CHATGPT'S MATHEMATICAL PROBLEM-SOLVING PERFORMANCE IN GRADE 11 GRAPH THEORY

Pham Thanh Dat[1], Goi Ngoc Tu*[2], Tran Khanh Anh[3], Ho Ngoc Bao Tram[4]

* Corresponding author
Email: goingoctu9at4@gmail.com

[1] Email: datpt@hcmue.edu.vn
[3] Email: Kaanh3851D@gmail.com
[4] Email: hongocbaotram2020@gmail.com
[1,2,3,4] Ho Chi Minh City University of Education
280 An Duong Vuong street, Cho Quan ward,
Ho Chi Minh City, Vietnam

Received: 19/5/2026
Revised: 21/6/2026
Accepted: 15/7/2026
Published: 20/9/2026

Abstract: Although ChatGPT's mathematical problem-solving performance has been extensively investigated across various domains, evaluating its specific proficiency in Graph Theory within the 2018 General Education Curriculum remains a notable research gap that warrants clarification. This study analyzes ChatGPT's problem-solving capabilities through the detailed assessment of 257 designated grade 11 Mathematics tasks derived from three currently used textbook series, while simultaneously identifying the model's underlying systematic errors. Empirical results clearly demonstrate significant performance variance across these instructional resources, highlighting five typical error categories the model frequently commits when processing graph data. Ultimately, these findings not only provide practical empirical evidence concerning the true capacity of artificial intelligence in modern mathematics education but also critically assist educators in clearly recognizing the tool's operational limitations. Consequently, this research establishes a solid foundation for proposing effective and appropriate integration strategies within the ongoing context of educational reforms.

Keywords: *ChatGPT, graph theory Maths, problem-solving performance, error.*

HIỆU SUẤT GIẢI TOÁN CỦA CHATGPT TRONG CHUYÊN ĐỀ LÍ THUYẾT ĐỒ THỊ LỚP 11

Phạm Thành Đạt[1], Goi Ngọc Tú*[2], Trần Khánh Anh[3], Hồ Ngọc Bảo Trâm[4]

* Tác giả liên hệ
Email: goingoctu9at4@gmail.com

[1] Email: datpt@hcmue.edu.vn
[3] Email: Kaanh3851D@gmail.com
[4] Email: hongocbaotram2020@gmail.com
[1,2,3,4] Trường Đại học Sư phạm Thành phố Hồ Chí Minh
280 An Dương Vương, phường Chợ Quán,
Thành phố Hồ Chí Minh, Việt Nam

Nhận bài: 19/5/2026
Chỉnh sửa xong: 21/6/2026
Chấp nhận đăng: 15/7/2026
Xuất bản: 20/9/2026

Tóm tắt: Mặc dù hiệu suất giải Toán của ChatGPT đã được khảo sát trên nhiều lĩnh vực, song việc đánh giá năng lực của mô hình này đối với nội dung Lí thuyết đồ thị trong Chương trình Giáo dục phổ thông 2018 vẫn còn là một khoảng trống cần được làm rõ. Nghiên cứu này tập trung phân tích khả năng giải quyết vấn đề của ChatGPT thông qua 257 nhiệm vụ thuộc chuyên đề Toán học lớp 11 từ ba bộ sách giáo khoa hiện hành, đồng thời nhận diện các sai sót mang tính hệ thống của mô hình. Kết quả thực nghiệm cho thấy, hiệu suất của ChatGPT có sự chênh lệch đáng kể giữa các bộ sách và chỉ ra 05 nhóm lỗi điển hình mà mô hình thường xuyên mắc phải khi xử lí dữ liệu đồ thị. Những phát hiện này không chỉ cung cấp minh chứng thực tiễn về năng lực của trí tuệ nhân tạo trong giáo dục Toán học mà còn giúp các nhà giáo dục nhận diện rõ giới hạn của công cụ, từ đó đưa ra những định hướng sử dụng hiệu quả và phù hợp trong bối cảnh đổi mới giáo dục phổ thông hiện nay.

Từ khóa: *ChatGPT, Lí thuyết đồ thị, hiệu suất giải Toán, lỗi sai.*

1. Đặt vấn đề

Trong những năm gần đây, trí tuệ nhân tạo đang được ứng dụng rộng rãi trong giáo dục (Díaz, 2024). Các nghiên cứu gần đây cho thấy các mô hình ngôn ngữ lớn có khả năng mô phỏng ngôn ngữ tự nhiên, thực hiện lập luận và hỗ trợ học tập một cách hiệu quả. Cụ thể, các hệ thống AI như ChatGPT, GPT-4o, o1-preview và Bard có thể đạt hoặc vượt mức điểm trung bình của học sinh trong các bài kiểm tra chuẩn hóa (Kaya & Yavuz, 2025; Oh, 2024; Phạm Tất Thành, 2024; Lê Anh Vinh và cộng sự, 2023). Tuy nhiên, các mô hình này vẫn tồn tại hạn chế trong việc xử lí các

bài toán đòi hỏi lập luận phức tạp, đồng thời thường mắc lỗi về khái niệm và suy luận logic (Oh, 2024). Khi xem xét sâu hơn về hiệu suất giải toán, các bằng chứng thực nghiệm cho thấy, năng lực của ChatGPT phụ thuộc mạnh mẽ vào thế hệ mô hình và độ phức tạp nhận thức của nhiệm vụ. Ở các bài kiểm tra chuẩn hóa phổ thông, các mô hình tiêu chuẩn như GPT-4o chỉ đạt mức hiệu suất trung bình - thấp với tỉ lệ chính xác khoảng 49.46%, trong khi dòng mô hình suy luận thế hệ mới như o1-preview có thể đạt tới 81.52%, tương đương với nhóm học sinh thuộc top đầu (Oh, 2024). Xu hướng này được ghi nhận trong các kì thi đòi hỏi tư duy cao tại Hoa Kì, nơi ChatGPT vượt trội hơn phần lớn học sinh ở cấp độ Tiểu học và Trung học cơ sở nhưng lại giảm sâu hiệu suất và không thể hiện sự nhạy bén có ý nghĩa thống kê khi đối mặt với các bài toán lớp 12 mang tính trừu tượng (Kaya & Yavuz, 2025). Tương tự, đối với các kì thi tuyển sinh đại học chuyên Toán có tính cạnh tranh cao tại Vương quốc Anh (như bài thi TMUA), phiên bản ChatGPT đời đầu bộc lộ độ chính xác rất hạn chế ở cả phần kiến thức nền tảng lẫn tư duy logic (Giannos & Delardas, 2023).

Bên cạnh đó, bản chất cấu trúc và dạng thức của bài toán quyết định lớn đến sự biến thiên của hiệu suất. Trong Toán học cao cấp như Vi tích phân, GPT-4 có thể đạt khoảng 65% độ chính xác đối với các bài toán có cấu trúc tiêu chuẩn nhưng tỉ lệ này ngay lập tức phân rã, chỉ còn 20% khi đối mặt với các dạng toán mang tính thách thức hoặc đòi hỏi mẹo giải toán của các kì thi học sinh giỏi (Gandolfi, 2025). Thậm chí, trong các tác vụ hỗ trợ giáo dục như tự động chấm điểm bài làm tự luận hay phát hiện các câu trả lời không mạch lạc của học sinh tiểu học, các mô hình ngôn ngữ lớn (LLM) vẫn cho thấy kết quả kém hơn so với các thuật toán học máy truyền thống do gặp rào cản lớn trước các câu hỏi có cấu trúc đệ quy hoặc hiện tượng mất tính nhất quán khi xử lí ngữ cảnh dài (Gandolfi, 2025). Đáng chú ý, mặc dù ChatGPT có tỉ lệ lỗi toán học ban đầu tương đối cao (khoảng 32%) khi tạo nội dung cứu trợ học tập cho học sinh, hiệu suất và độ chính xác của mô hình có thể được cải thiện tiệm cận mức tuyệt đối trong các phân môn như Đại số nhờ vào các kĩ thuật thúc đẩy nhận thức như tự kiểm tra tính nhất quán (Self-Consistency) (Pardos & Bhandari, 2024). Do đó, việc xây dựng các chuẩn mực đạo đức và cơ chế kiểm soát nội dung do AI tạo ra là cần thiết (Giannos & Delardas, 2023).

Mặc dù cùng tập trung đánh giá AI trong giáo dục, các nghiên cứu có sự khác biệt về phạm vi và phương pháp tiếp cận. Các nghiên cứu của Giannos & Delardas (2023) và Gandolfi (2025) chủ yếu phân tích năng lực lập luận và các giới hạn nhận thức của AI. Trong khi đó, Kaya & Yavuz (2025), Oh (2024) và Phạm Tất Thành (2024) tập trung so sánh hiệu suất giữa AI và học sinh trong các kì thi. Theo một hướng nghiên cứu khác, Lê Thái Bảo Thiên Trung và cộng sự (2025) nhấn mạnh khía cạnh ứng dụng thực tiễn, chỉ ra rằng, ChatGPT có thể góp phần nâng cao nhận thức và hứng thú của giáo viên Toán.

Trong các nghiên cứu đánh giá năng lực của trí tuệ nhân tạo, thuật ngữ "hiệu suất" (Performance) được hiểu là năng lực biểu hiện thực tế của mô hình thông qua mức độ hoàn thành các nhiệm vụ nhận thức cụ thể dựa trên hệ thống tiêu chuẩn định lượng (Giannos, 2023). Đối với lĩnh vực Giáo dục, hiệu suất của một mô hình ngôn ngữ lớn (LLM) không chỉ dừng lại ở tốc độ phản hồi mà tập trung vào độ chính xác, tính logic và khả năng thích ứng của câu trả lời trước các định dạng đề bài khác nhau (Lê Anh Vinh và cộng sự, 2023).

Kế thừa các tiếp cận trên, nghiên cứu này xác lập thuật ngữ "Hiệu suất giải toán" (Mathematical problem-solving performance) của ChatGPT là khả năng tiếp nhận, xử lí cấu trúc ngôn ngữ của bài toán, thực thi các bước lập luận logic, tính toán và kết xuất lời giải chính xác phù hợp với các tiêu chí đánh giá giáo dục hiện hành. ChatGPT thể hiện hiệu suất ra sao khi giải quyết các bài toán trong chủ đề này và nguyên nhân cốt lõi nào dẫn đến những nhiệm vụ mô hình xử lí sai? Để làm rõ vấn đề, nghiên cứu này hướng tới ba mục tiêu: 1) So sánh kết quả thực hiện giải toán của ChatGPT trên ba bộ sách giáo khoa thuộc Chương trình Giáo dục phổ thông 2018 đối với chủ đề Lí thuyết đồ thị; 2) Thống kê kết quả theo từng lỗi sai trong mỗi bộ sách; 3) Phân tích các nhiệm vụ minh họa tương ứng với từng loại lỗi sai.

2. Phương pháp nghiên cứu

2.1. Thiết kế nghiên cứu

Nghiên cứu được triển khai theo thiết kế mô tả - so sánh (Descriptive-comparative design), kết hợp giữa phân tích định lượng và định tính nhằm đánh giá hiệu suất giải Toán và nhận diện các lỗi sai mang tính hệ thống của ChatGPT trong chủ đề Lí thuyết đồ thị lớp 11. Đối tượng nghiên cứu là các phản hồi của mô hình ngôn ngữ lớn ChatGPT (phiên bản phổ thông, truy cập tháng 7 năm 2025) khi giải các nhiệm vụ toán học được trích từ ba bộ sách chuyên đề Toán 11 thuộc Chương trình Giáo dục phổ thông 2018. Đơn vị phân tích là nhiệm vụ Toán học độc lập (bao gồm bài tập, ví dụ minh họa và hoạt động học tập).

 https://doi.org/10.15625/2615-8957/1260911

Phương pháp này cho phép đo lường hiệu suất của ChatGPT dựa trên các chỉ số thống kê ở hai phương diện chính: *1) Phương diện nội dung*: Thống kê tỉ lệ hoàn thành chính xác các nhiệm vụ theo từng đơn vị bài học; *2) Phương diện yêu cầu nhận thức:* Phân tích sự biến thiên về hiệu suất của ChatGPT qua các mức độ nhận thức khác nhau của học sinh, từ đó xác lập các dữ liệu thống kê về tần suất xuất hiện sai số trong phản hồi của mô hình.

2.2. Quy trình nghiên cứu

Bước 1: Trích lục bài tập: Hệ thống hóa toàn bộ các nhiệm vụ Toán học thuộc chủ đề Lí thuyết đồ thị từ ba bộ sách giáo khoa chuyên đề Toán 11 (Chương trình Giáo dục phổ thông 2018). Mỗi nhiệm vụ đều được mã hóa chi tiết theo các tiêu chí: bộ sách, bài học và đơn vị kiến thức tương ứng. Quy trình thu thập dữ liệu được thực hiện thông qua phiên bản ChatGPTvới số lượng các nhiệm vụ toán học cụ thể như sau: Chân trời sáng tạo (108 nhiệm vụ); Cánh diều (66 nhiệm vụ); Kết nối tri thức với cuộc sống (83 nhiệm vụ). Dữ liệu nghiên cứu gồm 257 nhiệm vụ toán học thuộc chủ đề Lí thuyết đồ thị.

Bước 2: Chuẩn hóa dữ liệu đầu vào: Các nhiệm vụ được chuẩn hóa và nhập vào ChatGPT ở dạng văn bản. Đối với các bài toán có hình vẽ, dữ liệu trực quan được giữ nguyên nhằm đảm bảo tính toàn vẹn của ngữ cảnh Toán học.

Bước 3: Nhập liệu: Mỗi nhiệm vụ chỉ được nhập một lần, không cung cấp gợi ý hoặc dẫn dắt, nhằm đảm bảo tính khách quan và đo lường năng lực giải quyết vấn đề độc lập của mô hình.

Bước 4: Ghi nhận kết quả: Phản hồi của ChatGPT được lưu trữ nguyên văn và đối chiếu với đáp án chuẩn trong sách giáo khoa và nền tảng lí thuyết toán học để xác định tính chính xác. Dữ liệu được thu thập theo quy trình thống nhất, sử dụng cùng một phiên bản ChatGPT 4.5 (tháng 7 năm 2025).

2.3. Phương pháp phân tích dữ liệu

Phân tích định lượng: Đóng vai trò chủ đạo trong việc thống kê tỉ lệ giải đúng nhằm đo lường, so sánh hiệu suất của ChatGPT giữa các bộ sách dựa trên dữ liệu nghiên cứu gồm 257 nhiệm vụ Toán học.

Phân tích định tính: Tập trung nhận diện và phân tích các loại lỗi sai điển hình thường gặp của ChatGPT trong giải quyết các nhiệm vụ Toán học thuộc chủ đề Lí thuyết đồ thị; qua đó làm rõ bản chất của hiện tượng đứt gãy giữa lập luận logic của ChatGPT.

Trong quá trình chuẩn bị bản thảo này, các tác giả sử dụng ChatGPT như một đối tượng nghiên cứu và không sử dụng công cụ này trong thiết kế nghiên cứu. Phần mềm Zotero (phiên bản 8.0.4) được sử dụng để định dạng tài liệu tham khảo theo chuẩn APA. Các tác giả chịu hoàn toàn trách nhiệm về mọi sai sót trong nội dung (nếu có).

3. Kết quả nghiên cứu

3.1. Phân tích sơ bộ

Kết quả phân tích thực nghiệm cho thấy sự phân hóa có ý nghĩa về hiệu suất xử lí nhiệm vụ học thuật của mô hình ngôn ngữ lớn ChatGPT đối với ba bộ sách giáo khoa hiện hành (xem Bảng 1). Cụ thể, dữ liệu ghi nhận bộ sách Cánh diều đạt độ tương thích cao nhất với tỉ lệ chính xác 74,24% (49/66 nhiệm vụ), cho thấy khả năng xử lí tương đối tốt của mô hình đối với các nhiệm vụ Toán học. Mức hiệu suất này có thể được xem là tiệm cận năng lực khá giỏi, điều này tương tự với nghiên cứu của Oh (2024) (o1-preview đạt 81,52% và có hiệu suất tương đương nhóm người học trình độ cao). Bộ sách Kết nối tri thức với cuộc sống thể hiện mức độ đáp ứng trung bình đạt 65,06% (54/83 nhiệm vụ). Ngược lại, bộ sách Chân trời sáng tạo ghi nhận tỏ lệ phản hồi chính xác thấp nhất, chỉ dừng lại ở 45,37% (49/108 nhiệm vụ), phản ánh những hạn chế nội tại của mô hình trước các cấu trúc ngữ cảnh và yêu cầu sư phạm đặc thù. Xét trên bình diện tổng thể, tỉ lệ chính xác trung bình đạt 59,14% (152/257 nhiệm vụ), cho thấy năng lực xử lí ở mức trung bình khá nhưng vẫn tồn tại khoảng trống sai lệch gần 41%. Sự sai lệch này liên quan đến hiện tượng "Ảo giác AI", được định nghĩa là việc mô hình ngôn ngữ lớn tạo ra các nội dung sai lệch so với thực

Bảng 1: *Kết quả thực hiện của ChatGPT ở ba bộ sách*

	Tổng số nhiệm vụ	Số nhiệm vụ đúng	Số nhiệm vụ sai	Tỉ lệ đúng (%)
Chân trời sáng tạo	108	49	59	45,37%
Cánh diều	66	49	17	74,24%
Kết nối tri thức với cuộc sống	83	54	29	65,06%
Tổng	257	152	105	59,14%

Bảng 2: *Thống kê lỗi sai của ChatGPT ở ba bộ sách*

Lỗi	Chân trời sáng tạo	Cánh diều	Kết nối tri thức với cuộc sống	Tổng	Tỉ lệ (%)
L1: Nhận diện hình ảnh	48	10	18	76	72,38%
L2: Tạo lập hình ảnh	4	2	3	9	8,57%
L3: Nhận biết khái niệm	1	1	1	3	2,86%
L4: Đề xuất ví dụ	0	2	4	6	5,71%
L5: Lập luận giải thích	6	2	3	11	10,48%
Tổng	59	17	29	105	100%

tế khách quan hoặc mâu thuẫn với ngữ cảnh được cung cấp (Huang và cộng sự, 2025). Hệ quả sư phạm quan trọng rút ra là mặc dù ChatGPT bộc lộ tiềm năng trong việc hỗ trợ tra cứu và gợi mở ý tưởng, song sự hiện diện của sai số hệ thống và hiện tượng "ảo giác AI" đòi hỏi các phản hồi phải được kiểm chứng, thẩm định nghiêm ngặt với nguồn học liệu chuẩn tắc trước khi triển khai trong môi trường giáo dục chính quy.

3.2. Phân tích lỗi sai của ChatGPT trong từng bộ sách (Chân trời sáng tạo, Cánh diều, Kết nối tri thức với cuộc sống)

Phân tích 105 trường hợp ChatGPT thực hiện sai cho thấy có năm loại lỗi chính, bao gồm: Nhận diện hình ảnh, tạo lập hình ảnh, nhận biết khái niệm, đề xuất ví dụ và lập luận giải thích. Trong đó, lỗi nhận biết khái niệm thường xuất phát từ việc ChatGPT hiểu sai các định nghĩa trong lí thuyết đồ thị. Ngoài ra, lỗi lập luận và giải thích thường xuất hiện trong các bài toán đòi hỏi suy luận nhiều bước, khi mô hình sử dụng sai khái niệm, bỏ sót trường hợp hoặc đưa ra lập luận không phù hợp với yêu cầu của bài toán.

Kết quả phân tích cho thấy, phân bố lỗi của ChatGPT trong ba bộ sách chuyên đề là không đồng đều. Cụ thể, tỉ lệ các loại lỗi lần lượt là 72,38% (L_1), 8,57% (L_2), 2,86% (L_3), 5,71% (L_4) và 10,48% (L_5). Điều này cho thấy, phần lớn lỗi tập trung vào các tác vụ liên quan đến xử lí hình ảnh, trong khi các tác vụ xử lí ngôn ngữ tự nhiên có tỉ lệ lỗi tương đối thấp.

Bảng 2 cho thấy lỗi L_1 chiếm tỉ lệ cao nhất (72,38%), qua đó phản ánh đây là điểm yếu cốt lõi và phổ biến nhất của ChatGPT khi xử lí dữ liệu từ các bộ sách chuyên đề. Hạn chế này chủ yếu nằm ở khâu nhận diện hình ảnh, chiếm gần 75% tổng số lỗi. Ngược lại, lỗi L_3 có tỉ lệ thấp nhất (2,86%), cho thấy ChatGPT có khả năng nắm bắt và xử lí các khái niệm lí thuyết bằng văn bản khá tốt, với mức độ sai sót không đáng kể.

Dựa trên phân bố tỉ lệ lỗi, có thể phân loại các lỗi thành hai mức độ. Nhóm thấp (≤25%) bao gồm L_2, L_3, L_4 và L_5, chiếm 4/5 loại lỗi và tổng cộng 27,62% số lỗi. Điều này cho thấy các lỗi liên quan đến suy luận, hiểu khái niệm và đề xuất ví dụ xuất hiện với tần suất thấp và không chiếm tỉ trọng đáng kể trong tổng thể. Sự chênh lệch giữa tỉ lệ cao nhất (L_5: 10,48%) và thấp nhất (L_3: 2,86%) trong nhóm này là tương đối nhỏ (khoảng 7,62%), phản ánh mức độ phân tán không lớn.

Nhóm trung bình (25%–75%) chỉ bao gồm duy nhất lỗi L_1 (72,38%). Mặc dù nằm trong khoảng này, giá trị của L_1 tiệm cận ngưỡng trên (75%), cho thấy sự mất cân bằng rõ rệt trong phân bố lỗi. Các lỗi không phân bố đồng đều mà tập trung chủ yếu vào một loại lỗi duy nhất liên quan đến xử lí hình ảnh.

Kết quả thống kê các lỗi từ L_1 đến L_5 cho thấy sự khác biệt rõ rệt giữa hai nhóm năng lực của ChatGPT. Cụ thể, các lỗi liên quan đến xử lý thông tin trực quan (L_1 và L_2) chiếm tới 80,95% tổng số lỗi, trong khi các lỗi thuộc nhóm xử lý ngôn ngữ và suy luận (L_3, L_4 và L_5) chỉ chiếm 19,05%. Điều này cho thấy hạn chế chủ yếu của ChatGPT không nằm ở khả năng hiểu khái niệm hay suy luận logic, mà xuất phát từ việc tiếp nhận và biểu diễn thông tin hình ảnh. Đặc biệt, lỗi L_1 đóng vai trò trung tâm và có thể gây ra hiệu ứng lan truyền, dẫn đến các lỗi ở các bước suy luận và biểu diễn tiếp theo như L_2, L_4 và L_5. Ngược lại, tỉ lệ rất thấp của lỗi L_3 cho thấy mô hình có nền tảng vững chắc trong việc xử lí tri thức lí thuyết dưới dạng văn bản.

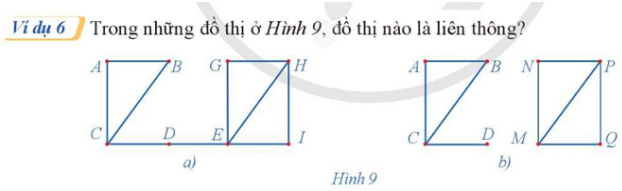
3.3. Phân tích các trường hợp điển hình minh họa cho các loại lỗi của ChatGPT

3.3.1. Lỗi sai 1: Nhận diện hình ảnh

Kết quả thống kê cho thấy lỗi L_1 chiếm tỉ lệ áp đảo với 72,38% trên tổng số lỗi sai ở cả ba bộ sách giáo khoa, minh chứng cho một hạn chế mang tính hệ thống của ChatGPT trong việc xử lí thông tin

 https://doi.org/10.15625/2615-8957/1260911

trực quan. Kết quả này không đơn thuần là một sai lệch kĩ thuật đơn lẻ mà còn chỉ ra điểm yếu về khả năng nhận diện hình ảnh của mô hình ngôn ngữ lớn khi tiếp cận các đối tượng toán học có cấu trúc trừu tượng như đồ thị. Để làm rõ bản chất của loại lỗi sai này ở ChatGPT, Ví dụ 6 ở Bài 1 trong bộ sách Cánh diều được phân tích như một trường hợp điển hình có giá trị thực nghiệm cao (xem Hình 1):



Hình 1: *Nội dung của Ví dụ 6 - Bài 1 Sách Cánh diều*

Hình 9a:

- Có 2 phần riêng biệt:
  - Một phần gồm các đỉnh: A, B, C, D (liên thông với nhau).
  - Phần còn lại gồm: E, G, H, I (liên thông với nhau).
- **Không có cạnh nối giữa hai phần này.**

→ **Không liên thông** vì không có đường đi giữa một số cặp đỉnh (ví dụ từ A đến E).

Hình 2: *Lời giải của ChatGPT*

Về mặt lí thuyết, đồ thị tại Hình 9a là một đồ thị liên thông nhờ sự hiện diện của cạnh nối DE đóng vai trò "cầu nối" giữa hai tập đỉnh {A, B, C, D} và {E, G, H, I}. Tuy nhiên, ChatGPT đã thực hiện một phép phân đoạn thị giác sai lệch khi cô lập hai cụm đỉnh này thành các thành phần rời rạc và kết luận đồ thị không liên thông. Sai phạm này không xuất phát từ việc thiếu hụt tri thức về định nghĩa Toán học mà từ sự thất bại trong việc thiết lập biểu diễn tri thức ban đầu. Việc không nhận diện được cạnh nối DE dẫn đến xác định sai tính liên thông của đồ thị, khiến toàn bộ chuỗi suy luận phía sau dù tuân thủ chặt chẽ và logic đều dẫn đến kết quả sai.

ChatGPT có xu hướng nhận diện các cụm đỉnh tạo thành cấu trúc hình học khép kín (như hình vuông, tam giác) là các đơn vị độc lập, nhưng bỏ sót các cạnh đơn lẻ. Nói cách khác, mô hình bị chi phối bởi các "vùng mật độ cao" thay vì thực hiện một quy trình quét toàn cục để xác định đầy đủ. Đáng chú ý hơn cả là sự xuất hiện của hiện tượng "giả logic", dù dữ liệu đầu vào bị nhận diện sai, ChatGPT vẫn triển khai lời giải theo một cấu trúc diễn dịch chuẩn mực: Phân tách tập đỉnh, kiểm tra sự tồn tại cạnh nối và áp dụng định nghĩa để đi đến kết luận.

3.3.2. Lỗi sai 2: Tạo lập hình ảnh

Kết quả thống kê cho thấy, lỗi L_2 tuy chiếm tần suất thấp hơn so với L_1 (chiếm tỉ lệ 8,57% trong tổng số lỗi sai) nhưng lại bộc lộ một nghịch lí năng lực đa phương thức đáng chú ý: Sự thiếu nhất quán trầm trọng giữa biểu diễn hình thái trực quan và lập luận toán học. Minh chứng điển hình qua Luyện tập 1 - Bài 1 Sách Cánh diều: "Có năm thành phố A,B,C,D,E sao cho hai thành phố bất kì trong chúng đều có đường nối với nhau. Sử dụng đồ thị để mô tả tình huống đó" (Cánh diều, tr.36).

1. Vẽ 5 điểm, mỗi điểm đại diện cho một thành phố A,B,C,D,E.
2. Nối mỗi cặp thành phố với một đoạn thẳng (đường nối), đảm bảo **tất cả các cặp đều được nối.**

Số cạnh của đồ thị hoàn chỉnh K_5:

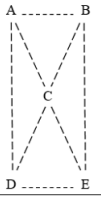
- Số cạnh là:

$$\binom{5}{2} = 10$$

tức là có 10 đường nối giữa các thành phố.

Minh họa:

Đây là sơ đồ thị đầy đủ với 5 đỉnh:



Hình 3: *Lời giải của ChatGPT*

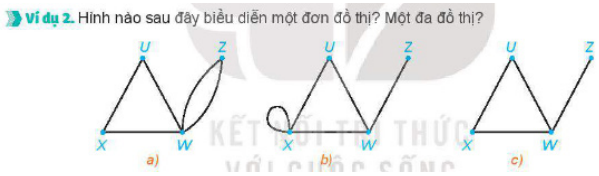
Có thể thấy, ChatGPT thể hiện năng lực tương đối tốt ở tầng suy luận kí hiệu khi xác lập đúng quy trình mô hình hóa đồ thị hoàn chỉnh và tính toán chính xác số cạnh bằng công thức tổ hợp $\binom{5}{2} = 10$ (xem Hình 3). Tuy nhiên, thực nghiệm cho thấy một sự "đứt gãy" hoàn toàn khi mô hình chuyển sang giai đoạn tạo đồ thị trực quan. Đồ thị ChatGPT đưa ra (*dạng ASCII*) không chỉ thiếu hụt về số lượng cạnh mà còn sai lệch về cấu trúc đồ thị, biến một đối tượng toán học chặt chẽ thành một biểu diễn hình học mang tính "ước lệ" và chắp vá.

Sâu xa hơn, sai phạm này phản ánh giới hạn của việc sử dụng kí hiệu tuyến tính (văn bản) để mô tả các cấu trúc phi tuyến tính (đồ thị). ChatGPT hiện tại chưa thực sự "vẽ" dựa trên các tiên đề toán học; thay vào đó, nó tái thiết lập các mẫu kí tự dựa trên xác suất phân phối từ ngữ.

Dưới góc độ sư phạm, lỗi L_2 dẫn đến sự mâu thuẫn giữa con số lí thuyết chính xác và hình ảnh minh họa sai lệch khiến người học rơi vào trạng thái hoang mang, làm suy giảm tính kiểm chứng trực quan vốn là giá trị cốt lõi của Lí thuyết đồ thị.

3.3.3. Lỗi sai 3: Nhận biết khái niệm

Mặc dù lỗi L_3 chiếm tỉ lệ thấp nhất (2,86%, tương ứng 3/105 nhiệm vụ) nhưng đây là loại lỗi nghiêm trọng nhất vì phản ánh trực tiếp sự hổng hổng về hệ thống tri thức chuyên ngành của mô hình. Trong chuyên đề Lí thuyết đồ thị Toán 11 với hệ thống khái



Hình 4: *Nội dung của Ví dụ 2 – Bài 8 – Sách Kết nối tri thức với cuộc sống*

Giải thích:

- **Đơn đồ thị** là đồ thị không có cạnh khép kín (loop) và không có hai đỉnh nối với nhau bằng nhiều cạnh (không có cạnh song song).
- **Đa đồ thị** có thể có cạnh song song hoặc cạnh khép kín.

Phân tích từng hình:

- **Hình a:** Có hai cạnh nối giữa W và Z → có cạnh song song → là đa đồ thị.
- **Hình b:** Có một cạnh nối đỉnh X và chính nó → có vòng khép kín (loop) → là đa đồ thị.
- **Hình c:** Mỗi cặp đỉnh chỉ nối bằng một cạnh, không có vòng khép kín → là đơn đồ thị.

Hình 5: *Lời giải của ChatGPT*

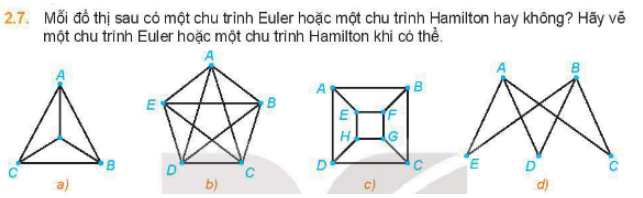
niệm trừu tượng và thuật ngữ khắt khe, ChatGPT bộc lộ rõ giới hạn về khả năng "thấu hiểu" bản chất toán học. Cụ thể, tại nhiệm vụ ở Hình 4.

Lời giải do ChatGPT đưa ra cho thấy mô hình đã phát biểu sai về khái niệm đa đồ thị (xem Hình 5). Cụ thể, theo định nghĩa, đa đồ thị là đồ thị không có khuyên và cho phép tồn tại nhiều cạnh nối giữa hai đỉnh. Tuy nhiên, ChatGPT lại cho rằng đa đồ thị có thể chứa khuyên (được mô tả là "cạnh khép kín"), dẫn đến sự sai lệch về trong quá trình thực hiện nhiệm vụ. Điều này chứng minh rằng, dù có khả năng xử lí ngôn ngữ ưu việt, ChatGPT vẫn chưa hoàn toàn làm chủ được các đơn vị tri thức chuyên sâu trong Chương trình Giáo dục phổ thông 2018, đòi hỏi sự thẩm định kĩ lưỡng từ phía người dạy khi sử dụng.

3.3.4. Lỗi sai 4: Đề xuất ví dụ

Dữ liệu thực nghiệm cho thấy một hạn chế đáng chú ý của ChatGPT trong việc đề xuất các ví dụ minh họa dựa trên tri thức lí thuyết đã cho. Mặc dù mô hình thể hiện khả năng phân tích yêu cầu nhiệm vụ khá tốt, nhưng lại gặp khó khăn trong giai đoạn cụ thể hóa tri thức thành các tình huống thực tế hoặc mô hình toán học tương ứng. Theo thống kê, lỗi L_4 chiếm tỉ lệ khiêm tốn 5,71% (6/105 nhiệm vụ) nhưng vẫn gây ảnh hưởng trực tiếp đến độ tin cậy của kết quả đầu ra. Cụ thể, trong bài tập tại Hình 6.

Xét kết quả ở câu b), ChatGPT đã thực hiện đầy đủ các bước kiểm tra sự tồn tại chu trình Euler của đồ thị cho trước; tuy nhiên, ví dụ được đề xuất lại không hợp lệ. Một số kết quả tương tự cũng xuất hiện ở vài bài tập khác (xem Hình 7). Điều này cho thấy khả năng vận dụng sáng tạo của AI vẫn còn khoảng cách xa so với năng lực giải quyết vấn đề



Hình 6: *Nội dung của Bài 2.7 – Bài 9 – Sách Kết nối tri thức với cuộc sống*

b)

- Đồ thị hoàn toàn K_5 (mọi đỉnh nối với mọi đỉnh khác)
- Số đỉnh: 5, mỗi đỉnh bậc 4 (chẵn)
- Tổng số cạnh: $\binom{5}{2} = 10$

Tất cả đỉnh đều bậc chẵn ⇒ Có thể có **chu trình Euler**

Đồ thị liên thông (vì là K_5)

⇒ **Có chu trình Euler**

Đồng thời vì là K_5 (hoàn toàn), cũng **có chu trình Hamilton**

Có cả **chu trình Euler** và **chu trình Hamilton**

Ví dụ:

- **Euler**: $A-B-D-E-B-C-D-E-A-A$
- **Hamilton**: $A-B-C-D-E-A$

Hình 7: *Lời giải của ChatGPT*

Toán học của con người, đặc biệt là trong việc kiến tạo dữ liệu mới từ các nguyên lí sẵn có.

3.3.5. Lỗi sai 5: Suy luận giải thích

Bài tập 5 - Bài 1 là nhiệm vụ tiêu biểu minh họa lỗi L_5 có nội dung như sau: "Một cuộc họp có 06 người tham dự. Hai người bất kì trong họ hoặc quen nhau hoặc không quen nhau. Chứng minh rằng có 03 người trong 06 người đó đôi một quen nhau hoặc đôi một không quen nhau" (Cánh diều, tr.43).

ChatGPT giải bài toán qua bốn bước, trong đó hai bước đầu được thực hiện chính xác. Ở bước 1, mô hình chuyển bài toán sang đồ thị bằng cách biểu diễn 06 người thành 06 đỉnh và mã hóa quan hệ bằng hai màu cạnh, đồng thời xác định đúng mục tiêu tìm tam giác đơn sắc. Ở Bước 2, khi xét đỉnh A có bậc 5, ChatGPT sử dụng giả thiết "không mất tính tổng quát" để suy ra A có ít nhất 3 cạnh cùng màu nối với B, C, D, cho thấy khả năng lập luận hợp lí (xem Hình 8).

Bước 1: Mô hình hóa bài toán

- Xem 6 người là 6 đỉnh của một đồ thị.
- Nối 2 đỉnh bằng **cạnh màu đỏ** nếu 2 người đó **quen nhau**.
- Nối 2 đỉnh bằng **cạnh màu xanh** nếu 2 người đó **không quen nhau**.
- Mỗi cặp người sẽ được nối bằng đúng **một** cạnh, hoặc đỏ hoặc xanh.

Mục tiêu: Chứng minh tồn tại **tam giác đơn sắc** (ba đỉnh mà tất cả các cạnh giữa chúng cùng màu).

Bước 2: Sử dụng nguyên lý Dirichlet (nguyên lý cái hộp)

Chọn một người, gọi là A.

A có 5 mối quan hệ với 5 người còn lại → tương ứng 5 cạnh xuất phát từ A, mỗi cạnh hoặc đỏ hoặc xanh.

Theo nguyên lý Dirichlet:

Trong 5 cạnh này, **ít nhất** sẽ có 3 cạnh cùng màu.

Không mất tính tổng quát, giả sử A quen **ít nhất 3 người**: B, C, D (cạnh màu đỏ)

Hình 8: *Lời giải của ChatGPT ở Bước 1 và Bước 2*

 https://doi.org/10.15625/2615-8957/1260911

Sai sót nghiêm trọng xuất hiện ở Bước 3 khi ChatGPT đưa ra kết luận không hợp lệ trong xử lí các cạnh khác màu (xem Hình 9). Từ giả thiết A – B, A – C, A – D cùng màu, mô hình suy luận sai rằng một cạnh khác màu giữa B, C, D có thể tạo tam giác đơn sắc với A. Thực tế, cấu trúc này không đảm bảo tính đồng nhất màu. Lỗi cho thấy mô hình không kiểm soát chặt điều kiện logic khi suy luận. Bên cạnh đó, ChatGPT còn lập luận chưa tối ưu khi yêu cầu nhiều cạnh cùng màu, trong khi chỉ cần một cạnh phù hợp là đủ để tạo tam giác đơn sắc. Điều này phản ánh hạn chế trong việc tối giản điều kiện logic.

Bước 3: Xét nhóm 3 người *B*, *C*, *D*

- Nếu **B quen C** (cạnh đỏ), **B quen D** (đỏ), và **C quen D** (đỏ), thì *A*, *B*, *C* tạo thành **tam giác quen nhau** (tam giác đỏ) → thỏa mãn đề.
- Nếu **không phải tất cả** các cạnh giữa *B*, *C*, *D* đều đỏ, tức là **có ít nhất một cạnh xanh**.
  Giả sử *B* và *C* **không quen nhau** (cạnh xanh).
  Khi đó *B*, *C*, *D* sẽ tạo ra một tam giác toàn xanh (nếu các cạnh còn lại cũng xanh) hoặc trong quá trình xét các cạnh sẽ tìm được tam giác toàn xanh.

Cách diễn đạt chuẩn hơn:

Trong nhóm *B*, *C*, *D*:

- Nếu có một cạnh xanh, thì 2 đỉnh của cạnh đó cùng với *A* tạo ra tam giác toàn xanh (vì cả hai không quen nhau, nhưng đều quen *A* – cần điều chỉnh ở bước này).

Hình 9: *Lời giải của ChatGPT ở Bước 3*

Ở Bước 4, ChatGPT áp dụng đúng lập luận đối xứng để xét trường hợp còn lại. Tuy nhiên, kết luận không còn giá trị do đã sai ở Bước 3 (xem Hình 10). Lỗi trước gây hiệu ứng lan truyền, khiến lập luận sau không hợp lệ, dù hình thức vẫn tuân theo cấu trúc chứng minh quen thuộc.

Bước 4: Kết luận

Bằng lập luận đối xứng, ta thấy:

- Nếu A quen ít nhất 3 người → hoặc có tam giác đỏ, hoặc có tam giác xanh.
- Nếu A không quen ít nhất 3 người → lập luận tương tự cho tam giác xanh.

Do đó, **luôn tồn tại 3 người đôi một quen nhau hoặc đôi một không quen nhau**.

Hình 10: *Lời giải của ChatGPT ở Bước 4*

4. Thảo luận

Các dạng lỗi mang tính hệ thống được phát hiện trong nghiên cứu này có sự giao thoa mật thiết với những nhận định về giới hạn của AI trong các bối cảnh giáo dục tương tự.

Khả năng nhận diện hình ảnh của ChatGPT trong chủ đề này bộc lộ những điểm nghẽn đáng kể, thể hiện qua sai lệch phổ biến của lỗi L_1 (72,38%). Nguyên nhân bắt nguồn từ đặc thù chủ đề Lí thuyết đồ thị, nơi các nhiệm vụ chủ yếu dựa trên dữ liệu trực quan. Lỗi L_1 đã phơi bày một "điểm nghẽn" dù đã là mô hình đa phương thức, khả năng giải mã và chuyển đổi thông tin từ hình ảnh sang cấu trúc logic toán học của ChatGPT vẫn tồn tại những hạn chế. Kết quả này tương đồng với nghiên cứu của Polverini và Gregorcic (2024) khi khảo sát năng lực AI trên đồ thị động học. Các tác giả chỉ ra ChatGPT có thể giải quyết và duy trì mạch suy luận chính xác nhưng thất bại ở bước "nhìn" đồ thị, đặc biệt là các giá trị số trên trục tọa độ và mối tương quan không gian.

Khoảng cách giữa tư duy ngôn ngữ và khả năng tạo lập hình ảnh trực quan được thể hiện rõ qua sự đứt gãy giữa năng lực lập luận kí hiệu và tạo hình ảnh trực quan trong lỗi L_2. (Ví dụ: trong *Luyện tập 1*, mặc dù ChatGPT lập luận logic chính xác khi xác lập đúng quy trình mô hình hóa đồ thị hoàn chỉnh, nhưng không thể biểu diễn tri thức đó thành một đồ thị trực quan tương ứng). Kết quả này tương đồng với nghiên cứu của Fergus (2023) khi đánh giá ChatGPT trong lĩnh vực Khoa học tự nhiên. Tác giả khẳng định AI có thể nắm vững từ vựng của đối tượng nhưng lại thiếu hụt tư duy không gian để biểu diễn cấu trúc chính xác. Theo Fergus (2023), ChatGPT mô tả lí thuyết trôi chảy hơn là xử lí các yêu cầu liên quan đến thông tin phi văn bản. Sự sai lệch đến từ bản chất dựa trên văn bản, không sở hữu hệ thống tri giác độc lập để kiểm chứng hình ảnh của mô hình. Đặc biệt, tính chất biến thiên của các phản hồi sai được Fergus (2023) nhấn mạnh cũng xuất hiện trong nghiên cứu này khi cùng một yêu cầu nhưng mô hình sinh ra các phiên bản lỗi khác nhau, cho thấy các hình vẽ minh họa của AI là sự sắp xếp kí tự, không dựa trên các quy tắc hình học hay logic không gian cố định.

Nhận thức của ChatGPT đối với các định nghĩa trong chủ đề này bộc lộ những hạn chế đáng chú ý. Mặc dù lỗi L_3 xuất hiện với tần suất thấp nhưng phản ánh mức độ biểu diễn tri thức nền tảng của mô hình. ChatGPT gặp khó khăn trong việc phân biệt các khái niệm (Ví dụ: Đa đồ thị). Sai lệch này cho thấy mô hình dễ nhầm lẫn giữa những khái niệm tương đồng về thuật ngữ nhưng khác biệt về bản chất, từ đó làm giảm độ chính xác trong lập luận và ảnh hưởng đến quá trình giải quyết. Theo Kortemeyer (2023), kiến thức của mô hình thiếu tính ổn định, nhầm lẫn các khái niệm học thuật cơ bản và trình độ thực tế chỉ tương đương người mới học. Hiện tượng "ảo giác khái niệm" khiến ChatGPT trình bày lập luận có vẻ hợp lí về ngôn ngữ nhưng lại thiếu nền tảng vững chắc.

Hạn chế về khả năng suy luận giải thích của ChatGPT trong chủ đề này được thể hiện qua việc lỗi L_5 xuất hiện phổ biến thứ hai chỉ sau lỗi L_1. Nguyên nhân do chủ đề chứa đựng nhiều nhiệm vụ suy luận ở mức độ phức tạp cao. Thực trạng này cho thấy hiệu suất của ChatGPT có mối liên hệ mật thiết với độ phức tạp của câu hỏi. Cụ thể, mô hình thường đạt độ chính xác cao ở các yêu cầu nhận diện hoặc áp dụng kiến thức trực tiếp, nhưng lại bộc lộ nhiều hạn chế khi giải quyết các bài toán nhiều bước. Đây là

minh chứng cho sự thiếu ổn định của ChatGPT trong việc duy trì chuỗi suy luận dài và đảm bảo tính nhất quán xuyên suốt lời giải, đặc biệt là các nhiệm vụ đòi hỏi lập luận chặt chẽ. Kết quả này tương đồng với các nhận định của Giannos & Delardas (2023) cũng như Lê Anh Vinh và cộng sự (2023). Các tác giả đều thống nhất rằng hiệu suất của ChatGPT chịu ảnh hưởng bởi độ phức tạp của nhiệm vụ. Khi đối mặt với các bài toán mang tính lập luận sâu, mô hình bộc lộ những sai lệch mang tính hệ thống. Sự suy giảm hiệu năng này khẳng định rằng, mặc dù ChatGPT có khả năng xử lí thông tin nhanh, nhưng việc giải quyết các bài toán phức tạp vẫn là thách thức lớn hiện nay.

5. Kết luận

Nghiên cứu này kết hợp phương pháp định lượng và định tính nhằm đánh giá hiệu suất giải toán của ChatGPT đối với các bài tập Lí thuyết đồ thị trong Chương trình Toán 11 theo Chương trình Giáo dục Phổ thông 2018. Kết quả cho thấy, năng lực giải quyết vấn đề của ChatGPT 4.5 ở mức độ trung bình khá với tỉ lệ chính xác 59,14%. Trong đó, lỗi nhận diện hình ảnh (L1) chiếm tỉ trọng áp đảo 72,38% phản ánh giới hạn khi xử lí các dữ liệu trực quan. Đặc biệt, nghiên cứu đã phân tích một biến số ít được khai thác ở các thực nghiệm trước đây là lỗi L4 tạo điểm nhấn mới trong phân loại lỗi sai của ChatGPT.

Về phương diện lí luận, thông qua hệ thống phân tích lỗi, nghiên cứu đã xác lập các minh chứng cho việc phân biệt giữa năng lực xử lí ngôn ngữ và tư duy logic hình thức - hai bình diện vốn thường bị đồng nhất trong các đánh giá cảm tính về AI. Về phương diện thực tiễn, nghiên cứu đề xuất tái định vị vai trò của ChatGPT trong lớp học không phải như công cụ thay thế tư duy, mà như phương tiện gợi mở và hỗ trợ kiến tạo tri thức dưới sự dẫn dắt của giáo viên. Theo đó, thiết kế dạy học cần ưu tiên các nhiệm vụ phân tích cấu trúc sâu và giải quyết vấn đề phức hợp, vốn là lĩnh vực AI còn bộc lộ nhiều hạn chế và cũng là nơi tư duy toán học đích thực được hình thành.

Tài liệu tham khảo

Díaz, B. & Nussbaum, M. (2024). Artificial intelligence for teaching and learning in schools: The need for pedagogical intelligence. *Computers & Education*, *217*, 105071. https://doi.org/10.1016/j.compedu.2024.105071.

Fergus, S., Botha, M. & Ostovar, M. (2023). Evaluating academic answers generated using ChatGPT. *Journal of Chemical Education*, *100*(4), pp.1672-1675. https://doi.org/10.1021/acs.jchemed.3c00087.

Gandolfi, A. (2025). GPT-4 in education: Evaluating aptness, reliability, and loss of coherence in solving calculus problems and grading submissions. *International Journal of Artificial Intelligence in Education*, *35*(1), pp.367-397. https://doi.org/10.1007/s40593-024-00403-3.

Giannos, P. & Delardas, O. (2023). Performance of ChatGPT on UK standardized admission tests: insights from the BMAT, TMUA, LNAT, and TSA examinations. *JMIR Medical Education*, *9*(1), Article e47737. https://doi.org/10.2196/47737.

Huang, L., Yu, W., Ma, W., Zhong, W., Feng, Z., Wang, H., ... & Liu, T. (2025). A survey on hallucination in large language models: Principles, taxonomy, challenges, and open questions. *ACM Transactions on Information Systems*, *43*(2), pp.1-55. https://doi.org/10.1145/3703155.

Kaya, D. & Yavuz, S. (2025). Can generative aI and ChatGPT break human supremacy in mathematics and reshape competence in cognitive-demanding problem-solving tasks?. *Journal of Intelligence*, *13*(4), p.43.

Kortemeyer, G. (2023). Could an artificial-intelligence agent pass an introductory physics course?. *Physical Review Physics Education Research*, *19*(1), p.010132. https://doi.org/10.1103/PhysRevPhysEducRes.19.010132.

Lê Anh Vinh, Bùi Thị Diễn, Lê Quang Quân & Vũ Văn Luân. (2023). Khả năng thực hiện bài kiểm tra định kì môn Toán và môn Ngữ văn cấp Trung học của công cụ ChatGPT: Kết quả nghiên cứu và một số khuyến nghị ban đầu. *Tạp chí Khoa học Giáo dục Việt Nam*, *19*(2), tr.1-10.

Le Thai Bao Thien Trung, Nguyen Minh Dat, Tang Minh Dung, Trinh Van Thanh & Nguyen Minh Nhut. (2025). Promoting teachers' use of ChatGPT: A case study on generating real-world problems in 10th grade algebra instruction. *HCMUE Journal of Science*, *22*(3), pp.500-511.

Oh, S. (2024). Evaluating Mathematical Problem-Solving Abilities of Generative AI Models: Performance Analysis of o1-preview and gpt-4o Using the Korean College Scholastic Ability Test. *IEEE Access*, *13*, pp.1227-1235.

Pardos, Z. A. & Bhandari, S. (2024). ChatGPT-generated help produces learning gains equivalent to human tutor-authored help on mathematics skills. *PLOS ONE*, *19*(5), e0304013. https://doi.org/10.1371/journal.pone.0304013.

Phạm Tất Thành. (2024). Đánh giá kết quả phản hồi của Chatbot Bard trong thực hiện bài thi môn Địa lí kì thi tốt nghiệp Trung học Phổ thông Quốc gia Việt Nam từ năm 2019 đến năm 2023. *Tạp chí Giáo dục, 24*(4), tr.35-40.

Polverini, G. & Gregorcic, B. (2024). Performance of ChatGPT on the test of understanding graphs in kinematics. *Physical Review Physics Education Research, 20*(1), 010109. *https://doi.org/10.1103/PhysRevPhysEducRes.20.010109.*

 https://doi.org/10.15625/2615-8957/1260911