

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

Lê Anh Vinh¹, Bùi Thị Diễm^{*2},
Lê Quang Quân³, Vũ Văn Luân⁴

¹ Email: vinhla@vnies.edu.vn

* Tác giả liên hệ

² Email: dienbt@vnies.edu.vn

³ Email: quanlq@vnies.edu.vn

Viện Khoa học Giáo dục Việt Nam

101 Trần Hưng Đạo, Hoàn Kiếm,

Hà Nội, Việt Nam

⁴ Email: vuvanluanftu.k50@gmail.com

Công ty Cổ phần Công nghệ Giáo dục xanh

Số 3, ngõ 814, đường Láng, Đống Đa,

Hà Nội, Việt Nam

TÓM TẮT: ChatGPT đang trở thành chủ đề nóng trên quy mô toàn cầu bởi tính mới mẻ và khả năng xử lý vượt trội. Công cụ ChatGPT làm gia tăng kì vọng vào những cải tiến mang tính đột phá mà trí tuệ nhân tạo có thể mang đến cho đời sống con người, trong đó có lĩnh vực giáo dục. Bài viết này nghiên cứu kết quả thực hiện bài kiểm tra định kì môn Ngữ văn và Toán cấp Trung học, lớp 9 và lớp 12 của ChatGPT trong mối tương quan với kết quả kiểm tra thực tế của học sinh. Nghiên cứu chỉ ra rằng, ChatGPT có khả năng thực hiện các bài kiểm tra ở mức độ nhất định nhưng chất lượng câu trả lời không ổn định và phụ thuộc vào nhiều yếu tố. Nghiên cứu cung cấp một góc nhìn chi tiết về khả năng của ChatGPT trong kiểm tra, đánh giá, giúp các nhà quản lí giáo dục, giáo viên, học sinh... có cơ sở để đưa ra các phương án sử dụng công cụ này một cách phù hợp và hiệu quả.

TỪ KHÓA: ChatGPT, trí tuệ nhân tạo, kiểm tra đánh giá, môn Toán, môn Ngữ văn.

→ Nhận bài 15/02/2023 → Nhận bài đã chỉnh sửa 20/02/2023 → Duyệt đăng 25/02/2023.

DOI: <https://doi.org/10.15625/2615-8957/12310201>

1. Đặt vấn đề

Các cuộc Cách mạng công nghiệp đã tạo ra sự thay đổi lớn trên nhiều phương diện, mang đến những đổi thay tích cực cho đời sống xã hội, trong đó có giáo dục. Những thành tựu của công nghệ thông tin đã góp phần nâng cao hiệu quả dạy và học, cải thiện thành tích và hứng thú học tập cho học sinh [1]. Đặc biệt, sự ra đời của trí tuệ nhân tạo (Artificial intelligence - AI) đã thúc đẩy các phương pháp dạy học tích cực, các hình thức đánh giá khách quan, qua đó tăng cường trải nghiệm học tập cá nhân hoá, học tập thích ứng [2]. Tuy nhiên, bên cạnh các lợi ích, nghiên cứu cũng chỉ ra những thách thức của ứng dụng trí tuệ nhân tạo trong giáo dục như sự phụ thuộc vào công nghệ, vấn đề an toàn số, dữ liệu thiếu hoàn thiện và chưa được xác thực [3]. Chính vì vậy, các nghiên cứu ứng dụng và đánh giá hiệu quả của công nghệ giáo dục có vai trò quan trọng trong việc đưa ra những định hướng phù hợp.

Trong số những công cụ trí tuệ nhân tạo mới nhất hiện nay, ChatGPT (Generative Pre-trained Transformer) nổi lên như một xu hướng mới, thu hút sự quan tâm toàn cầu bởi tính mới mẻ và khả năng xử lý vượt trội. ChatGPT là một mô hình ngôn ngữ lớn được huấn luyện bằng các phương pháp học sâu (deep learning), được OpenAI phát triển từ năm 2018 [4]. Mô hình này được huấn luyện từ một lượng lớn dữ liệu văn bản trên Internet, với mục tiêu là tạo ra một công cụ đa năng có

thể giải quyết nhiều vấn đề bằng ngôn ngữ tự nhiên [5]. ChatGPT được đánh giá có khả năng tương tác và trả lời thông minh, dễ dàng tích hợp vào các ứng dụng và linh hoạt sử dụng trên nhiều nền tảng khác nhau [6]. Tuy nhiên, ứng dụng này có những hạn chế nhất định liên quan đến độ chính xác, mức độ cập nhật của thông tin và quyền riêng tư [4]. Từ cuối năm 2022, ChatGPT đã trở thành đối tượng nghiên cứu trong rất nhiều lĩnh vực để tìm hiểu về khả năng ứng dụng của công cụ này [5], [7].

Trong xu thế trên, các nền giáo dục trên thế giới cũng rất quan tâm đến tác động của ChatGPT đến quá trình dạy học và quản lí giáo dục. Các nhà giáo dục, các nhà nghiên cứu bắt đầu chú ý tìm hiểu về cách thức ứng dụng, hiệu quả cũng như thách thức mà công cụ này mang lại. Theo Markel và cộng sự (2023), ChatGPT có thể được sử dụng để phát triển nền tảng tập huấn giáo viên hiệu quả [8]. Kwon (2023) chỉ ra các ứng dụng trí tuệ nhân tạo như ChatGPT có thể là phương tiện dạy học ngôn ngữ [9]. Phillips và cộng sự (2022) nhấn mạnh vai trò của ChatGPT như công cụ đánh giá hiệu quả [10], Gilson và cộng sự (2023) chỉ ra thành tích đáng kể của ChatGPT trong kì thi Y học của Mỹ [5]. Cụ thể, ChatGPT trả lời đúng trên 60% câu hỏi trong đề thi Y khoa, bằng với điểm đạt của một sinh viên Y khoa năm thứ ba [5], thậm chí có thể vượt qua kì thi Luật và Kinh doanh với mức điểm trung bình của

trường đại học Mĩ [11]. Tuy nhiên, với đặc thù của giáo dục là đào tạo con người có phẩm chất, năng lực thì sự vượt trội của ChatGPT cũng đẩy lên những mối lo ngại lớn về nguy cơ gian lận, làm ảnh hưởng đến chất lượng đầu ra của giáo dục [12]. Thực tế đã phát hiện những trường hợp gian lận như người học sử dụng ChatGPT trong các kì thi [13]. Tại Việt Nam, nhận thức rõ được cơ hội và thách thức của các ứng dụng AI nói chung và ứng dụng ChatGPT nói riêng, Bộ Giáo dục và Đào tạo đã tổ chức tọa đàm “*ChatGPT, trí tuệ nhân tạo - lợi ích và thách thức đối với giáo dục*” [14]. Trong tọa đàm, những người tham gia đã thảo luận và chia sẻ về đặc điểm, ảnh hưởng của AI, ChatGPT, bàn về các chiến lược hành động trong tương lai của ngành Giáo dục.

Cơ hội và thách thức của ChatGPT trong giáo dục vẫn đang là một vấn đề cần được giải mã khi mà những nghiên cứu về khả năng của ChatGPT vẫn còn hạn chế không chỉ trong nước mà cả trên thế giới. Với những lí do ở trên, nhóm tác giả đã nghiên cứu khả năng thực hiện các bài kiểm tra định kì của ChatGPT với các tiêu chí về độ chính xác, tính phù hợp và hiệu quả trong môn Ngữ văn và môn Toán lớp 9 và lớp 12 trong mối tương quan với kết quả kiểm tra thực tế của học sinh. Kết quả nghiên cứu là cơ sở để đưa ra những khuyến nghị ban đầu cho các nhà quản lí, giáo viên và các đối tượng quan tâm trong việc ứng dụng và quản lí việc sử dụng ChatGPT trong giáo dục.

2. Nội dung nghiên cứu

2.1. Phương pháp nghiên cứu

- *Đối tượng nghiên cứu:* Công cụ AI - ChatGPT (phiên bản tháng 11 năm 2022).

- *Dữ liệu đầu vào:* Đề kiểm tra định kì môn Toán - Ngữ văn cuối học kì I (năm học 2022 - 2023) lớp 9 và lớp 12 của một trường phổ thông liên cấp tại Hà Nội. Nhóm nghiên cứu chọn môn Ngữ văn và môn Toán là hai môn đại diện cho nhóm môn khoa học xã hội và tự nhiên với hình thức kiểm tra tự luận và trắc nghiệm. Đề kiểm tra định kì được xây dựng theo hướng dẫn của Bộ Giáo dục và Đào tạo. Ma trận mạch nội dung và kĩ năng của các đề kiểm tra định kì được mô tả như bảng dưới đây (xem Bảng 1 và Bảng 2).

172 học sinh lớp 9 và 161 học sinh lớp 12 thực hiện đánh giá chất lượng cuối học kì I theo đề kiểm tra định kì của Phòng Giáo dục và Đào tạo (lớp 9) và của Trường liên cấp (lớp 12). Học sinh kiểm tra tập trung, không sử dụng công cụ ChatGPT hoặc bất kỳ thiết bị điện tử nào (trừ máy tính cầm tay cho môn Toán). Bài làm của học sinh sau đó được chấm điểm và phân tích với các giá trị điểm trung bình, phổ điểm.

- *Quy trình nghiên cứu:*

- Đối với mỗi đề kiểm tra, nhóm nghiên cứu thiết kế thành 2 nhóm lệnh hỏi, mục đích tìm hiểu khả năng xử lí của ChatGPT đối với các cách hỏi khác nhau. Nhóm lệnh 1: Những câu hỏi được trích nguyên văn từ đề.

Bảng 1: Ma trận xây dựng đề kiểm tra môn Ngữ văn lớp 9 và lớp 12 (Đơn vị %)

Lớp	Mạch nội dung, kĩ năng			Mức độ nhận thức			
	Đọc hiểu	Nghị luận xã hội	Nghị luận văn học	Nhận biết	Thông hiểu	Vận dụng	Vận dụng cao
Lớp 9	45	20	35	20	40	30	10
Lớp 12	30	20	50	20	40	30	10

Bảng 2: Ma trận xây dựng đề kiểm tra môn Toán lớp 9 và lớp 12 (Đơn vị %)

Lớp 9	Mức độ nhận thức				Tổng	Lớp 12	Mức độ nhận thức				Tổng
	Nhận biết	Thông hiểu	Vận dụng	Vận dụng cao			Nhận biết	Thông hiểu	Vận dụng	Vận dụng cao	
Căn thức và biến đổi đại số	5	20	5	5	35	Ứng dụng đạo hàm để khảo sát và vẽ đồ thị hàm số	10	22	6	2	40
Hàm số và đồ thị	12.5	5	7.5		25	Hàm số lũy thừa - Hàm số mũ - Hàm số logarit	6	12	2		20
Tỉ số lượng giác	10				10	Khối đa diện	6	6	2		14
Đường tròn		20	10		30	Mặt nón, mặt trụ, mặt cầu	14	4	2	6	26
Tổng	27.5	45	22.5	5	100	Tổng	36	44	12	8	100

Nhóm lệnh 2: Câu hỏi trong đề nhưng có điều chỉnh theo hướng hỏi chi tiết hoặc bổ sung gợi ý.

- Riêng đối với môn Toán lớp 12, nội dung đề kiểm tra xuất hiện một số yêu cầu đặc biệt gây bất lợi cho ChatGPT như quan sát các hình minh họa hoặc cần thực hiện vẽ hình để giải quyết bài toán. Do đó, đối với môn Toán lớp 12, nhóm nghiên cứu thực hiện hai lượt đánh giá với phiên bản đề kiểm tra đầy đủ 50 câu và phiên bản 31 câu (bỏ đi 19 câu hỏi gây bất lợi cho ChatGPT bao gồm 11 câu hỏi gắn với hình minh họa cho trước và 8 câu hỏi mà học sinh cần thực hiện vẽ hình để tính toán).

- Sử dụng 20 tài khoản ChatGPT, dán các lệnh hỏi vào ChatGPT để thu thập câu trả lời. Mỗi tài khoản thực hiện cả 2 nhóm lệnh đề môn Ngữ văn và Toán lớp 9 và lớp 12.

- Mỗi bài làm của ChatGPT được chấm bởi 3 chuyên gia độc lập để đảm bảo kết quả khách quan.

- So sánh kết quả bài thi của học sinh và ChatGPT theo các tiêu chí điểm trung bình, phổ điểm. Bài làm bằng ChatGPT được phân tích sâu về mức độ chính xác/phù hợp của câu trả lời theo các mạch nội dung và mức độ nhận thức.

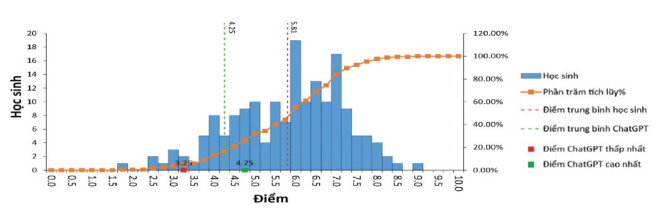
Nghiên cứu sử dụng phương pháp định lượng để thống kê mô tả tần suất, tỉ lệ, giá trị trung bình với sự hỗ trợ phần mềm thống kê SPSS, kết hợp với phương pháp định tính trong việc đưa ra các phân tích, lí giải dựa trên cơ sở các dữ liệu thu thập được.

2.2. Kết quả nghiên cứu

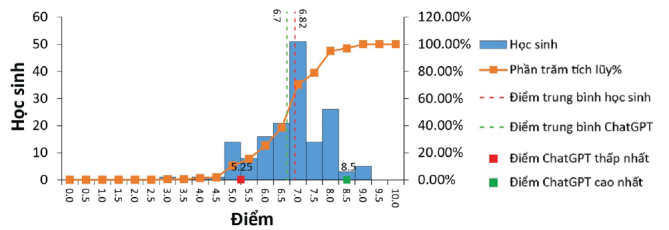
2.2.1. Kết quả thực hiện bài kiểm tra môn Toán và môn Ngữ văn lớp 9, lớp 12 của ChatGPT

Nghiên cứu cho thấy, kết quả thực hiện bài kiểm tra định kì môn Ngữ văn và môn Toán của ChatGPT đạt mức tương đối. Theo đó, trong môn Ngữ văn, điểm trung bình của ChatGPT thấp hơn so với mức điểm trung bình của học sinh. Cụ thể, đối với môn Ngữ văn lớp 9 (xem Hình 1) điểm trung bình của học sinh và ChatGPT lần lượt là 5.81 và 4.25 với hiệu số chênh lệch là 1.56, trong khi với môn Ngữ văn lớp 12 (xem Hình 2) là 6.82 và 6.7 với hiệu số chênh lệch chỉ là 0.12. Điểm cao nhất môn Ngữ văn lớp 9 và lớp 12 của học sinh lần lượt là 9 và 9, trong khi của ChatGPT là 4.75 và 8.5. Điểm thấp nhất thì có xu hướng ngược lại, hiệu số chênh lệch là -1.5 và -2.25 đối với môn Ngữ văn lớp 9 và môn Ngữ văn lớp 12 (xem Hình 1 và Hình 2).

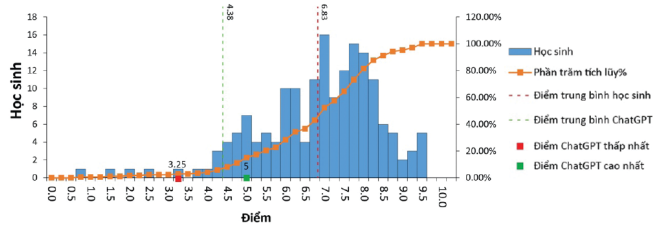
Đối với đề kiểm tra môn Toán lớp 9, ChatGPT đạt kết quả trung bình 4.38 điểm, thấp hơn so với điểm trung bình của học sinh (6.82 điểm) với hiệu số là 2.45. Nguyên nhân dẫn đến sự chênh lệch này là do ChatGPT gặp bất lợi khi không thể giải đúng các bài tập về Hình học (là bài tập mà trên thực tế học sinh cần vẽ hình để đưa ra lời giải) hay ChatGPT không có khả năng



Hình 1: Thống kê điểm số của học sinh và ChatGPT trong kiểm tra định kì môn Ngữ văn lớp 9



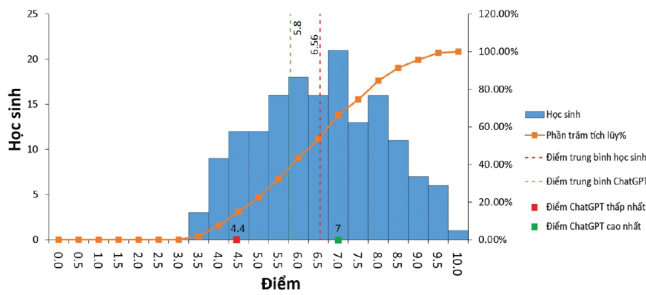
Hình 2: Thống kê điểm số bài làm của học sinh và ChatGPT trong kiểm tra định kì môn Ngữ văn lớp 12



Hình 3: Thống kê điểm số bài làm của học sinh và ChatGPT trong kiểm tra định kì môn Toán lớp 9

thực hiện các phép biến đổi đại số phức tạp cũng như thể hiện chưa tốt việc phải trình bày các bước giải theo cách hợp lí, chặt chẽ để đảm bảo yêu cầu của một đề kiểm tra tự luận. Cũng giống môn Ngữ văn, điểm cao nhất của ChatGPT thấp hơn điểm số cao nhất của học sinh (5 so với 9.5), nhưng khi xét đến điểm số thấp nhất thì kết quả thay đổi ngược lại (3.25 so với 0.75).

Đối với môn Toán lớp 12, từ đề thi ban đầu, nhóm nghiên cứu xây dựng gói câu lệnh 1, giữ nguyên các nội dung, biểu diễn các công thức Toán học phức tạp bằng ngôn ngữ TeX để máy có thể hiểu được. Sau khi chạy gói câu lệnh, kết quả cho thấy, ChatGPT có thể đạt điểm số cao nhất là 5.6, thấp nhất là 4.4 và mức điểm số trung bình là 4.93. Kết quả này thấp hơn nhiều so với mức điểm trung bình của các học sinh (6.56 điểm, dao động từ 3.6 đến 10) (xem Hình 4). Bên cạnh những khó khăn mà ChatGPT gặp phải như với đề kiểm tra môn Toán lớp 9 và lớp 12, ChatGPT còn gặp bất lợi khi chưa được tích hợp các chức năng thuận tiện cho việc nhập - xuất hình vẽ. ChatGPT gần như không có cơ sở thông tin nào để thực hiện các câu yêu cầu phải quan sát để tìm thông tin từ hình minh họa, là các bảng biến thiên hoặc đồ thị hàm số.



Hình 4: Thống kê điểm số bài làm của học sinh và ChatGPT trong kiểm tra định kì môn Toán lớp 12

Kết quả ở gói lệnh thứ 2, sau khi đã loại 19 câu không thích ứng với ChatGPT, công cụ này đạt điểm trung bình là 3.8 (dao động từ 3.2 đến 4.2 điểm) trên mức điểm tối đa 6.2. Nghiên cứu giả định rằng, ChatGPT có khả năng giải được 19 câu hỏi còn lại với tỉ lệ đúng tương tự. Khi đó, điểm số trung bình mà ChatGPT có thể đạt được là: $3.8 + 2.33 = 6.13$. Như vậy, ngay cả khi loại bỏ điểm yếu của ChatGPT về mặt thao tác với hình ảnh thì điểm trung bình của học sinh vẫn cao hơn của ChatGPT nhưng sự chênh lệch đã giảm đáng kể, chỉ còn 0.43 điểm.

Kết quả nghiên cứu cho thấy, ChatGPT có khả năng thực hiện các bài kiểm tra môn Toán và môn Ngữ văn với mức điểm thấp hơn mức điểm trung bình của học sinh. Xét theo phổ điểm, điểm cao nhất đối với ChatGPT thấp hơn hoặc thấp hơn nhiều so với điểm cao nhất của học sinh. Ngược lại, ChatGPT có điểm thấp nhất cao hơn so với điểm thấp nhất của học sinh. Có nhiều nguyên nhân dẫn đến kết quả này. Cụ thể, ChatGPT có khả năng tổng hợp, khái quát dữ liệu lớn, ít trùng lặp để đưa ra câu trả lời nên giá trị điểm trung bình của ChatGPT có thể đạt được ngưỡng điểm tương đối. Điểm cao nhất của ChatGPT không cao bằng so với điểm cao nhất của học sinh bởi những hạn chế của ChatGPT trong khả năng sáng tạo và mức độ chính xác. Điểm thấp nhất của ChatGPT luôn cao hơn so với điểm thấp nhất của học sinh vì thực tế, tác vụ được thực hiện trên ChatGPT luôn đảm bảo trả lời đủ tất cả các câu hỏi, trong khi đó, có thể xảy ra trường hợp học sinh bỏ bài, bỏ câu hoặc một số học sinh gặp khó khăn đặc biệt trong quá trình làm bài thực tế.

2.2.2. Mức độ chính xác và hợp lí trong bài làm của ChatGPT qua các mạch nội dung và mức độ nhận thức

Kết quả môn Ngữ văn ở lớp 9 và lớp 12 được thực hiện trên ChatGPT cũng có những khác biệt tương đối, phụ thuộc vào mức độ dễ và khó của đề và sự thích ứng của công cụ với đề kiểm tra. Dựa vào sự phân bố phổ điểm và điểm trung bình trên kết quả bài làm của học sinh (xem Hình 1 và Hình 2), có thể thấy đề Ngữ văn lớp 9 và lớp 12 có mức độ khó - dễ khác nhau. Hiệu số chênh lệch giữa kết quả thực hiện bài kiểm tra trên ChatGPT cũng cho thấy những điểm mạnh và điểm yếu của công cụ này trong việc xử lý các câu hỏi. Cụ thể, phân tích sâu cho thấy, trong đề Ngữ văn lớp 12, các câu hỏi thường thiên về mặt nội dung văn bản (Với thông tin được cung cấp trong phần ngữ liệu), đưa ra các lí giải, suy ngẫm về các vấn đề văn học và xã hội, trong khi ở đề Ngữ văn lớp 9, các câu hỏi có chứa nhiều đơn vị kiến thức về tiếng Việt, phát hiện các kiến thức về đặc điểm nghệ thuật của văn bản kết hợp với cách hỏi chùng (một câu có nhiều ý hỏi), yêu cầu liên hệ văn bản ngoài nên việc xử lý các câu hỏi này bằng AI gặp nhiều khó khăn hơn. Open AI (2022) cũng khẳng định rằng, ChatGPT có nhiều hạn chế về độ chính xác của thông tin trong câu trả lời, hạn chế trong việc tổng hợp thông tin trong các bối cảnh khác nhau hoặc mang tính địa phương. Ngoài ra, những dữ liệu nền tảng của ChatGPT hiện tại chỉ được cập nhật đến năm 2021 [4].

Trong môn Ngữ văn, nghiên cứu chi tiết từng bài làm của ChatGPT cho thấy, câu trả lời có tỉ lệ chính xác và độ hợp lí khác nhau ở các mức độ nhận biết, thông hiểu, vận dụng và vận dụng cao. Mức độ chính xác/hợp lí trong câu trả lời ở các mạch nội dung, kĩ năng Ngữ văn (đọc hiểu, nghị luận xã hội, nghị luận văn học) cũng cho thấy sự chênh lệch đáng kể như thể hiện ở Bảng 3.

Cụ thể, phân đọc hiểu, với các đơn vị kiến thức liên quan tiếng Việt, lí luận văn học như xác định phương thức biểu đạt, hình thức ngôn ngữ... ChatGPT phần lớn đều trả lời sai. ChatGPT xử lí kém ở các câu nhận biết về đặc điểm hình thức và nghệ thuật, nhưng lại xử lí tốt ở các câu nhận biết về nội dung, có chứa thông tin trong phần đọc hiểu với tỉ lệ trả lời đúng lên đến gần 100%. Nhìn chung, trong việc xử lí các đề thi môn Ngữ văn, ChatGPT khá mạnh trong phân lập luận, diễn giải

Bảng 3: Mức độ chính xác/hợp lí trong bài làm của ChatGPT qua các mạch nội dung và mức độ nhận thức trong bài kiểm tra định kì môn Ngữ văn (Đơn vị: %)

Mạch nội dung			Mức độ nhận thức			
Đọc hiểu	Nghị luận xã hội	Nghị luận văn học	Nhận biết	Thông hiểu	Vận dụng thấp	Vận dụng cao
50	70	45	50	65	60	45

ý nghĩa, nội dung mang tính quan điểm, lập luận như viết bài nghị luận xã hội, phân tích văn học, trong khi những nội dung yêu cầu tính xác thực của thông tin thì chưa tốt, ví dụ tên tác giả, tác phẩm, thể loại.

Trong môn Toán lớp 9, ChatGPT có khả năng làm tốt các câu hỏi nhận biết, thông hiểu thuộc các nội dung: Căn thức và biến đổi đại số; Hàm số và đồ thị; Tỉ số lượng giác. Ở mức vận dụng, ChatGPT thực hiện tương đối tốt câu hỏi thuộc nội dung Hàm số và đồ thị, bởi nội dung của bài toán đó có thể dễ dàng được mô hình hóa và giải theo các bước, quy trình có sẵn. Tuy nhiên, ở câu hỏi về Hình học (là nhiệm vụ mà học sinh trên thực tế cần vẽ hình để giải) hay các câu hỏi cần thực hiện biến đổi đại số phức tạp thì ChatGPT gần như chưa thể thực hiện được (xem Bảng 4).

Ở môn Toán lớp 12, mức độ chính xác/hợp lý của ChatGPT ở các mạch nội dung và mức độ nhận thức cũng cho kết quả tương tự như ở môn Toán lớp 9 (xem Bảng 5). ChatGPT giải tương đối tốt các câu hỏi ở mức độ nhận biết và thông hiểu của các mạch nội dung về Giá trị lớn nhất và nhỏ nhất của hàm số; Đường tiệm cận; Khái niệm về thể tích của khối đa diện; Mặt nón, mặt trụ, mặt cầu. Nội dung của các câu hỏi ở nhóm này tập trung vào các khái niệm hoặc các bài toán nhỏ có thể được giải quyết bằng cách vận dụng các công thức có sẵn. Tuy nhiên, với các bài đòi hỏi thực hiện kỹ năng biến đổi qua nhiều bước hoặc vận dụng tổng hợp các kiến thức thì ChatGPT chưa làm tốt. Trong một số trường hợp, ChatGPT nêu được các nội dung kiến thức liên quan và thuật toán để giải bài nhưng không thể tự vận dụng để đưa ra kết quả đúng.

Tất nhiên, mức độ hợp lý trong câu trả lời của ChatGPT vẫn phụ thuộc nhiều vào cách thức đặt lệnh cho máy. Các lệnh cụ thể mang tính điều hướng sẽ dẫn đến tỉ lệ câu trả lời chính xác/hợp lý cao hơn (xem Bảng 6). Như

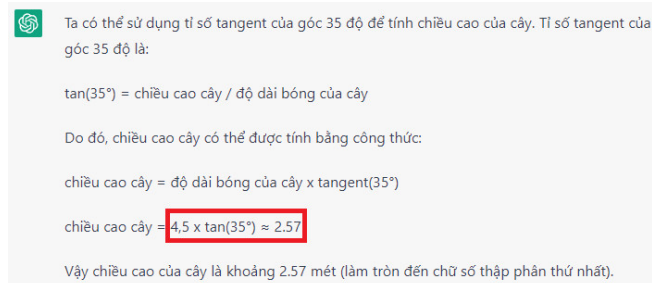
vậy, ngoại trừ những phạm vi ChatGPT (2023) chưa xử lý tốt, nếu các lệnh hỏi được đặt hợp lý, rõ ràng, mang tính điều hướng đúng thì kết quả đạt được cao hơn hẳn, đặc biệt đối với những mạch nội dung hoặc dạng câu hỏi thuộc về thể mạnh của ChatGPT.

Như vậy, thành tích của ChatGPT không độc lập mà phụ thuộc vào tính chủ động, vốn kiến thức và kỹ năng của người hỏi. Ví dụ, trong môn Toán dưới đây, khi người dùng thay đổi cách diễn đạt câu hỏi, cung cấp thêm các kiến thức nền thì mức độ chính xác của ChatGPT cải thiện đáng kể. Bảng 7 minh họa cách điều chỉnh lệnh hỏi theo hướng bổ sung diễn giải cho các hình minh họa (đồ thị, bảng biến thiên) và giới thiệu tóm tắt các nội dung lý thuyết liên quan.

Điểm số trung bình của gói câu lệnh có sự khác biệt

Bảng 6: Điểm trung bình của gói lệnh 1 và gói lệnh 2 trong thực hiện bài kiểm tra định kì môn Toán và Ngữ văn trên ChatGPT

Môn	Lệnh 1	Lệnh 2	Trung bình
Ngữ văn 9	4	4.5	4.25
Ngữ văn 12	5.08	8.32	6.7
Toán 9	4.17	4.58	4.38
Toán 12	4.93	6.67	5.8



Hình 5: Minh họa phần trả lời của ChatGPT trong đề kiểm tra môn Toán lớp 9

Bảng 4: Mức độ chính xác/hợp lý trong bài làm của ChatGPT qua các mạch nội dung và mức độ nhận thức trong bài kiểm tra định kì môn Toán lớp 9 (đơn vị: %)

Mạch nội dung	Mức độ năng lực						
	Căn thức và biến đổi đại số	Hàm số và đồ thị	Tỉ số lượng giác	Đường tròn	Nhận biết	Thông hiểu	Vận dụng
52.38	75	62.5	1.39	77.28	29.63	35.18	25

Bảng 5: Mức độ chính xác/hợp lý trong bài làm của ChatGPT qua các mạch nội dung và mức độ nhận thức trong bài kiểm tra định kì môn Toán lớp 12 (Đơn vị: %)

Mạch nội dung	Mức độ năng lực						
	Ứng dụng đạo hàm để khảo sát và vẽ đồ thị hàm số	Hàm số lũy thừa - Hàm số mũ - Hàm số logarit	Khối đa diện	Mặt nón, mặt trụ, mặt cầu	Nhận biết	Thông hiểu	Vận dụng
55.38	45.45	46.22	56.73	82.39	46.21	36.08	41.75

Bảng 7: Ví dụ về cách thay đổi gói lệnh cho đề kiểm tra môn Toán lớp 12

Đề kiểm tra	Gói câu lệnh 1	Gói câu lệnh 2																		
<p>Cho hàm số có bảng biến thiên như hình vẽ. Hàm số đã cho nghịch biến trên khoảng nào dưới đây?</p> <table border="1" style="margin-left: auto; margin-right: auto;"> <tr> <td>x</td> <td>$-\infty$</td> <td>-2</td> <td>0</td> <td>2</td> <td>$+\infty$</td> </tr> <tr> <td>$f'(x)$</td> <td>$-$</td> <td>0</td> <td>$+$</td> <td>0</td> <td>$-$</td> </tr> <tr> <td>$f(x)$</td> <td>$+\infty$</td> <td>1</td> <td>3</td> <td>1</td> <td>$+\infty$</td> </tr> </table> <p>A. (0; 2) B. (-2; 0) C. (0; $+\infty$) D. (2; $+\infty$)</p>	x	$-\infty$	-2	0	2	$+\infty$	$f'(x)$	$-$	0	$+$	0	$-$	$f(x)$	$+\infty$	1	3	1	$+\infty$	<p>Cho hàm số có bảng biến thiên như hình vẽ. Hàm số đã cho nghịch biến trên khoảng nào dưới đây?</p> <p>A. (0; 2) B. (-2; 0) C. (0; $+\infty$) D. (2; $+\infty$)</p>	<p>Cho hàm số $y = f(x)$ có bảng biến thiên được mô tả như dưới đây:</p> <ul style="list-style-type: none"> - $f'(x) = 0$ tại các điểm $x = -2, x = 0, x = 2$ - Khi x chạy từ $-\infty$ tới -2 thì $f'(x)$ mang dấu âm và $f(x)$ có giá trị giảm dần từ $+\infty$ tới 1. - Khi x chạy từ -2 tới 0 thì $f'(x)$ mang dấu dương và $f(x)$ có giá trị tăng dần từ 1 tới 3. - Khi x chạy từ 0 tới 2 thì $f'(x)$ mang dấu âm và $f(x)$ có giá trị giảm dần từ 3 tới 1. - Khi x chạy từ 2 tới $+\infty$ thì $f'(x)$ mang dấu dương và $f(x)$ có giá trị tăng từ 1 tới $+\infty$. <p>Hàm số đã cho nghịch biến trên khoảng nào dưới đây?</p> <p>A. (0; 2) B. (-2; 0) C. (0; $+\infty$) D. (2; $+\infty$)</p>
x	$-\infty$	-2	0	2	$+\infty$															
$f'(x)$	$-$	0	$+$	0	$-$															
$f(x)$	$+\infty$	1	3	1	$+\infty$															
<p>Điểm trong của khối đa diện là:</p> <p>A. Điểm thuộc khối đa diện. B. Điểm thuộc hình đa diện. C. Điểm thuộc khối đa diện nhưng không thuộc hình đa diện giới hạn khối đa diện ấy. D. Điểm không thuộc hình đa diện.</p>	<p>Điểm trong của khối đa diện là:</p> <p>A. Điểm thuộc khối đa diện. B. Điểm thuộc hình đa diện. C. Điểm thuộc khối đa diện nhưng không thuộc hình đa diện giới hạn khối đa diện ấy. D. Điểm không thuộc hình đa diện.</p>	<p>Biết rằng: Một hình H cùng với các điểm nằm trong hình H được gọi là khối đa diện giới hạn bởi hình H.</p> <p>Điểm trong của khối đa diện là:</p> <p>A. Điểm thuộc khối đa diện. B. Điểm thuộc hình đa diện. C. Điểm thuộc khối đa diện nhưng không thuộc hình đa diện giới hạn khối đa diện ấy. D. Điểm không thuộc hình đa diện.</p>																		

lớn, gói 1 (4.62 điểm) và gói 2 (7.44 điểm). Như vậy, ChatGPT có thể làm tốt công việc đưa ra các gợi ý, hướng dẫn để học sinh hình thành ý tưởng phục vụ cho việc giải Toán. Tuy nhiên, để khai thác ChatGPT đạt hiệu quả ở góc độ này, người học cần nắm được các kiến thức nền tảng và phải cẩn trọng khi sử dụng các kết quả do ChatGPT đưa ra. Học sinh chỉ nên sử dụng ChatGPT như một nguồn tham khảo và các em cần có kỹ năng phân biệt, kiểm chứng lại các kết quả biến đổi, suy luận hay tính toán mà hệ thống đưa ra, bởi thực tế cho thấy rằng, ChatGPT chưa thực sự thể hiện tốt các kỹ năng đó. Điền hình cho nhận xét này, chúng tôi nhận thấy, ChatGPT thường mắc lỗi ở các phép biến đổi cơ bản và cả các bước tính toán, Hình 5 là một ví dụ như vậy.

Thông thường, các máy tính toán sẽ luôn đưa ra kết quả gần như chính xác tuyệt đối, độ tin cậy cao. Tuy nhiên, ChatGPT có thể sử dụng cơ chế tính toán khác, không ổn định, dẫn đến việc vẫn có các lỗi sai về tính toán cơ bản xuất hiện với tần suất không nhỏ. Bên cạnh đó, khi đặt các câu hỏi môn Toán cho ChatGPT, chúng ta có thể nhận được lời giải với các nội dung kiến thức vượt quá nội dung hướng đến, chẳng hạn như ChatGPT sử dụng đạo hàm hay các lý thuyết cao cấp hơn để giải các bài Toán lớp 9. Do vậy, khi sử dụng ChatGPT để tìm gợi ý, học sinh nên nêu ra một số nội dung kiến thức liên quan để công cụ này hướng đến việc tìm lời giải xoay quanh các nội dung kiến thức đó. Đối với môn Ngữ văn, ChatGPT thường mắc cả những lỗi sai cơ bản như nhầm tên tác giả và tác phẩm, viết đoạn văn

Tác phẩm đó là "Lão Hạc" của tác giả Tô Hoài. Trong tác phẩm này, tác giả đã sử dụng các chi tiết về ngoại hình của nhân vật Lão Hạc để khắc họa tâm lý của ông, từ đó tạo nên một hình ảnh đầy sức mạnh và sâu sắc về một người đàn ông già có cô độc và đầy những mâu thuẫn tâm lý. Tác phẩm "Lão Hạc" đã trở thành một tác phẩm văn học kinh điển và là một trong những tác phẩm được giảng dạy nhiều nhất trong chương trình Ngữ văn Trung học cơ sở.

Tác phẩm trong chương trình Ngữ văn Trung học cơ sở mà khắc họa tâm lý của nhân vật qua ngoại hình rất thành công là "Chiếc lá cuối cùng" của nhà văn O. Henry. Tác phẩm này miêu tả sự thay đổi trong ngoại hình của nhân vật Johnsy và cảm xúc của cô sau khi mắc bệnh. O. Henry là bút danh của tác giả William Sydney Porter, một nhà văn người Mỹ, được biết đến với những truyện ngắn với cách kể chuyện hài hước và khéo léo.

Trong chương trình Ngữ văn Trung học cơ sở, có một tác phẩm khắc họa tâm lý của nhân vật qua ngoại hình rất thành công đó là "Chí Phèo" của tác giả Nam Cao.

Trong tác phẩm này, nhân vật Chí Phèo được miêu tả với ngoại hình giàu tâm lý, khắc họa chi tiết từng nếp nhăn trên mặt, từng gợn tóc bạc trên đầu và từng cử chỉ khép khiếm khi đi lại. Từ những chi tiết này, độc giả có thể cảm nhận được sự đau khổ, chịu đựng và bất hạnh của Chí Phèo, từ đó độc giả được đưa vào tâm lý của nhân vật và đồng cảm với anh ta.

Tác phẩm trong chương trình Ngữ văn Trung học cơ sở khắc họa tâm lý của nhân vật qua ngoại hình rất thành công và được nhiều người yêu thích là truyện "Lão Hạc" của tác giả Nam Cao. Trong truyện, nhân vật chính là Lão Hạc được miêu tả chi tiết về ngoại hình, từ đó giúp độc giả hiểu rõ hơn về tâm lý, tính cách của nhân vật này.

Hình 6: Minh họa phần trả lời của ChatGPT trong đề kiểm tra môn Ngữ văn lớp 9

thành bài văn và ngược lại, không đảm bảo các yêu cầu nêu ra trong lệnh hỏi, nội dung phân tích sơ sài (xem Hình 6). Việc sai những kiến thức cơ bản như vậy, nếu người dùng hoàn toàn phụ thuộc vào ChatGPT thì sẽ gây ra những ảnh hưởng không tốt đến nhận thức của người học.

2.2.3. Điểm mạnh và điểm hạn chế của ChatGPT trong thực hiện bài kiểm tra qua một số mẫu minh họa bài làm

Như đã đề cập ở trên, ChatGPT có những nhược điểm mang tính hệ thống là không ổn định và ngẫu nhiên. Các câu trả lời của ChatGPT có xác suất sai hoặc chưa hợp lí tương đối cao. Điều này có thể do mỗi lần sử dụng thuật toán, ChatGPT thường tổng hợp và khái quát thông tin ở nhiều nguồn khác nhau nên nhiều khi không chính xác. Những ví dụ dưới đây minh họa cho các phần làm tốt và chưa tốt của ChatGPT trong các bài kiểm tra môn Ngữ văn và môn Toán.

Khi được đặt lệnh cho câu hỏi: “*Trong Chương trình Ngữ văn Trung học cơ sở có một tác phẩm khắc họa tâm lí của nhân vật qua ngoại hình rất thành công. Đó là tác phẩm nào? Tác giả là ai?*” (Ngữ văn 9), các tài khoản ChatGPT đưa ra các câu trả lời khác nhau, trong đó có câu trả lời đúng, đúng một phần hoặc sai hoàn toàn. Các lỗi sai được phát hiện như kể tên tác phẩm ở cấp học khác, ghép nhầm tên tác giả và tác phẩm hoặc lựa chọn chưa phù hợp.

Ngược lại, cũng có phần trả lời thể hiện khả năng vượt trội của ChatGPT trong việc khái quát hoá kiến thức, phân tích, tổng hợp và diễn đạt. Hình 7 là ví dụ về khả năng xử lí câu nghị luận xã hội trong đề Ngữ văn (xem Hình 7).

Trong phần trả lời cho câu hỏi nghị luận xã hội ở trên, ChatGPT đưa ra giải thích khá phù hợp, thể hiện quan điểm tương đối đa dạng, sâu sắc. Tuy nhiên, những đánh giá được đưa ra thường mang tính phổ quát, chung chung, thay vì gắn liền với trải nghiệm của người viết. Một ví dụ khác dưới đây cho thấy khả năng giới thiệu, phân tích, tổng hợp tốt của ChatGPT trong viết mở bài và kết bài của đề Nghị luận văn học Ngữ văn 12: “*Phân tích vẻ đẹp hình tượng sông Đà trong đoạn trích (được cung cấp). Từ đó, nhận xét về cái tôi tài hoa trong tùy bút Nguyễn Tuân*” (xem Hình 8).

Trong phần mở bài và kết bài tương chừng như đầy đủ và hoàn thiện của ChatGPT vẫn thấy xuất hiện lỗi diễn đạt và lỗi về tính chính xác của thông tin, ví dụ như nhầm tên tập tùy bút “*Sông Đà*” thành “*Con sông Đà*”.

Đối với môn Toán, ChatGPT cũng thể hiện được khả năng trình bày tự luận tương đối tốt đối với các câu hỏi ở mức độ nhận biết, thông hiểu. Các bước giải có thể được trình bày tương đối rõ ràng, chặt chẽ như được thể hiện ở Hình 9.

Mặc dù vậy, ChatGPT vẫn còn gặp nhiều hạn chế

ANH/Chị hiểu thế nào về phương châm sống: “Việc mình làm được thì đừng để người khác”.

Phương châm sống “Việc mình làm được thì đừng để người khác” thể hiện tinh thần tự lập, trách nhiệm cá nhân và lòng tự tôn của mỗi người. Đó là một cách nhìn nhận cuộc sống và trách nhiệm của bản thân, cho rằng mỗi người có khả năng, tài năng và kinh nghiệm để thực hiện những việc mình đang làm một cách tốt nhất.

Phương châm sống này cũng có thể thể hiện sự độc lập, chủ động và sáng tạo trong suy nghĩ và hành động của mỗi người. Nó khuyến khích mỗi người tự tin và khát khao đóng góp vào xã hội và tránh hoàn toàn phụ thuộc vào người khác.

Tuy nhiên, phương châm này cũng cần được đánh giá một cách cân nhắc và đúng đắn. Nếu quá chú trọng vào việc làm một mình, có thể dẫn đến sự cô độc, thiếu sự hợp tác và giao lưu với người khác. Một số việc cần sự hỗ trợ và giúp đỡ của người khác để có thể thực hiện tốt hơn.

Tóm lại, phương châm “Việc mình làm được thì đừng để người khác” có thể là một nguồn động lực và sự khát khao vươn lên trong cuộc sống của mỗi người, tuy nhiên cần được áp dụng đúng mức và đánh giá một cách khách quan để có hiệu quả tốt nhất.

Hình 7: Minh họa phần trả lời của ChatGPT trong đề kiểm tra môn Ngữ văn lớp 12

Nguyễn Tuân là một trong những tác giả văn học Việt Nam nổi tiếng với tài hoa văn chương đặc biệt, sắc bén và tinh tế. Tác phẩm “Người lái đò sông Đà” của ông là một trong những tác phẩm đặc sắc và được yêu thích nhất của văn học Việt Nam. Đoạn trích dưới đây từ tác phẩm này đã rất tài tình và ấn tượng khi miêu tả vẻ đẹp của sông Đà một cách chân thật và sâu sắc. Ta có thể thấy được sự tài hoa của Nguyễn Tuân trong việc sử dụng ngôn ngữ và hình ảnh để tạo nên những tác phẩm văn học tuyệt vời. Trong bài viết này, chúng ta sẽ phân tích vẻ đẹp hình tượng sông Đà và đánh giá tài hoa của Nguyễn Tuân trong việc sáng tác văn học.

Tóm lại, qua tùy bút “Con Sông Đà”, Nguyễn Tuân đã cho thấy rõ cái tôi tài hoa của mình thông qua khả năng tạo hình ảnh sống động, miêu tả chân thật và truyền tải cảm xúc sâu sắc. Tác phẩm của ông không chỉ tạo được ấn tượng sâu sắc với người đọc mà còn khẳng định vị thế của ông là một trong những tác giả văn học tài hoa của Việt Nam. Trong đoạn trích này, Nguyễn Tuân đã tài hoa thể hiện vẻ đẹp của con sông Đà một cách sâu sắc và tinh tế. Tác phẩm “Người lái đò sông Đà” đã tạo ra một bức tranh sống động về cảnh quan và văn hóa dân tộc ở vùng sông núi phía Bắc. Với một lối viết tươi mới, đầy cảm xúc và nhân văn, đoạn trích này đã thể hiện được sự gợi cảm và tinh tế của tác giả, từ đó gợi lại cho người đọc những cảm nhận sâu sắc về vẻ đẹp của con sông Đà.

Hình 8: Minh họa phần trả lời của ChatGPT trong đề kiểm tra môn Ngữ văn lớp 12

Cho hàm số

$$y = x^\alpha (\alpha \in \mathbb{R})$$

. Trong các khẳng định sau, khẳng định nào là đúng:

A. Đồ thị hàm số luôn đi qua điểm (0;0) và (1;1).
 B. Đồ thị hàm số luôn đi qua điểm (1;1).
 C. Hàm số luôn nghịch biến trên tập xác định của nó.
 D. Hàm số luôn đồng biến trên tập xác định của nó.

Trong số các khẳng định được đưa ra, khẳng định đúng là:

B. Đồ thị hàm số luôn đi qua điểm (1;1).

Giải thích:

A. Khẳng định này không đúng vì nếu $\alpha \leq 0$ thì hàm số sẽ không đi qua điểm (0;0) vì nó không xác định ở $x = 0$.

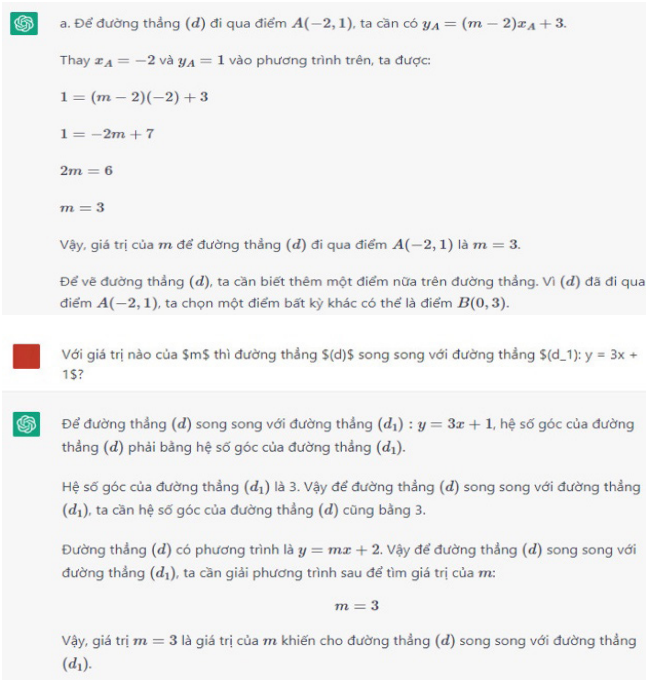
C. Khẳng định này không đúng vì nếu $\alpha < 0$ thì hàm số sẽ không nghịch biến trên tập xác định của nó vì khi x thay đổi từ 0 đến 1 thì hàm số tăng lên.

D. Khẳng định này cũng không đúng vì khi $\alpha < 0$ thì hàm số sẽ đồng biến trên tập xác định của nó vì khi x tăng thì giá trị của hàm số giảm.

B. Khẳng định này đúng vì khi $x = 1$, ta có $y = 1^\alpha = 1$. Do đó, đồ thị của hàm số sẽ đi qua điểm (1;1) với mọi giá trị của α .

Hình 9: Minh họa phần trả lời của ChatGPT trong đề kiểm tra môn Toán lớp 12

trong quy trình giải bài tự luận, có thể do chưa được huấn luyện trước đó. Chẳng hạn, khi thực hiện giải các bài toán liên quan đến ẩn số, ChatGPT không chú ý đến các bước cần kiểm tra, đối chiếu với điều kiện xác định; hay khi giải bài toán liên quan đến tính chất song song của hai đường thẳng trên mặt phẳng tọa độ, ChatGPT chỉ quan tâm đến điều kiện cần (hệ số góc bằng nhau) mà không thực hiện kiểm tra với điều kiện đủ để tránh trường hợp hai đường thẳng trùng nhau (xem Hình 10).



Hình 10: Minh họa phần trả lời của ChatGPT trong đề kiểm tra môn Toán lớp 9

Như vậy, xác suất ChatGPT cung cấp câu trả lời đúng sai, hợp lý hay chưa hợp lý phụ thuộc vào tính chất câu hỏi, khả năng đưa lệnh hỏi và tính ngẫu nhiên trong các lần sinh kết quả của máy.

2.3. Thảo luận

Kết quả nghiên cứu ở trên chỉ ra rằng, ChatGPT là công cụ có khả năng nhất định trong việc thực hiện các bài kiểm tra hoặc đề thi. Kết quả của nghiên cứu này đồng thuận với những nghiên cứu gần đây của nhóm nghiên cứu Gilson (2023) và Cotton (2023) [5], [13]. Điểm trung bình của ChatGPT thấp hơn so với điểm trung bình của học sinh. Kết quả này tương đối đồng thuận với kết quả thực nghiệm về Chat GPT trong mối tương quan với học sinh lớp 6 của Singapore. Tuy nhiên, thành tích của ChatGPT trong nghiên cứu này cao hơn so với nghiên cứu được thực hiện ở Singapore (Toán 16/110 điểm; Tiếng Anh 11/20 điểm, Khoa học 21/100 điểm) [15].

Thành tích của ChatGPT không ổn định về mức độ

chính xác/phù hợp của câu trả lời. Mỗi lần nhân bản, ChatGPT lại đưa ra những câu trả lời khác với độ chính xác khác nhau. Điều này cũng phù hợp với những lí giải được nêu ra trong các nghiên cứu gần đây [5], [15]. Với đề Ngữ văn, ChatGPT có nhiều nguy cơ xác định sai trọng tâm đề dẫn đến hiểu sai đề. Chính vì thế, người học cần có kiến thức và kĩ năng, hiểu vấn đề mình đang hỏi để điều hướng và đánh giá được tính xác thực và hợp lí của câu trả lời. Những câu hỏi ChatGPT xử lí tốt là những câu có chứa sẵn nội dung trong phần đọc hiểu, hoặc những câu trả lời mang tính diễn ngôn, đưa ra quan điểm, lập luận. Ngược lại, những câu đòi hỏi nhận biết và phân tích các kiến thức tiếng Việt, kiến thức lí luận văn học hoặc những câu phức hợp thì ChatGPT chưa thực sự làm tốt. Những câu nghiêng về việc phát hiện và phân tích nội dung có tỉ lệ đúng cao hơn những câu như phát hiện biện pháp và đặc điểm nghệ thuật. Có nghĩa là, những nội dung mang tính phổ quát, chung chung được máy khái quát hoá hiệu quả hơn.

Đối với các nhiệm vụ trong đề kiểm tra Toán, ChatGPT giải tương đối tốt các câu hỏi ở mức độ nhận biết và thông hiểu. Tuy nhiên, với các bài đòi hỏi những kĩ năng phức tạp hoặc linh hoạt hơn với nhiều bước hoặc yêu cầu khả năng tổng hợp kiến thức thì ChatGPT chưa làm tốt. Đôi khi, ChatGPT nêu được các nội dung kiến thức liên quan và thuật toán nhưng không có khả năng vận dụng để đưa ra câu trả lời phù hợp. Bên cạnh đó, đối với các câu hỏi về hình học hoặc các câu cần biến đổi đại số phức tạp thì ChatGPT chưa xử lí được. Chính vì thế, nên coi ChatGPT như một công cụ tham khảo, hỗ trợ ý tưởng thay vì tin tưởng hoặc phụ thuộc hoàn toàn vào công cụ này. Nói chung, người dùng cần có sự điều hướng và chọn lọc, cũng như khả năng đánh giá và phán đoán trong mỗi vấn đề đưa ra tương tác với ChatGPT dựa trên những kiến thức và kĩ năng của bản thân.

Như vậy, trong kiểm tra đánh giá, dù công cụ ChatGPT có thể hỗ trợ ở mức độ nhất định nhưng để đạt được kết quả cao vẫn chủ yếu là do thực lực của người học. Kết quả nghiên cứu cũng cho thấy, công cụ ChatGPT có lợi thế khái quát và tổng hợp, nhân bản không giới hạn, đưa ra câu trả lời phong phú cho cùng một câu hỏi, có thể hữu ích trong việc tính toán độ tin cậy của đề, dự đoán các khả năng làm bài của học sinh. Với những điểm mạnh của ChatGPT trong khả năng diễn đạt, lập luận, ChatGPT có thể được sử dụng là công cụ hỗ trợ khả năng viết luận, đọc hiểu nội dung, công cụ tham khảo hoặc kiểm tra, đánh giá năng lực viết của học sinh. Điểm này trùng hợp với phát hiện được chỉ ra từ nghiên cứu của Basic và cộng sự (2023) [16]. Đối với môn Toán, có thể sử dụng ChatGPT trong việc hỗ trợ giải

thích, tìm kiếm ý tưởng, cách thức hoặc đối chiếu để thực hiện các nhiệm vụ Toán hiệu quả.

3. Kết luận

ChatGPT được kì vọng tạo ra đột phá trong nhiều lĩnh vực. Với đặc trưng của lĩnh vực Giáo dục, việc ứng dụng ChatGPT cần thận trọng, một mặt kế thừa những điểm mạnh để nâng cao chất lượng dạy và học, mặt khác cần nghiên cứu, đưa ra các giải pháp, hành động để phát huy ưu điểm và giảm thiểu các tác động tiêu cực trong giáo dục (nếu có). Chính vì thế, các nghiên cứu thử nghiệm, đánh giá rất cần thiết trong thời điểm này.

Kết quả thử nghiệm công cụ ChatGPT cho thấy, ChatGPT có khả năng thực hiện các bài kiểm tra Ngữ văn và Toán nhưng với điểm trung bình thấp hơn bài làm thực tế của học sinh. Độ chênh lệch điểm trung bình của học sinh và ChatGPT phụ thuộc vào môn học và mức độ thích ứng của công cụ đối với đề kiểm tra. Chất lượng câu trả lời của ChatGPT như tính chính xác và mức độ phù hợp, không ổn định khi thử trên các tài khoản hoặc các lần sinh kết quả khác nhau. Phản hồi từ ChatGPT có thể chứa những lỗi sai kiến thức cơ bản, sai quy trình hoặc diễn đạt. Ngoài ra, chất lượng câu trả lời của ChatGPT phụ thuộc nhiều bởi mức độ cụ thể, rõ ràng và tính điều hướng của lệnh hỏi. Điều này có nghĩa nếu phụ thuộc vào công cụ thì sẽ gây ra những hệ lụy không nhỏ về nhận thức và phát triển năng lực của

người dùng. Nhưng nếu sử dụng ChatGPT như là một công cụ hỗ trợ việc tự học thì sẽ hữu ích. Nhìn chung, người dùng cần chủ động, cân trọng đánh giá mức độ đúng sai và tính phù hợp của các câu trả lời do máy đưa ra.

Nghiên cứu chỉ ra rằng, việc sử dụng ChatGPT có hiệu quả tích cực trong việc tạo ra những bài kiểm tra chất lượng bằng việc sử dụng công cụ này để tính toán trước về khả năng trả lời của học sinh. Nghiên cứu này giúp cho các nhà giáo dục, giáo viên có thêm căn cứ để xây dựng đề kiểm tra, đánh giá phù hợp trong bối cảnh bùng nổ các công nghệ AI hỗ trợ học tập. Từ đây, giáo viên cũng có định hướng sử dụng ChatGPT như một phương tiện dạy học tích cực hoặc có những giải pháp phòng tránh tiêu cực, gian lận trong hoạt động kiểm tra, đánh giá. Đối với học sinh, cần hiểu bản chất của công cụ để xác định mục đích và cách sử dụng phù hợp, tránh tình trạng lạm dụng, phụ thuộc.

Cuối cùng, các tác giả hi vọng các nghiên cứu về ChatGPT trong giáo dục sẽ được thúc đẩy hơn nữa. Đây mới là một trong những nghiên cứu đầu tiên về ChatGPT ở Việt Nam nên chỉ phản ánh một góc độ nhất định về khả năng của công cụ này trong thời điểm hiện tại. Do đó, rất cần thêm nhiều nghiên cứu về các phương diện khác nhau trong việc ứng dụng công cụ ChatGPT để khai thác được tiềm năng, tận dụng cơ hội và hạn chế các thách thức mà công cụ mang lại.

Tài liệu tham khảo

- [1] Das, K. (2019), *The role and impact of ICT in improving the quality of education: An overview*, International Journal of Innovative Studies in Sociology and Humanities, 4(6), 97-103.
- [2] Xie, H., Chu, H. C., Hwang, G. J., & Wang, C. C. (2019), *Trends and development in technology-enhanced adaptive/personalized learning: A systematic review of journal publications from 2007 to 2017*, Computers & Education, 140, 103599.
- [3] Qin, H., & Wang, G. (2022, January), *Benefits, challenges and solutions of artificial intelligence applied in education*, In 2022 11th International Conference on Educational and Information Technology (ICEIT), pp.62-66, IEEE.
- [4] OpenAI, (2023), *ChatGPT: optimizing language models for dialogue*, 2022 Nov 30, URL: <https://openai.com/blog/chatgpt/> [accessed 2022-1-22].
- [5] Gilson, A., Safranek, C. W., Huang, T., Socrates, V., Chi, L., Taylor, R. A., & Chartash, D. (2023), *How does ChatGPT perform on the United States medical licensing examination? The implications of large language models for medical education and knowledge assessment*, JMIR Medical Education, 9(1), e45312.
- [6] Gordijn, B., & Have, H. T. (2023), *ChatGPT: evolution or revolution?* Medicine, Health Care and Philosophy, 1-2.
- [7] George, A. S., & George, A. H. (2023), *A Review of ChatGPT AI's Impact on Several Business Sectors*, Partners Universal International Innovation Journal, 1(1), 9-23.
- [8] Markel, J. M., Opferman, S. G., Landay, J. A., & Piech, C. (2023), *GPTeach: Interactive TA Training with GPT Based Students*, arXiv:2302.04818 [cs.CY].
- [9] Kwon, T. (2023), *Interfaces for Personalized Language Learning with Generative Language Models* (Doctoral dissertation, Columbia University).
- [10] Phillips, T., Saleh, A., Glazewski, K. D., Hmelo-Silver, C. E., Mott, B., & Lester, J. C. (2022), *Exploring the use of GPT-3 as a tool for evaluating text-based collaborative discourse*, Companion Proceedings of the 12th, 54.
- [11] Eaton, S. E., Brennan, R., Wiens, J., & McDermott, B. (2023, January 25), *Artificial intelligence and academic integrity: The ethics of teaching and learning with algorithmic writing technologies*, <https://prism.ualgary.ca/handle/1880/115769>.
- [12] Samantha Murphy Kelly, (2022), *ChatGPT passes exams from law and business schools*, Retrieved on 15th February 2022, <https://edition.cnn.com/2023/01/26/tech/chatgpt-passes-exams/index.html>.
- [13] Cotton, D., Cotton, P., & Shipway, J. R. (2023, January 10), *Chatting and Cheating*, Ensuring academic integrity in the era of ChatGPT, <https://doi.org/10.35542/osf.io/mrz8h>.
- [14] Bộ Giáo dục và Đào tạo, (2023), *Toạ đàm "ChatGPT, Trí*

tuệ nhân tạo - Lợi ích và thách thức đối với giáo dục”, Truy xuất ngày 14 tháng 02 năm 2022, <https://moet.gov.vn/tintuc/Pages/tin-tong-hop.aspx?ItemID=8407>.

- [15] Mehul Reuben Das, (2023), *Not smarter than a 6th grader: ChatGPT fails Singapore's 6th-grade maths and science exams*, Retrieved on 10th February 2022, <https://>

www.firstpost.com/world/chatgpt-fails-singapore-6th-grade-maths-and-science-exams-12189482.html.

- [16] Basic, Z., Banovac, A., Kruzic, I., & Jerkovic, I, (2023), *Better by you, better than me, chatgpt3 as writing assistance in students essays*, arXiv preprint arXiv:2302.04536.

PERFORMANCE OF CHATGPT IN CONDUCTING END-TERM TESTS IN MATHEMATICS AND VIETNAMESE-LITERATURE AT SECONDARY SCHOOL LEVEL: RESEARCH RESULTS AND SOME INITIAL RECOMMENDATIONS

Le Anh Vinh¹, Bui Thi Dien*²,
Le Quang Quan³, Vu Van Luan⁴

¹ Email: vinhla@vnies.edu.vn

* Corresponding author

² Email: dienbt@vnies.edu.vn

³ Email: quanlq@vnies.edu.vn

The Vietnam National Institute of Educational Sciences
101 Tran Hung Dao, Hoan Kiem, Hanoi, Vietnam

⁴ Email: vuvanluanftu.k50@gmail.com

Green Education Technology Joint Stock Company
No.3, Land 814, Lang street, Dong Da, Hanoi, Vietnam

ABSTRACT: *ChatGPT has garnered attention worldwide due to its novelty and exceptional processing capabilities. This tool is expected to be applied in many fields, including education. This study evaluates the results of ChatGPT in conducting end-of-term tests for secondary school Vietnamese Literature and Mathematics subjects in comparison with the actual test results of students in grade 9 and 12. The research shows that ChatGPT is capable of performing tests, but the quality of answers is unstable and depends on many factors. The study provides a specific perspective, providing useful information for educational administrators, teachers, students, which can serve as a foundation to use this tool appropriately and effectively.*

KEYWORDS: ChatGPT, artificial intelligence, assessment, Mathematics, Vietnamese Literature.