDPI, truy cập vào tháng 5 16, 2026, https://www.mdpi.com/2227-7102/15/3/343
52. Báo Cáo Bài 6 | PDF - Scribd, truy cập vào tháng 5 16, 2026, https://www.scribd.com/document/985836096/BAO-CAO-BAI-6
53. Using AI-Generated Content in an IEEE Article and its Review, truy cập vào tháng 5 16, 2026, https://ieee-aess.org/using-ai-generated-content-ieee-article-and-its-review
54. Generative Artificial Intelligence: Promoting academic integrity | education.vic.gov.au, truy cập vào tháng 5 16, 2026, https://www2.education.vic.gov.au/pal/generative-artificial-intelligence/guidance/promoting-academic-integrity
55. Student Academic Integrity - VinUni Policy - VinUniversity, truy cập vào tháng 5 16, 2026, https://policy.vinuni.edu.vn/all-policies/student-academic-integrity/
56. 40+ Infographic Ideas, Examples & Templates for 2026 - Venngage, truy cập vào tháng 5 16, 2026, https://venngage.com/blog/infographic-ideas/
57. 40 of the best infographics to inspire you - Canva, truy cập vào tháng 5 16, 2026, https://www.canva.com/learn/best-infographics/
58. 100+ good topics for infographics - Infogram, truy cập vào tháng 5 16, 2026, https://infogram.com/blog/topics-for-infographics/
59. Dự án Infographic: Sử dụng AI Có Trách Nhiệm Trong Học Tập - Studocu, truy cập vào tháng 5 16, 2026, https://www.studocu.vn/vn/document/truong-dai-hoc-kinh-te/tai-chinh-doanh-nghiep/du-an-infographic-su-dung-ai-co-trach-nhiem-trong-hoc-tap/150265592
60. Ai ethics infographic Images - Free Download on Freepik, truy cập vào tháng 5 16, 2026, https://www.freepik.com/free-photos-vectors/ai-ethics-infographic
61. Designing Infographics: Visual Representations for Enhancing Education, Communication, and Scientific Research - PMC, truy cập vào tháng 5 16, 2026, https://pmc.ncbi.nlm.nih.gov/articles/PMC9274103/
62. Ai Responsibility: Over 8194 Royalty-Free Licensable Stock Illustrations & Drawings, truy cập vào tháng 5 16, 2026, https://www.shutterstock.com/search/ai-responsibility?image_type=illustration
63. Responsible Ai Icon royalty-free images - Shutterstock, truy cập vào tháng 5 16, 2026, https://www.shutterstock.com/search/responsible-ai-icon
64. Responsible ai Icons & Symbols - Flaticon, truy cập vào tháng 5 16, 2026, https://www.flaticon.com/free-icons/responsible-ai