

Sự “phẳng hóa” chiều sâu nội tâm và hiện tượng lệch chuẩn văn hóa trong tái hiện nhân vật văn học Việt Nam qua trí tuệ nhân tạo

The “Flattening” of Psychological Depth and Cultural Deviation in the Recreation of Vietnamese Literary Characters through AI

Hoàng Thị Hương^{a*}, Nguyễn Thị Thanh Thu^a
Hoang Thi Huong^{a*}, Nguyen Thi Thanh Thu^a

^aTrường Ngôn ngữ - Xã hội Nhân văn, Đại học Duy Tân, Đà Nẵng, Việt Nam

^aSchool of Languages, Humanities and Social Science, Duy Tan University, Da Nang, 550000, Vietnam

(Ngày nhận bài: 25/5/2026, ngày phản biện xong: 03/6/2026, ngày chấp nhận đăng: 18/6/2026)

Tóm tắt

Nghiên cứu này tập trung phân tích sự biến đổi của hình tượng nhân vật văn học Việt Nam hiện đại khi được tái cấu trúc từ mã ngôn từ sang mã thị giác thông qua trí tuệ nhân tạo (AI) tạo sinh. Kết hợp phương pháp thực nghiệm (kỹ nghệ câu lệnh trên ChatGPT, Gemini, Grok) và khảo sát định lượng 150 mẫu, kết quả chỉ ra AI gặp giới hạn lớn trong việc giải mã văn hóa, dẫn đến hiện tượng “phẳng hoá” chiều sâu tâm lý và “ảo giác văn hóa” (Tây hóa bối cảnh). Từ đó, nghiên cứu đề xuất định hướng sử dụng AI như một công cụ phản biện trong giảng dạy văn học thay vì công cụ minh họa thuần túy.

Từ khoá: trí tuệ nhân tạo, phẳng hóa, ảo giác văn hóa, văn học Việt Nam hiện đại, thiên kiến thuật toán

Abstract

This study focuses on analyzing the transformation of modern Vietnamese literary character imagery when restructured from linguistic to visual code through Artificial Intelligence (AI) generation. Combining experimental methods (prompt engineering on ChatGPT, Gemini, Grok) and quantitative surveys of 150 samples, the study indicates that AI faces significant limitations in decoding culture, leading to the “flattening” of psychological depth and “cultural illusions” (Westernization of context). Therefore, the study proposes using AI as a critical thinking tool in literature teaching rather than a purely illustrative tool.

Keywords: artificial intelligence, flattening, cultural hallucination, modern Vietnamese literature, algorithmic bias

1. Đặt vấn đề

Sự bùng nổ của các mô hình ngôn ngữ lớn (LLMs) và AI tạo sinh trong những năm gần đây đã mở ra khả năng để máy tính tham gia ngày càng sâu vào quá trình sáng tạo và tái tạo hình tượng nghệ thuật. Không chỉ dừng ở việc xử lý văn bản, AI hiện nay còn có thể tạo ra hình ảnh,

video và hoạt hình ngắn, các công cụ trí tuệ nhân tạo tổng hợp như Synthesia có thể chuyển văn bản thành video có nhân vật ảo thuyết minh, mô phỏng lại một trích đoạn kịch hoặc tình huống truyện [6]. Dựa trên dữ liệu văn học, từ đó tác động trực tiếp đến cách con người tiếp cận và cảm thụ tác phẩm. Đối với văn học Việt Nam

*Tác giả liên hệ: Hoàng Thị Hương
Email: hoangthihuong@duytan.edu.vn

hiện đại, đặc biệt là thể loại truyện ngắn, nhân vật không chỉ mang câu chuyện cá nhân mà còn gắn chặt với văn hóa, bối cảnh lịch sử và tâm lý xã hội đặc thù của dân tộc. Những hình tượng như Chí Phèo (trong tác phẩm *Chí Phèo* của Nam Cao), Tràng hay người vợ (trong tác phẩm *Vợ nhặt* của Kim Lân) đều được xây dựng trong mối quan hệ sâu sắc với không gian làng quê, đời sống nghèo đói và các biến động lịch sử của xã hội Việt Nam.

Việc AI tái hiện các hình tượng này không đơn thuần là vấn đề kỹ thuật tạo hình mà còn liên quan đến khả năng hiểu và chuyển tải chiều sâu văn hóa - tư tưởng của tác phẩm. Mặc dù ứng dụng AI trong giáo dục đang trở thành xu thế nổi bật, phần lớn các nghiên cứu hiện nay chủ yếu tập trung vào vai trò hỗ trợ của AI như tạo học liệu, minh họa nội dung hay hỗ trợ giáo viên trong công tác giảng dạy và đánh giá, kiểm tra đạo văn, phân tích tính mạch lạc trong lập luận - những yếu tố mà trước đây chỉ có giáo viên con người mới thực hiện được [1; 50]. Tuy nhiên, những nghiên cứu này thường xem AI như một công cụ trung tính, chưa chú ý đầy đủ đến nguy cơ sai lệch trong quá trình AI diễn giải và tái tạo văn học. Đặc biệt, vẫn còn thiếu các nghiên cứu đi sâu phân tích mức độ biến đổi, chuyển nghĩa và thiên kiến văn hóa của hình tượng văn học khi đi qua lăng kính thuật toán.

Trong nhiều trường hợp, AI có thể đơn giản hóa nhân vật, điện ảnh hóa bối cảnh hoặc áp đặt các mô hình thẩm mỹ phổ quát, khiến giá trị hiện thực và chiều sâu nhân văn của tác phẩm bị suy giảm. Xuất phát từ khoảng trống đó, nghiên cứu này tập trung khảo sát một số truyện ngắn Việt Nam hiện đại nhằm đánh giá những giới hạn của AI trong việc tái hiện hình tượng văn học. Đồng thời, nghiên cứu cũng hướng đến việc phân tích các biểu hiện thiên kiến dữ liệu, hiện tượng lệch chuẩn văn hóa và sự đứt gãy bối cảnh lịch sử khi AI tham gia vào quá trình minh họa tác phẩm. Qua đó, nghiên cứu không chỉ góp phần nhận

diện những thách thức của AI trong lĩnh vực giáo dục văn học mà còn đề xuất cách tiếp cận AI như một đối tượng phản biện, giúp người học phát triển tư duy độc lập, năng lực phản biện và khả năng đánh giá công nghệ trong bối cảnh chuyển đổi số hiện nay.

2. Phương pháp nghiên cứu

2.1. Phương pháp phân tích - tổng hợp

Bóc tách thi pháp nhân vật (ngoại hình, tâm lý, bối cảnh) từ nguyên tác thành các bộ từ khóa đầu vào cho thuật toán.

2.2. Phương pháp so sánh - đối chiếu: Đối chiếu song song giữa nhân vật trong văn bản của tác giả và hình ảnh sản phẩm từ AI để làm nổi bật sự biến đổi, sai lệch

2.3. Phương pháp thực nghiệm - trải nghiệm có sáng tạo

Thực hiện quy trình kỹ nghệ câu lệnh (Prompt Engineering) trên ba nền tảng ChatGPT, Gemini và Grok và kết hợp thang đo Rubric để đo lường hình ảnh đầu ra một cách khoa học. Quá trình sử dụng này được thực hiện dựa trên công thức kỹ nghệ câu lệnh để thực nghiệm và quan sát khả năng giải mã văn học của máy tính, từ đó thu thập các kết quả (văn bản/hình ảnh) phục vụ cho việc phân tích.

2.4. Phương pháp định lượng (khảo sát)

Thực hiện khảo sát bằng bảng hỏi (Google Forms) để thu thập ý kiến của người học về các công cụ AI. Khảo sát từ ngày 21/04/2026 đến 20/05/2026 với 150 người tham gia (chủ yếu là sinh viên). Ngoài ra, sử dụng các công cụ AI gồm ChatGPT-5.5, Gemini-Nano Banana 2 và Grok 4.2. Sau khi thu thập, dữ liệu được mã hóa; các phiếu khảo sát không hợp lệ được loại bỏ. Các biến định lượng được mã hóa theo thang đo Likert 5 mức độ (từ 1 = Hoàn toàn không đồng ý đến 5 = Hoàn toàn đồng ý). Dữ liệu sau đó được tổng hợp và xử lý bằng các phương pháp thống kê mô tả như tần suất, tỷ lệ phần trăm, giá trị trung bình và độ lệch chuẩn để phân tích xu

hướng đánh giá của người tham gia đối với các công cụ AI.

3. Kết quả nghiên cứu và thảo luận

3.1. Hiện tượng “phẳng hóa” chiều sâu tâm lý và sự suy giảm giá trị nhân văn

Bộ từ khóa sử dụng trong thực nghiệm không được lựa chọn ngẫu nhiên mà được xây dựng từ quá trình phân tích văn bản gốc của các truyện ngắn trong phạm vi nghiên cứu. Thống kê những chi tiết miêu tả nhân vật như ngoại hình, hành động, ngôn ngữ, tâm lý, tính cách, môi trường sống và phong cách nghệ thuật của nhà văn. Đây là các yếu tố cốt lõi của thi pháp nhân vật, góp phần tạo nên bản sắc và khả năng nhận diện hình tượng văn học. Hệ thống từ khóa được chuyển hóa từ ngôn ngữ văn học sang ngôn ngữ lệnh (prompt language) để AI có thể xử lý và trực quan hóa thành hình ảnh. Việc đánh giá tính hợp lý của từ khóa dựa trên ba tiêu chí: phản ánh đúng đặc điểm nhân vật trong tác phẩm, có tính điển hình và khả năng nhận diện cao, đồng thời đủ cụ thể để AI tạo thành hình ảnh thị giác. Độ chính xác của hình ảnh AI được đối chiếu trực

tiếp với văn bản gốc thay vì dựa vào tranh minh họa hay cảm nhận chủ quan. Quy trình thử nghiệm prompt được thực hiện theo hướng hệ thống hóa thay vì điều chỉnh cảm tính. Prompt được xây dựng theo cấu trúc gồm: chủ thể nhân vật, ngoại hình, hành động, bối cảnh, cảm xúc, phong cách, thiết lập kỹ thuật và chất lượng đầu ra. Với cấu trúc này, người dùng chỉ cần thay đổi các yếu tố nội dung cụ thể như nhân vật hay bối cảnh, trong khi vẫn giữ nguyên khung câu lệnh, giúp hạn chế sai lệch và tối ưu hóa hiệu quả tạo ảnh của AI.

Ví dụ, prompt Chí Phèo: Hãy tạo giúp tôi hình ảnh nhân vật Chí Phèo trong tác phẩm “Chí Phèo” của Nam Cao, với đầu trọc, khuôn mặt đầy sẹo, ánh mắt dữ tợn và đáng sợ lưu manh, mặc quần áo nông dân cũ bẩn, đang chửi bới, uống rượu và cầm dao, trong bối cảnh làng quê Việt Nam nghèo trước Cách mạng với không khí u ám, tối tăm, thể hiện sự đau khổ, tuyệt vọng và khao khát lương thiện, phong cách điện ảnh hiện thực, ánh sáng tương phản mạnh, góc máy thấp, siêu chi tiết, nét rõ.

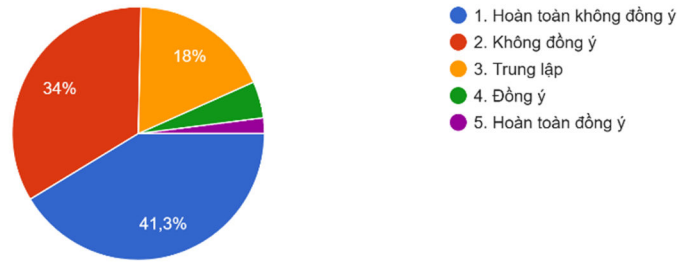
Kết quả như Hình 1:



Hình 1. Hình ảnh nhân vật Chí Phèo do AI tạo ra

Thực nghiệm cho thấy AI chỉ bắt được các biểu hiện ngoại hình mà thất bại trong việc truyền tải sự giằng xé nội tâm đa tầng của nhân vật. Kết quả khảo sát 150 mẫu cũng cố nhận định này. Với câu hỏi: “Bạn nghĩ rằng AI có thể tạo ra những mâu thuẫn nội tâm sâu sắc và phức tạp cho nhân vật một cách chân thực?”, có tới 41,3%

người tham gia hoàn toàn không đồng ý, 34% không đồng ý (tổng cộng 75,3%) và 18% giữ thái độ trung lập. Điều này cho thấy người tham gia khảo sát đánh giá khá thấp khả năng của AI trong việc tái hiện những xung đột tinh thần nhiều tầng bậc vốn là yếu tố quan trọng tạo nên chiều sâu của nhân vật văn học.



Hình 2. Kết quả khảo sát về khả năng tái hiện mâu thuẫn nội tâm nhân vật của AI

(Nguồn: Tác giả khảo sát và tổng hợp)

Cụ thể trạng thái “say rượu” của Chí Phèo thường được AI thể hiện khá trực diện qua ánh mắt lơ đãng, khuôn mặt dữ dằn, nhiều vết sẹo, tư thế buông thả và đang cầm chai rượu. Người xem dễ dàng nhận ra đây là hình ảnh của một kẻ nghiện rượu sống trong men say. Nhưng cái “say” trong nguyên tác Nam Cao không chỉ là trạng thái sinh lý, mà còn là biểu hiện của bi kịch bị tha hóa, nỗi đau bị tước quyền làm người và tiếng kêu tuyệt vọng của một con người cùng đường. Khi chuyển sang hình ảnh AI, chiều sâu ấy thường bị giản lược thành biểu cảm mạnh về thị giác. Quan trọng hơn, AI không thể biểu đạt được cái đau đớn cốt lõi của Chí Phèo là bi kịch bị tước đoạt quyền làm người. Sự thất bại của AI không chỉ nằm ở việc không thể hiện đủ cảm xúc, mà sâu hơn là không thể biểu đạt những phạm trù trừu tượng như giai cấp, tha hóa và quyền làm người. Là những yếu tố không tồn tại dưới dạng hình ảnh trực tiếp, mà chỉ có thể được nhận thức thông qua quá trình đọc, suy tư, và thấu cảm kèm nhận thức sâu sắc về lịch sử và văn hoá. Đây chính là nguyên nhân dẫn đến hiện tượng “phẳng hóa” chiều sâu nội tâm nhân vật, tức là nhân vật khi được AI tái hiện thường chỉ còn lại lớp vỏ bên ngoài, trong khi những tầng ý

nghĩa tâm lý và tư tưởng vốn làm nên giá trị của hình tượng văn học lại bị giản lược đáng kể. Qua đó càng khẳng định kỹ nghệ câu lệnh dù tốt đến đâu cũng khó giúp AI chạm tới những chiều sâu tâm lý.

Đối với prompt Vợ nhặt: Hãy tạo hình ảnh nhân vật Tràng trong tác phẩm “Vợ nhặt” của Kim Lân. Nhân vật có thân hình thô kệch, khuôn mặt chất phác, dáng đi nặng nề và mặc quần áo nông dân cũ. Tràng đang kéo xe và dẫn vợ về nhà trong bối cảnh xóm ngụ cư thời nạn đói năm 1945 với không khí vắng lặng, tiêu điều. Hình ảnh cần thể hiện tính cách hiền lành, nhân hậu và có trách nhiệm. Phong cách điện ảnh hiện thực, ánh sáng tự nhiên dịu, góc máy ngang tầm mắt, hình ảnh rõ nét. Đồng thời nhân vật nữ tên Thị, với dáng vẻ tiêu tụy, gầy gò, khuôn mặt hốc hác và quần áo rách rưới, đang ăn uống vội vã và đứng cạnh Tràng, trong bối cảnh nạn đói năm 1945 với không gian nghèo nàn, xơ xác, thể hiện sự chao chát ban đầu nhưng ẩn chứa nét hiền hậu, phong cách điện ảnh hiện thực, ánh sáng tự nhiên yếu, góc máy ngang tầm mắt, siêu chi tiết, nét rõ.

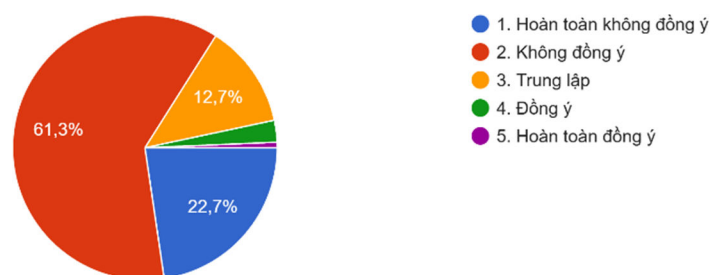
Kết quả như Hình 3:



Hình 3. Hình ảnh nhân vật Tráng và Thị do AI tạo ra (Nguồn: ChatGPT, Gemini, Grok)

AI cũng có khả năng nhận diện những trạng thái như “đói” và “ngơ ngác”. Nhân vật người vợ (Thị) thường được khắc họa với thân hình gầy gò, khuôn mặt hốc hác, ánh mắt mệt mỏi, từ đó gợi liên tưởng đến nạn đói. Nhưng cái đói trong nguyên tác của Kim Lân không chỉ là thiếu ăn, mà là sự ám ảnh kéo dài khiến con người biến đổi cả nhân phẩm, hành vi và tâm lý. Nhưng hình ảnh AI chủ yếu biểu đạt cái đói qua ngoại hình nên chưa truyền tải hết sức nặng tinh thần của

hoàn cảnh lịch sử. Với Tràng, AI thường tái hiện nét mặt bối rối, ngạc nhiên trước tình huống của cô nhặt vợ. Trong nguyên tác, trạng thái “ngơ ngác” của Tràng là sự hòa trộn giữa bất ngờ, vui mừng, lo lắng và hy vọng mong manh về mái ấm gia đình. Đây là diễn biến tâm lý nhiều tầng bậc, thay đổi liên tục theo hoàn cảnh. Trong khi đó, AI thường chỉ chọn một khoảnh khắc cảm xúc rõ nét nhất để thể hiện, khiến trạng thái nội tâm bị đơn giản hóa.



Hình 4. Kết quả khảo sát về mức độ tin tưởng của người tham gia đối với khả năng AI tạo ra chiều sâu tâm lý nhân vật

(Nguồn: Tác giả khảo sát và tổng hợp)

Theo kết quả khảo sát có tới 22,7% người tham gia hoàn toàn không đồng ý và 61,3% không đồng ý, chiếm tổng cộng 84%. Số liệu này phản ánh đa số người khảo sát chưa tin tưởng vào khả năng của AI trong việc giữ được chiều sâu tâm lý cốt lõi của nhân vật và phần lớn người dùng lo ngại sự “phẳng hóa” này sẽ làm mờ đi ý đồ nghệ thuật ban đầu của tác giả.

Từ những kết quả thực nghiệm cho thấy, hiện tượng “phẳng hóa” nội tâm còn kéo theo sự suy giảm giá trị nhân văn của tác phẩm. Những vấn đề lớn mà văn học Việt Nam đặt ra như quyền làm người, phẩm giá con người, khát vọng hạnh phúc hay bi kịch của người lao động nghèo đều khó được thể hiện đầy đủ bằng hình ảnh do AI

tạo ra. Quan niệm của tác giả cho rằng AI có thể tái hiện hình dáng nhân vật nhưng chưa đủ khả năng truyền tải trọn vẹn những giá trị tư tưởng và thông điệp nhân đạo mà nhà văn muốn gửi gắm. Vì vậy, khi hình tượng văn học được chuyển từ ngôn từ sang hình ảnh AI, các giá trị nhân văn cốt lõi của tác phẩm có nguy cơ bị suy giảm.

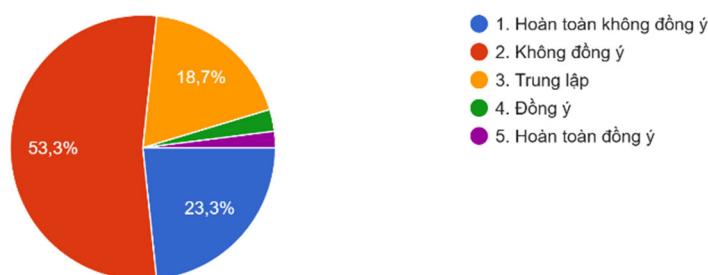
Để làm rõ hơn những khác biệt giữa hình tượng nhân vật trong nguyên tác văn học và hình ảnh do AI tạo ra, tác giả tiến hành đối chiếu theo các tiêu chí thuộc thi pháp nhân vật như ngoại hình, hành động, tâm lý, ý nghĩa xã hội, bối cảnh văn hóa - lịch sử và giá trị nhân văn.

Bảng 1. Đối chiếu hình tượng nhân vật trong nguyên tác và hình ảnh do AI tạo ra

(Nguồn: Tổng hợp từ kết quả thực nghiệm của tác giả.)

Tiêu chí	Trong tác phẩm văn học	Trong hình ảnh AI
Ngoại hình	Được mô tả cụ thể, mang dấu ấn số phận nhân vật	Tái hiện khá chính xác các đặc điểm nhận diện
Tâm lý	Nhiều tầng bậc cảm xúc, diễn biến phức tạp	Chủ yếu thể hiện một trạng thái cảm xúc nổi bật
Ý nghĩa xã hội	Gắn với hoàn cảnh lịch sử, thân phận và bi kịch con người	Khó thể hiện các tầng nghĩa xã hội và nhân văn
Bối cảnh	Mang đậm đặc trưng văn hóa, lịch sử Việt Nam	Có xu hướng điện ảnh hóa, thẩm mỹ hóa
Giá trị nhân văn	Thể hiện tư tưởng và thông điệp của tác giả	Thường bị giản lược thành các dấu hiệu trực quan

Bảng 1 cho thấy AI có khả năng tái hiện khá hiệu quả các đặc điểm ngoại hình và hành động mang tính trực quan của nhân vật, nhưng gặp nhiều hạn chế khi thể hiện chiều sâu tâm lý, ý nghĩa xã hội và giá trị nhân văn. Mức độ tương đồng giữa hình ảnh AI và nguyên tác chủ yếu dừng lại ở lớp biểu hiện bề mặt, trong khi các tầng nghĩa tư tưởng gắn với bối cảnh lịch sử, văn hóa và thân phận con người thường bị giản lược hoặc biến dạng. Điều này góp phần lý giải hiện tượng “phẳng hóa” nhân vật và “đứt gãy” bối cảnh văn hóa khi ứng dụng AI vào minh họa tác phẩm văn học.



Hình 5. Kết quả khảo sát về mức độ đánh giá khả năng tái hiện bối cảnh của AI so với nguyên tác

Với kết quả khảo sát trên, có tới 76,6% người tham gia lựa chọn mức không đồng ý/hoàn toàn không đồng ý, cho thấy phần lớn người dùng chưa tin tưởng vào khả năng của AI trong việc đặt nhân vật vào đúng hoàn cảnh lịch sử, xã hội và logic hành động của tác phẩm.

3.2. Hiện tượng lệch chuẩn văn hoá và sự đứt gãy bối cảnh lịch sử

Nếu hiện tượng “phẳng hóa” làm mất đi chiều sâu tâm lý của nhân vật thì một hạn chế khác của AI là sự sai lệch trong việc tái hiện bối cảnh văn hóa và lịch sử. Nhân vật văn học luôn gắn liền với môi trường văn hóa và hoàn cảnh lịch sử đã sản sinh ra họ, việc tái hiện nhân vật chỉ thực sự chính xác khi đồng thời tái hiện đúng không gian văn hóa và bối cảnh xã hội của tác phẩm.

(Nguồn: Tác giả khảo sát và tổng hợp)

Các không gian nghệ thuật mang đậm dấu ấn nông thôn Việt Nam như làng Vũ Đại trong *Chí Phèo*, xóm ngụ cư trong *Vợ nhặt* thường bị AI “thẩm mỹ hóa” và “điện ảnh hóa” theo những khuôn mẫu thị giác phổ biến trên thế giới. Thay vì tái hiện đúng không gian làng quê Việt Nam

thì AI lại tạo ra các khung cảnh mang màu sắc cổ điển phương Tây, với kiến trúc mơ hồ, ánh sáng điện ảnh và bảng màu cũ kỹ, thiên về đen trắng. Những hình ảnh này dễ gợi liên tưởng đến phim nghệ thuật châu Âu hơn là đời sống xã hội Việt Nam những năm đầu thế kỷ XX trong nguyên tác văn học. Sự lệch chuẩn ấy không chỉ làm thay đổi diện mạo không gian mà còn kéo theo sự biến đổi trong cách tiếp nhận nhân vật và nội dung tác phẩm. Trong văn học hiện thực Việt Nam, bối cảnh xã hội giữ vai trò đặc biệt quan trọng vì nó gắn chặt với số phận nhân vật. Chí Phèo không chỉ là một cá nhân tha hóa mà còn là sản phẩm của làng Vũ Đại. Tràng trong *Vợ nhặt* không chỉ là một người đàn ông nghèo mà còn là hình ảnh đại diện cho nạn đói năm 1945 với khung cảnh xác xơ, đói khát bao trùm. Khi AI làm sai lệch bối cảnh lịch sử và văn hóa, hành động, tâm lý và biểu cảm của nhân vật cũng trở nên thiếu thuyết phục, mất đi nền tảng xã hội vốn tạo nên chiều sâu tư tưởng của tác phẩm. Người học vì thế có thể hiểu sai và tiếp nhận văn học theo hướng đơn giản hóa, thiên về hình ảnh bề mặt hơn là bản chất hiện thực mà tác giả muốn phản ánh.

Hiện tượng này trước hết có thể được lý giải từ góc độ khách quan là thiên kiến dữ liệu (data bias) khi một thuật toán đưa ra các kết quả thiên lệch một cách hệ thống, thường do những giả định sai lầm trong quá trình xây dựng [5]. Các mô hình AI hiện nay được huấn luyện trên kho dữ liệu hình ảnh toàn cầu, trong đó dữ liệu về văn hóa phương Tây chiếm tỉ lệ áp đảo so với hình ảnh về văn hóa Việt Nam truyền thống. Tuy nhiên, sự lệch chuẩn này không chỉ bắt nguồn từ thuật toán mà còn chịu sự chi phối mạnh mẽ bởi yếu tố chủ quan từ phía người dùng thông qua kỹ nghệ câu lệnh. Trong quá trình tương tác, nếu người dùng thiết lập các câu lệnh mang tính khái quát, thiếu vắng những tín hiệu văn hóa đặc thù của Việt Nam (như nhà tranh, sân đất, áo nâu sòng, hay không gian sinh hoạt đặc trưng của làng quê Bắc Bộ), hệ thống sẽ thiếu đi cơ sở dữ

liệu để khu biệt bối cảnh. Khi đối mặt với những khoảng trống thông tin do câu lệnh mơ hồ tạo ra, AI sẽ tự động lấp đầy bằng các khuôn mẫu thẩm mỹ phổ quát mang tính quốc tế. Sự cộng hưởng giữa thiên kiến dữ liệu sẵn có và sự thiếu hụt các chỉ dấu văn hóa bản địa từ người dùng đã dẫn đến hiện tượng đồng nhất hóa văn hóa, vô tình tạo ra những nội dung sai lệch và làm mất đi tính nguyên bản của di sản văn hóa [7] khi một thuật toán đưa ra các kết quả thiên lệch một cách hệ thống, thường do những giả định sai lầm trong quá trình xây dựng [5]. Các mô hình AI hiện nay được huấn luyện trên kho dữ liệu hình ảnh toàn cầu, trong đó dữ liệu về văn hóa phương Tây chiếm tỉ lệ cao hơn so với hình ảnh về văn hóa Việt Nam truyền thống. Do đó khi tiếp nhận yêu cầu tạo hình cho những bối cảnh như làng quê nghèo, nông thôn cũ hay người nông dân khổ cực, AI thường ưu tiên những mẫu hình quen thuộc mang tính quốc tế. Điều này dẫn đến hiện tượng đồng nhất hóa văn hóa khi vô tình tạo ra những nội dung sai lệch hoặc làm mất đi tính nguyên bản của di sản văn hóa [7]. Theo nghiên cứu Liefvooghe et al, “việc thuần túy gắn mác khuôn mặt nhân tạo là do trí tuệ nhân tạo (AI) tạo ra - hoàn toàn không dựa trên sự khác biệt trực quan nào - vẫn làm khởi phát những thiên kiến hệ thống đối với mức độ tin cậy” [2]. Nhận định này cho thấy hình ảnh do AI tạo ra không chỉ chịu tác động của dữ liệu và thuật toán, mà còn gắn với những kỳ vọng thẩm mỹ và nhận thức xã hội về tính hoàn hảo của hình ảnh nhân tạo, phải đi theo hướng tối ưu hóa hình ảnh theo tiêu chí thẩm mỹ hiện đại, khiến các bối cảnh nghèo đói, khắc nghiệt trong văn học hiện thực bị “làm đẹp” quá mức. Những khung cảnh có thể trở nên sạch sẽ, ánh sáng được xử lý nghệ thuật hơn, gương mặt nhân vật ít dấu vết khổ đau hơn so với mô tả trong nguyên tác. Điều này vô tình làm suy giảm giá trị hiện thực và tinh thần phê phán xã hội của tác phẩm. Văn học hiện thực vốn hướng đến việc phản ánh chân thực những bất công, nghèo đói và bi kịch con người, nhưng qua

bộ lọc của AI, nhiều hình ảnh lại trở nên lãng mạn hóa và mang tính trình diễn thị giác.

Quan niệm của tác giả cho rằng hiện tượng lệch chuẩn văn hóa luôn gắn liền với sự đứt gãy bối cảnh lịch sử. Chí Phèo không chỉ là một con người cá biệt mà còn là sản phẩm của xã hội nông thôn Việt Nam trước Cách mạng. Tràng và người vợ nhặt không chỉ là những cá nhân nghèo khổ mà còn là nạn nhân của nạn đói năm 1945. Khi AI tái hiện sai không gian văn hóa và làm

mờ các dấu ấn lịch sử, nhân vật cũng bị tách khỏi nền tảng xã hội đã tạo nên số phận của họ. Hệ quả là người xem có thể tiếp nhận tác phẩm theo hướng đơn giản hóa, chú ý nhiều đến hình ảnh bề mặt mà không nhận thức đầy đủ ý nghĩa hiện thực và nhân văn mà nhà văn muốn truyền tải.

Dựa trên bộ bốn tiêu chí định tính, nghiên cứu tiến hành chấm điểm và so sánh mức độ tái hiện hình ảnh của ba chatbot AI; kết quả được tổng hợp trong Bảng 2.

Bảng 2. So sánh điểm đánh giá định tính của ba chatbot AI theo các tiêu chí tái hiện hình ảnh nhân vật văn học

Tiêu chí đánh giá	Chat GPT	Gemini	Grok
Bám sát ngoại hình	4	3	3
Bối cảnh văn hóa	4	2	2
Chiều sâu biểu cảm (choáng váng/tuyệt vọng)	3	2	3
Không bị thiên kiến/Tây hóa	3	2	1
Tổng điểm trung bình	3,5	2,25	2,25

(Nguồn: Tác giả)

Dựa trên bộ tiêu chí đánh giá, tác giả tiến hành đối chiếu kết quả tạo hình ảnh của ba chatbot AI, kết quả cho thấy ChatGPT đạt điểm trung bình cao nhất với 3,5/5, trong khi Gemini và Grok cùng đạt 2,25/5, phản ánh sự khác biệt rõ rệt trong khả năng tái hiện nhân vật văn học.

- ChatGPT nổi bật nhất ở khả năng bám sát ngoại hình nhân vật và tái hiện bối cảnh văn hóa, đều đạt 4 điểm. Các hình ảnh do ChatGPT tạo ra nhìn chung thể hiện khá đúng đặc điểm nhân vật, đồng thời gợi được không khí nông thôn Việt Nam với trang phục, không gian và màu sắc gần với bối cảnh văn học. Nhưng mô hình vẫn còn một số hạn chế nhất định, như có xu hướng hiện đại hóa và điện ảnh hóa hình ảnh, khiến một vài chi tiết chưa hoàn toàn giữ được tinh thần nguyên bản của tác phẩm.

- Gemini đạt mức trung bình ở tiêu chí ngoại hình với 3 điểm, nhưng chỉ đạt 2 điểm ở tiêu chí bối cảnh văn hóa và chiều sâu biểu cảm. Hình ảnh do Gemini tạo ra tương đối đúng về mặt hình

thức nhưng còn khá chung chung, thiếu cảm xúc và chưa thể hiện rõ nét đặc trưng lịch sử - văn hóa Việt Nam. Điều này khiến nhân vật chưa tạo được cảm giác chân thực và gần gũi với nguyên tác văn học.

- Grok có ưu thế hơn Gemini ở khả năng thể hiện cảm xúc nhân vật, đặc biệt là những trạng thái như buồn bã hay tuyệt vọng. Tuy nhiên, mô hình này lại gặp hạn chế lớn trong việc tái hiện bối cảnh văn hóa và kiểm soát yếu tố thiên kiến. Hình ảnh thường mang phong cách phương Tây, từ đường nét khuôn mặt đến cách xây dựng không gian, làm giảm tính đặc trưng của nhân vật Việt Nam.

Sự khác biệt giữa ba mô hình AI có thể được lý giải từ dữ liệu huấn luyện và cách xử lý ngôn ngữ của từng hệ thống. Đối với ChatGPT có lợi thế nhờ được huấn luyện trên nguồn dữ liệu đa dạng về ngôn ngữ và văn hóa, trong đó có nhiều nội dung liên quan đến châu Á và Việt Nam. Nhờ đó, mô hình có khả năng hiểu ngữ cảnh tốt

hơn, không chỉ dựa vào từ khóa mà còn biết liên kết giữa mô tả văn học với yếu tố hình ảnh phù hợp. Điều này giúp hình ảnh tạo ra có tính nhất quán hơn về nhân vật, không gian và cảm xúc. Ngược lại, Gemini và Grok có thể vẫn chịu ảnh hưởng từ nguồn dữ liệu thiên về phương Tây và thiếu dữ liệu chuyên sâu về văn hóa Việt Nam, khi tạo hình ảnh các mô hình này dễ rơi vào xu hướng “Tây hóa”, khiến nhân vật và bối cảnh mất đi nét đặc trưng bản địa. Một số mô hình còn ưu tiên yếu tố thẩm mỹ hiện đại hoặc phong cách điện ảnh, dẫn đến việc hình ảnh đẹp mắt nhưng chưa thực sự phù hợp với tinh thần của tác phẩm văn học Việt Nam.

3.3. Định hướng ứng dụng AI trong dạy học văn học

Từ những hạn chế đó, nghiên cứu khẳng định rằng không nên sử dụng sản phẩm của AI chỉ như một công cụ minh họa thuần túy trong giáo dục. Thay vào đó, AI cần được khai thác như một công cụ phản biện trong quá trình dạy học.

Thứ nhất, cần lựa chọn tác phẩm phù hợp với mục tiêu giảng dạy, ưu tiên các truyện tiêu biểu có giá trị nội dung và nghệ thuật rõ nét như Chí Phèo, Vợ nhặt,... Sau đó, giáo viên tổ chức cần xác định rõ mục tiêu sử dụng AI như có thể hỗ trợ tóm tắt, xây dựng sơ đồ nhân vật và gợi ý câu hỏi thảo luận, nhưng kết quả cần được đối chiếu với nguyên tác để tránh sai lệch nội dung. Sau đó, giáo viên tổ chức cần xác định rõ mục tiêu sử dụng AI như hỗ trợ tìm hiểu nhân vật, phân tích bối cảnh xã hội, khám phá nghệ thuật tự sự hay quảng bá giá trị của tác phẩm đến công chúng trẻ. Ở giai đoạn này, AI có thể hỗ trợ tóm tắt văn bản, xây dựng sơ đồ nhân vật, gợi ý câu hỏi thảo luận hoặc đề xuất nhiều hướng tiếp cận khác nhau. Tuy nhiên, toàn bộ dữ liệu đầu ra cần được đối chiếu với nguyên tác để tránh hiện tượng giản lược nội dung hoặc sai lệch thông tin.

Thứ hai, thay vì coi AI là công cụ minh họa nên khai thác AI như một công cụ phản biện trong quá trình dạy học. Giáo viên có thể yêu cầu

AI tạo hình ảnh nhân vật, bối cảnh hay tình huống truyện, sau đó hướng dẫn học sinh so sánh với nguyên tác để phát hiện những điểm chưa chính xác, phiến diện và thiếu chiều sâu nghệ thuật. Ví dụ, khi dạy Chí Phèo, AI có thể tái hiện nhân vật theo kiểu dữ tợn, điện ảnh hóa nhưng chưa thể hiện hết bi kịch bị tha hóa và khát vọng được làm người lương thiện. Thông qua hoạt động đối chiếu này, học sinh sẽ nhận ra rằng hình tượng văn học luôn sâu sắc và đa tầng hơn sản phẩm mô phỏng của máy móc. AI vì thế trở thành chất liệu để kích thích tư duy phản biện, thay vì thay thế việc đọc hiểu tác phẩm.

Thứ ba, sau khi tiếp cận sản phẩm do AI tạo ra, giáo viên có thể chia thành các nhóm để thảo luận các vấn đề như AI đã hiểu sai điều gì? Vì sao AI chưa thể hiện được tâm lý nhân vật? Ngôn ngữ văn học khác với hình ảnh trực quan ở điểm nào? Nếu chỉnh sửa lại prompt, cần bổ sung yếu tố gì để gần với tinh thần nguyên tác hơn? Những câu hỏi đó sẽ giúp người học chuyển từ vị thế tiếp nhận thụ động sang chủ động kiến tạo tri thức. Không chỉ hiểu nội dung truyện ngắn mà còn rèn luyện năng lực đọc hiểu, lập luận, phản biện, hợp tác nhóm và khả năng đánh giá công nghệ trong đời sống hiện đại.

Những lợi ích trong giảng dạy và quảng bá văn học, việc ứng dụng AI cũng đặt ra nhiều vấn đề về bản quyền số và đạo đức công nghệ. Khi sử dụng AI để tái hiện nhân vật, bối cảnh hay xây dựng nội dung từ các truyện ngắn Việt Nam, người dùng cần tuân thủ các nguyên tắc “đảm bảo an toàn dữ liệu cá nhân, tôn trọng quyền tác giả, bản quyền sở hữu trí tuệ” [4; 34-45]. Không nên sao chép nguyên văn nội dung và khai thác trái phép các sản phẩm văn học. Trong bối cảnh chuyển đổi số, bản quyền không chỉ áp dụng với văn bản mà còn mở rộng sang dữ liệu số, hình ảnh, âm thanh và các sản phẩm do AI hỗ trợ tạo ra. Vì vậy, việc phát triển và sử dụng AI cần gắn với trách nhiệm xã hội, tính minh bạch và sự tôn trọng giá trị con người. Ngoài ra cần bảo đảm

tính minh bạch trong quá trình sử dụng AI. Các hình ảnh, video, bản tóm tắt hay nội dung phân tích có sử dụng AI nên được chú thích rõ nguồn gốc công nghệ để tránh nhầm lẫn giữa sáng tạo của con người và sản phẩm do máy hỗ trợ. AI không nên thay thế hoàn toàn tư duy và khả năng cảm thụ văn học của người học mà chỉ đóng vai trò gợi ý, phản biện và hỗ trợ mở rộng ý tưởng. Cần chú ý bảo vệ bản sắc văn hóa dân tộc khi ứng dụng AI vào văn học Việt Nam. Do được huấn luyện trên nguồn dữ liệu toàn cầu, AI có thể vô tình áp đặt các khuôn mẫu ngoại lai, làm sai lệch không gian văn hóa và tinh thần tác phẩm. Vì vậy, các sản phẩm AI liên quan đến truyền ngữ Việt Nam cần được kiểm chứng bởi giáo viên hoặc người có chuyên môn nhằm bảo đảm tính xác thực văn hóa. Cần khẳng định rằng AI phải phục vụ lợi ích của con người, xã hội và sự phát triển bền vững còn công nghệ chỉ nên là phương tiện hỗ trợ, không phải chủ thể thay thế sáng tạo.

4. Kết luận

Quá trình chuyển đổi từ ngôn ngữ văn học sang hình ảnh do AI tạo ra đã mở ra nhiều cơ hội mới trong việc tiếp cận và minh họa tác phẩm nghệ thuật. Nhưng sự chuyển đổi này cũng bộc lộ khoảng cách lớn giữa lớp vỏ hình ảnh bên ngoài và chiều sâu nội tại của văn học. Nếu ngôn từ văn học khơi gợi trí tưởng tượng và cảm xúc cá nhân của người đọc thì AI lại có xu hướng biến những hình tượng ấy thành các khuôn mẫu trực quan cố định. Nhiều nhân vật văn học khi được AI tái hiện thường bị “phẳng hóa”, chỉ còn nổi bật ở ngoại hình mà thiếu đi chiều sâu tâm lý, tính cách và những giằng xé nội tâm vốn làm nên giá trị của tác phẩm và tạo ra hiện tượng “ảo giác văn hóa” do chịu ảnh hưởng từ thiên kiến dữ liệu và thuật toán. Khi nguồn dữ liệu huấn luyện chủ yếu phản ánh một số nền văn hóa phổ biến, hình ảnh do AI tạo ra dễ mang tính áp đặt và sai lệch đối với các nền văn hóa khác. Những chi tiết đặc trưng về con người, trang phục hay

bối cảnh có thể bị đồng nhất hóa, làm mất đi bản sắc riêng của tác phẩm văn học. Việc phụ thuộc quá nhiều vào AI trong quá trình tiếp nhận văn học cũng có nguy cơ làm suy giảm khả năng tưởng tượng và tư duy độc lập của con người. Một trong những giá trị lớn nhất của văn học là cho phép mỗi người đọc tự hình dung thế giới nghệ thuật theo cảm nhận riêng, khi AI cung cấp sẵn hình ảnh trực quan, người đọc dễ tiếp nhận thụ động thay vì chủ động suy ngẫm và liên tưởng. Để ứng dụng công nghệ một cách bền vững, con người cần giữ vai trò là trung tâm, nên xem AI chỉ là một công cụ hỗ trợ chứ không thể thay thế hoàn toàn yếu tố của con người trong dạy học [3; 178]. Đồng thời, việc sử dụng AI cần tuân thủ các nguyên tắc, không nên sao chép nguyên văn nội dung, sử dụng trái phép hình tượng văn học hoặc khai thác thương mại khi chưa được sự cho phép của chủ sở hữu quyền tác giả. Quan trọng hơn, con người cần sử dụng công cụ số để rèn luyện tư duy phản biện và khả năng sáng tạo, thay vì phụ thuộc hoàn toàn vào sự diễn giải của máy móc.

Từ kết quả của đề tài, các nghiên cứu sau có thể mở rộng theo nhiều hướng khác nhau nhằm làm rõ hơn mối quan hệ giữa văn học và trí tuệ nhân tạo. Có thể mở rộng phạm vi khảo sát sang nhiều thể loại khác như tiểu thuyết, thơ, kịch hoặc văn học dân gian để đánh giá khả năng tái hiện của AI trong nhiều hình thức biểu đạt nghệ thuật khác nhau. Việc nghiên cứu thêm các nhân vật mang tính biểu tượng văn hóa đặc thù cũng sẽ giúp nhận diện rõ hơn mức độ ảnh hưởng của thiên kiến thuật toán. Các nghiên cứu tiếp theo có thể so sánh nhiều mô hình AI mới hơn để đánh giá sự thay đổi về chất lượng tái hiện hình tượng nhân vật theo quá trình phát triển công nghệ. Có thể kết hợp thêm các phương pháp phân tích dữ liệu, xử lý hình ảnh, đánh giá người dùng trên quy mô lớn để tăng tính khách quan cho kết quả nghiên cứu. Ngoài hướng nghiên cứu học thuật, đề tài cũng có thể được phát triển theo hướng ứng dụng thực tiễn trong giáo dục và

truyền thông số. Việc xây dựng hệ thống AI hỗ trợ giảng dạy văn học Việt Nam theo hướng tôn trọng bản sắc văn hóa và tăng khả năng phản biện của người học là một hướng đi đáng được quan tâm trong tương lai. Cần có thêm những nghiên cứu chuyên sâu về đạo đức AI, bản quyền số và tác động của AI đối với khả năng cảm thụ văn học của con người, đặc biệt trong môi trường giáo dục hiện đại.

Tài liệu tham khảo

- [1] Hiền, N.T.T (2025). “Ứng dụng trí tuệ nhân tạo (AI) trong học tập và giảng dạy”. *Tạp chí Tâm Lý - Giáo Dục*, 31 (06).
- [2] Liefoghe, B., Oliveira, M., Leisten, L.M., Hoogers, E., Aarts, H., Hortensius (2023). “R.: Are natural faces merely labelled as artificial trusted less?”. *Collabra: Psychology* 9 (1), 73066 (2023).
- [3] Phương, T.T.H, Ngân, C.T.K (2024). “Vận dụng Trí tuệ nhân tạo (AI) trong hoạt động dạy học Ngữ văn ở trường phổ thông”. *Tạp chí Thiết bị Giáo dục*, số 308.
- [4] Thuận, B.M. (2024). “Đạo Đức của người sử dụng trí tuệ nhân tạo trong nền giáo dục 5.0”. *Tạp chí Khoa học Giáo dục Kỹ thuật*.
- [5] Đại học Duy Tân (n.d). *Thiên kiến AI*. Truy cập 22/3/2026, từ: <https://is.duytan.edu.vn/study-corner/thien-kien-ai-ng7tr>
- [6] ETP TESOL (2025). *Ứng dụng trí tuệ nhân tạo trong dạy học Ngữ văn: Tương lai của cảm thụ văn học*. Truy cập 24/3/2026, từ: <https://etp-tesol.edu.vn/ung-dung-ai-trong-day-hoc-ngu-van/>
- [7] Phạm Việt Long (2025). *Trí tuệ nhân tạo (AI) với việc bảo tồn và phát huy di sản văn hóa*, Truy cập 25/03/2026, từ: <https://phamvietlong.vn/tri-tue-nhan-tao-ai-voi-viec-bao-ton-va-phat-huy-di-san-van-hoa>