

ĐẠO ĐỨC TRONG SỬ DỤNG TRÍ TUỆ NHÂN TẠO TẠO SINH CỦA SINH VIÊN ĐẠI HỌC: NGHIÊN CỨU TRƯỜNG HỢP TẠI TRƯỜNG ĐẠI HỌC KHOA HỌC - ĐẠI HỌC THÁI NGUYÊN

ETHICS IN GENERATIVE ARTIFICIAL INTELLIGENCE USE AMONG UNIVERSITY STUDENTS: A CASE STUDY AT THE UNIVERSITY OF SCIENCES, THAI NGUYEN UNIVERSITY

Nguyễn Thị Trà My,
Tạ Thị Thảo,
Phạm Chiến Thắng⁺,
Phùng Phương Nga,
Vũ Thị Hạnh,
Ma Thị Thanh Hiếu,
Dương Thị Huyền

Trường Đại học Khoa học - Đại học Thái Nguyên
⁺Tác giả liên hệ • Email: thangpc@tnus.edu.vn

Article history

Received: 22/7/2026

Accepted: 02/8/2026

Published: 05/10/2026

Keywords

Ethical use of generative AI,
Generative AI in learning,
university students, academic
integrity, AI use disclosure

ABSTRACT

Generative artificial intelligence (GenAI) is reshaping university learning while raising concerns about academic integrity, accountability, privacy, bias, learner autonomy, and equitable access. This study examines how students in a bounded institutional case identify, judge, prioritise, and act on ethical issues related to GenAI use. A qualitative case study was conducted with 28 students at the University of Sciences, Thai Nguyen University. Six themes were developed: students viewed GenAI as a support tool while retaining final responsibility; the boundary between legitimate assistance and misconduct depended on whether GenAI displaced the targeted thinking; disclosure was endorsed but inconsistently practised; verification weakened under time pressure; privacy, bias, and unequal access remained ethical blind spots; and students expected task-specific guidance. The findings represent participants' accounts rather than direct observations of behaviour. The awareness-action gap was shaped by motivation, AI literacy, peer norms, assessment design, policy clarity, and moral disengagement. Universities should adopt a conditionally permissive, task-specific approach, require proportionate disclosure, strengthen verification and data-protection competencies, and document learning processes.

1. Mở đầu

Sự phổ biến của các hệ thống trí tuệ nhân tạo (AI) tạo sinh đang làm thay đổi cách sinh viên tìm kiếm thông tin, xây dựng ý tưởng, viết, lập trình, dịch thuật và hoàn thành nhiệm vụ học tập. Các khả năng này tạo cơ hội hỗ trợ phản hồi và cá nhân hóa, nhưng đồng thời làm mờ ranh giới giữa trợ giúp học tập và việc thay thế năng lực cần được đánh giá; phát sinh rủi ro về thông tin sai, nguồn trích dẫn không tồn tại, quyền riêng tư, thiên lệch và bất bình đẳng tiếp cận (Chan và Hu, 2023; Miao và Holmes, 2023; Tlili và cộng sự, 2023). Vì vậy, vấn đề trọng tâm không còn là cấm hay cho phép AI một cách tuyệt đối, mà là xác định mục đích, mức độ sử dụng, nghĩa vụ khai báo và cơ chế chịu trách nhiệm phù hợp với từng nhiệm vụ.

Ở cấp chuẩn mực, UNESCO (2021) đặt đạo đức AI trên nền tảng quyền con người, công bằng, minh bạch, quyền riêng tư và giám sát của con người. Khung năng lực AI dành cho HS, sinh viên tiếp tục nhấn mạnh tư duy lấy con người làm trung tâm, đạo đức AI, hiểu biết kỹ thuật và khả năng thiết kế có trách nhiệm (Miao và cộng sự, 2024). Tại Việt Nam, Thông tư số 02/2025/TT-BGDĐT quy định Khung năng lực số cho người học; Luật Bảo vệ dữ liệu cá nhân số 91/2025/QH15 có hiệu lực từ ngày 01/01/2026 tạo thêm cơ sở pháp lý cho yêu cầu sử dụng dữ liệu an toàn và có trách nhiệm (Bộ GD-ĐT, 2025; Quốc hội, 2025). Tuy nhiên, các nguyên tắc vĩ mô chỉ trở thành hành vi khi được chuyển hóa thành chỉ dẫn rõ ở cấp học phần và nhiệm vụ.

Nghiên cứu quốc tế cho thấy sinh viên nhìn nhận tích cực các lợi ích của AI tạo sinh nhưng đồng thời lo ngại về độ chính xác, quyền riêng tư, đạo đức và ảnh hưởng đến sự phát triển cá nhân (Chan và Hu, 2023). Đối với khai báo

sử dụng AI, Gonsalves (2025) phát hiện sự không tuân thủ có liên quan đến nỗi sợ hậu quả học tập, hướng dẫn mờ hồ, thực thi thiếu nhất quán và ảnh hưởng bạn bè. Những kết quả này gợi ý rằng nhận thức đạo đức không tự động chuyển thành hành động; hành vi còn phụ thuộc vào cấu trúc đánh giá và môi trường xã hội.

Tại Việt Nam, xu hướng nghiên cứu về việc sinh viên sử dụng AI tạo sinh đang phát triển nhanh trong những năm gần đây. Le và Tran (2024) ghi nhận sinh viên có nhận thức tương đối tích cực về tính hữu ích và dễ sử dụng của ChatGPT trong hỗ trợ bài tập. Nguyen và Goto (2024), qua việc khảo sát 1.386 sinh viên tại bốn trường thành viên của Đại học Thái Nguyên cũng cho thấy tỉ lệ thừa nhận gian lận bằng AI qua câu hỏi trực tiếp là 9,6%, trong khi ước tính bằng kỹ thuật list experiment là 23,7%; chênh lệch này cảnh báo thiên lệch mong muốn xã hội trong dữ liệu tự báo cáo. Tran và cộng sự (2025), qua khảo sát 269 sinh viên và phỏng vấn 15 người đã chỉ ra khoảng cách giữa nhận thức về đạo văn, hiểu biết quy định trích dẫn và thực hành khai báo ChatGPT. Ở góc nhìn giảng viên ngoại ngữ, Cong-Lem và cộng sự (2024) cũng nhấn mạnh nguy cơ phụ thuộc, gian lận và nhu cầu điều chỉnh chính sách, đánh giá...

Như vậy, khoảng trống không còn nằm ở việc “chưa có nghiên cứu” về sinh viên Việt Nam sử dụng AI. Vấn đề còn thiếu chính là những nghiên cứu giải thích định tính, tìm hiểu về cách người học chuyển từ nhận diện vấn đề đạo đức sang phán đoán, ưu tiên giá trị và hành động trong các tình huống cụ thể; đặc biệt khi liên chính học thuật tương tác với áp lực thời gian, chuẩn mực bạn bè, năng lực AI, quy định môn học, quyền riêng tư, thiên lệch và khả năng tiếp cận. Nghiên cứu này đóng góp bằng cách tích hợp mô hình bốn thành tố của Rest (1986), lý thuyết vô hiệu hóa đạo đức của Bandura (1999) và góc nhìn xã hội - kỹ thuật trong một nghiên cứu trường hợp có ranh giới tổ chức.

Trường Đại học Khoa học - Đại học Thái Nguyên được lựa chọn vì trường hợp này cho phép khảo sát sự đa dạng trải nghiệm trong một cơ sở đa ngành. Tính đến tháng 7/2026, Nhà trường đang đào tạo 23 ngành đại học, 10 chuyên ngành thạc sĩ và 6 chuyên ngành tiến sĩ, với hơn 10.000 người học (TNUS, 2026). Việc lựa chọn đồng thời dựa trên khả năng tiếp cận nghiên cứu. Do đó, trường hợp không được xem là đại diện thống kê cho các trường đại học Việt Nam. Nghiên cứu trả lời ba câu hỏi quan trọng: (1) Sinh viên nhận thức thế nào về cơ hội và rủi ro đạo đức khi sử dụng AI tạo sinh trong học tập? (2) Sinh viên xác lập ranh giới giữa sự hỗ trợ hợp lệ và hành vi vi phạm liên chính học thuật ra sao? (3) Những điều kiện cá nhân, công nghệ và giáo dục nào thúc đẩy hoặc cản trở hành vi sử dụng AI có trách nhiệm?

2. Phương pháp nghiên cứu

2.1. Thiết kế nghiên cứu và khung phân tích

Nghiên cứu sử dụng thiết kế nghiên cứu trường hợp định tính theo định hướng diễn giải. “Trường hợp” được giới hạn ở trải nghiệm đạo đức trong sử dụng AI tạo sinh của sinh viên hệ đại học chính quy tại Trường Đại học Khoa học - Đại học Thái Nguyên; ranh giới trải nghiệm được xác định bằng tiêu chí đã sử dụng ít nhất một công cụ AI trong học tập trong sáu tháng gần nhất tại thời điểm phỏng vấn. Thiết kế này phù hợp với mục tiêu giải thích cách người học tạo nghĩa và đưa ra quyết định trong quan hệ với loại nhiệm vụ, hình thức đánh giá, ngành học, bạn bè, giảng viên và điều kiện tiếp cận công nghệ. Mô hình bốn thành tố của Rest (1986) được dùng làm trục phân tích gồm: nhận diện vấn đề đạo đức; phán đoán phương án có thể biện minh; ưu tiên giá trị đạo đức trước lợi ích cạnh tranh; và năng lực thực thi ý định. Lý thuyết vô hiệu hóa đạo đức của Bandura (1999) bổ sung cơ chế giải thích những cách cá nhân tạm thời làm suy yếu tự giám sát, chẳng hạn phân tán trách nhiệm, giảm nhẹ hậu quả hoặc chuyển trách nhiệm. Các khái niệm này là lăng kính định hướng, không phải giả định rằng mọi sử dụng AI đều là sai phạm. Góc nhìn xã hội - kỹ thuật đặt quyết định cá nhân trong quan hệ với đặc tính công cụ, bản miễn phí/trả phí, mục tiêu đánh giá, chỉ dẫn của giảng viên, chuẩn mực bạn bè và nguồn lực nhà trường.

Bảng 1. Khung phân tích của nghiên cứu

Cấu phần lý thuyết	Câu hỏi phân tích dữ liệu
Nhận diện	Sinh viên nhận ra ai hoặc giá trị nào có thể bị ảnh hưởng trong quá trình sử dụng AI?
Phán đoán	Sinh viên dựa vào tiêu chí nào để xác định đúng và sai khi sử dụng AI cho mục đích học tập?
Ưu tiên	Giá trị đạo đức có được sinh viên ưu tiên trước điểm số, thời gian và sự tiện lợi khi dùng AI?
Thực thi	Sinh viên làm gì để đảm bảo việc sử dụng AI là đúng đạo đức theo quan điểm cá nhân của họ?
Vô hiệu hóa đạo đức	Sinh viên hợp thức hóa hành vi sử dụng AI “bắt chấp” quy định và hướng dẫn của giảng viên/nhà trường của mình bằng cách nào?
Bối cảnh xã hội - công nghệ	Những điều kiện nào có thể làm tăng hoặc giảm khả năng sử dụng AI có trách nhiệm?

2.2. Bối cảnh, người tham gia và chiến lược chọn mẫu

Dữ liệu được thu thập từ tháng 02 đến tháng 5/2026. Lời mời tham gia nghiên cứu được phát hành công khai qua các kênh thông tin nội bộ và diễn đàn sinh viên của Nhà trường. Để kiểm soát và giảm thiểu áp lực từ quan hệ quyền lực giữa giảng viên, CBQL và người học, quy trình tiếp cận ban đầu được thực hiện thông qua một trợ lý nghiên cứu độc lập. Người tham gia được đảm bảo tuyệt đối về việc đồng ý hay từ chối tham gia, cũng như mọi quan điểm được

chia sẻ trong phỏng vấn, hoàn toàn không liên quan và không ảnh hưởng đến điểm số, kết quả học tập hay đánh giá rèn luyện của các em tại Trường.

Mẫu gồm 28 sinh viên được chọn có chủ đích nhằm tạo độ biến thiên theo năm học, nhóm ngành, giới tính và mức độ sử dụng AI. Tiêu chí đưa vào gồm: *đang học đại học chính quy; đã sử dụng ít nhất một công cụ AI trong hoạt động học tập trong sáu tháng gần nhất; và đồng ý trao đổi về trải nghiệm sử dụng*. Người không đáp ứng một trong các tiêu chí trên được loại khỏi mẫu. Việc tuyển chọn nhằm thu nhận những trường hợp giàu thông tin, không nhằm ước lượng tỉ lệ cho quần thể. Quy mô 28 người tham gia được biện minh theo logic sức mạnh thông tin: câu hỏi nghiên cứu có phạm vi tập trung; mẫu có trải nghiệm trực tiếp với AI nhưng vẫn bảo đảm biến thiên; khung lý thuyết tương đối xác định; phỏng vấn có chiều sâu 45-60 phút; và chiến lược phân tích hướng đến diễn giải liên chủ đề (Malterud và cộng sự, 2016). Nhằm đảm bảo tính đại diện, độ bão hòa dữ liệu được kiểm soát chéo trên cả 4 nhóm ngành đào tạo (Báo chí, Luật, Khoa học tự nhiên - công nghệ, Ngôn ngữ). Trong quá trình phân tích, các chủ đề cốt lõi bắt đầu lặp lại và không xuất hiện thêm mã mới đáng kể từ sau cuộc phỏng vấn thứ 20.

2.3. Công cụ phỏng vấn

Hướng dẫn phỏng vấn bán cấu trúc được tổ chức quanh trải nghiệm dùng công cụ và nhiệm vụ, tiêu chí xác định đúng - sai, khai báo, kiểm chứng, dữ liệu nhập vào AI, cảm nhận về công bằng và kì vọng đối với giảng viên/nhà trường. Người phỏng vấn có thể thay đổi thứ tự và dùng câu hỏi gợi mở để làm rõ lí do, điều kiện và trường hợp trái chiều. Bảng 3 trình bày nhóm nội dung, không thay thế bản hướng dẫn phỏng vấn nguyên văn.

Bảng 3. Ảnh xạ nhóm câu hỏi phỏng vấn với khung phân tích

Câu phân	Nhóm câu hỏi gợi mở
Nhận diện	Trong tình huống này, ai hoặc giá trị nào có thể bị ảnh hưởng? Rủi ro nào dễ/khó nhận ra?
Phán đoán	Khi nào AI là hỗ trợ hợp lệ, khi nào trở thành gian lận? Vì sao?
Ưu tiên	Nếu thời hạn, điểm số hoặc bạn bè gây áp lực, điều gì quyết định lựa chọn?
Thực thi	Bạn kiểm chứng, khai báo và bảo vệ dữ liệu như thế nào? Điều gì cản trở?
Vô hiệu hóa	Sinh viên thường dùng lí do nào để biện minh cho việc không khai báo hoặc sao chép?
Bối cảnh	Quy định môn học, giá công cụ, chất lượng đầu ra và thái độ giảng viên ảnh hưởng thế nào?

2.4. Thu thập dữ liệu và đạo đức nghiên cứu

Trước khi tiến hành phỏng vấn, tất cả sinh viên đều được cung cấp một Phiếu thông tin nghiên cứu nêu rõ mục đích, quy trình, và quyền rút lui bất kì lúc nào mà không cần lí do. Người tham gia đã kí Giấy đồng thuận tham gia tự nguyện. Toàn bộ bản ghi âm và văn bản gõ băng được mã hóa ẩn danh ngay lập tức, lưu trữ trên hệ thống máy chủ có mật khẩu bảo vệ và chỉ các thành viên cốt cán của nhóm nghiên cứu mới có quyền truy cập. Dữ liệu gốc sẽ được hủy bỏ hoàn toàn sau 3 năm kể từ ngày công bố bài báo.

Các cuộc phỏng vấn bán cấu trúc được thực hiện trực tiếp, kéo dài 45-60 phút. Bản ghi âm được gõ băng để phục vụ phân tích. Trong bản thảo, người tham gia được mã hóa từ SV01 đến SV28; trích dẫn chỉ kèm năm học và nhóm ngành ở mức cần thiết cho diễn giải, không nêu tên hay thông tin định danh trực tiếp. Tiêu chí tham gia yêu cầu sinh viên đồng ý trao đổi về trải nghiệm sử dụng AI.

2.5. Phân tích dữ liệu, phản tư và độ tin cậy

Dữ liệu được phân tích bằng phương pháp Phân tích chủ đề phản tư theo sáu pha: làm quen với dữ liệu; tạo mã ban đầu; phát triển chủ đề; rà soát chủ đề; xác định và đặt tên chủ đề; và viết báo cáo (Braun và Clarke, 2022). Quá trình mã hóa kết hợp giữa suy diễn (dựa trên các khái niệm định hướng của Rest và Bandura) và quy nạp (từ cách người tham gia mô tả vùng xám, áp lực thời gian, khai báo, kiểm chứng và công bằng tiếp cận). NVivo 15 được sử dụng để tổ chức dữ liệu, với nguyên tắc phần mềm không thay thế quyết định diễn giải của nhà nghiên cứu.

Quá trình phân tích do hai thành viên thực hiện độc lập, sau đó thảo luận đối chiếu để đào sâu ý nghĩa thay vì chỉ tìm kiếm một bộ mã thống nhất mang tính cơ học. Nhóm nghiên cứu ý thức rõ vị thế của mình là những người đang trực tiếp tham gia công tác giảng dạy và quản lí đào tạo. Do đó, các tác giả liên tục duy trì nhật kí phản tư để tự kiểm duyệt, tránh áp đặt lăng kính chuẩn mực của người dạy lên tường thuật trải nghiệm của sinh viên. Quá trình rà soát đặc biệt chú ý đến tính nhất quán và các ca trái chiều (ví dụ: cùng một thao tác AI nhưng được xem là hợp lệ hay sai

phạm tùy thuộc vào mục tiêu đánh giá). Kết quả được báo cáo theo chuẩn minh bạch của nghiên cứu định tính, phân biệt rõ lời người tham gia, diễn giải của tác giả và khuyến nghị (O'Brien và cộng sự, 2014).

3. Kết quả nghiên cứu

Phân tích phát triển sáu chủ đề có quan hệ với nhau. Các từ như “một số”, “nhiều người tham gia” và “mẫu hình gợi ý” được dùng để mô tả dữ liệu trong mẫu, không phải tỉ lệ của quần thể. Bảng 4 hệ thống hóa mã phụ, bằng chứng đại diện, trường hợp đối chiếu và ý nghĩa đạo đức; các trích dẫn nguyên văn được giữ nguyên từ bản thảo gốc.

Bảng 4. Ma trận bằng chứng cho sáu chủ đề

Chủ đề	Mã phụ	Bằng chứng/đối chiếu	Ý nghĩa đạo đức
1. AI là công cụ hỗ trợ	Gợi ý; giải thích; sửa diễn đạt; trách nhiệm cuối cùng	SV06 nhấn mạnh phải tự lập luận và giải thích được bài nộp; đối chiếu: trách nhiệm ít được mở rộng đến dữ liệu và thiên lệch.	Trách nhiệm cá nhân được thừa nhận nhưng còn hẹp.
2. Ranh giới hỗ trợ -sai phạm	Mức thay thế tư duy; mục tiêu đánh giá; vùng xám	SV14 phân biệt nhiệm vụ kiểm tra khả năng viết với nhiệm vụ phân tích/dịch. Cùng thao tác có thể được đánh giá khác theo bối cảnh.	Đạo đức phụ thuộc mục tiêu học tập, không chỉ phụ thuộc công cụ.
3. Minh bạch nhưng thiếu nhất quán	Ngưỡng khai báo; sợ bất lợi; chuẩn mực bạn bè	SV21 sẵn sàng khai báo nhưng lo bị “soi kĩ hơn” nếu bạn khác không khai báo.	Minh bạch cần quy tắc chung và cơ chế bảo vệ công bằng.
4. Kiểm chứng có chọn lọc	Truy nguồn; độ tin cậy bề mặt; áp lực thời gian	SV09 từng gặp nguồn không tồn tại, sau đó tự tìm tài liệu; khi gặp, câu trả lời “trông rất khoa học” vẫn dễ được chấp nhận.	Ý định kiểm chứng suy yếu nếu thiếu thời gian và kĩ năng.
5. Điểm mù: dữ liệu, thiên lệch, tiếp cận	Quyền riêng tư; tiếng Việt; bản miễn phí/trả phí	SV25 cảm nhận đầu ra khác nhau giữa bản trả phí và miễn phí. Nghiên cứu không kiểm định chất lượng kĩ thuật của các phiên bản.	Nhận thức về công bằng và bảo vệ dữ liệu chưa ổn định.
6. Nhu cầu quy tắc theo nhiệm vụ	Không cấm tuyệt đối; mức cho phép; mẫu khai báo	SV03 đề nghị mỗi môn/bài tập ghi rõ được dùng AI đến đâu; nếu chỉ nói “có trách nhiệm” thì mỗi người hiểu khác nhau.	Quản trị phải chuyển nguyên tắc chung thành chỉ dẫn có thể thực hiện.

3.1. AI là công cụ hỗ trợ nhưng trách nhiệm vẫn thuộc về người học

Người tham gia thường mô tả AI như “trợ lý”, “người gợi ý” hoặc công cụ tìm kiếm. SV06 (năm 2, nhóm Ngôn ngữ) cho biết: “*Em có thể hỏi AI để tìm hướng làm bài luận, nhưng phần lập luận phải là của em. Nếu nộp một bài luận mà mình không giải thích được thì vẫn là lỗi của mình, không thể nói do máy.*” Phát biểu này thể hiện việc giữ quyền tác giả và trách nhiệm cuối cùng ở người học. Tuy nhiên, trách nhiệm chủ yếu được hiểu là đọc lại và tránh thông tin sai trong bài nộp; quyền riêng tư, thiên lệch hoặc hệ quả của dữ liệu đầu vào ít xuất hiện nếu người phỏng vấn chưa gợi mở.

3.2. Ranh giới đạo đức phụ thuộc vào mức độ AI thay thế năng lực cần được đánh giá

Người tham gia không coi mọi sử dụng AI là sai. Tiêu chí nổi bật là AI có thay thế năng lực mà nhiệm vụ cần đo lường hay không. SV14 (năm 3, nhóm Khoa học tự nhiên - công nghệ) giải thích: “*Nếu đề bài nhằm kiểm tra khả năng viết mà AI viết lại hết thì điều đó không còn phản ánh năng lực của em. Nhưng nếu môn học chỉ cần phân tích dữ liệu, dịch vài câu tiếng Anh thì có thể chấp nhận được.*” Theo chia sẻ của sinh viên, sao chép toàn bộ câu trả lời, tạo bài rồi chỉ sửa vài từ, hoặc dùng AI trong bài kiểm tra khi không được phép thường bị xem là sai phạm. Nghiên cứu không phân tích văn bản chính sách của từng học phần, vì vậy đây là nhận định của người tham gia, không phải kết luận về mức độ cấm trên toàn Trường.

3.3. Khai báo được ủng hộ về nguyên tắc nhưng thiếu điều kiện thực thi

Nhiều người tham gia đồng ý rằng cần minh bạch khi AI đóng góp đáng kể, nhưng chưa xác định được ngưỡng khai báo, vị trí khai báo và cách mô tả mức can thiệp. SV21 (năm 4, nhóm Báo chí - Truyền thông) nói: “*Em sẵn sàng ghi là đã dùng AI để lên dàn ý, nhưng phải có quy định chung cho những hành vi này. Nếu chỉ mình em ghi còn các bạn không ghi thì em sợ bài của mình bị soi kĩ hơn và đánh giá thấp hơn các bạn khác.*” Khi hướng dẫn chưa rõ, người học có xu hướng dựa vào chuẩn mực bạn bè hoặc giữ im lặng. Vì vậy, dữ liệu cho thấy khoảng cách giữa ủng hộ minh bạch và thực hành khai báo có liên quan đến nỗi sợ bất lợi và sự thiếu nhất quán của quy tắc.

3.4. Kiểm chứng thông tin mang tính chọn lọc và suy giảm dưới áp lực thời gian

Người tham gia thường ưu tiên kiểm tra số liệu, nội dung pháp luật và tài liệu tham khảo, nhưng ít kiểm tra những diễn giải có văn phong hợp lí. SV09 (năm 2, nhóm Khoa học xã hội) chia sẻ: “*Có lần AI cho tên bài báo rất đúng chủ đề nhưng em tìm không thấy. Từ đó em chỉ lấy từ khóa rồi tự tìm tài liệu. Nhưng những lúc gấp thì đúng là dễ tin vì câu trả lời trông rất khoa học.*” Các chiến lược được kể lại gồm truy tìm nguồn gốc, đối chiếu Google Scholar, đọc tài liệu và hỏi giảng viên. Tuy nhiên, thời hạn nộp bài làm giảm độ sâu kiểm chứng; tính trôi chảy của đầu ra trở thành tín hiệu thay thế cho độ tin cậy.

Bên cạnh đó, khi thảo luận về hoạt động kiểm chứng và tính liêm chính, cần lưu ý hạn chế nội tại của dữ liệu thực nghiệm. Dữ liệu dựa trên trải nghiệm tự thuật có thể chịu tác động của thiên lệch mong muốn xã hội, trong đó sinh viên có xu hướng che giấu hoặc giảm nhẹ các hành vi vi phạm nhạy cảm (Nguyen và Goto, 2024). Sự thiếu vắng đối chiếu chéo với lịch sử thao tác AI thực tế (prompt history) của người học là một điểm cần được thận trọng khi diễn giải kết quả.

3.5. Quyền riêng tư, thiên lệch và công bằng tiếp cận còn là những điểm mù

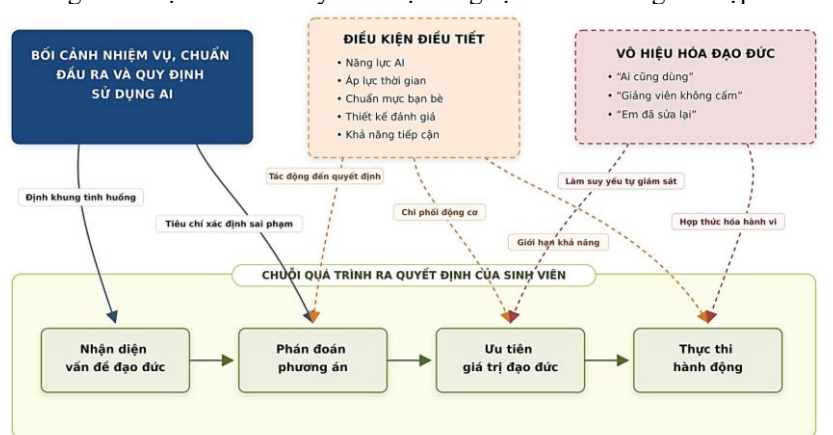
Một số người tham gia ít quan tâm hoặc không biết công cụ thu thập, lưu giữ và sử dụng dữ liệu đầu vào như thế nào. Ngay cả khi biết có điều khoản dữ liệu, sự tiện lợi vẫn thường được ưu tiên. Kết quả này được diễn đạt như nhận thức tự báo cáo, không phải đánh giá pháp lý hay kiểm toán thực hành dữ liệu của nền tảng. Về công bằng tiếp cận, SV25 (năm 1, nhóm Luật) nhận xét: “Nếu một bạn dùng bản trả phí, còn bạn khác chỉ có bản miễn phí thì kết quả mà AI cung cấp sẽ không giống nhau.” Đây là cảm nhận của người tham gia về chênh lệch chất lượng và cơ hội tiếp cận; nghiên cứu không kiểm định độ chính xác kỹ thuật giữa các phiên bản. Tương tự, một số sinh viên cảm nhận đầu ra tiếng Việt còn chung chung, thiếu bối cảnh địa phương hoặc có lỗi diễn đạt. Bài viết không chuyển các cảm nhận này thành kết luận khách quan về toàn bộ công cụ AI.

3.6. Sinh viên mong muốn quy tắc cụ thể theo từng nhiệm vụ

Người tham gia nhìn chung không đề xuất cấm tuyệt đối; họ muốn nhà trường có nguyên tắc chung và giảng viên xác định mức sử dụng ở từng bài. SV03 (năm 1, nhóm Khoa học xã hội) cho biết: “Điều cần nhất là mỗi môn học hoặc bài tập cần ghi rõ được dùng AI hỗ trợ đến đâu. Nếu chỉ nói dùng có đạo đức hay trách nhiệm thì mỗi người hiểu một kiểu.” Các đề xuất được kể lại gồm thang mức sử dụng, mẫu khai báo, hướng dẫn trích dẫn, danh mục dữ liệu không được nhập, quy trình kiểm chứng và ví dụ phân biệt sử dụng hợp lệ với vi phạm.

3.7. Phân tích liên chủ đề: từ nhận diện đến hành động

Đặt sáu chủ đề vào mô hình Rest cho thấy nhận diện đạo đức mạnh nhất với hành vi hữu hình như sao chép toàn bộ, dùng AI làm bài kiểm tra hoặc bịa nguồn. Những rủi ro ít hữu hình hơn như dữ liệu cá nhân, thiên lệch, công bằng tiếp cận và suy giảm năng lực tự học thường chỉ nổi lên sau câu hỏi gợi mở. Nếu vấn đề chưa được nhận diện như một vấn đề đạo đức, chuỗi phán đoán và hành động khó được kích hoạt. Phán đoán của sinh viên dựa trên ba tiêu chí lặp lại: mức độ AI thay thế năng lực cần đánh giá, hướng dẫn của giảng viên và mức độ minh bạch. Một mẫu hình gợi ý trong dữ liệu là sinh viên năm cuối nhấn mạnh tự đánh giá và kiểm chứng nguồn, trong khi sinh viên năm thứ nhất lại chủ yếu có xu hướng dựa vào quy định (cấm hoặc không cấm sử dụng AI) của giảng viên. Đặc biệt, phân tích đối chiếu chéo còn cho thấy sự khác biệt về nhạy cảm đạo đức giữa các nhóm ngành: sinh viên ngành Luật có xu hướng nhạy cảm hơn về các rủi ro bảo mật dữ liệu và vi phạm bản quyền, trong khi sinh viên khối Khoa học tự nhiên - công nghệ thường nhìn nhận AI thuần túy như một công cụ tối ưu hóa logic và lập trình.



Hình 1. Mô hình diễn giải hành vi đạo đức trong sử dụng AI tạo sinh của sinh viên

Khoảng cách phán đoán - hành động xuất hiện khi sinh viên biết một thực hành có rủi ro nhưng vẫn ưu tiên thời hạn, điểm số, sự tiện lợi hoặc tránh bất lợi so với bạn bè. Các diễn đạt “ai cũng dùng”, “giảng viên không nói” và “em đã sửa lại” tương ứng với phân tán trách nhiệm, chuyển trách nhiệm và giảm nhẹ mức can thiệp. Theo Bandura (1999), các cơ chế này không xóa chuẩn mực đạo đức mà tạm thời làm suy yếu tự giám sát. Tổng hợp cho thấy chuỗi giải thích bắt đầu từ bối cảnh nhiệm vụ và quy định; tiếp đến là nhận diện, phán đoán, ưu tiên và thực thi. Năng lực AI, áp lực thời gian, chuẩn mực bạn bè, thiết kế đánh giá và khả năng tiếp cận điều tiết nhiều điểm trong chuỗi; vô hiệu hóa đạo đức làm suy yếu mối liên hệ giữa phán đoán và hành động.

3.8. Thảo luận

3.8.1. Ranh giới đạo đức cần được xác định từ mục tiêu đánh giá

Phát hiện trung tâm là sinh viên không phân loại đạo đức chỉ theo việc “có dùng AI hay không”, mà theo mức độ AI thay thế năng lực cần được đánh giá. Kết quả này phù hợp với quan điểm thiết kế đánh giá theo mức độ cho phép: cùng một công cụ có thể bị cấm, được hỗ trợ có giới hạn hoặc được tích hợp sâu tùy chuẩn đầu ra (Furze và cộng sự, 2024; Perkins và cộng sự, 2024). Vì vậy, một lệnh cấm đồng nhất dễ bỏ qua mục tiêu học tập, trong khi cho phép chung chung tạo vùng xám. Đóng góp của nghiên cứu là nổi tiêu chí “thay thế tư duy” do sinh viên nêu với chuỗi phán đoán - hành động của Rest.

3.8.2. Minh bạch chỉ khả thi khi người học được bảo vệ khỏi bất lợi không công bằng

Chủ đề khai báo cho thấy sinh viên ủng hộ minh bạch nhưng do dự khi quy tắc không áp dụng đồng đều. Kết quả tương thích với Tran và cộng sự (2025), trong đó việc thừa nhận và trích dẫn ChatGPT phụ thuộc yêu cầu giảng viên và hiểu biết chính sách; đồng thời tương thích với Gonsalves (2025) về vai trò của sợ hậu quả, hướng dẫn mơ hồ và chuẩn mực bạn bè. Do đó, khai báo không nên chỉ là mệnh lệnh đạo đức. Nhà trường cần quy định ngưỡng khai báo, mẫu ngắn gọn, cách đánh giá phần việc của sinh viên và bảo đảm rằng sử dụng hợp lệ đã khai báo không bị mặc định là gian lận.

3.8.3. Năng lực AI là năng lực đạo đức có thể thực thi

Việc biết rằng AI “có thể sai” chưa đủ để tạo hành vi có trách nhiệm. Dữ liệu cho thấy năng lực thực thi ít nhất gồm bốn nhóm: truy nguồn và đánh giá nguồn; phân loại dữ liệu trước khi nhập; ghi nhận mức đóng góp của AI; và duy trì rà soát đầu ra dưới áp lực thời gian. Khung năng lực AI của UNESCO và Khung năng lực số cho người học tại Việt Nam cung cấp nền tảng để chuyển các yêu cầu này thành mục tiêu đào tạo (Bộ GD-ĐT, 2025; Miao và cộng sự, 2024). Quyền riêng tư, thiên lệch và công bằng tiếp cận phải được dạy như phần cốt lõi, không phải phụ lục của chống đạo văn.

3.8.4. Hàm ý quản trị: cho phép có điều kiện theo từng nhiệm vụ

Các khuyến nghị được truy ngược trực tiếp từ sáu chủ đề và cơ chế giải thích: (1) Cấp trường: ban hành nguyên tắc chung về trách nhiệm, khai báo, dữ liệu, công bằng và xử lý vi phạm; tránh đồng nhất mọi sử dụng AI với gian lận. Khuyến nghị xuất phát từ các chủ đề 3, 5 và 6. (2) Cấp học phần/nhiệm vụ: ghi rõ AI bị cấm, được dùng có giới hạn hay được khuyến khích, căn cứ vào chuẩn đầu ra và phần năng lực cần đánh giá. Khuyến nghị giải quyết vùng xám ở chủ đề 2 và nhu cầu hướng dẫn ở chủ đề 6. (3) Kiểm tra, đánh giá: Yêu cầu khai báo công cụ, mục đích, phần nội dung được hỗ trợ và cách kiểm chứng. Cần thiết kế bảng kiểm đánh giá (rubric) quá trình bắt buộc sinh viên nộp kèm lịch sử câu lệnh (prompt history), đồng thời tích hợp trọng số điểm cho vòng thi vấn đáp để bảo vệ tính nguyên bản của ý tưởng. (4) Phát triển năng lực: tổ chức mô-đun bắt buộc về kiểm chứng, dữ liệu cá nhân, bản quyền, thiên lệch, công bằng và sử dụng AI để hỗ trợ thay vì thay thế tư duy. Khuyến nghị bắt nguồn từ các chủ đề 1, 4 và 5.

4. Kết luận và bình luận

Từ trải nghiệm tự báo cáo của 28 sinh viên trong một trường hợp cụ thể, nghiên cứu đưa ra ba đóng góp. Một là, ranh giới đạo đức của AI tạo sinh không chỉ nằm ở việc có sử dụng công cụ, mà ở mức độ công cụ thay thế năng lực cần được đánh giá và mức minh bạch của người học. Hai là, khoảng cách giữa phán đoán và hành động được giải thích bằng sự tương tác giữa động cơ, năng lực AI, áp lực thời gian, chuẩn mực bạn bè, thiết kế đánh giá, quy định và vô hiệu hóa đạo đức. Ba là, nhà trường nên chuyển từ lựa chọn cấm/cho phép chung sang cơ chế cho phép có điều kiện theo từng nhiệm vụ, đi kèm khai báo tương xứng, bảo vệ dữ liệu, kiểm chứng và lưu dấu quá trình học tập. Các kết luận có giá trị diễn giải và chuyển giao thận trọng sang bối cảnh tương đồng, không có ý nghĩa khái quát thống kê cho toàn bộ sinh viên đại học Việt Nam.

Về lý luận, nghiên cứu mở rộng thảo luận từ gian lận/đạo văn sang bản đồ đạo đức gồm trách nhiệm, khai báo, độ tin cậy, quyền riêng tư, thiên lệch, công bằng tiếp cận và quyền tự chủ học tập. Sự tích hợp Rest - Bandura - xã hội kỹ thuật giúp giải thích hiện tượng nhận thức đúng vẫn có thể không dẫn đến hành động khi động cơ, kỹ năng và môi trường không hỗ trợ. Về thực tiễn, mô hình “cho phép có điều kiện theo nhiệm vụ” chuyển nguyên tắc đạo đức thành các điểm can thiệp ở cấp trường, học phần, đánh giá và năng lực.

Tuy có đóng góp nhưng nghiên cứu này vẫn còn 05 giới hạn: (1) đây là nghiên cứu trường hợp tại một trường và không đại diện thống kê cho sinh viên Việt Nam; (2) dữ liệu tự báo cáo có thể chịu tác động của thiên lệch mong muốn xã hội; (3) chưa quan sát được hành vi sử dụng AI thực tế hoặc phân tích sản phẩm học tập; (4) chưa kết hợp dữ liệu từ giảng viên, cán bộ quản lý, đề bài, bảng kiểm đánh giá (rubric) và chính sách chính thức; (5) các đối chiếu theo năm học/ngành chỉ mang tính gợi ý và chưa đủ chiều sâu để kết luận khác biệt nhóm. Những hạn chế này gợi mở hướng nghiên cứu tiếp theo nên phát triển theo hướng kết hợp phỏng vấn đa bên, phân tích văn bản, nhật ký sử dụng AI, dấu vết quá trình và nghiên cứu theo thời gian để kiểm định khả năng chuyển giao của mô hình.

Tuyên bố về vai trò của các tác giả: Nguyễn Thị Trà My, Phạm Chiến Thắng: Lên ý tưởng nghiên cứu, xác định phương pháp và công cụ nghiên cứu; Tạ Thị Thảo, Phùng Phương Nga: Thu thập, Phân tích dữ liệu; Phạm Chiến Thắng: Trực quan hoá dữ liệu và viết bản thảo. Phạm Chiến Thắng: Giám sát, chỉ đạo quá trình nghiên cứu; Dương Thị Huyền, Ma Thị Thanh Hiếu, Vũ Thị Hạnh: Viết bản thảo, sửa chữa bản thảo.

Tuyên bố về GenAI và Quyền tác giả: Trong quá trình chuẩn bị bản thảo này, các tác giả không sử dụng công cụ AI nào.

Tuyên bố về xung đột lợi ích: Các tác giả tuyên bố không có xung đột lợi ích.

Thông tin tài trợ: Nghiên cứu này được tài trợ bởi Trường Đại học Khoa học - Đại học Thái Nguyên theo mã số nhóm nghiên cứu: NNCM.DHKH.2025.05.

Tài liệu tham khảo

- Bandura, A. (1999). Moral disengagement in the perpetration of inhumanities. *Personality and Social Psychology Review*, 3(3), 193-209. https://doi.org/10.1207/S15327957PSPR0303_3
- Braun, V., & Clarke, V. (2022). *Thematic analysis: A practical guide*. SAGE.
- Bộ GD-ĐT (2025). *Thông tư số 02/2025/TT-BGDĐT ngày 24/01/2025 quy định Khung năng lực số cho người học*.
- Chan, C. K. Y., & Hu, W. (2023). Students' voices on generative AI: perceptions, benefits, and challenges in higher education. *International Journal of Educational Technology in Higher Education*, 20(1).
- Cong-Lem, N., Tran, T. T., & Nguyen, T. T. (2024). Academic integrity in the age of generative AI: Perceptions and responses of Vietnamese EFL teachers. *Teaching English with Technology*, 24(1), 28-47.
- Furze, L., Perkins, M., Roe, J., & MacVaugh, J. (2024). The AI Assessment Scale (AIAS) in action: A pilot implementation of GenAI-supported assessment. *Australasian Journal of Educational Technology*, 40(4), 38-55. <https://doi.org/10.14742/ajet.9434>
- Gonsalves, C. (2025). Addressing student non-compliance in AI use declarations: Implications for academic integrity and assessment in higher education. *Assessment & Evaluation in Higher Education*, 50(4), 592-606. <https://doi.org/10.1080/02602938.2024.2415654>
- Le, M. T. T., & Tran, K. V. (2024). Vietnamese university students' perceptions of ChatGPT for homework assistance. *International Journal of Interactive Mobile Technologies*, 18(15), 190-204. <https://doi.org/10.3991/ijim.v18i15.48819>
- Malterud, K., Siersma, V. D., & Guassora, A. D. (2016). Sample size in qualitative interview studies: Guided by information power. *Qualitative Health Research*, 26(13), 1753-1760. <https://doi.org/10.1177/1049732315617444>
- Miao, F., & Holmes, W. (2023). *Guidance for generative AI in education and research*. UNESCO.
- Miao, F., Shiohira, K., & Lao, N. (2024). *AI competency framework for students*. UNESCO.
- Nguyen, H. M., & Goto, D. (2024). Unmasking academic cheating behavior in the artificial intelligence era: Evidence from Vietnamese undergraduates. *Education and Information Technologies*, 29, 15999-16025. <https://doi.org/10.1007/s10639-024-12495-4>
- O'Brien, B. C., Harris, I. B., Beckman, T. J., Reed, D. A., & Cook, D. A. (2014). Standards for reporting qualitative research: A synthesis of recommendations. *Academic Medicine*, 89(9), 1245-1251. <https://doi.org/10.1097/ACM.0000000000000388>
- Perkins, M., Furze, L., Roe, J., & MacVaugh, J. (2024). The Artificial Intelligence Assessment Scale (AIAS): A framework for ethical integration of generative AI in educational assessment. *Journal of University Teaching & Learning Practice*, 21(6). <https://doi.org/10.53761/q3azde36>
- Quốc hội (2025). *Luật Bảo vệ dữ liệu cá nhân*. Luật số 91/2025/QH15, ban hành ngày 26/06/2025.
- Rest, J. R. (1986). *Moral development: Advances in research and theory*. Praeger.
- Tran, K. V., Le, T. T. M., & Pham, A. T. (2025). Exploring Vietnamese students' plagiarism awareness and actual practices through using of ChatGPT as a learning tool. *International Journal for Educational Integrity*, 21, 32. <https://doi.org/10.1007/s40979-025-00207-5>
- TNUS (2026). *Lịch sử hình thành và phát triển*. <https://tnus.edu.vn/gioi-thieu/lich-su-hinh-thanh-va-phat-trien>
- Tlili, A., Shehata, B., Adarkwah, M. A., Bozkurt, A., Hickey, D. T., Huang, R., & Agyemang, B. (2023). What if the devil is my guardian angel: ChatGPT as a case study of using chatbots in education. *Smart Learning Environments*, 10, 15. <https://doi.org/10.1186/s40561-023-00237-x>
- UNESCO (2021). *Recommendation on the ethics of artificial intelligence*. <https://unesdoc.unesco.org/ark:/48223/pf0000381137>