

ĐÁNH GIÁ HIỆU SUẤT GIẢI TOÁN XÁC SUẤT CỦA CHATGPT (PHIÊN BẢN GPT-5) DỰA TRÊN CÁC BỘ SÁCH TOÁN 12 THEO CHƯƠNG TRÌNH 2018

EVALUATING THE MATHEMATICAL PROBLEM-SOLVING PERFORMANCE OF CHATGPT (GPT-5) ON PROBABILITY TOPICS BASED ON GRADE 12 MATHEMATICS TEXTBOOKS UNDER THE 2018 CURRICULUM

Phạm Thành Đạt,
Lê Thái Bảo Thiên Trung,
Lê Hoàng Minh Quân⁺,
Hoàng Thị Minh Thư

Trường Đại học Sư phạm Thành phố Hồ Chí Minh
+ Tác giả liên hệ • Email: lehoangminhquan03@gmail.com

Article history

Received: 28/01/2026

Accepted: 16/3/2026

Published: 29/4/2026

Keywords

ChatGPT, probability,
mathematical reasoning,
problem solving,
performance, large language
models

ABSTRACT

Conditional probability is a central topic in the Grade 12 Mathematics curriculum, demanding a high level of logical reasoning. Amid the rapid advancement of large language models in education, this study evaluates the accuracy and reasoning quality of GPT-5 on Grade 12 probability problems through a mixed-methods approach - combining quantitative and qualitative analysis - applied to a dataset of 513 problems drawn from current Grade 12 Mathematics textbooks. Accuracy rates were statistically quantified, while qualitative analysis identified characteristic reasoning patterns and error types. The findings indicate that GPT-5 correctly solved 84.01% of the textbook problems. The model demonstrated proficiency in formula application but exhibited notable difficulties in reasoning, particularly in tasks requiring the construction and interpretation of tree diagrams. These findings provide teachers with empirical evidence regarding the reliability of GPT-5 as a instructional tool. The study recommends extending the evaluation to other mathematical topics, investigating student-ChatGPT interactions in authentic learning contexts, and conducting controlled experimental studies to establish causal evidence for the pedagogical integration of AI in mathematics education.

1. Mở đầu

Sự phát triển nhanh chóng của các mô hình ngôn ngữ lớn (Large Language Models - LLMs) trong những năm gần đây đã tạo ra chuyển biến đáng kể trong lĩnh vực giáo dục toán học. Trong giai đoạn trước, các hệ thống tính toán kí hiệu như Maple, Mathematica hay WolframAlpha đã được ứng dụng rộng rãi nhờ khả năng xử lí biểu thức toán học với độ chính xác cao, tuy nhiên, những hệ thống này hoạt động theo cơ chế thực thi quy tắc cố định, thiếu khả năng giải thích bằng ngôn ngữ tự nhiên và định hướng tư duy cho người học (Kasneji và cộng sự, 2023). Sự xuất hiện của ChatGPT đánh dấu một bước ngoặt: các hệ thống AI không chỉ cung cấp kết quả mà còn có khả năng trình bày quá trình lập luận theo cách gần với diễn đạt tự nhiên của con người, mở ra tiềm năng ứng dụng đáng kể trong hỗ trợ học tập toán học. Nhiều nghiên cứu đã chỉ ra rằng năng lực toán học của LLMs vẫn còn những giới hạn quan trọng và mang tính hệ thống. Chúng thường gặp khó khăn khi giải các bài toán đòi hỏi lập luận nhiều bước, đặc biệt là các bài toán có lời văn, hơn thế độ chính xác của những mô hình suy giảm rõ rệt khi chuyển từ các bài toán cơ bản sang nâng cao (Hendrycks và cộng sự, 2021). Thông qua bộ dữ liệu MATH gồm 12.500 bài toán thi đấu, Hendrycks và cộng sự (2021) đã chứng minh rằng ngay cả các mô hình Transformer lớn cũng chỉ đạt độ chính xác từ 3% đến 6,9% trên toàn bộ tập dữ liệu, trong khi mức cao nhất ở phân bậc dễ nhất cũng chỉ đạt 15% - cho thấy khoảng cách lớn giữa quy mô mô hình và năng lực lập luận toán học thực sự. Frieder và cộng sự (2024) mở rộng đánh giá lên cấp độ đại học và sau đại học, ghi nhận rằng ChatGPT có thể được dùng hiệu quả như một công cụ tra cứu tri thức toán học và hỗ trợ ở mức đại cương nhưng thất bại rõ rệt ở các bài toán đòi hỏi chứng minh và tính toán kí hiệu phức tạp ở bậc nâng cao. Ở bậc phổ thông, Pardos và Bhandari (2024) đã chứng minh rằng gợi ý được tạo bởi ChatGPT mang lại kết quả học tập tương đương với gợi ý do gia sư người thật soạn thảo trên bốn chủ đề toán học, song cũng nhấn mạnh rằng hiệu quả này phụ thuộc lớn vào mức độ phù hợp của nội dung với cấu trúc bài toán. Nhiều nghiên cứu ở Việt Nam, cũng đã chỉ ra rằng năng lực toán học của LLMs vẫn còn tồn tại những hạn chế đáng

kê. Nghiên cứu của Lê Anh Vinh và cộng sự (2023) chỉ ra rằng khả năng thực hiện của ChatGPT đối với bài kiểm tra năng lực Toán và Ngữ Văn của HS hai khối lớp 9 và 12 đạt mức trung bình (từ 5,7 đến 6,8 điểm). Kết quả nghiên cứu của Đào Xuân Quy và cộng sự (2024) cũng cho thấy năng lực giải toán của ChatGPT có sự khác biệt rõ rệt theo từng chủ đề, trong đó mô hình chủ yếu xử lý tốt các câu hỏi ở mức độ nhận biết.

Tuy vậy, trong lĩnh vực xác suất, những nội dung đòi hỏi mô hình hóa cấu trúc toán học và lí luận đa tầng, các mô hình LLMs thể hiện các lỗi suy luận đặc trưng và có xu hướng bị ảnh hưởng bởi thiên kiến nhận thức tương tự con người trong quá trình xử lý thông tin không đầy đủ (Hagendorff và cộng sự, 2023). Những hạn chế trên phản ánh một đặc điểm cốt lõi của LLMs: các mô hình này hoạt động chủ yếu dựa trên cơ chế tái tạo mẫu thống kê từ dữ liệu huấn luyện thay vì nắm bắt cấu trúc khái niệm toán học theo nghĩa thực sự (Hendrycks và cộng sự, 2021; Frieder và cộng sự, 2024). Trong bối cảnh đó, nghiên cứu ứng dụng ChatGPT trong dạy học toán học đang phát triển nhanh chóng, song tập trung chủ yếu ở hai hướng: khảo sát thái độ, nhận thức của GV và HS và đánh giá hiệu suất tổng quát trên các bộ kiểm tra chuẩn hóa đa môn. Kasneci và cộng sự (2023) lập luận rằng ChatGPT có tiềm năng đóng vai trò hệ thống hỗ trợ học tập thích ứng, cung cấp phản hồi từng bước theo tiến trình giải quyết vấn đề, một đặc tính mang giá trị sư phạm đáng kể. Từ một góc nhìn thận trọng hơn, từ nghiên cứu của Kasneci và cộng sự (2023) cũng đã cảnh báo nguy cơ lạm dụng các công cụ AI trong học tập có thể làm suy giảm khả năng tư duy độc lập của người học và đặt ra những thách thức nghiêm trọng cho vấn đề liêm chính học thuật trong môi trường giáo dục số.

Từ tổng quan trên, có thể nhận thấy “khoảng trống” nghiên cứu hiện nay nằm ở sự thiếu vắng các công trình đánh giá chuyên sâu hiệu suất của ChatGPT trên một chủ đề toán học cụ thể trong bối cảnh chương trình giáo dục phổ thông Việt Nam. Trong khi các nghiên cứu hiện có tập trung vào đánh giá tổng thể hoặc so sánh đa môn, việc phân tích chi tiết chất lượng lập luận của mô hình trong chủ đề Xác suất vốn đòi hỏi sự kết hợp giữa tư duy tổ hợp, lí luận điều kiện và mô hình hóa ngữ cảnh vẫn chưa được nghiên cứu một cách hệ thống. Trong bối cảnh trên, việc ChatGPT được nghiên cứu như một công cụ tiềm năng hỗ trợ quá trình dạy học là cần thiết. Cụ thể, xem xét ChatGPT giải quyết các bài toán xác suất trong SGK Toán 12 theo Chương trình giáo dục phổ thông 2018 một cách chính xác về kết quả và phù hợp về phương pháp giải trong chất lượng lập luận giúp mang lại giá trị sư phạm trong việc hỗ trợ học tập đối với các em HS ở giai đoạn cuối của quá trình đào tạo phổ thông. Những thực tế này thúc đẩy chúng tôi đặt ra câu hỏi nghiên cứu: Độ chính xác và chất lượng lập luận của ChatGPT (phiên bản GPT-5) trong việc giải các bài toán thuộc chủ đề Xác suất ở lớp 12 theo Chương trình giáo dục phổ thông 2018 là như thế nào?

2. Kết quả nghiên cứu

2.1. Phương pháp nghiên cứu

Để trả lời cho câu hỏi nghiên cứu trên chúng tôi sử dụng kết hợp phương pháp nghiên cứu định lượng và định tính để có thể đánh giá hiệu suất giải toán của ChatGPT (phiên bản GPT-5). Chúng tôi tổng hợp tất cả các bài toán thuộc chủ đề Xác suất trong bốn bộ sách Toán hiện hành. Sau đó, sẽ chuẩn hóa chúng theo định dạng tex nhằm đảm bảo tính chính xác về mặt kí hiệu toán học trong câu hỏi. Đồng thời, các câu trả lời nhận được từ GPT-5 cũng sẽ được chúng tôi yêu cầu để xuất hiện dưới dạng tex, cụ thể như sau: “*Tôi là một HS lớp 12 ở Việt Nam, hãy giải bài toán này cho tôi: [Bài toán] và xuất câu trả lời thành định dạng tex*”. Từ kết quả thống kê được, chúng tôi tiến hành phân tích lời giải từ những câu trả lời mà GPT-5 đưa ra và đánh giá mức độ hợp lí của lời giải, thông qua việc đối sánh với lời giải trong SGK và sách GV để kiểm chứng. Chất lượng lập luận trong lời giải của GPT-5 được phân tích về độ phù hợp, tính logic và đầy đủ đối với đối tượng là HS trung học phổ thông. Dựa trên cơ sở đó, chúng tôi tiến hành phân tích định tính để có được kết quả sau cùng, từ đó đánh giá câu trả lời của GPT-5.

Phương pháp nghiên cứu định lượng và định tính được chúng tôi kết hợp sử dụng để có thể thống kê lại kết quả giải toán của GPT-5. Từ kết quả thống kê được, chúng tôi tiến hành phân tích lời giải từ những câu trả lời mà GPT-5 đưa ra và đánh giá mức độ hợp lí của lời giải. Tất cả câu trả lời đều được đánh giá để có thể xác định được các lỗi sai của GPT-5 trong giải toán và từ đó kiểm tra lại kết quả thống kê.

Chúng tôi tiến hành thử nghiệm GPT-5 trên tổng số 513 nhiệm vụ thuộc chủ đề Xác suất trong cả bốn bộ SGK Toán lớp 12 hiện hành. Ở đây chúng tôi quan tâm đến các nhiệm vụ xuất hiện trong một bài tập, có nghĩa rằng một bài tập có thể có nhiều nhiệm vụ. Ví dụ, như trong *bài tập 6.3 (trang 70 trong sách Kết nối tri thức với cuộc sống)* xác định có 2 nhiệm vụ. Nhiệm vụ thứ nhất là “*Tính xác suất để tổng số chấm xuất hiện trên hai con xúc xắc bằng 7 biết rằng ít nhất có một con xúc xắc xuất hiện mặt 5 chấm*”; nhiệm vụ thứ hai là “*Tính xác suất để có ít nhất có một con xúc xắc xuất hiện mặt 5 chấm nếu biết rằng tổng số chấm xuất hiện trên hai con xúc xắc bằng 7*”.

6.3. Gieo hai con xúc xắc cân đối, đồng chất. Tính xác suất để:

- a) Tổng số chấm xuất hiện trên hai con xúc xắc bằng 7 nếu biết rằng ít nhất có một con xúc xắc xuất hiện mặt 5 chấm;
- b) Có ít nhất có một con xúc xắc xuất hiện mặt 5 chấm nếu biết rằng tổng số chấm xuất hiện trên hai con xúc xắc bằng 7.

Hình 1. SGK Kết nối tri thức với cuộc sống, bài 18, bài tập 6.3, trang 70

Từ 513 nhiệm vụ đã tổng hợp chúng tôi đã thống kê cụ thể các nhiệm vụ như sau:

Bảng 1. Thống kê số nhiệm vụ theo bài học của các bộ sách

		Cánh điều	Kết nối tri thức với cuộc sống	Cùng khám phá	Chân trời sáng tạo
Bài 1	SGK	22	31	26	35
	SBT	36	15	18	29
Bài 2	SGK	24	30	25	24
	SBT	25	9	18	15
Bài tập cuối chương	SGK	9	13	13	17
	SBT	11	16	20	30
Tổng		127	114	120	152

Ghi chú: Ở bộ sách Kết nối tri thức với cuộc sống, Bài 1 và Bài 2 tương ứng với Bài 18 và Bài 19.

Lời giải do GPT-5 tạo ra được nhóm nghiên cứu đánh giá theo bốn tiêu chí: kết quả đúng với lời giải hợp lý, kết quả đúng với lời giải không hợp lý, kết quả sai với lời giải hợp lý và kết quả sai với lời giải không hợp lý. Các phân tích được dựa trên lời giải trong sách GV và sách bài tập ban hành theo từng bộ SGK. Kết quả cho thấy có 435 nhiệm vụ được giải đúng với lời giải hợp lý, 34 nhiệm vụ có kết quả đúng nhưng lời giải chưa hợp lý, 4 nhiệm vụ có kết quả sai nhưng lời giải có vẻ hợp lý và 40 nhiệm vụ có kết quả sai kèm theo lời giải không hợp lý.

Bảng 2. Đánh giá hiệu suất giải toán về chủ đề Xác suất của ChatGPT trong 4 bộ sách hiện hành

Số lượng câu hỏi		Đúng		Sai	
		Hợp lý	Không hợp lý	Hợp lý	Không hợp lý
Cánh điều	127	118 (92,9%)	0 (0%)	1 (0,8%)	8 (6,3%)
Kết nối tri thức với cuộc sống	114	81 (71,1%)	22 (19,3%)	2 (1,7%)	9 (7,9%)
Cùng khám phá	120	110 (91,7%)	3 (2,5%)	0 (0%)	7 (5,8%)
Chân trời sáng tạo	152	123 (80,9%)	9 (5,9%)	1 (0,7%)	19 (12,5%)

Có thể thấy rằng tỉ lệ giải quyết đúng các bài toán của GPT-5 chiếm tỉ trọng lớn, tuy nhiên lại không đồng đều giữa các bộ sách. Những kết quả này cho thấy GPT-5 chưa thể giải quyết một cách đầy đủ và ổn định mọi dạng bài toán thuộc chủ đề Xác suất.

2.2. Phân tích kết quả của từng bộ sách

2.2.1. Sách Cánh điều

Nhìn chung, GPT-5 giải quyết khá tốt các bài tập mang tính tổng hợp và các bài tập trong Sách bài tập. Rõ ràng, ở phần Bài tập cuối chương, GPT-5 đạt độ chính xác tuyệt đối 100%. Bên cạnh đó, các lỗi sai của GPT-5 thường rơi vào các bài tập luyện tập sau hoạt động hình thành kiến thức. Tỉ lệ trả lời đúng lần lượt là 89,7% và 93,9% tương ứng với Bài 1 và Bài 2 ở cả 2 sách. Lí giải cho điều này là các bài toán luyện tập thường yêu cầu HS phải tự mô hình hóa từ một tình huống thực tế sang ngôn ngữ toán học và GPT-5 đã thất bại ở bước chuyển đổi này hơn là sai ở bước tính toán công thức. Qua đây cho ta thấy khả năng giải các bài toán chủ đề Xác suất của GPT-5 đã phản ánh một quy luật học tập: thành thạo kĩ năng để đạt được hơn là thấu hiểu bản chất. GPT-5 đã thành thạo kĩ năng tính toán và áp dụng công thức, tuy nhiên vẫn đang chật vật học cách hiểu thế giới thực để đưa vào toán học.

Bảng 3. Đánh giá hiệu suất giải toán về chủ đề Xác suất của ChatGPT trong sách Cánh điều

	Số lượng	Đúng		Sai		
		Hợp lý	Không hợp lý	Hợp lý	Không hợp lý	
Bài 1	SGK	22	19 (86,36%)	0 (0%)	0 (0%)	3 (13,64%)
	SBT	36	33 (91,67%)	0 (0%)	0 (0%)	3 (8,33%)
Bài 2	SGK	24	22 (91,67%)	0 (0%)	1 (4,17%)	1 (4,17%)
	SBT	25	24 (96%)	0 (0%)	0 (0%)	1 (4%)

Bài tập cuối chương	SGK	9	9 (100%)	0 (0%)	0 (0%)	0 (0%)
	SBT	11	11 (100%)	0 (0%)	0 (0%)	0 (0%)

2.2.2. Sách Kết nối tri thức với cuộc sống

Dựa vào việc thống kê bộ sách Kết nối tri thức và cuộc sống có thể thấy rằng hiệu suất giải quyết chính xác hoàn toàn bài toán nằm ở mức ổn định (từ 53,% đến 77,7%). Điều này phản ánh được khả năng của GPT-5 trong việc giải quyết các bài toán xác suất ở lớp 12. Tuy nhiên, lời giải của GPT-5 còn chưa hợp lý nhất là những nội dung hình thành kiến thức ở bài 18 và 19 trong SGK, cho thấy rằng GPT-5 chưa thể thay thế được vai trò của GV trong việc hình thành kiến thức cho HS về xác suất, và lập luận của GPT-5 còn nhiều hạn chế, chưa phù hợp cho HS THPT.

Bảng 4. Đánh giá hiệu suất giải toán về chủ đề Xác suất của ChatGPT trong sách Kết nối tri thức với cuộc sống

		Số lượng	Đúng		Sai	
			Hợp lý	Không hợp lý	Hợp lý	Không hợp lý
Bài 18	SGK	31	23(74,2%)	4 (12,9%)	0 (0%)	4 (12,9%)
	SBT	15	11 (73,3%)	3 (20%)	0 (0%)	1 (6,7%)
Bài 19	SGK	30	21(70%)	4 (13,3%)	1 (3,4%)	4 (13,3%)
	SBT	9	7 (77,7%)	2 (22,2%)	0 (0%)	0 (0%)
Bài tập cuối chương	SGK	13	7 (53,8%)	5 (38,5%)	1 (7,7%)	0 (0%)
	SBT	16	11 (68,8%)	5 (31,2%)	0 (0%)	0 (0%)

2.2.3. Sách Cùng khám phá

Từ kết quả phân tích cho thấy hiệu suất thực hiện các hoạt động, ví dụ minh họa và bài tập về chủ đề Xác suất của GPT-5 trong sách Cùng khám phá khá cao. Cụ thể, tỉ lệ câu trả lời đúng và lập luận hợp lý lần lượt là 76,92%, 92%, 94,44% và 100% tương ứng với bài 1, 2 và bài cuối chương trong SGK và sách bài tập. Các con số này cho thấy phần lớn yêu cầu đặt ra trong từng bài toán về chủ đề Xác suất khá chính xác, đầy đủ và hợp lý. Khi xét trên toàn bộ chủ đề, tỉ lệ câu trả lời đúng và lập luận hợp lý đạt khoảng 90,84% trong đó tỉ lệ câu trả lời đúng không hợp lý là 11,54% và câu trả lời sai và không hợp lý là 8,37%. Điều đó cho thấy GPT-5 có khả năng xử lý khá hiệu quả các dạng nhiệm vụ liên quan đến Xác suất trong chương trình Toán lớp 12, trong so sánh kết quả giữa các bài cho thấy có sự dao động nhất định trong hiệu suất. Cụ thể, GPT-5 thể hiện sự tích lũy kinh nghiệm dần qua các bài học. Các bài tập 1 và 2 trong sách bài tập ghi nhận tỉ lệ đúng và hợp lý ở mức tuyệt đối, đạt 100%; trong khi bài 1 và 2 ở SGK ghi nhận kết quả tương đối lần lượt với 76,92%, 94,44%. Bài tập cuối chương trong SGK đạt tỉ lệ đúng, hợp lý 92% và tỉ lệ đã được tăng lên ở bài tập cuối chương trong sách bài tập là 94,44%. Khoảng chênh lệch giữa các bài nằm ở mức 10,85%, phản ánh một xu hướng tích cực: GPT-5 đặc biệt hiệu quả ở các bài tập mang tính tổng hợp và ôn tập cuối chương. Đáng chú ý, tất cả các bài đều đạt tỉ lệ đúng trên 75%, qua đó cho thấy tính ổn định cũng như hiệu quả cao của GPT-5 khi được sử dụng để giải quyết các bài toán thuộc chủ đề Xác suất trong bối cảnh dạy học Toán ở trường phổ thông.

Bảng 5. Đánh giá hiệu suất giải toán về chủ đề Xác suất của ChatGPT trong sách Cùng khám phá

		Số lượng	Đúng		Sai	
			Hợp lý	Không hợp lý	Hợp lý	Không hợp lý
Bài 1	SGK	26	20 (76,92%)	3 (11,54%)	0 (0%)	3 (11,54%)
	SBT	18	17 (94,44%)	0 (0%)	0 (0%)	1 (5,56%)
Bài 2	SGK	25	23 (92%)	0 (0%)	0 (0%)	2 (8%)
	SBT	18	18 (100%)	0 (0%)	0 (0%)	0 (0%)
Bài tập cuối chương	SGK	13	13 (100%)	0 (0%)	0 (0%)	0 (0%)
	SBT	20	19 (94,44%)	0 (0%)	0 (0%)	1 (5,56%)

2.2.4. Sách Chân trời sáng tạo

Kết quả phân tích cho thấy GPT-5 thể hiện mức độ hoàn thành các hoạt động và bài tập trong chủ đề Xác suất và Thống kê của sách Chân trời sáng tạo ở mức khá tốt. Cụ thể, tỉ lệ trả lời đúng và có lập luận hợp lý lần lượt là 76,56%, 76,92% và 100% tương ứng với bài 1, bài 2 và các bài tập cuối chương (tổng hợp từ cả SGK và Sách bài tập). Các con số này cho thấy phần lớn các yêu cầu đặt ra trong từng bài đã được GPT-5 đáp ứng một cách tương đối đầy đủ và chính xác. Khi xét trên toàn bộ chủ đề, tỉ lệ đúng trung bình đạt tới 84%, trong khi tỉ lệ câu trả lời sai hoặc có lập luận không hợp lý chỉ chiếm 16%. Điều này cho thấy GPT-5 có khả năng xử lý khá hiệu quả các dạng nhiệm vụ liên

quan đến Xác suất trong chương trình toán lớp 12 trong So sánh kết quả giữa các bài cho thấy có sự dao động nhất định trong hiệu suất. Cụ thể, GPT-5 thể hiện sự tích lũy kinh nghiệm dần qua các bài học. Các bài tập cuối chương ghi nhận tỉ lệ đúng tuyệt đối, đạt 100%, trong khi bài 1 ghi nhận kết quả thấp nhất với 76.56%. Khoảng chênh lệch giữa các bài nằm ở mức 23.44%, phản ánh một xu hướng tích cực: GPT-5 đặc biệt hiệu quả ở các bài tập mang tính tổng hợp và ôn tập cuối chương. Đáng chú ý, tất cả các bài đều đạt tỉ lệ đúng trên 75%, qua đó cho thấy tính ổn định cũng như hiệu quả cao của GPT-5 khi được sử dụng để giải quyết các bài toán thuộc chủ đề Xác suất và Thống kê trong bối cảnh dạy học Toán ở trường phổ thông.

Bảng 6. Đánh hiệu suất giải toán về chủ đề Xác suất của ChatGPT trong sách Chân trời sáng tạo

	Số lượng	Đúng		Sai	
		Hợp lí	Không hợp lí	Hợp lí	Không hợp lí
Bài 1	SGK	35	25 (71,43%)	4 (11,43%)	6 (17,14%)
	SBT	29	24 (82,76%)	0 (0%)	5 (17,24%)
Bài 2	SGK	24	17 (70,83%)	2 (8,33%)	5 (20,83%)
	SBT	15	13 (86,67%)	1 (6,67%)	0 (0%)
Bài tập cuối chương	SGK	17	17 (100%)	0 (0%)	0 (0%)
	SBT	30	30 (100%)	0 (0%)	0 (0%)

2.2.5. Các lỗi sai điển hình

Dựa trên những tổng hợp về lời giải, nghiên cứu phân loại 4 dạng lỗi đặc trưng của GPT-5 khi giải toán xác suất: (1) Lỗi tính toán; (2) Lỗi suy luận; (3) Vượt quá khả năng của ChatGPT; (4) Không giải quyết trọn vẹn bài toán. Cụ thể như sau:

Dạng 1: Lỗi tính toán

Câu 3. Trong mỗi ý a), b), c), d) ở câu 3, 4, chọn phương án: đúng (D) hoặc sai (S).

Lớp 12A có 40 học sinh. Trong một buổi kiểm tra định kì, số học sinh của lớp được chia thành hai phòng như sau:

Học sinh	Phòng 1	Phòng 2
Nam	11	8
Nữ	9	12

Chọn ngẫu nhiên một học sinh của lớp 12A.

Xét các biến cố:

A: "Học sinh được chọn ở phòng 2".

B: "Học sinh được chọn là học sinh nữ".

a) Biến cố học sinh được chọn là học sinh nữ ở phòng 2 là $A \cap B$.

b) $P(A \cap B) \neq \frac{3}{10}$

c) $P(B) = \frac{21}{40}$

d) $P(A|B) = \frac{3}{7}$

Lời giải

Tổng số học sinh của lớp: 40.

Số học sinh phòng 2: $8 + 12 = 20 \Rightarrow P(A) = \frac{20}{40} = \frac{1}{2}$.

Số học sinh nữ: $9 + 12 = 21 \Rightarrow P(B) = \frac{21}{40}$.

Học sinh nữ ở phòng 2: $12 \Rightarrow P(A \cap B) = \frac{12}{40} = \frac{3}{10}$.

a) "Nữ ở phòng 2" đúng là giao của A và B. $\Rightarrow D$.

b) $P(A \cap B) = \frac{3}{10}$, nên mệnh đề " $\neq \frac{3}{10}$ " là sai. $\Rightarrow S$.

c) $P(B) = \frac{21}{40}$ như tính trên. $\Rightarrow D$.

d) $P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{12/40}{21/40} = \frac{12}{21} = \frac{4}{7} \neq \frac{3}{7} \Rightarrow S$.

Hình 2. Sách bài tập Cánh diều Toán 12, Bài 1, Bài tập 3, trang 87

GPT-5 đưa ra lời giải chính xác về kiến thức và lập luận, tuy nhiên xuất hiện lỗi ở việc tính toán các giá trị của bài toán. Có thể thấy việc tính toán của GPT-5 rất nhanh chóng tuy nhiên vẫn có thể mắc một số lỗi như tính toán sai, thay sai số liệu hay lỗi làm tròn. Điều này sẽ dễ bị bỏ qua nếu như người đọc không thực sự tính toán lại bài toán.

Dạng 2: Lỗi suy luận

b) Gọi B là biến cố "Bình bị loại". Bình bị loại nếu không qua vòng 1 hoặc không qua vòng 2. Ta có:

$$B = \overline{A_1 \cap A_2} = \overline{A_1} \cup (A_1 \cap \overline{A_2}).$$

Xác suất để Bình qua vòng 1 nhưng không qua vòng 2 là:

$$P(A_1 \cap \overline{A_2}) = P(A_1) \cdot P(\overline{A_2}|A_1) = 0,9 \cdot (1 - 0,8) = 0,9 \cdot 0,2 = 0,18.$$

Vậy xác suất yêu cầu là:

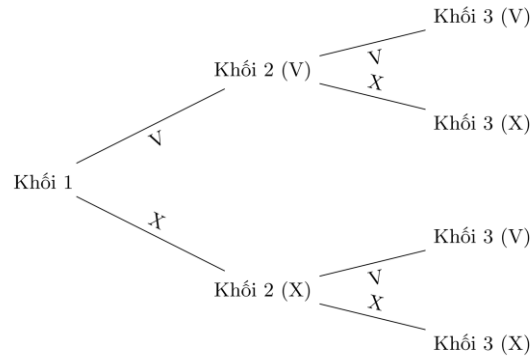
$$\boxed{0,18}.$$

Hình 3. Sách bài tập Kết nối tri thức với cuộc sống Toán 12, bài 18, ví dụ 2, trang 44

GPT-5 sai sót trong việc lập luận các biến cố để đưa ra chính xác trường hợp biến cố cần tính toán. Trong trường hợp này, GPT-5 đã suy luận sai, thay vì tính xác suất có điều kiện thì chỉ tính xác suất giao. Lỗi này phản ánh việc

GPT-5 chưa thực sự hiểu ngữ cảnh toán học mà chỉ đang lắp ghép các khuôn mẫu. Đây là lỗi nghiêm trọng, liên quan đến bản chất toán học.

Dạng 3: Vượt quá khả năng của GPT-5



Hình 4. SGK Cánh diều Toán 12, Bài 1, Hoạt động 2, trang 94

GPT-5 không thể hoàn thành trọn vẹn yêu cầu của bài toán. Ở bài toán này, mặc dù kết quả tính toán của GPT-5 có thể chính xác tuy nhiên GPT-5 gặp khó khăn lớn khi mô hình hóa trực quan sơ đồ cây cho bài toán. Cụ thể, sơ đồ đã biểu diễn thiếu không gian mẫu (bỏ sót hoàn toàn thao tác rẽ nhánh cho khối gỗ thứ ba) và thể hiện sự lúng túng, thiếu nhất quán trong việc gán nhãn các nút, nhánh, cho thấy khả năng minh họa còn hạn chế. Có thể thấy rằng GPT-5 chỉ coi sơ đồ cây là một công cụ trực quan hóa bài toán chứ không phải là một kỹ thuật để có thể giải quyết những bài toán xác suất có điều kiện hay liên kết đến những điều kiện phức tạp hơn.

Dạng 4: Không giải quyết trọn vẹn bài toán

3. Mỗi bạn học sinh trong lớp của Minh lựa chọn học một trong hai ngoại ngữ là tiếng Anh hoặc tiếng Nhật. Xác suất chọn tiếng Anh của mỗi bạn học sinh là 0,6 và của mỗi bạn học sinh nam là 0,7. Lớp của Minh có 25 bạn nữ và 20 bạn nam. Chọn ra ngẫu nhiên một bạn trong lớp.

Sử dụng sơ đồ hình cây, tính xác suất của các biến cố:

A: "Bạn được chọn là nam và học tiếng Nhật";

B: "Bạn được chọn là nữ và học tiếng Anh".

Hình 5. SGK Chân trời sáng tạo Toán 12, bài 1, câu 3, trang 75

Việc GPT-5 không phản hồi với các bài toán được đưa ra có thể phản ánh được rằng có thể GPT-5 không thể phân tích và hiểu được dữ liệu của bài toán được đưa ra.

3. Kết luận

Nghiên cứu này cung cấp bằng chứng hệ thống về năng lực và giới hạn của ChatGPT trong việc giải các bài toán xác suất thuộc chương trình Toán lớp 12 theo Chương trình giáo dục phổ thông 2018. Với tỉ lệ chính xác tổng thể đạt 84,01%, mô hình cho thấy tiềm năng ứng dụng trong hỗ trợ học tập, song hiệu suất phân hoá rõ rệt theo mức độ nhận thức: ChatGPT xử lý tốt các bài toán áp dụng trực tiếp công thức xác suất cơ bản nhưng sụt giảm đáng kể ở các bài toán đòi hỏi nhận diện biến cố trong ngữ cảnh thực tiễn, lập luận xác suất có điều kiện và tư duy nhiều bước. Đặc biệt, ngay cả khi đưa ra được đáp án đúng, lời giải của ChatGPT thường không phản ánh đúng quá trình lập luận toán học cần thiết, một hạn chế xuất phát từ bản chất của mô hình là tái tạo mẫu thống kê từ dữ liệu huấn luyện thay vì nắm bắt cấu trúc khái niệm toán học theo nghĩa thực sự (Hendrycks và cộng sự, 2021; Frieder và cộng sự, 2024). Những phát hiện này nhất quán với xu hướng được ghi nhận trong tài liệu quốc tế rằng ChatGPT thực hiện tốt ở các nhiệm vụ thủ tục nhưng sụt giảm rõ rệt khi bài toán đòi hỏi chiều sâu khái niệm và lập luận nhiều bước (Kaya và Yavuz, 2025; Turmuzi và cộng sự, 2026), đồng thời bổ sung bằng chứng gắn với một chủ đề toán học cụ thể trong bối cảnh chương trình phổ thông Việt Nam. Hơn thế, kết quả nghiên cứu cung cấp thêm bằng chứng cho phép đề xuất khung lý thuyết trong đó ChatGPT nên đóng vai trò trợ giảng ảo cá nhân hoá học tập hay công cụ rèn luyện tư duy phân biện mà theo Kasneci và cộng sự (2023), Turmuzi và cộng sự (2026) hai chức năng này có thể bổ sung cho nhau như thế nào dưới sự điều tiết có chủ đích của GV. Cụ thể, chính những lỗi lập luận có hệ thống của mô hình trong chủ đề Xác suất có thể được khai thác như một tác nhân nhận thức: khi HS được yêu cầu phát hiện, phân tích và sửa chữa sai sót trong lời giải do ChatGPT tạo ra, các em buộc phải tường minh hoá quá trình lập luận của bản thân, từ đó củng cố hiểu biết khái niệm với việc chỉ thực hành giải bài thông thường (Lo, 2023).

Kết quả được giới hạn trong chủ đề Xác suất lớp 12 và chưa thể khái quát hoá trực tiếp sang các chủ đề toán học khác; hiệu suất của mô hình đồng thời chịu ảnh hưởng đáng kể bởi cách thức thiết kế câu lệnh. Quan trọng hơn, do trí tuệ nhân tạo đang thay đổi theo từng tháng, kết quả đánh giá mô hình tại thời điểm nghiên cứu có thể không còn đại diện khi các phiên bản cập nhật mới hơn được phát hành đặt ra câu hỏi về tính bền vững của các kết luận theo thời gian (Lo, 2023). Từ những hạn chế này, nghiên cứu đề xuất ba hướng tiếp theo: mở rộng đánh giá sang các chủ đề toán học khác trong chương trình 2018 nhằm xây dựng bức tranh toàn diện hơn về năng lực của mô hình; nghiên cứu tương tác trực tiếp giữa HS và ChatGPT trong môi trường học tập số; và thực hiện các nghiên cứu thực nghiệm có nhóm đối chứng về hiệu quả của các chiến lược tích hợp ChatGPT, từ đó cung cấp bằng chứng nhân quả cho các khuyến nghị sư phạm thay vì chỉ dừng lại ở mức gợi ý dựa trên phân tích hiệu suất mô hình.

Tài liệu tham khảo

- Đào Xuân Quy, Ngô Bắc Biên, Phan Xuân Dũng, Lê Ngọc Bích (2024). Đánh Giá ChatGPT trong việc giải Đề thi minh họa kì thi tốt nghiệp trung học phổ thông 2023 tại Việt Nam. *Tạp chí Khoa học Giáo dục Kỹ thuật, Trường Đại học Sư phạm Kỹ thuật Thành phố Hồ Chí Minh*, 19(2), 1-11. <https://doi.org/10.54644/jte.2024.1365>
- Frieder, S., Pinchetti, L., Chevalier, A., Griffiths, R.-R., Salvatori, T., Lukasiewicz, T., Petersen, P. C., & Berner, J. (2024). Mathematical capabilities of ChatGPT. *Transactions on Machine Learning Research*. <https://openreview.net/forum?id=0oOtdrLILW>
- Hagendorff, T., Fabi, S., & Kosinski, M. (2023). Human-like intuitive behavior and reasoning biases emerged in large language models but disappeared in ChatGPT. *Nature Computational Science*, 3, 833-838. <https://doi.org/10.1038/s43588-023-00527-x>
- Hendrycks, D., Burns, C., Kadavath, S., Arora, A., Basart, S., Tang, E., Song, D., & Steinhardt, J. (2021). Measuring mathematical problem solving with the MATH dataset. *Advances in Neural Information Processing Systems*, 34 (NeurIPS 2021 Datasets and Benchmarks Track).
- Kasneji, E., Sessler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F.,... & Kasneji, G. (2023). ChatGPT for Good? On Opportunities and Challenges of Large Language Models for Education. *Learning and Individual Differences*, 103, 102274. <https://doi.org/10.1016/j.lindif.2023.102274>
- 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), 43. <https://doi.org/10.3390/jintelligence13040043>
- Lê Anh Vinh, Bùi Thị Diễm, 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), 1-10.
- Lo, C. K. (2023). What is the impact of ChatGPT on education? A rapid review of the literature. *Education Sciences*, 13(4), 410. <https://doi.org/10.3390/educsci13040410>
- 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>
- Turmuzi, M., Azmi, S., & Kertiyani, N. M. I. (2026). ChatGPT in school mathematics education: A systematic review of opportunities, challenges, and pedagogical implications. *Teaching and Teacher Education*, 170, 105286.