

KIỂM TRA, ĐÁNH GIÁ Ở BẬC ĐẠI HỌC: TÁI THIẾT KẾ TRONG KỶ NGUYÊN AI TẠO SINH

TẠ LÊ MỸ HẠNH

Khoa Ngoại ngữ, Đại học Bách khoa Hà Nội

Nhận bài ngày 12/4/2026. Sửa chữa xong 20/5/2026. Duyệt đăng 27/5/2026.

Abstract

The rise of generative artificial intelligence is posing unprecedented challenges to higher education assessment systems. AI's ability to outperform students in traditional examinations without being detected has exposed the inadequacies of current competency measurement methods, which remain overly focused on monitoring and detecting AI use rather than improving assessment design itself. This article synthesises the latest research to clarify the challenges that generative AI presents to conventional assessment methods. At the same time, it also proposes theoretical frameworks and necessary conditions for redesigning assessment that emphasises core competencies which AI cannot replace.

Keywords: Academic integrity, assessment, generative AI, higher education, redesign.

1. Đặt vấn đề

Năm 2023, ChatGPT đạt điểm cao hơn sinh viên (SV) y khoa trong bài kiểm tra giải phẫu 40 câu tại một trường đại học Thổ Nhĩ Kỳ; cùng năm đó, các giám khảo độc lập tại Anh chấm điểm xuất sắc cho một bài luận vật lý do ChatGPT viết mà không phát hiện bất kỳ dấu hiệu đạo văn nào (Evangelista, 2025; Xia và cộng sự, 2024). Những sự kiện này đặt ra một câu hỏi đáng lo ngại hơn cả vấn đề gian lận trong kiểm tra, đánh giá (KT-ĐG). Đó là nếu AI có thể vượt qua SV trong các bài đánh giá mà chúng ta thiết kế, vậy rồi cuộc những bài đánh giá đó đang đo lường điều gì?. Sự xuất hiện của AI tạo sinh, đặc biệt là các mô hình ngôn ngữ lớn như GPT-4o, Claude và Gemini đã mang đến những năng lực vượt xa việc tóm tắt văn bản đơn giản, bao gồm những năng lực như phân tích ngôn ngữ, sáng tác và giải quyết vấn đề phức tạp. Những năng lực này có thể được phân thành bốn nhóm dựa trên khả năng tiếp cận, khả năng cá nhân hóa, khả năng tự động hóa và khả năng tương tác (Lye & Lim, 2024). UNESCO khuyến cáo tốc độ phát triển của AI tạo sinh đã vượt quá khả năng thích ứng của các khung pháp lý quốc gia, khiến phần lớn các cơ sở giáo dục không kịp đưa ra những hướng dẫn rõ ràng về mặt qui định hay thể chế (Miao & Holmes, 2023). Xia và cộng sự (2024) cho rằng, trong tất cả các khía cạnh của giáo dục đại học (GDĐH) chịu ảnh hưởng bởi AI tạo sinh, KT-ĐG là lĩnh vực gây tranh cãi nhất bởi nó gắn trực tiếp với mục tiêu học tập, thực hành giảng dạy và chứng nhận năng lực. Tuy nhiên, theo Kickbusch và cộng sự (2025), phần lớn các phản ứng mang tính thể chế hiện nay vẫn tập trung vào việc giám sát và phát hiện gian lận, trong khi vấn đề cốt lõi lại nằm ở những hạn chế trong thiết kế của các hoạt động KT-ĐG.

Bài viết phân tích các nghiên cứu gần đây về KT-ĐG ở bậc đại học trong bối cảnh AI tạo sinh theo bốn hướng: xác lập phạm vi thách thức mà AI tạo sinh đặt ra cho KT-ĐG truyền thống; phân tích các khung lý thuyết được đề xuất cho việc tái thiết kế KT-ĐG và xem xét các điều kiện thể chế và chuyên môn cần thiết cho quá trình tái thiết kế KT-ĐG gia trong thời đại AI.

2. Nội dung nghiên cứu

Bài viết được triển khai theo ba phần chính. Phần thứ nhất, phân tích những thách thức mà AI tạo sinh đặt ra đối với các phương thức KT-ĐG truyền thống trong GDĐH, qua đó làm rõ những hạn chế và lỗ hổng mang tính cấu trúc của các hình thức KT-ĐG hiện tại. Phần thứ hai, bài viết trình bày sơ lược các

Email: hanh.talemy@hust.edu.vn

khung lý thuyết giúp định hướng quá trình tái thiết kế KT-ĐG trong bối cảnh AI ngày càng được tích hợp vào hoạt động học tập. Thứ ba, bài viết thảo luận các điều kiện mang tính định chế cần thiết để thúc đẩy cải cách KT-ĐG một cách bền vững.

2.1. Thách thức của AI tạo sinh đối với kiểm tra, đánh giá truyền thống

Sự xuất hiện của AI tạo sinh đã đặt ra những thách thức căn bản đối với các phương thức KT-ĐG truyền thống. Điều này không chỉ liên quan đến việc SV có thể sử dụng AI để hoàn thành bài tập mà quan trọng hơn, điều này đang làm bộc lộ những lỗ hổng có tính hệ thống về cấu trúc của các bài KT-ĐG hiện nay.

2.1.1. Bằng chứng thực tế về năng lực của AI tạo sinh

Mối đe dọa mà AI tạo sinh gây ra đối với đánh giá truyền thống không chỉ còn là giả thuyết. Một loại những bằng chứng từ nhiều lĩnh vực cho thấy các mô hình ngôn ngữ lớn có khả năng đạt hoặc vượt ngưỡng điểm đạt trong nhiều bài KT-ĐG chuyên ngành. Một đánh giá hệ thống dựa trên 753 hồ sơ ghi nhận ChatGPT đã vượt qua Kỳ thi Cấp phép Hành nghề Y khoa Hoa Kỳ (USMLE) và có điểm số vượt qua mức điểm số của bậc đại học trong bài kiểm tra giải phẫu gồm 40 câu. Một đánh giá phạm vi (scoping review) phân tích 32 nghiên cứu thực nghiệm trên 11 lĩnh vực cho thấy ChatGPT tạo ra câu trả lời thuộc lĩnh vực kinh tế học được chuyên gia xếp loại “đạt”, giải các bài tập kỹ thuật ở mức chấp nhận được và đặc biệt, tạo ra một bài luận vật lý được 05 giám khảo độc lập chấm điểm xuất sắc (Xia và cộng sự, 2024). Một điều đáng lưu ý là các bài viết do AI tạo ra này không bị phát hiện bởi các công cụ kiểm tra AI thông thường. Những kết quả này có ý nghĩa quan trọng không phải vì chúng xác lập sự ưu việt về nhận thức của AI so với con người mà vì chúng làm bộc lộ những lỗ hổng về cấu trúc của các phương thức KT-ĐG quá chú trọng vào việc ghi nhớ và tóm tắt - những việc AI có thể thực hiện một cách dễ dàng. Những bằng chứng từ nhiều ngành, lĩnh vực nêu trên cho thấy lỗ hổng này mang tính hệ thống chứ không phải chỉ là ngẫu nhiên.

2.1.2. Sự thất bại của các biện pháp ứng phó dựa trên việc phát hiện gian lận

Phản ứng phổ biến nhất của các cơ sở giáo dục là triển khai các phần mềm phát hiện AI nhưng biện pháp này đang vấp phải những chỉ trích nghiêm trọng cả về kỹ thuật lẫn sư phạm. Về mặt kỹ thuật, việc truy vết nội dung AI được nhận định là gần như không thể (Xia và cộng sự, 2024). Các nền tảng phổ biến như Turnitin, ZeroGPT hay GPTZero thường xuyên mắc sai số nghiêm trọng ở cả hai chiều: sai số âm tính bỏ lọt hành vi sử dụng AI tinh vi, trong khi sai số dương tính dẫn đến xử phạt nhầm SV. Theo Evangelista (2025), cả hai loại sai số này đều thiếu độ chuẩn xác để làm căn cứ pháp lý cho qui trình kỷ luật học thuật. Đáng quan ngại hơn, các công cụ phát hiện AI thường tồn tại sự thiên lệch đối với người học Tiếng Anh không phải bản ngữ do cách viết của họ đôi khi có thể bị nhận diện nhầm là văn bản do AI tạo ra, từ đó đặt ra những lo ngại nghiêm trọng về tính công bằng trong KT-ĐG (Ley & Lim, 2024). Về mặt sư phạm, cách tiếp cận này biến quá trình KT-ĐG thành mối quan hệ đối đầu, trực tiếp xói mòn niềm tin giữa người dạy và người học - một nền tảng cốt lõi của một nền sư phạm hiệu quả (Kickbusch và cộng sự, 2025). Từ thực trạng đó, các nghiên cứu dẫn đi đến đồng thuận là mô hình phát hiện và trừng phạt không còn khả thi, và các cơ sở giáo dục cần chuyển dịch chiến lược sang thiết kế các phương thức KT-ĐG sáng tạo thay vì chỉ tập trung vào tăng cường giám sát việc sử dụng AI (Ncube và cộng sự, 2025).

2.1.3. Vấn đề về thiết kế hệ thống, không phải vấn đề đạo đức

Việc SV sử dụng AI để hoàn thành các bài KT-ĐG không nên được nhìn nhận đơn thuần là một vấn đề đạo đức cá nhân mà thực chất là hệ quả của các đặc tính cấu trúc của những bài KT-ĐG được thiết kế kém hiệu quả (Evangelista, 2025). Khi một nhiệm vụ học tập không đòi hỏi sự hiểu biết sâu, không gắn với các mục tiêu học tập rõ ràng và cũng không yêu cầu khả năng phán đoán trong bối cảnh cụ thể, AI vô tình sẽ trở thành một “lối tắt hợp lý” cho SV. Các GV cũng thừa nhận rằng nhiều dạng bài tập hiện nay quá dễ để AI hoàn thành, điều này củng cố quan điểm cho rằng đây là vấn đề liên quan đến thiết kế hệ thống hơn là những sai phạm mang tính đạo đức và nhân cách của SV (Khlaif và cộng sự, 2025). Evangelista (2025) nhấn mạnh rằng tình trạng này nếu kéo dài còn có thể dẫn đến nguy cơ lệ thuộc nhận thức, gây tổn hại đến tiến trình phát triển trí tuệ lâu dài của SV.

Sự tái định nghĩa này đòi hỏi việc chuyển trọng tâm can thiệp từ giám sát hành vi của SV sang cải tiến thiết kế KT-ĐG, từ thực thi kỷ luật sang đổi mới tư phạm. Điều này hoàn toàn tương thích với lời kêu gọi của UNESCO về việc xem xét lại không chỉ phương pháp KT-ĐG mà cả những giá trị thực sự cần được đánh giá trong bối cảnh AI ngày càng có khả năng thực hiện nhiều nhiệm vụ nhận thức phức tạp (Miao & Holmes, 2023). Thay vì tăng cường giám sát bài làm của SV, mục tiêu cần hướng tới là kiến tạo những phương thức KT-ĐG trong đó việc giám sát không còn cần thiết (Lye & Lim, 2024; Kickbusch và cộng sự, 2025).

2.2. Một số khung lý thuyết cho tái thiết kế kiểm tra, đánh giá

Trong bối cảnh AI tạo sinh ngày càng tham gia sâu vào hoạt động học thuật, các nhà nghiên cứu ngày càng nhấn mạnh nhu cầu tái thiết kế KT-ĐG thay vì tập trung vào phát hiện và kiểm soát việc sử dụng AI. Lập trường này được thể hiện qua nhận định rằng tính xác thực (authenticity) không thể được tạo ra chỉ bằng việc giám sát mà phải được tái thiết kế (Kickbusch và cộng sự, 2025) và việc chạy đua về các công nghệ phát hiện AI là vừa thiếu hiệu quả vừa chưa giải quyết đúng bản chất vấn đề vì việc cần thiết là chuyển đổi triết lý đánh giá (Lye & Lim, 2024). Có thể nhận diện bốn chuyển dịch cấu trúc đang định hình các khung KT-ĐG mới, bao gồm: chuyển trọng tâm đánh giá từ sản phẩm sang quá trình, từ sở hữu tri thức sang ứng dụng và phán đoán, từ phát hiện gian lận sang tái thiết kế và từ làm bài độc lập sang hợp tác có phân biện giữa con người và AI.

2.2.1. Chuyển từ sản phẩm sang quá trình

Awadallah Alkhouk và Khlaif (2024) đề xuất mô hình đánh giá cả quá trình lẫn sản phẩm, khác với logic truyền thống chỉ tập trung đánh giá sản phẩm cuối cùng. Theo mô hình này, phạm vi đánh giá được mở rộng để bao gồm chất lượng câu lệnh (prompt) SV xây dựng khi tương tác với AI, các quyết định phân biện trong quá trình chấp nhận hay chỉnh sửa nội dung AI tạo ra và mức độ sản phẩm phản ánh quyền sở hữu trí tuệ thực sự của người học (Awadallah Alkhouk & Khlaif, 2024). Hướng tiếp cận đó được Kickbusch và cộng sự (2025) cụ thể hóa qua nguyên tắc đảm bảo tính minh bạch của quá trình: SV nộp kèm báo cáo ghi lại toàn bộ quá trình tương tác với AI, bao gồm các câu lệnh đã dùng và lý do chấp nhận hay từ chối từng phản hồi của AI. Theo lập luận của nhóm tác giả, một SV có thể giải thích được tại sao họ từ chối gợi ý của AI thể hiện mức độ học tập sâu hơn người chỉ nộp sản phẩm hoàn chỉnh mà không để lại dấu vết phân biện nào.

2.2.2. Từ sở hữu tri thức sang ứng dụng và phán đoán

Thang phân loại Bloom sửa đổi (Bloom's Revised Taxonomy) là khung phân tích được sử dụng phổ biến nhất để phân biệt các nhiệm vụ KT-ĐG theo mức độ dễ bị AI thay thế. Các nhiệm vụ nhận thức bậc thấp như hiểu và ghi nhớ là những nhiệm vụ mà AI tạo sinh thực hiện dễ dàng nhất, trong khi các nhiệm vụ bậc cao như phân tích, đánh giá và sáng tạo có thể chống AI tốt hơn vì những nhiệm vụ này đòi hỏi khả năng phán đoán theo ngữ cảnh, suy luận, và tính sáng tạo. Logic này được tích hợp trực tiếp vào Thang kiểm tra đánh giá kháng AI (AI-Resistance Assessment Scale - AIAS) của Awadallah Alkhouk và Khlaif (2024), và được Khlaif và cộng sự (2025) khuyến nghị như một nguyên tắc thiết kế quan trọng.

Kickbusch và cộng sự (2025) đưa sự chuyển dịch này xa hơn thông qua nguyên tắc đánh giá năng lực phán đoán. Theo nguyên tắc này, trọng tâm không còn là SV "biết gì" mà là SV "quyết định như thế nào" khi vận dụng tri thức trong bối cảnh AI. Một ví dụ minh họa điển hình là nhiệm vụ tự đánh giá trong y khoa, trong đó SV được yêu cầu viết về một tình huống đạo đức họ trực tiếp trải qua trong thực hành lâm sàng, bao gồm phân tích xung đột, trình bày lập luận cá nhân và liên hệ với sự phát triển nghề nghiệp. Nhiệm vụ này giúp kháng AI một cách hiệu quả vì nó đòi hỏi trải nghiệm cá nhân và năng lực phán đoán vượt ra ngoài thao tác truy xuất thông tin đơn thuần. Tuy nhiên, có một thực tế là giới hạn năng lực của AI tạo sinh không phải là cố định. Các phiên bản mới ngày càng thực hiện tốt hơn những nhiệm vụ trước đây được xem là đòi hỏi việc phân tích sâu và SV vẫn có thể tạo ra câu trả lời dường như đạt được mức độ nhận thức bậc cao bằng cách tiếp nhận câu trả lời của AI mà không cần phân biện. Do đó, thang đo Bloom là công cụ định hướng hữu ích nhưng không nên chỉ áp dụng một cách máy móc. Bối cảnh xã hội và điều kiện đánh giá cụ thể là những yếu tố quan trọng không kém cấp

độ nhận thức của nhiệm vụ KT-ĐG (Lye & Lim, 2024).

2.2.3. Từ phát hiện gian lận sang tái thiết kế kiểm tra, đánh giá

Lye và Lim (2024) đề xuất khung Chống-Tránh-Áp dụng (Against-Avoid-Adopt, AAA), cung cấp định hướng ra quyết định ngay từ giai đoạn thiết kế nhiệm vụ KT-ĐG giá thay vì chỉ phản ứng sau khi vi phạm xảy ra. Nhóm Chống bao gồm các nhiệm vụ mà AI không thể hoàn thành vì những nhiệm vụ này đòi hỏi trải nghiệm sống thực tế hoặc tri thức địa phương. Nhóm Tránh hướng đến giảm thiểu điều kiện khiến việc sử dụng AI như một lối tắt trở nên hấp dẫn hơn, thông qua hồ sơ học tập, đánh giá thực hành hoặc nộp bài thành nhiều giai đoạn. Nhóm Áp dụng chủ động tích hợp AI vào đánh giá, yêu cầu SV giải trình về những lựa chọn của mình khi sử dụng nội dung do AI tạo ra. Đóng góp quan trọng nhất của khung AAA là thay đổi cách lý giải nguyên nhân lạm dụng AI. Cụ thể, thay vì xem đây là vấn đề đạo đức cá nhân, khung này chỉ ra rằng nguyên nhân nằm ở chính cấu trúc của nhiệm vụ đánh giá - yếu tố có thể tạo điều kiện cho hành vi vi phạm. Điều này giúp chuyển trọng tâm từ kiểm soát hành vi sang thiết kế lại hệ thống một cách căn bản hơn. Khung AAA cũng đồng thời nhận diện các rủi ro khác của việc tích hợp AI, bao gồm hiện tượng “ảo giác” của AI - hiện tượng đặc biệt nguy hiểm trong các ngành y khoa và luật hay sự thiên lệch mang tính hệ thống do dữ liệu huấn luyện mất cân bằng cùng với nguy cơ suy giảm năng lực tư duy khi SV chấp nhận kết quả do AI tạo ra mà không kiểm chứng lại.

2.2.4. Từ làm bài độc lập sang hợp tác có phản biện giữa người và AI

Kickbusch và cộng sự (2025) đặt lại vấn đề về khái niệm tính xác thực (authenticity) trong KT-ĐG. Thay vì xem xác thực là thuộc tính của sản phẩm, tức bài luận chỉ xác thực khi không có sự hỗ trợ của AI, nhóm tác giả đề xuất nên hiểu tính xác thực như thuộc tính của quá trình tương tác với tri thức và công cụ. Một bài KT-ĐG được xem là xác thực khi việc dùng AI được dự đoán trước, công khai và được xem xét một cách phản biện. Từ đó, nguyên tắc sử dụng AI có phản biện và đạo đức được hình thành: nhiệm vụ KT-ĐG phải buộc SV chủ động phân tích, chỉnh sửa và kiểm chứng nội dung AI tạo ra thay vì chỉ tiếp nhận một cách thụ động.

Nghiên cứu định tính tiến hành trên 61 GV thuộc nhiều lĩnh vực xác nhận rằng việc giúp SV sẵn sàng cho môi trường nghề nghiệp tích hợp AI là một trong ba động lực chính thúc đẩy tái thiết kế KT-ĐG. Điều này phù hợp với lập luận của Kickbusch và cộng sự rằng AI đã trở thành công cụ tiêu chuẩn trong nhiều ngành nghề, do vậy, các hình thức KT-ĐG cấm đoán việc sử dụng AI sẽ khiến SV bị tách rời khỏi môi trường làm việc họ sẽ gia nhập.

2.3. Điều kiện định chế cho cải cách kiểm tra, đánh giá bền vững

2.3.1. Khung chính sách quốc gia và toàn cầu

Cải cách KT-ĐG không thể chỉ dựa trên sự đổi mới đơn lẻ của từng GV mà đòi hỏi các điều kiện tiên quyết về chính sách định chế và cơ sở hạ tầng phát triển chuyên môn. Khung chính sách quốc tế đầu tiên về AI tạo sinh của UNESCO đã kêu gọi các quốc gia thực hiện các biện pháp bảo vệ tức thời về quyền riêng tư dữ liệu, đồng thời xây dựng chiến lược quốc gia dài hạn nhằm nâng cao năng lực con người (Miao & Holmes, 2023). Trong bối cảnh này, việc đánh giá và xác nhận kết quả học tập cần được tái định nghĩa không chỉ ở cấp độ quy định nội bộ mà còn ở ý nghĩa xã hội của bằng cấp, đặc biệt là trong việc giải quyết hồ ngăn cách kỹ thuật và sự phân bố không đều của hạ tầng AI để đảm bảo tính công bằng. Ở cấp độ quốc gia, các hướng dẫn từ những cơ quan kiểm định chất lượng như QAA (Anh) hay TEQSA (Úc) đóng vai trò là khung tham chiếu quan trọng hỗ trợ các cơ sở giáo dục trong quá trình thích ứng này.

2.3.2. Thực thi chính sách tại các cơ sở giáo dục

Các nguyên tắc chung của UNESCO được cụ thể hóa thông qua nhiều mô hình thực tiễn tại các đại học hàng đầu. Điển hình như Đại học Stanford (Mỹ) đã chuyển đổi sang mô hình đánh giá dựa trên dự án (project-based), yêu cầu việc ứng dụng kiến thức vào thực tiễn. Đại học Duke (Mỹ) thành lập hội đồng đạo đức AI để duy trì khung tiêu chuẩn chung hay Đại học Sydney (Úc) tăng cường các hình thức thi vấn đáp và làm việc nhóm - những cơ chế có khả năng chống lại về mặt cấu trúc đối với sự thay thế

của AI. Tuy nhiên, nhiều nghiên cứu cho thấy hiệu quả của chính sách thường phụ thuộc vào sự nhất quán và các can thiệp hệ thống như xây dựng lộ trình kỹ năng số (digital literacy), tránh tình trạng thiếu đồng bộ giữa các khoa chuyên môn trong thực tiễn triển khai (Ncube và cộng sự, 2025; Khlaif và cộng sự, 2025; Awadallah Alkhouk & Khlaif, 2024).

2.3.3. Phát triển chuyên môn cho đội ngũ giảng viên

Việc thiết kế các bài KT-ĐG có khả năng “kháng AI” đòi hỏi sự giao thoa giữa ba mảng năng lực cốt lõi: năng lực AI (AI literacy), năng lực đánh giá (assessment literacy) và năng lực số (digital literacy). Các mô hình đào tạo hiệu quả cần chuyển từ thảo luận lý thuyết sang các hoạt động thiết kế thực hành, giúp GV tạo ra và tinh chỉnh các nguyên mẫu đánh giá ngay trong quá trình đào tạo. Dù vậy, tiến trình này vẫn đối mặt với nhiều rào cản như áp lực thời gian, lo ngại về sự bất bình đẳng trong tiếp cận công cụ của SV và sự phản kháng đối với thay đổi sự phạm nội bộ, đòi hỏi các chương trình phát triển chuyên môn phải mang tính bền vững, dựa theo bối cảnh và được hỗ trợ mạnh mẽ từ phía định chế thay vì chỉ là các sự kiện đào tạo ngắn hạn (Khlaif và cộng sự, 2025).

3. Kết luận

Như vậy, việc AI tạo sinh có thể đạt điểm cao hơn SV trong các kỳ thi không nên được coi là minh chứng cho sự vượt trội của máy móc mà là chỉ dấu cho thấy các bài KT-ĐG hiện tại chỉ đang đo lường những gì AI có thể tái tạo chứ chưa đánh giá được năng lực phán đoán, tư duy đạo đức, ứng dụng theo bối cảnh và tư duy phản biện, vốn là những yếu tố cốt lõi còn nằm ngoài khả năng mô phỏng của AI. Bài viết chỉ ra bốn chuyển dịch cấu trúc trong triết lý đánh giá: từ sản phẩm sang quá trình, từ sở hữu sang ứng dụng kiến thức, từ phát hiện gian lận sang tái thiết kế cấu trúc, và từ làm việc độc lập sang cộng tác giữa người và AI. Các chuyển dịch này được hỗ trợ bởi các công cụ hỗ trợ như nguyên tắc AAA, thang đo AIAS, mô hình chuyển từ sản phẩm sang quá trình, và thang đo Bloom cải tiến (Lye & Lim, 2024; Kickbusch và cộng sự, 2025; Khlaif và cộng sự, 2025). Tuy nhiên, cơ sở thực chứng của các khung hỗ trợ này vẫn còn chưa đầy đủ. Các mô hình đánh giá hiện có chủ yếu dựa trên phân tích lý thuyết, tập trung vào một số ngành và khu vực địa lý nhất định, trong khi tiếng nói của SV - chủ thể trực tiếp của đánh giá vẫn chưa được khai thác một cách hệ thống. Vì vậy, hướng nghiên cứu tiếp theo cần ưu tiên các thử nghiệm có đối chứng, mở rộng phạm vi ngành học cũng như phạm vi địa lý, đồng thời đưa quan điểm người học vào trung tâm của thiết kế KT-ĐG. Sự trỗi dậy của AI tạo sinh không để KT-ĐG trở nên lỗi thời mà làm bộc lộ những bất cập của những phương KT-ĐG hiện tại. Thách thức của GDĐH là thiết kế những bài KT-ĐG chứng minh được những giá trị không thể thay thế của con người, thay vì chỉ yêu cầu SV thực hiện những tác vụ mà máy móc có thể làm tốt hơn.

Tài liệu tham khảo

- Awadallah Alkhouk, W., & Khlaif, Z. N. (2024). *AI-resistant assessments in higher education: Practical insights from faculty training workshops*. *Frontiers in Education*, 9, Article 1499495. <https://doi.org/10.3389/educ.2024.1499495>.
- Evangelista, E. D. L. (2025). *Ensuring academic integrity in the age of ChatGPT: Rethinking exam design, assessment strategies, and ethical AI policies in higher education*. *Contemporary Educational Technology*, 17(1), Article ep559. <https://doi.org/10.30935/cedtech/15775>.
- Khlaif, Z. N., Alkhouk, W. A., Salama, N., & Abu Eideh, B. (2025). *Redesigning assessments AI-enhanced learning: A framework for educators in the generative AI era*. *Education Sciences*, 15(2), Article 174. <https://doi.org/10.3390/educsci15020174>.
- Kickbusch, S., Ashford-Rowe, K., Kemp, A., Boreland, J., & Huijsers, H. (2025). *Beyond detection: Redesigning authentic assessment in an AI-mediated world*. *Education Sciences*, 15(11), Article 1537. <https://doi.org/10.3390/educsci15111537>.
- Lye, C. Y., & Lim, L. (2024). *Generative artificial intelligence in tertiary education: Assessment redesign principles and considerations*. *Education Sciences*, 14(6), Article 569. <https://doi.org/10.3390/educsci14060569>.
- Miao, F., & Holmes, W. (2023). *Guidance for generative AI in education and research*. UNESCO. <https://doi.org/10.54675/EWBV4364>.
- Ncube, P. D. N., Dzvapatsva, G. P., Matobobo, C., & Ranga, M. M. (2025). *Redefining student assessment in AI-infused learning environments: A systematic review of challenges and strategies for academic integrity*. *AI & Ethics*. Advance online publication. <https://doi.org/10.1007/s43681-025-00871-w>.
- Xia, Q., Weng, X., Ouyang, F., Lin, T. J., & Chiu, T. K. F. (2024). *A scoping review on how generative artificial intelligence transforms assessment in higher education*. *International Journal of Educational Technology in Higher Education*, 21(1), Article 40. <https://doi.org/10.1186/s41239-024-00468-z>.