TÁC ĐỘNG CỦA KHUNG HƯỚNG DẪN MINH BẠCH AI ĐẾN HÀNH VI VÀ CHẤT LƯỢNG VIẾT LUẬN HỌC THUẬT CỦA SINH VIÊN NĂM THỨ HAI KHOA TIẾNG ANH, TRƯỜNG ĐẠI HỌC MỞ HÀ NỘI

Vũ Diệu Thúy[1]
Email: vdthuy_foe@hou.edu.vn, ORCID: 0009-0007-4622-4295

Ngày tòa soạn nhận được bài báo: 15/03/2026. Ngày phản biện đánh giá: 15/05/2026.
Ngày bài báo được duyệt đăng: 01/06/2026
DOI: 10.59266/houjs.2026.1273

Tóm tắt: *Nghiên cứu đánh giá tác động của Khung hướng dẫn Minh bạch AI đến hành vi và chất lượng viết luận của sinh viên năm hai Khoa Tiếng Anh, Trường Đại học Mở Hà Nội. Được thiết kế như công cụ tạo ma sát nhận thức nhằm bảo vệ liêm chính học thuật, khung can thiệp được đánh giá qua phương pháp hỗn hợp (tiền thực nghiệm trên 30 sinh viên và phân tích nhật ký AI). Kết quả cho thấy 83,3% người học áp dụng chiến lược Lắp ghép (Bricolage), biến công nghệ thành giàn giáo nhận thức thay vì ủy thác hoàn toàn. Điểm số cải thiện có ý nghĩa thống kê ($p<0,001$): hình thức tăng nhẹ (+0,35), trong khi minh chứng (+0,98) và phản biện (+0,40) tăng mạnh. Nghiên cứu đề xuất chuyển dịch sang đánh giá quá trình nhằm phát triển Năng lực AI (Critical AI Literacy) bền vững cho người học.*

Từ khóa: *khung minh bạch AI, đánh giá quá trình, năng lực AI, trí tuệ nhân tạo tạo sinh, viết học thuật*

I. Đặt vấn đề

Sự trỗi dậy của các Mô hình ngôn ngữ lớn (LLMs) đã tạo ra đứt gãy phương pháp luận trong viết luận học thuật, đẩy sinh viên vào cạm bẫy trút bỏ gánh nặng nhận thức. Thách thức này đặc biệt nghiêm trọng với sinh viên năm hai Khoa Tiếng Anh, Trường Đại học Mở Hà Nội - nhóm đối tượng đang chuyển tiếp sang kỹ năng viết luận phức tạp nhưng thường lạm dụng công nghệ để bù đắp thiếu hụt tư duy phản biện. Trước sự vô hiệu của các biện pháp cấm đoán, nghiên cứu đề xuất Khung hướng dẫn Minh bạch AI. Khác với những quy định đơn thuần, khung này là một thiết kế sư phạm nhằm tái thiết lập ma sát nhận thức, buộc người học chậm lại để minh bạch hóa tương tác và kiểm chứng thông tin. Mục tiêu nghiên cứu là giải mã chiến lược tương tác và đo lường tác động lên chất lượng bài luận, từ đó biến LLMs từ mối đe dọa liêm chính thành hệ thống giàn giáo nhận thức.

[1] Trường Đại học Mở Hà Nội, Hà Nội, Việt Nam

II. Cơ sở lý thuyết

2.1. Sự chuyển đổi trong loại hình công nghệ giáo dục: Từ các công cụ sử dụng giáo trình đóng đến mô hình tạo sinh

Trước sự bùng nổ của Mô hình ngôn ngữ lớn (LLMs), công nghệ giáo dục chủ yếu mang tính công cụ hóa với các lộ trình khép kín (Web 2.0). Một số thảo luận gần đây về giảng dạy viết học thuật tại Việt Nam cho thấy việc giảng dạy tư duy bậc cao như Viết học thuật vẫn chủ yếu dựa vào phương pháp truyền thống, thiếu vắng sự hiện diện của Trí tuệ nhân tạo tạo sinh (GenAI). Tuy nhiên, GenAI đã tạo ra một sự đứt gãy phương pháp luận trong phân môn Viết ngoại ngữ. Vượt khỏi vai trò kiểm tra lỗi tĩnh, GenAI đang trở thành gia sư ngôn ngữ cá nhân, chuyển dịch việc học từ tiêu thụ sang đồng kiến tạo (Moorhouse & Wong, 2025). Dù vậy, môi trường tương tác mở thiếu định hướng này cũng đang đặt ra những thách thức nghiêm trọng về liêm chính và năng lực học thuật.

2.2. Cơ sở tâm lý học của việc lạm dụng GenAI

Sự lạm dụng LLMs trong môi trường học thuật có thể được lý giải thông qua sự giao thoa giữa Mô hình Chấp nhận Công nghệ (TAM) của Davis (1989) và Thuyết Vùng phát triển gần (ZPD) của Vygotsky (1978). Theo TAM, Nhận thức về tính dễ sử dụng (PEoU) cực cao của LLMs đã kích thích mạnh mẽ ý định hành vi của sinh viên. Khi đóng vai trò Kẻ hiểu biết hơn (MKO) trong thuyết ZPD, LLMs cung cấp các phản hồi tức thì nhưng thiếu đi ma sát nhận thức (cognitive friction) cần thiết. Sự cộng hưởng giữa tác động của MKO và chỉ số PEoU cực đại này dẫn đến hiện tượng quá tải giàn giáo (over-scaffolding), khiến người học dễ dàng trút bỏ gánh nặng nhận thức (cognitive offloading) (Farrokhnia và cộng sự, 2024) thay vì rèn luyện tư duy nội tại. Cần lưu ý, Ma sát nhận thức là một thiết kế sư phạm có chủ đích nhằm tạo kích thích tư duy, khác biệt hoàn toàn với Gánh nặng nhận thức (Cognitive Load) là tổng nỗ lực tâm trí để xử lý thông tin.

2.3. Hệ lụy của việc lạm dụng GenAI đối với các giá trị cốt lõi của bài viết học thuật

Để đánh giá tác động của LLMs, nghiên cứu đối chiếu với 7 giá trị cốt lõi theo van Niekerk và cộng sự. (2025). Nhờ cơ chế dự đoán từ vựng (Jurafsky & Martin, 2024), LLMs mô phỏng tốt các tiêu chí hình thức: cấu trúc (C1), tính súc tích (C5), khách quan (C6) và giọng điệu (C7). Tuy nhiên, công cụ này thường thất bại trong các tiêu chí về mặt chiều sâu: minh chứng xác thực (C2), tư duy phản biện (C3) và tính cân bằng (C4). Do không thấu hiểu sự thật, LLMs dễ sinh ra ảo giác, dẫn đến ngụy tạo số liệu và trích dẫn sai lệch (Ji & cộng sự, 2023). Sự khiếm khuyết này khẳng định tính trơn tru của thuật toán không thể thay thế sự xác thực trong tư duy con người.

2.4. Tái định hình Năng lực Trí tuệ Nhân tạo (Critical AI Literacy)

Để đối trọng rủi ro xói mòn tư duy, giới học thuật kêu gọi phát triển Năng lực AI (Critical AI Literacy). Theo Bhusal (2025), năng lực AI phê phán đòi hỏi người học không xem ChatGPT như một nguồn thông tin trung lập, mà cần phản tư, kiểm chứng và can thiệp chủ động vào các đầu ra của công cụ. Biểu hiện thực tiễn của năng lực này chính là chiến lược

Lắp ghép (Bricolage), đây là quá trình người học chọn lọc, tái kết hợp và điều chỉnh các nguồn lực sẵn có vào khung ý tưởng của bản thân, dựa trên cách tiếp cận của Lévi-Strauss (1966). Dưới góc độ sư phạm, chiến lược này biến LLMs thành một giàn giáo nhận thức (Vygotsky, 1978). Đây là bước đệm cần thiết giúp sinh viên rèn luyện năng lực tư duy phản biện và làm chủ quy trình viết luận trước khi có thể độc lập thực hiện các nhiệm vụ học thuật mà không cần sự hỗ trợ của thuật toán.

2.5. Khoảng trống nghiên cứu và Sự cần thiết của Khung hướng dẫn Minh bạch AI

Trước những sự thay đổi trong hướng tiếp cận dạy và học viết do LLMs gây ra, Khung hướng dẫn Minh bạch AI được đề xuất như một thiết kế sư phạm nhằm chuyển hóa TAM và ZPD thành hành động thực tiễn. Thay vì là quy định đơn thuần, khung này tái thiết lập ma sát nhận thức trong quy trình viết bằng cách yêu cầu người học minh bạch hóa nhật ký tương tác và giải trình các bước kiểm chứng. Nhờ đó, trọng tâm đánh giá được chuyển từ kiểm soát sản phẩm cuối cùng sang đánh giá quá trình (Cotton & cộng sự, 2024). Cơ chế này đối trọng với hiện tượng quá tải giàn giáo và trút bỏ gánh nặng nhận thức, đồng thời tạo điều kiện cho chiến lược Lắp ghép phát huy hiệu quả. Khung minh bạch cũng giúp giảng viên quan sát hành vi và tư duy của sinh viên, qua đó định hướng LLMs thành giàn giáo nhận thức nhằm bảo vệ liêm chính học thuật và phát triển Năng lực AI bền vững.

2.6. Câu hỏi nghiên cứu

1. (RQ1) Khung hướng dẫn Minh bạch AI ảnh hưởng như thế nào đến chiến lược tương tác với LLMs của sinh viên năm hai ngành Ngôn ngữ Anh, Trường Đại học Mở Hà Nội?

2. (RQ2) Qua so sánh pre-test và post-test, khung hướng dẫn Minh bạch AI tác động như thế nào đến chất lượng bài luận học thuật của sinh viên, đặc biệt với tiêu chí về minh chứng xác thực (C2) và tư duy phản biện (C3)?

III. Phương pháp nghiên cứu

Để trả lời cho các câu hỏi nghiên cứu đã đề ra, phần này trình bày chi tiết về phương pháp luận được áp dụng. Cụ thể, nội dung bao gồm việc thiết lập mô hình nghiên cứu, mô tả khách thể tham gia, trình bày chi tiết quy trình can thiệp sư phạm tích hợp Khung hướng dẫn Minh bạch AI, và cuối cùng là các công cụ cùng phương pháp phân tích dữ liệu.

3.1. Thiết kế và khách thể nghiên cứu

Nghiên cứu sử dụng phương pháp hỗn hợp, kết hợp thiết kế tiền thực nghiệm kiểm tra trước - sau một nhóm với phân tích định tính hồ sơ sử dụng AI. Dữ liệu định lượng được dùng để đo sự thay đổi chất lượng bài viết, trong khi dữ liệu định tính từ bản khai báo và nhật ký câu lệnh giúp làm rõ chiến lược tương tác với LLMs của sinh viên. Khách thể nghiên cứu gồm 30 sinh viên năm hai ngành Ngôn ngữ Anh, Trường Đại học Mở Hà Nội, thuộc lớp Viết 4 Nhóm 8 K31. Nhóm này được lựa chọn theo phương pháp chọn mẫu thuận tiện có chủ đích vì đã hoàn thành các học phần Viết 1-3, có trình độ đầu vào tương đương B2 theo CEFR và đang chuyển sang giai đoạn rèn luyện kỹ năng viết luận học thuật phức tạp.

3.2. Quy trình can thiệp sư phạm

Quá trình thực nghiệm được thiết kế dựa trên mô hình Viết theo quy trình (Process Writing) có tích hợp đánh giá định hướng AI (AI-Integrated Assessment). Trọng tâm can thiệp sư phạm là việc áp dụng Khung hướng dẫn Minh bạch AI (AI Use Disclosure Protocol) nhằm tái thiết lập "ma sát nhận thức". Khung này được cấu trúc dựa trên 3 nguyên tắc cốt lõi:

1. Phân định phạm vi sử dụng (Allowed/Not allowed): Sinh viên được phép dùng AI để lên ý tưởng, phản biện lập luận (critical reviewer) hoặc hỗ trợ ngôn ngữ cục bộ. Tuy nhiên, nghiêm cấm việc dùng AI để ngụy tạo bằng chứng/trích dẫn hoặc ủy thác AI soạn thảo thay các phần trọng tâm (ghostwriting) mà người học không thể tự giải thích lại lập luận được đưa ra.

2. Quy tắc kiểm chứng kép (Double-check rule): Mọi thông tin sự kiện, số liệu do AI gợi ý phải được đối chiếu với ít nhất 02 nguồn học thuật tin cậy. AI chỉ được xem là công cụ gợi ý tìm nguồn, không phải là bằng chứng học thuật.

3. Yêu cầu minh chứng quy trình (Process Evidence): Sinh viên bắt buộc nộp kèm một Bản khai báo (Disclosure statement) dài 80-150 từ và tối thiểu 03-06 ảnh chụp màn hình lịch sử câu lệnh (prompt logs) minh chứng cho quá trình tương tác thực tế.

Quy trình thực nghiệm diễn ra trong 4 tuần:

- Tuần 1 - Pre-test: Viết bài luận tại lớp (40 phút, tối thiểu 250 từ) không có sự hỗ trợ của công nghệ để thiết lập điểm chuẩn năng lực nội tại.

- Tuần 2 - Thiết lập khung: Giảng viên giới thiệu Protocol và phân định ranh giới đạo đức.

- Tuần 3 - Huấn luyện năng lực AI: Hướng dẫn kỹ thuật lệnh lặp (Iterative prompting) và kiểm chứng chéo để chống ảo giác AI.

- Tuần 4 - Post-test: Viết bài luận thứ hai tại nhà (tối thiểu 250 từ) có sự hỗ trợ của GenAI và nộp kèm Minh chứng sử dụng AI theo đúng Protocol.

3.3. Công cụ thu thập và phương pháp phân tích dữ liệu

Dữ liệu định tính được thu thập từ Minh chứng sử dụng AI, gồm bản khai báo và ảnh chụp nhật ký câu lệnh, sau đó được phân tích nội dung và mã hóa mở để xác định các chiến lược tương tác với LLMs như ủy thác toàn phần, lên ý tưởng, đồng kiến tạo, kiểm chứng thông tin và chỉnh sửa giọng văn cá nhân. Dữ liệu định lượng gồm bài viết pre-test và post-test, được chấm theo barem tinh chỉnh từ IELTS Writing Task 2 trên thang điểm 9.0. Điểm Hình thức được tính từ Mạch lạc, Từ vựng và Ngữ pháp; Điểm Nội dung gồm Minh chứng xác thực (C2) và Tư duy phản biện (C3). Dữ liệu điểm số được xử lý bằng SPSS 26.0 với phép kiểm định T-test bắt cặp. Việc đối chiếu dữ liệu định tính và định lượng giúp tăng độ tin cậy trong diễn giải kết quả.

IV. Kết quả nghiên cứu và thảo luận

4.1. Sự phân hóa hành vi tương tác với AI: Chiến lược "Lắp ghép" đa tầng (RQ1)

Dữ liệu định tính từ 30 bộ Minh chứng sử dụng AI cho thấy Khung hướng dẫn Minh bạch tạo ra sự phân hóa rõ trong

hành vi tương tác với LLMs. Chỉ 16,7% sinh viên (n=5) có xu hướng lệ thuộc thuật toán hoặc ủy thác phần lớn quá trình viết cho AI. Ngược lại, 83,3% sinh viên (n=25) áp dụng chiến lược Lắp ghép (Bricolage), tức chủ động chọn lọc, điều chỉnh và tích hợp gợi ý của AI vào khung ý tưởng cá nhân. Chiến lược này gồm ba bước chính: (1) khởi tạo và xây dựng dàn ý, trong đó sinh viên kết hợp ý tưởng ban đầu với gợi ý của AI; (2) đồng kiến tạo, khi sinh viên dùng AI để cải thiện độ trôi chảy và cấu trúc học thuật của bản nháp; và (3) gọt đẽo và hậu kiểm (Chiseling), khi sinh viên loại bỏ cách diễn đạt không phù hợp, kiểm chứng thông tin và diễn đạt lại bằng giọng văn cá nhân.

4.2. Tác động của Khung Minh bạch đến chất lượng bài luận học thuật (RQ2)

Để định lượng hóa tác động của các hành vi tương tác trên đối với chất lượng bài viết, nghiên cứu tiến hành kiểm định Paired-samples T-test cho 30 sinh viên. Kết quả phân tích (Bảng 1) cho thấy sự gia tăng có ý nghĩa thống kê ở cả ba tiêu chí đánh giá.

Bảng 1. Đối chiếu điểm trung bình các tiêu chí cốt lõi (N=30)

Tiêu chí đánh giá	Pre-test (Mean)	Post-test (Mean)	Mức độ chênh lệch	Mức ý nghĩa (p-value)
Điểm Hình thức (C1, C5, C7)	6,28	6,63	+ 0,35	p < 0,001
Điểm Minh chứng (C2)	5,35	6,33	+ 0,98	p < 0,001
Điểm Phản biện (C3)	5,95	6,35	+ 0,40	p < 0,001

Ghi chú: C1, C5, C7 thuộc nhóm Hình thức; C2 = minh chứng xác thực; C3 = tư duy phản biện. Giá trị p được tính bằng kiểm định T-test bắt cặp. (Nguồn: Dữ liệu khảo sát của tác giả)

Kết quả định lượng ở Bảng 1 tương thích với phát hiện định tính tại mục 4.1. Mức tăng khiêm tốn ở Điểm Hình thức (+0,35, p<0,001) cho thấy LLMs chủ yếu hỗ trợ cải thiện độ trôi chảy và tính chuẩn mực của văn bản, nhưng không làm thay đổi quá mạnh bề mặt ngôn ngữ. Điểm Phản biện (C3) tăng ổn định (+0,40, p<0,001), cho thấy sinh viên đã sử dụng AI như công cụ gợi mở và kiểm tra lập luận thay vì thay thế hoàn toàn tư duy cá nhân. Đáng chú ý, mức tăng mạnh nhất thuộc về tiêu chí Minh chứng (C2: +0,98, p<0,001). Dưới quy tắc kiểm chứng kép, sinh viên có điều kiện chọn lọc, đối chiếu và chuyển hóa gợi ý từ LLMs thành minh chứng phù hợp hơn, qua đó góp phần hạn chế nguy cơ sử dụng thông tin sai lệch hoặc ảo giác của công cụ.

4.3. Thảo luận

4.3.1. Sự hình thành năng lực AI phê phán thông qua chiến lược "Lắp ghép" (Thảo luận cho RQ1)

Kết quả nghiên cứu cho thấy dưới tác động của Khung hướng dẫn Minh bạch, phần lớn sinh viên không ủy thác hoàn toàn quá trình viết cho LLMs mà lựa chọn chiến lược Lắp ghép (Bricolage). Điều này phản ánh vai trò của ma sát nhận thức trong việc buộc người học chậm lại, kiểm chứng thông tin và duy trì quyền kiểm soát văn bản. Đặc biệt, bước gọt đẽo và hậu kiểm cho thấy sinh viên không chỉ tiếp nhận gợi ý từ AI mà còn chỉnh sửa, loại bỏ hoặc diễn đạt lại các phần chưa phù hợp với lập luận và giọng văn cá nhân. Đây là biểu hiện của Năng lực AI phê phán, khi người học sử dụng

công nghệ như công cụ hỗ trợ thay vì thay thế tư duy. Phát hiện này phù hợp với cách hiểu về Bricolage như một quá trình tái kết hợp linh hoạt các nguồn lực sẵn có, đồng thời củng cố yêu cầu chuyển từ đánh giá sản phẩm sang đánh giá quá trình trong bối cảnh GenAI.

4.3.2. Hiệu quả của Khung Minh bạch trong việc nâng cấp chất lượng tư duy (Thảo luận cho RQ2)

Mức tăng có ý nghĩa thống kê ở tiêu chí Minh chứng (C2: +0,98, p<0,001) và Phản biện (C3: +0,40, p<0,001) cho thấy Khung hướng dẫn Minh bạch AI đã hỗ trợ sinh viên cải thiện các thành tố nội dung của bài viết. Dưới quy tắc kiểm chứng kép, người học buộc phải đối chiếu gợi ý của LLMs với các nguồn học thuật, từ đó giảm nguy cơ sử dụng thông tin sai lệch và tăng chất lượng minh chứng. Đồng thời, việc yêu cầu nộp nhật ký câu lệnh giúp quá trình tương tác với AI trở nên có thể quan sát và đánh giá, phù hợp với định hướng đánh giá quá trình của Cotton và cộng sự (2024). Từ góc nhìn ZPD, LLMs có thể được xem như một hình thức giàn giáo nhận thức, nhưng hiệu quả của nó phụ thuộc vào mức độ người học kiểm chứng, điều chỉnh và làm chủ quyết định viết.

V. Kết luận và kiến nghị

5.1. Tổng kết các phát hiện chính

Nghiên cứu khẳng định tác động của Khung hướng dẫn Minh bạch AI qua hai phát hiện chính. Thứ nhất, hành vi người học chuyển từ tiếp nhận thụ động sang tương tác có kiểm soát với LLMs. Việc 83,3% sinh viên áp dụng chiến lược Lắp ghép (Bricolage) qua ba bước khởi tạo, đồng kiến tạo, gọt đẽo và hậu kiểm cho thấy ma sát nhận thức đã góp phần hạn chế việc trút bỏ gánh nặng tư duy. Thứ hai, chất lượng bài viết cải thiện rõ nhất ở tiêu chí Minh chứng (C2: +0,98) và Phản biện (C3: +0,40), cho thấy GenAI có thể hỗ trợ quá trình viết nếu được đặt trong cơ chế kiểm chứng và giải trình minh bạch.

5.2. Ý nghĩa và đóng góp của nghiên cứu

Nghiên cứu cung cấp bằng chứng cho thấy GenAI không nhất thiết đe dọa liêm chính học thuật nếu có thiết kế sư phạm phù hợp. Về mặt lý luận, nghiên cứu góp phần làm rõ vai trò của Khung hướng dẫn Minh bạch AI như một cơ chế tạo ma sát nhận thức, giúp chuyển hóa các khung lý thuyết như TAM, ZPD và Năng lực AI phê phán vào thực tiễn dạy học viết học thuật. Về mặt thực tiễn, khung này hỗ trợ giảng viên chuyển trọng tâm từ kiểm soát sản phẩm cuối cùng sang quan sát và đánh giá quá trình sử dụng công nghệ của người học. Đồng thời, nghiên cứu gợi ý định hướng sinh viên dùng LLMs để lên ý tưởng, kiểm chứng thông tin và phát triển lập luận, thay vì thay thế tư duy và viết độc lập.

5.3. Hạn chế và hướng nghiên cứu tiếp theo

Mặc dù đạt kết quả khả quan, nghiên cứu vẫn có một số hạn chế. Trước hết, thiết kế tiền thực nghiệm kiểm tra trước - sau một nhóm chưa loại trừ hoàn toàn các yếu tố ảnh hưởng đến tính giá trị nội tại, như tác động lịch sử, sự trưởng thành của người học hoặc hiệu ứng luyện tập. Việc đối chiếu dữ liệu điểm số với nhật ký sử dụng AI giúp củng cố diễn giải kết quả, nhưng chưa thay thế được nhóm đối chứng. Bên cạnh đó, mẫu nghiên cứu gồm 30 sinh viên tại một cơ sở đào tạo duy nhất

nên khả năng khái quát hóa còn hạn chế. Các nghiên cứu tiếp theo cần mở rộng mẫu, bổ sung nhóm đối chứng và nghiên cứu dọc để đánh giá độ bền vững của năng lực viết và tư duy phản biện khi sinh viên giảm hoặc ngừng dùng AI.

Tài liệu tham khảo

Bhusal, P. C. (2025). Fostering critical AI literacy through a decolonial use of ChatGPT in ESL/EFL classrooms. *Advances in Mobile Learning Educational Research, 5*(2), 1472-1487. https://doi.org/10.25082/AMLER.2025.02.005

Cotton, D. R. E., Cotton, P. A., & Shipway, J. R. (2024). Chatting and cheating: Ensuring academic integrity in the era of ChatGPT. *Innovations in Education and Teaching International, 61*(2), 228-239. https://doi.org/10.1080/14703297.2023.2190148.

Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly, 13*(3), 319-340. https://doi.org/10.2307/249008.

Farrokhnia, M., Banihashem, S. K., Noroozi, O., & Wals, A. (2024). A SWOT analysis of ChatGPT: Implications for educational practice and research. *Innovations in Education and Teaching International, 61*(3), 460-474. https://doi.org/10.1080/14703297.2023.2195846.

Ji, Z., Lee, N., Frieske, R., Yu, T., Su, D., Xu, Y., Ishii, E., Bang, Y. J., Chen, D., Dai, W., Chan, H. S., Madotto, A., & Fung, P. (2023). Survey of hallucination in natural language generation. *ACM Computing Surveys, 55*(12), Article 248, 1-38. https://doi.org/10.1145/3571730.

Jurafsky, D., & Martin, J. H. (2024). *Speech and language processing* (3rd ed. draft). Stanford University.

Lévi-Strauss, C. (1966). *The savage mind*. University of Chicago Press.

Moorhouse, B. L., & Wong, K. M. (2025). *Generative artificial intelligence and language teaching*. Cambridge University Press.

van Niekerk, J., Delport, P. M. J., & Sutherland, I. (2025). Addressing the use of generative AI in academic writing. *Computers and Education: Artificial Intelligence, 8*, Article 100342. https://doi.org/10.1016/j.caeai.2024.100342.

Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Harvard University Press.

THE IMPACT OF THE AI USE DISCLOSURE PROTOCOL ON ACADEMIC WRITING BEHAVIOR AND QUALITY OF UNIVERSITY STUDENTS

Vu Dieu Thuy[1]

Abstract: *This study evaluates the impact of the AI Use Disclosure Protocol on the academic essay-writing behavior and quality of second-year English majors at Hanoi Open University. Designed as a tool for generating cognitive friction to safeguard academic integrity, the intervention was examined using a mixed-methods approach, combining a one-group pretest-posttest design with 30 students and AI prompt-log analysis. The findings show that 83.3% of the learners adopted the Bricolage strategy, using AI as cognitive scaffolding rather than fully delegating the writing task. Writing scores improved significantly ($p < 0.001$): linguistic form increased slightly (+0.35), while evidence use (+0.98) and critical thinking (+0.40) showed greater improvement. The study, therefore, proposes a shift toward process-based assessment to foster sustainable Critical AI Literacy among learners.*

Keywords: *academic writing, AI disclosure protocol, critical AI literacy, generative AI, process-based assessment*

[1] Hanoi Open University, Hanoi, Vietnam