

AI VÀ TRÁCH NHIỆM CHUYÊN MÔN CỦA GIẢNG VIÊN TRONG KIỂM TRA, ĐÁNH GIÁ Ở GIÁO DỤC ĐẠI HỌC: TIẾP CẬN THEO MÔ HÌNH HUMAN-IN-THE-LOOP

Lê Thị Ngân

Khoa Cơ khí chế tạo, Trường Đại học Công nghệ Kỹ thuật Vinh

Email: lethingankt31@gmail.com

Tóm tắt: Sự phát triển của trí tuệ nhân tạo (AI), đặc biệt là AI tạo sinh (Generative AI), đang làm thay đổi hoạt động kiểm tra, đánh giá trong giáo dục đại học và đặt ra những thách thức về tính trung thực học thuật, minh bạch và trách nhiệm trong ra quyết định. Nghiên cứu này phân tích sự thay đổi trách nhiệm chuyên môn của giảng viên trong kiểm tra, đánh giá dưới góc tiếp cận Human-in-the-loop (HITL). Trên cơ sở tổng hợp và phân tích các nghiên cứu về AI trong giáo dục, đạo đức, đánh giá và trách nhiệm giải trình, bài viết xây dựng chuỗi tương tác gồm bốn tầng: (1) AI-generated content; (2) AI-assisted evaluation; (3) Teacher professional judgment; và (4) Accountability. Kết quả cho thấy AI có thể hỗ trợ nhiều khâu của hoạt động đánh giá nhưng không thay thế phán đoán chuyên môn của giảng viên. Ngược lại, sự tham gia của AI làm gia tăng yêu cầu kiểm soát đầu ra, thẩm định độ tin cậy, bảo đảm tính minh bạch và trách nhiệm giải trình đối với quyết định đánh giá. Nghiên cứu qua đó làm rõ vai trò của giảng viên như chủ thể chuyên môn kết nối công nghệ với đạo đức nghề nghiệp và trách nhiệm giải trình trong bối cảnh AI ngày càng tham gia sâu vào giáo dục đại học.

Từ khóa: Trí tuệ nhân tạo (AI); AI tạo sinh; Human-in-the-loop; đạo đức AI; trách nhiệm giải trình.

AI AND LECTURERS' PROFESSIONAL RESPONSIBILITY IN HIGHER EDUCATION ASSESSMENT: A HUMAN-IN-THE-LOOP APPROACH

Abstract: The development of artificial intelligence (AI), particularly generative AI (GenAI), is transforming assessment in higher education and creating new challenges related to academic integrity, transparency, and accountability in decision-making. This study examines changes in lecturers' professional responsibility in assessment from a Human-in-the-loop (HITL) perspective. Based on a synthesis and analysis of studies on AI in education, ethics, assessment, and accountability, the article develops a four-stage interaction chain: (1) AI-generated content; (2) AI-assisted evaluation; (3) Teacher professional judgment; and (4) Accountability. The findings indicate that AI can support multiple stages of the assessment process but cannot replace lecturers' professional judgment. Instead, the involvement of AI increases the need for lecturers to review AI outputs, assess their reliability, and ensure transparency and accountability in assessment decisions. The study thus clarifies the role of lecturers as professional actors who connect technology with professional ethics and accountability as AI becomes increasingly integrated into higher education.

Keywords: Artificial intelligence (AI); Generative AI; Human-in-the-loop; AI ethics; Accountability.

Nhận bài: 14/09/2026

Phản biện: 15/10/2026

Duyệt đăng: 17/10/2026

I. ĐẶT VẤN ĐỀ

Sự phát triển nhanh chóng của trí tuệ nhân tạo (Artificial Intelligence – AI), đặc biệt là AI tạo sinh (Generative AI – GenAI), đang làm thay đổi giáo dục đại học. GenAI mở rộng khả năng hỗ trợ dạy học và kiểm tra, đánh giá, nhưng đồng thời đặt ra các vấn đề về tính trung thực học thuật, minh bạch, công bằng và trách nhiệm. UNESCO nhấn mạnh việc ứng dụng GenAI trong giáo dục cần theo cách tiếp cận lấy con người làm trung tâm, trong đó các giá trị và trách nhiệm của con người giữ vai trò nền tảng (Miao & Holmes, 2023). AI không chỉ được người học sử dụng để tạo sản phẩm học tập mà còn ngày càng tham gia vào thiết kế, chấm và phân tích kết quả đánh giá.

AI có thể hỗ trợ tạo câu hỏi, phân tích câu trả lời, cung cấp phản hồi hoặc đề xuất kết quả đánh giá, nhưng đầu ra của AI không đồng nghĩa với một quyết định chuyên môn có thể được chấp

nhận tự động. Việc sử dụng AI trong giáo dục vì vậy cần gắn với đạo đức, vai trò của con người và khả năng kiểm soát hệ thống (Holmes et al., 2022; Amershi et al., 2014). Trong đánh giá đại học, yêu cầu này đặc biệt quan trọng vì quyết định đánh giá ảnh hưởng trực tiếp đến kết quả và quyền lợi của người học. Đồng thời, việc xác định hành vi sử dụng GenAI không thể chỉ dựa vào công cụ phát hiện AI mà cần xem xét nhiều loại bằng chứng và bối cảnh cụ thể (Munoz et al., 2026; Wright, 2026).

Trong bối cảnh đó, mô hình Human-in-the-loop (Con người trong vòng lặp kiểm soát – HITL) cung cấp cách tiếp cận phù hợp để xem xét mối quan hệ giữa AI và giảng viên trong đánh giá. HITL nhấn mạnh sự tương tác giữa khả năng xử lý của AI và vai trò kiểm soát, diễn giải, phán đoán của con người. Cách tiếp cận này đặc biệt có ý nghĩa khi GenAI vừa mở rộng cơ hội vừa tạo

ra các rủi ro đối với tính toàn vẹn của đánh giá (Kofinas et al., 2025; Peters & Angelov, 2025). Tuy nhiên, các nghiên cứu hiện nay chủ yếu tập trung vào khía cạnh kỹ thuật, lợi ích công nghệ hoặc rủi ro về tính trung thực học thuật, trong khi sự chuyển dịch vai trò và trách nhiệm chuyên môn của giảng viên khi AI tham gia vào đánh giá, đặc biệt dưới góc tiếp cận HITL, vẫn chưa được làm rõ đầy đủ.

Trên cơ sở đó, bài viết phân tích sự thay đổi trách nhiệm chuyên môn của giảng viên trong kiểm tra, đánh giá ở giáo dục đại học theo mô hình HITL và xây dựng chuỗi tương tác gồm bốn tầng: (1) AI-generated content; (2) AI-assisted evaluation; (3) Teacher professional judgment; và (4) Accountability. Qua đó, nghiên cứu làm rõ vai trò của giảng viên trong kiểm soát đầu ra AI, thực hiện phán đoán chuyên môn và chịu trách nhiệm giải trình đối với quyết định đánh giá.

II. PHƯƠNG PHÁP NGHIÊN CỨU

Bài viết sử dụng phương pháp nghiên cứu định tính, dựa trên cách tiếp cận tổng hợp, phân tích tài liệu (document analysis) và lược khảo tài liệu (literature review). Nguồn tài liệu được thu thập là các bài báo khoa học trên các cơ sở dữ liệu học thuật uy tín quốc tế, các báo cáo hướng dẫn của UNESCO và các nghiên cứu chuyên sâu cập nhật về AI trong giáo dục, đạo đức AI, tính trung thực học thuật và mô hình Human-in-the-loop (HITL).

Quá trình nghiên cứu được triển khai qua ba bước cơ bản: (1) Khảo lược và hệ thống hóa cơ sở lý thuyết về mức độ can thiệp của AI tạo sinh vào quy trình kiểm tra, đánh giá tại giáo dục đại học; (2) Ứng dụng lăng kính của mô hình HITL để đối chiếu, phân tích sự thay đổi trong cấu trúc tương tác giữa AI và giảng viên so với quy trình truyền thống; (3) Tổng hợp, suy luận logic để xây dựng chuỗi tương tác gồm 4 tầng (AI-generated content; AI-assisted evaluation; Teacher professional judgment; Accountability) nhằm xác định rõ giới hạn trách nhiệm. Thông qua phương pháp này, nghiên cứu không chỉ nhận diện vấn đề mà còn đề xuất được khung trách nhiệm chuyên môn một cách có hệ thống, phù hợp với yêu cầu thực tiễn của quản trị giáo dục hiện đại.

III. CƠ SỞ LÝ LUẬN

3.1. Trí tuệ nhân tạo và AI tạo sinh trong kiểm tra, đánh giá

Trí tuệ nhân tạo (AI) trong giáo dục được sử dụng để thực hiện hoặc hỗ trợ các nhiệm vụ như xử lý thông tin, nhận diện mẫu, dự đoán và tạo nội

dung. Trong giáo dục đại học, AI đã được ứng dụng vào cá nhân hóa học tập, phân tích dữ liệu, phản hồi và đánh giá. Tuy nhiên, các hệ thống AI có mức độ tự động hóa và chức năng khác nhau; do đó, việc sử dụng AI cần gắn với mục đích sử phạm và bối cảnh giáo dục cụ thể (Holmes et al., 2022).

Sự xuất hiện của AI tạo sinh (GenAI) mở rộng đáng kể phạm vi ứng dụng này. Người học có thể sử dụng AI để hỗ trợ tạo sản phẩm đánh giá, trong khi giảng viên có thể sử dụng AI để xây dựng câu hỏi, phân tích bài làm hoặc hỗ trợ phản hồi. Điều này làm thay đổi mối quan hệ giữa nhiệm vụ đánh giá, sản phẩm của người học và bằng chứng về năng lực thực tế (Lodge et al., 2023; Kofinas et al., 2025). Vì vậy, đánh giá không chỉ cần xem xét kết quả cuối cùng mà còn phải chú ý đến quá trình tạo lập, xác minh và diễn giải kết quả. Việc xác định phạm vi AI được tham gia và quyền kiểm soát của con người trở thành yêu cầu quan trọng nhằm duy trì tính toàn vẹn của đánh giá trong môi trường số (Dawson, 2021; Peters & Angelov, 2025).

3.2. Mô hình Human-in-the-loop trong giáo dục

Human-in-the-loop (HITL) là cách tiếp cận trong đó con người giữ vai trò kiểm soát, tương tác và ra quyết định trong hệ thống có sự tham gia của AI. Thay vì để AI hoạt động hoàn toàn tự động, HITL nhấn mạnh sự kết hợp giữa khả năng xử lý của máy và phán đoán của con người. Amershi et al. (2014) cho rằng vai trò của con người trong hệ thống học máy bao gồm tương tác, giám sát và điều chỉnh hệ thống trong quá trình hoạt động.

Trong giáo dục đại học, HITL đặc biệt phù hợp với các hoạt động có ảnh hưởng trực tiếp đến người học như kiểm tra, đánh giá. AI có thể hỗ trợ xử lý lượng dữ liệu lớn hoặc thực hiện các nhiệm vụ lặp lại, nhưng kết quả cần được con người kiểm tra và đặt trong bối cảnh giáo dục cụ thể. Holmes et al. (2022) nhấn mạnh rằng các vấn đề đạo đức của AI trong giáo dục không thể được giải quyết chỉ bằng công nghệ mà cần sự tham gia của các chủ thể giáo dục. Vì vậy, HITL vừa là cơ chế duy trì quyền kiểm soát của con người, vừa góp phần bảo đảm tính minh bạch và trách nhiệm trong quyết định giáo dục. Cách tiếp cận này phù hợp với định hướng lấy con người làm trung tâm trong quản trị AI và giáo dục (Floridi & Cowls, 2019; Miao & Holmes, 2023).

3.3. Phán đoán chuyên môn của giảng viên

Phán đoán chuyên môn của giảng viên là quá trình vận dụng kiến thức, kinh nghiệm sử phạm,

tiêu chí đánh giá và bối cảnh để nhận định chất lượng hoặc mức độ đạt chuẩn của người học. Khi AI tham gia vào đánh giá, chức năng này không mất đi mà chuyển mạnh hơn sang kiểm tra, diễn giải và thẩm định đầu ra của hệ thống. AI có thể nhận diện mẫu và đưa ra đề xuất nhanh, nhưng không thể mặc nhiên xác định kết quả đó phù hợp với mục tiêu và bối cảnh đánh giá. Vì vậy, giảng viên cần có khả năng phát hiện sai lệch, kiểm tra tính hợp lệ và quyết định khi nào nên chấp nhận, điều chỉnh hoặc bác bỏ đầu ra AI (Holmes et al., 2022; Luo et al., 2025).

Vai trò này đặc biệt quan trọng trong các trường hợp nghi vấn sử dụng GenAI, bởi đặc điểm giống nội dung do AI tạo ra không tự động chứng minh vi phạm học thuật. Quyết định cần dựa trên nhiều loại bằng chứng và bối cảnh cụ thể (Munoz et al., 2026; Wright, 2026). Do đó, phán đoán chuyên môn vẫn là cơ sở để diễn giải bằng chứng và đưa ra quyết định phù hợp.

3.4. Trách nhiệm giải trình trong đánh giá có sự tham gia của AI

Trách nhiệm giải trình (accountability) trong đánh giá là yêu cầu chủ thể ra quyết định có khả năng giải thích và chịu trách nhiệm đối với quyết định của mình. Khi AI tham gia vào quy trình đánh giá, yêu cầu này trở nên phức tạp hơn, đòi hỏi tính minh bạch, khả năng truy xuất và kiểm chứng quá trình ra quyết định (Floridi & Cowls, 2019; Miao & Holmes, 2023).

Trong mô hình HITL, trách nhiệm được phân bổ theo từng khâu: AI cung cấp đầu ra; giảng viên kiểm tra, diễn giải và đưa ra quyết định; cơ sở giáo dục quy định điều kiện và giới hạn sử dụng. Holmes et al. (2022) nhấn mạnh yêu cầu minh bạch và khả năng xem xét lại các quyết định có sự tham gia của AI trong giáo dục, đặc biệt khi chúng ảnh hưởng đến kết quả học tập của người học. Vì vậy, AI có thể mở rộng năng lực xử lý nhưng không tự động thay thế phán đoán chuyên môn và trách nhiệm của giảng viên. Trên cơ sở đó, HITL được sử dụng để phân tích sự phân công giữa công nghệ và con người qua chuỗi AI-generated content → AI-assisted evaluation → Teacher professional judgment → Accountability.

IV. PHÂN TÍCH TRÁCH NHIỆM CHUYÊN MÔN CỦA GIẢNG VIÊN TRONG MÔ HÌNH HUMAN-IN-THE-LOOP

4.1. Từ AI-generated content đến yêu cầu kiểm soát bằng chứng đánh giá

Sự xuất hiện của GenAI làm phức tạp mối quan hệ giữa sản phẩm học tập và năng lực thực

tế của người học. Khi AI có thể tạo ra sản phẩm chất lượng cao, sản phẩm cuối cùng không nhất thiết phản ánh trực tiếp năng lực tự thân mà có thể là kết quả tương tác giữa người học và AI (Lodge et al., 2023; Kofinas et al., 2025).

Tuy nhiên, không phải mọi hình thức sử dụng AI đều là vi phạm học thuật. Việc phát hiện AI chỉ là một nguồn thông tin và không thể thay thế phán đoán về tính trung thực học thuật (Wright, 2026). Vì vậy, giảng viên cần xem xét cả sản phẩm, quá trình thực hiện và khả năng giải thích của người học. Thiết kế lại nhiệm vụ đánh giá cũng có vai trò quan trọng trong bảo đảm tính xác thực và toàn vẹn của đánh giá (Dawson, 2021; Peters & Angelov, 2025).

4.2. AI-assisted evaluation: từ tự động hóa sang hỗ trợ quyết định

AI có thể hỗ trợ tạo câu hỏi, xây dựng hoặc kiểm tra rubric, phân loại câu trả lời, cung cấp phản hồi và tổng hợp kết quả, qua đó giảm các công việc lặp lại và hỗ trợ giảng viên trong quá trình đánh giá (Luo et al., 2025). Tuy nhiên, khả năng xử lý nhanh không đồng nghĩa với độ chính xác tuyệt đối. AI có thể hiểu sai yêu cầu, bỏ qua bối cảnh hoặc đưa ra kết quả không phù hợp với mục tiêu học tập.

Do đó, đầu ra AI cần được xem là thông tin hoặc đề xuất hỗ trợ, không phải kết luận đánh giá mặc định. Khi AI đảm nhiệm nhiều thao tác kỹ thuật, trách nhiệm của giảng viên chuyển sang lựa chọn công cụ, xác định phạm vi sử dụng, kiểm tra đầu ra và nhận diện các trường hợp AI không đáng tin cậy. Theo cách tiếp cận HITL, mức độ kiểm soát của con người cần tương ứng với mức độ rủi ro của quyết định, đặc biệt đối với những quyết định ảnh hưởng trực tiếp đến kết quả học tập hoặc tính trung thực học thuật (Amershi et al., 2014; Holmes et al., 2022).

4.3. Phán đoán chuyên môn của giảng viên: sự chuyển dịch chứ không phải biến mất

Trong đánh giá có sự hỗ trợ của AI, phán đoán chuyên môn của giảng viên là mắt xích nối đầu ra công nghệ với quyết định giáo dục. AI có khả năng xử lý và nhận diện mẫu, nhưng giảng viên mới là chủ thể đặt kết quả trong mục tiêu môn học và bối cảnh sư phạm. Vì vậy, AI không loại bỏ phán đoán chuyên môn mà làm thay đổi cách thức thực hiện phán đoán đó.

Vai trò của giảng viên tập trung vào thẩm định đầu ra AI, diễn giải trong bối cảnh cụ thể và đưa ra quyết định cuối cùng. Điều này đặc biệt quan trọng khi có tranh chấp hoặc nghi vấn vi phạm

học thuật, bởi công cụ phát hiện AI chỉ cung cấp chỉ báo chứ không thể mặc nhiên trở thành bằng chứng quyết định. Munoz et al. (2026) cho thấy sức mạnh của các loại bằng chứng trong nghi vấn sử dụng GenAI là khác nhau. Do đó, AI càng tham gia sâu vào đánh giá thì yêu cầu đối với năng lực thẩm định, nhận diện giới hạn và sai lệch của giảng viên càng cao (Luo et al., 2025; Sharma, 2026).

4.4. Trách nhiệm giải trình khi quyết định có sự tham gia của AI

Khi AI tham gia vào các quyết định ảnh hưởng trực tiếp đến kết quả học tập, trách nhiệm giải trình trở nên đặc biệt quan trọng. Người học cần có khả năng biết cơ sở của quyết định, yêu cầu giải thích và xem xét lại khi cần thiết. Vì vậy, hệ thống đánh giá có AI hỗ trợ phải bảo đảm khả năng giải thích và truy xuất quá trình ra quyết định (Floridi & Cowls, 2019; Miao & Holmes, 2023).

Trong mô hình HITL, trách nhiệm trước hết gắn với chủ thể đưa ra quyết định cuối cùng. Việc sử dụng AI để chấm hoặc phân tích không chuyển giao hoàn toàn trách nhiệm cho hệ thống; giảng viên vẫn phải có khả năng giải thích tiêu chí, dữ liệu, vai trò của AI và những điều chỉnh đối với đề xuất của hệ thống. Đồng thời, cơ sở giáo dục cần quy định phạm vi sử dụng, yêu cầu kiểm tra của con người, cơ chế lưu vết và xem xét lại quyết định. Theo đó, accountability không chỉ là xác định ai chịu trách nhiệm mà còn là khả năng giải thích, kiểm chứng và can thiệp vào quyết định có sự tham gia của AI (Floridi & Cowls, 2019; Holmes et al., 2022).

4.5. Mối quan hệ giữa bốn tầng và sự thay đổi trách nhiệm của giảng viên

Từ các phân tích trên, trách nhiệm chuyên môn của giảng viên được khái quát qua chuỗi: AI-generated content → AI-assisted evaluation → Teacher professional judgment → Accountability. Chất lượng của bằng chứng đầu vào ảnh hưởng đến đầu ra AI; đầu ra này cần được giảng viên thẩm định và diễn giải trước khi trở thành quyết định; quyết định cuối cùng phải có khả năng giải thích và xem xét lại.

Sự thay đổi cốt lõi không nằm ở việc AI thay thế giảng viên mà ở sự chuyển dịch trách nhiệm từ thực hiện sang giám sát, thẩm định, phán đoán và giải trình. Trong mô hình HITL, giảng viên vẫn giữ quyền kiểm soát đối với quyết định cuối cùng (Selwyn, 2019).

V. HÀM Ý VÀ ĐỀ XUẤT KHUNG TRÁCH NHIỆM CHUYÊN MÔN CỦA GIẢNG VIÊN

5.1. Khung trách nhiệm chuyên môn của giảng viên trong quy trình đánh giá có AI

Trên cơ sở chuỗi phân tích bốn tầng ở Mục IV, trách nhiệm của giảng viên trong mô hình Human-in-the-loop được chuyển hóa thành khung vận hành gồm năm bước: (1) xác định mục đích và giới hạn sử dụng AI; (2) kiểm soát đầu ra; (3) thực hiện phán đoán chuyên môn; (4) đưa ra và giải thích quyết định; và (5) rà soát, điều chỉnh sau đánh giá.

Trước hết, giảng viên cần xác định phạm vi và mức độ sử dụng AI; những quyết định ảnh hưởng trực tiếp đến kết quả học tập hoặc xác định vi phạm học thuật cần mức độ kiểm soát cao hơn của con người (Holmes et al., 2022; Miao & Holmes, 2023). Đầu ra AI chỉ nên được xem là thông tin hỗ trợ và phải được kiểm tra về tính chính xác, phù hợp với tiêu chí và bối cảnh đánh giá. Kết quả từ công cụ phát hiện GenAI cũng chỉ là một nguồn bằng chứng cần được kiểm chứng (Munoz et al., 2026; Wright, 2026).

Quyết định cuối cùng phải dựa trên tiêu chí đã xác định và có khả năng giải thích. Đồng thời, trách nhiệm không kết thúc khi công bố kết quả mà còn bao gồm rà soát sai sót, phản hồi của người học và điều chỉnh quy trình ở các lần đánh giá tiếp theo. Theo đó, HITL được xem như một vòng phản hồi liên tục, trong đó con người duy trì vai trò giám sát và điều chỉnh (Amershi et al., 2014).

5.2. Yêu cầu về năng lực của giảng viên

Việc tích hợp AI vào đánh giá đặt ra yêu cầu về một tổ hợp năng lực gồm năng lực chuyên môn và đánh giá, năng lực AI, năng lực đạo đức và năng lực giải trình. Năng lực chuyên môn vẫn là nền tảng để xây dựng tiêu chí, lựa chọn bằng chứng và xác định mức độ đạt chuẩn của người học.

Bên cạnh đó, năng lực AI không chỉ là thao tác kỹ thuật mà còn bao gồm hiểu giới hạn của hệ thống, nhận diện đầu ra không đáng tin cậy và lựa chọn mức độ sử dụng phù hợp (Luo et al., 2025). Năng lực đạo đức liên quan đến khả năng nhận diện thiên lệch, bảo đảm minh bạch, quyền riêng tư, công bằng và quyền lợi của người học (Holmes et al., 2022; Floridi & Cowls, 2019). Cuối cùng, năng lực giải trình đòi hỏi giảng viên có thể giải thích căn cứ của quyết định và vai trò của AI trong quá trình đánh giá, thay vì chấp nhận máy móc đầu ra của hệ thống.

5.3. Hàm ý đối với cơ sở giáo dục đại học

Trách nhiệm chuyên môn của giảng viên không thể được bảo đảm chỉ bằng yêu cầu cá nhân sử dụng AI có trách nhiệm mà cần được hỗ trợ bởi quy định và cơ chế quản trị của cơ sở giáo dục. Trước hết, cần quy định rõ phạm vi sử dụng AI trong đánh giá, bao gồm các hình thức được phép, phải khai báo và không được phép.

Đồng thời, cần bảo đảm sự giám sát và can thiệp của con người đối với các quyết định có ảnh hưởng lớn đến người học, xác định rõ chủ thể chịu trách nhiệm cuối cùng và có cơ chế lưu vết, giải thích và rà soát quyết định. Đào tạo giảng viên cũng cần tập trung vào khả năng kiểm định đầu ra, nhận diện giới hạn của AI và thiết kế nhiệm vụ đánh giá phù hợp, thay vì chỉ hướng dẫn sử dụng công cụ. Điều này phù hợp với xu hướng

chuyển từ phát hiện và ngăn chặn AI sang thiết kế lại đánh giá nhằm bảo đảm tính xác thực và toàn vẹn học thuật (Kofinas et al., 2025; Peters & Angelov, 2025).

5.4. Từ kiểm soát AI sang quản trị trách nhiệm

Mô hình HITL cho thấy vấn đề cốt lõi không phải lựa chọn giữa “sử dụng AI” và “không sử dụng AI”, mà là quản trị cách AI tham gia vào quá trình đánh giá. Nếu quá chú trọng kiểm soát, cơ sở giáo dục có thể bỏ qua lợi ích của AI; nếu quá chú trọng tự động hóa, nguy cơ phụ thuộc vào hệ thống và suy giảm phán đoán chuyên môn có thể gia tăng.

Theo đó, trách nhiệm được phân bổ theo chức năng: AI hỗ trợ xử lý và cung cấp đầu ra; giảng viên kiểm chứng, diễn giải và phán đoán; cơ sở giáo dục thiết lập quy định, giám sát và cơ chế giải trình. Khung đề xuất được thể hiện như sau:

Bảng 1. Khung trách nhiệm chuyên môn của giảng viên trong quy trình đánh giá có sự tham gia của AI

Giai đoạn	Vai trò của AI	Trách nhiệm của giảng viên	Yêu cầu chính
Xác định nhiệm vụ	Hỗ trợ đề xuất	Xác định mục tiêu và giới hạn	Phù hợp mục tiêu đánh giá
Tạo/xử lý nội dung	Tạo hoặc xử lý dữ liệu	Kiểm tra đầu ra	Chính xác và phù hợp
Đánh giá	Đề xuất kết quả/ phản hồi	Thẩm định và điều chỉnh	Phán đoán chuyên môn
Ra quyết định	Cung cấp thông tin hỗ trợ	Quyết định cuối cùng	Minh bạch, công bằng
Sau đánh giá	Hỗ trợ phân tích phản hồi	Rà soát và điều chỉnh	Cải tiến liên tục

Nguồn: Tác giả tổng hợp và đề xuất

Khung trên nhấn mạnh rằng HITL không chỉ yêu cầu có con người tham gia mà còn phải bảo đảm con người có quyền kiểm soát thực chất, năng lực thẩm định và trách nhiệm giải trình tại các điểm quyết định quan trọng.

VI. KẾT LUẬN

AI tạo sinh đang làm thay đổi hoạt động kiểm tra, đánh giá trong giáo dục đại học và đặt ra yêu cầu xem xét lại trách nhiệm chuyên môn của giảng viên. AI có thể hỗ trợ tạo nội dung, xử lý dữ liệu, phân tích bài làm và cung cấp phản hồi, nhưng đầu ra của AI không thể mặc nhiên trở thành căn cứ cuối cùng của quyết định đánh giá.

Trên cơ sở mô hình Human-in-the-loop, bài viết khái quát chuỗi AI-generated content →

AI-assisted evaluation → Teacher professional judgment → Accountability, qua đó cho thấy trách nhiệm của giảng viên không giảm khi AI được sử dụng mà chuyển từ trực tiếp thực hiện sang lựa chọn công cụ, kiểm soát đầu ra, thẩm định bằng chứng, phán đoán và giải trình quyết định. Vì vậy, ứng dụng AI trong đánh giá cần được tiếp cận theo hướng quản trị trách nhiệm, với quy định rõ về phạm vi sử dụng, giám sát của con người, tính minh bạch và cơ chế xem xét lại. Đồng thời, giảng viên cần được phát triển năng lực chuyên môn, đánh giá và AI nhằm khai thác lợi ích của công nghệ mà vẫn bảo đảm tính công bằng, minh bạch và toàn vẹn của hoạt động đánh giá.

TÀI LIỆU THAM KHẢO

Amershi, S., Cakmak, M., Knox, W. B., & Kulesza, T. (2014). Power to the people: The role of humans in interactive machine learning. *AI Magazine*, 35(4), 105–120. <https://doi.org/10.1609/aimag.v35i4.2513>

Dawson, P. (2021). *Defending assessment security in a digital world: Preventing e-cheating and supporting academic integrity in higher education*. Routledge. <https://doi.org/10.4324/9780429324178>

Floridi, L., & Cows, J. (2019). A unified framework of five principles for AI in society. *Harvard Data Science Review*,

1(1). <https://doi.org/10.1162/99608f92.8cd550d1>

Holmes, W., Porayska-Pomsta, K., Holstein, K., Sutherland, E., Baker, T., Buckingham Shum, S., Santos, O. C., Rodrigo, M. T., Cukurova, M., Bittencourt, I. I., & Koedinger, K. R. (2022). Ethics of AI in education: Towards a community-wide framework. *International Journal of Artificial Intelligence in Education*, 32(3), 504–526. <https://doi.org/10.1007/s40593-021-00239-1>

Kofinas, A. K., Tsay, C.-H., & Pike, D. (2025). The impact of generative AI on academic integrity of authentic assessments within a higher education context. *British Journal of Educational Technology*, 56(6), 2522–2549. <https://doi.org/10.1111/bjet.13585>

Lodge, J. M., Thompson, K., & Corrin, L. (2023). Mapping out a research agenda for generative artificial intelligence in tertiary education. *Australasian Journal of Educational Technology*, 39(1), 1–8. <https://doi.org/10.14742/ajet.8695>

Luo, J., Zheng, C., Yin, J., & Teo, H. H. (2025). Design and assessment of AI-based learning tools in higher education: A systematic review. *International Journal of Educational Technology in Higher Education*, 22, Article 42. <https://doi.org/10.1186/s41239-025-00540-2>

Miao, F., & Holmes, W. (2023). Guidance for generative AI in education and research. UNESCO. <https://doi.org/10.54675/EWZM9535>

Munoz, A., Hinchcliff, M., Langfield, C., & Rogerson, A. (2026). How strong is the evidence in generative AI-related academic misconduct allegations? A mixed-methods analysis. *International Journal for Educational Integrity*, 22, Article 26. <https://doi.org/10.1007/s40979-026-00235-9>

Peters, M., & Angelov, D. (2025). Redefining assessment tasks to promote students' creativity and integrity in the age of generative artificial intelligence. *International Journal for Educational Integrity*, 21, Article 25. <https://doi.org/10.1007/s40979-025-00201-x>

Selwyn, N. (2019). *Should robots replace teachers? AI and the future of education*. Polity Press.

Sharma, S. (2026). Educational integrity in GenAI-augmented assessment: Making judgement visible. *International Journal for Educational Integrity*, 22, Article 6. <https://doi.org/10.1007/s40979-026-00216-y>

Wright, C. (2026). Transcription is not generation: Distinguishing non-generative AI tool use from academic misconduct in higher education assessment. *International Journal for Educational Integrity*, 22, Article 24. <https://doi.org/10.1007/s40979-026-00234-w>