Contents lists available at ScienceDirect

Currents in Pharmacy Teaching and Learning

Is ChatGPT reliable and accurate in answering pharmacotherapy-related inquiries in both Turkish and English?

Nur Ozturk [a,b,*], Irem Yakak [b], Melih Buğra Ağ [b,c], Nilay Aksoy [a]

[a] Altinbas University, School of Pharmacy, Department of Clinical Pharmacy, Istanbul, Turkey
[b] Istanbul Medipol University, Graduate School of Health Sciences, Clinical Pharmacy PhD Program, Istanbul, Turkey
[c] Istanbul Medipol University, School of Pharmacy, Department of Clinical Pharmacy, Istanbul, Turkey

ARTICLE INFO

ABSTRACT

Introduction: Artificial intelligence (AI), particularly ChatGPT, is becoming more and more prevalent in the healthcare field for tasks such as disease diagnosis and medical record analysis. The objective of this study is to evaluate the proficiency and accuracy of ChatGPT in different domains of clinical pharmacy cases and queries.

Methods: The study NAPLEX® Review Questions, 4th edition, pertaining to 10 different chronic conditions compared ChatGPT's responses to pharmacotherapy cases and questions obtained from McGraw Hill's, alongside the answers provided by the book's authors. The proportion of correct responses was collected and analyzed using the Statistical Package for the Social Sciences (SPSS) version 29.

Results: When tested in English, ChatGPT had substantially higher mean scores than when tested in Turkish. The average accurate score for English and Turkish was 0.41 ± 0.49 and 0.32 ± 0.46, respectively, $p = 0.18$. Responses to queries beginning with "Which of the following is correct?" are considerably more precise than those beginning with "Mark all the incorrect answers?" 0.66 ± 0.47 as opposed to 0.16 ± 0.36; $p = 0.01$ in English language and 0.50 ± 0.50 as opposed to 0.14 ± 0.34; $p < 0.05$ in Turkish language.

Conclusion: ChatGPT displayed a moderate level of accuracy while responding to English inquiries, but it displayed a slight level of accuracy when responding to Turkish inquiries, contingent upon the question format. Improving the accuracy of ChatGPT in languages other than English requires the incorporation of several components. The integration of the English version of ChatGPT into clinical practice has the potential to improve the effectiveness, precision, and standard of patient care provision by supplementing personal expertise and professional judgment. However, it is crucial to utilize technology as an adjunct and not a replacement for human decision-making and critical thinking.

Introduction

Advances in machine learning and artificial intelligence (AI) have created several opportunities for research and entrepreneurship. The medical industry is seeing an increase in the usage of AI for tasks such as illness diagnosis and medical record analysis.[1]

---

* Corresponding author at: Altinbas University, School of Pharmacy, Department of Clinical Pharmacy, Istanbul, Turkey.
E-mail addresses: nur.ozturk@altinbas.edu.tr (N. Ozturk), iremsubasi777@gmail.com (I. Yakak), melih.ag@medipol.edu.tr (M.B. Ağ), nilay.aksoy@altinbas.edu.tr (N. Aksoy).

ChatGPT is short for "chat generative pretrained transformer. "The name refers to a unique language framework developed by OpenAI (OpenAI, Inc.). The introduction of ChatGPT has had a significant impact on the sector of education. ChatGPT, being an advanced language model, can create responses that resemble human-like ones, making it a helpful tool for students and instructors to obtain answers to various inquiries and questions.[2] Moreover, ChatGPT provides medical knowledge to the patients in diverse clinical fields such as cardiology, endocrinology and gynecology.[3] It is believed that ChatGPT can improve the learning experience and help students in grasping complicated ideas faster by offering immediate and precise responses to their questions.[4]

Gilson's utilization of ChatGPT during the United Medical Licensing Examination revealed the model's ability to attain passing scores comparable to those achieved by a third-year medical student in the United States. In addition, they highlighted ChatGPT's ability to provide sensible and educative replies, showing that it may be utilized as a dynamic resource for enhancing students' learning in medical education.[5] According to Huh, students might receive assistance in comprehending complex subjects by utilizing ChatGPT to ask questions and receive explanations. In addition to the fact that ChatGPT offers lessons and assistance with assignments, its knowledge and explanations during parasitology exams were found to be insufficient compared to those provided by Korean medical students.[6] Questions on pharmacokinetics, action mechanisms, clinical usage, side effects, interactions, contraindications, and tests of knowledge were administered using ChatGPT by Nisar et al. from a pharmacology textbook. The findings demonstrated that the precision and pertinence of ChatGPT responses were satisfactory for students to employ in self-study, however enhancements were needed to answer the intricate cases.[7] In their study, Wang et al. conducted a comparative analysis of the efficacy of ChatGPT in successfully completing the Taiwanese pharmacist licensing examination which includes questions regarding pharmacology, pharmaceutical chemistry, pharmacognosy, pharmacy administration, clinical pharmacy and therapeutics, dispensing pharmacy and pharmacy law. They tested its performance in both Chinese and English languages. Their findings indicated that ChatGPT achieved a greater accuracy rate on the English examination compared to the Chinese exam in all areas, however the extent of this difference varied depending on the specific subject.[8]

The use of AI in clinical pharmacy has the ability to substantially improve both operational and therapeutic efficiency. Typically, clinical pharmacists have a vital function in guaranteeing patient safety, optimizing drug therapy, and delivering care that focuses on the patient. They have the expertise and proficiency necessary to provide clinical pharmacy services to the patients and healthcare providers.[9] Systems driven by artificial intelligence have shown interest in assisting clinical pharmacists in resolving complex situations involving chronic illnesses.[10]

By evaluating ChatGPT's responses to 10 distinct chronic disease cases, this study aimed to assess the reliability and accuracy of ChatGPT across a range of pharmacotherapy cases and queries. Additionally, Our university teaches pharmacy in English, but most other universities educate in Turkish. We anticipate that some students may write questions in English and others in Turkish. Our goal is to determine the difference in answers between the two languages.

Methods

Data source

The dataset utilized in this study comprises a collection of one hundred inquiries extracted from the fourth edition of McGraw Hill's NAPLEX® Review Questions, sourced from the Access Pharmacy database within Altınbaş University's electronic library. The study employed ChatGPT version 4, the most recent version of the ChatGPT model, which was publicly released on March 23, 2023.

Study design

A total of one hundred questions were distributed throughout ten sections, with each section comprising ten multiple-choice questions pertaining to a distinct chronic disease query. The selection of queries was conducted in order to assess the proficiency and aptitude of ChatGPT in dealing with different chronic diseases data. The number of questions included in the study was determined by calculating the sample size, which was based on the population size of 1400 questions, a margin of error of 5%, and a confidence level of 95%.

The queries provided pertained to the following chronic conditions: chronic heart failure, inflammatory bowel disease, cystic fibrosis, thyroid disorders, diabetes mellitus, osteoarthritis, asthma, chronic kidney diseases, hypertension, and dyslipidemia. The chronic diseases that were selected for testing were chosen in order to address the curriculum of the pharmacotherapy lectures at our university.

Each question was asked to ChatGPT in English and Turkish language with 2 different formats:

The first question format employed in this study was "which of the following is correct.?" in both the English and Turkish languages. This format consisted of a single right choice.

The second question structure had been modified to include the instruction "mark all the incorrect answers..." in both English and Turkish languages, so accommodating the possibility of multiple correct options. The task of translating all questions from English to Turkish was undertaken by two authors (M.A, I·Y) who are very proficient in both the Turkish and English languages. Subsequently, their translation was subjected to comparison, and any inconsistencies were verified by a third author (N. O). The 'New Chat' function was utilized to input each question into ChatGPT. The queries and responses produced by ChatGPT were recorded in both the English and Turkish languages.

The responses generated by ChatGPT were evaluated by comparing them to the standard answers. If the response from ChatGPT matched the standard correct answer, it was assigned a score of 1. Conversely, if the response was incorrect and did not align with the

standard correct answer, it was assigned a score of 0. The potential scores for each chronic condition range from 0 to 10. The overall values for all diseases were derived by calculating the mean values of the 10 different diseases and evaluating the percentage of accuracy. Answers with a correct rate below 50% are often deemed inaccurate. A slightly accurate range is between 50% and 60%, a moderately accurate range is between 60% and 80%, a highly accurate range is between 80% and 95%, and an exceptionally accurate range is >95%.

Statistical analysis

Data analysis was conducted utilizing the Statistical Package for the Social Sciences (SPSS) version 29. Descriptive data and chi-square tests were used. The data is reported as the mean value together with the standard deviation (SD). In order to make the difference more comparable the effect size was calculated by using Cohen's d. Cohen distinguished between small (d = 0.2), medium (d = 0.5), and large (d > 0.8) impact sizes.

Results

For English language the correct answer rate of "Which of the following is correct?" was 65% and 16% for "Mark all the incorrect answers?" question format. The correct answer rate of "Which of the following is correct?" question format in the Turkish language was 50% and 14% for the question format "Mark all the incorrect answers?"

For English language the overall mean score of "Which of the following is correct?" questions, was (0.66 ± 0.47) and it was (0.16 ± 0.36) $p = 0.01$, *d* effect = 1.19 for "Mark all the incorrect answers?". While for the Turkish language it was (0.50 ± 0.50) to (0.14 ± 0.34) $p < 0.05$, *d* effect = 0.84 for Turkish language. Regarding the question format "Which of the following is correct? Cystic fibrosis got the highest score (1.00±0.00) in English language and chronic kidney diseases (0.80±0.42) in Turkish language. Moreover Cystic fibrosis recorded the largest significant impact regarding the same previous question and "Mark all the incorrect answers" question. Table 1 compare the differences between the languages in the same identical question format. In view of comparing the different question format in the English language, there a difference between the mean scores in chronic heart failure, inflammatory bowel disease, diabetes miletus, asthma, chronic kidney disease and hypertension $p < 0.05$ and the effect size was >0.8 in them as well.

For the Turkish language different question format, there was a different between the chronic kidney diseases, hypertension and the overall disease scores $p < 0.05$ and *d*-effect>0.8. Table 2 compare the differences between the question format in the same language.

Discussion

The popularity of artificial intelligence (AI) applications, particularly Chat-GPT, has experienced significant increase in recent decades due to its extensive utilization in both academic and industrial settings.[11] Concerns over ChatGPT and similar AI algorithms' accuracy and usefulness in healthcare have divided opinion. The application of ChatGPT in the areas of patient care and pharmaceutical practice has elicited significant apprehensions. This study assesses the proficiency and accuracy of ChatGPT in addressing inquiries related to several chronic diseases in two different languages, employing two different question formats for each language.

Overall, the average scores of ChatGPT were much higher in English language compared to Turkish language. This highlights the linguistic challenge faced by ChatGPT in comprehending Turkish, hence restricting its usage within communities for whom Turkish is their mother language. Similarly a study conducted by Wang et al., Studied the performance of ChatGPT in the Taiwanese Pharmacist Licensing Examination in 2023 that was conducted in Chinese and English. They found that the mean score harvested by ChatGPT in the English examination was 4.3% higher than the mean score that was harvested in the Chinese test in the first step of the examination. Moreover, in the second step, the score gained in English examination was 25.3% higher than the average Chinese test score. It is possible that translating mistakes are to blame for the language-specific performance differences seen in ChatGPT.[8]

Table 1
The comparison between the English and Turkish responses for the identical question format.

| | Which of the following is correct?
English against Turkish | | | | | | Mark all the incorrect answers?
English against Turkish | | | | | |
| | English | | Turkish | | *P* | d-effect | English | | Turkish | | P | d-effect |
	Mean	Sd	Mean	Sd			Mean	Sd	Mean	Sd		
Chronic heart Failure	0.50	0.52	0.40	0.51	0.67	0.19	0.10	0.31	0.10	0.31	1	0
Inflammatory bowel Disease	0.70	0.48	0.60	0.51	0.66	0.20	0.10	0.31	0.30	0.48	0.29	0.49
Cystic Fibrosis	1.00	0.00	0.60	0.51	0.02	1.10	0.10	0.31	0.60	0.51	0.02	1.18
Thyroid	0.60	0.51	0.40	0.51	0.39	0.39	0.10	0.31	0.00	0.00	0.34	0.45
Diabetes Miletus	0.50	0.52	0.60	0.51	0.67	0.19	0.10	0.31	0.10	0.31	1	0
Osteoarthritis	0.60	0.51	0.40	0.51	0.39	0.39	0.10	0.31	0.10	0.31	1	0
Asthma	0.70	0.48	0.40	0.51	0.19	0.60	0.30	0.48	0.10	0.31	0.29	0.49
Chronic kidney diseases	0.90	0.31	0.80	0.42	0.55	0.27	0.30	0.48	0.10	0.31	0.29	0.49
Hypertension	0.50	0.52	0.30	0.48	0.38	0.39	0.10	0.31	0.00	0.00	0.34	0.45
Dyslipidemia	0.60	0.51	0.50	0.52	0.67	0.19	0.30	0.48	0.20	0.42	0.62	0.22
All diseases	0.66	0.47	0.50	0.50	0.47	0.32	0.16	0.36	0.14	0.34	0.90	0.05

Table 2

The comparison between the two different question format in the same language.

	In English language Which of the following is correct? <u>Against</u> Mark all the incorrect answers?				*P*	d-effect	In Turkish language Which of the following is correct? <u>Against</u> Mark all the incorrect answers?				*P*	d-effect
	Mean	Sd	Mean	Sd			Mean	Sd	Mean	Sd		
Chronic heart Failure	0.50	0.52	0.10	0.31	0.05	0.05	0.40	0.51	0.10	0.31	0.13	0.71
Inflammatory bowel Disease	0.70	0.48	0.10	0.31	<0.001	1.48	0.60	0.51	0.30	0.48	0.19	0.60
Cystic Fibrosis	1.00	0.00	0.10	0.31	1	4.10	0.60	0.51	0.60	0.51	0.21	0
Thyroid	0.60	0.51	0.10	0.31	0.02	1.18	0.40	0.51	0.00	0.00	0.03	1.10
Diabetes Miletus	0.50	0.52	0.10	0.31	0.05	0.93	0.60	0.51	0.10	0.31	0.02	1.18
Osteoarthritis	0.60	0.51	0.10	0.31	0.02	1.18	0.40	0.51	0.10	0.31	0.13	0.71
Asthma	0.70	0.48	0.30	0.48	0.08	0.83	0.40	0.51	0.10	0.31	0.13	0.71
Chronic kidney diseases	0.90	0.31	0.30	0.48	<0.001	1.48	0.80	0.42	0.10	0.31	<0.001	1.89
Hypertension	0.50	0.52	0.10	0.31	0.05	0.93	0.30	0.48	0.00	0.00	0.08	0.88
Dyslipidemia	0.60	0.51	0.30	0.48	0.19	0.60	0.50	0.52	0.20	0.42	0.17	0.63
All diseases	0.66	0.47	0.16	0.36	0.01	1.19	0.50	0.50	0.14	0.34	0.08	0.84

Concerning the question format, ChatGPT achieved better mean scores in "Which of the following is correct?" questions compared to the scores obtained in "Mark all the incorrect answers?" questions. This suggests that chatGPT experiences uncertainty when faced with several options that need selection. It serves as a reminder that evaluating a pharmacist's proficiency should not solely rely on multiple-choice questions, but rather encompass a broader array of assessment methodologies.

Furthermore, our findings indicate that even when the model gives inaccurate answers, the responses consistently provide a coherent justification for the chosen answer, which instills confidence in the student regarding the selected answer.

While our investigation demonstrates that ChatGPT is an unreliable resource for patients and cardiologists seeking information about hypertension, Kenya et al. evaluated the accuracy of ChatGPT in answering clinical questions based on the Japanese Society of Hypertension Guidelines and discovered that ChatGPT could possibly be a useful tool for physicians in hypertension management.[12] Another research assessed the efficacy of ChatGPT in the field of hypertension determined that ChatGPT demonstrated impressive accuracy in answering frequently asked questions concerning hypertension.[13]

.In a study conducted by Meo et al., they found that ChatGPT was able to answer only 52 questions out of 100 questions in diabetes and endocrinology.[14] These results agrees with our findings that ChatGPT is not qualified enough to cover the wide areas of information in medicine and pharmacotherapy especially in diabetes Miletus, chronic heart Failure, thyroid and hypertension.

On the other hand, Roosan et al., investigated the ability of ChatGPT to perform skills needed by a pharmacist as identifying drug interactions and recommending alternatives through real patients' cases. They found that ChatGPT correctly solved 39 out of 39 (100%) patient cases that ranged from simple cases to complex ones. In addition ChatGPT was able to correctly recommend alternatives and detect drug-drug interactions in 97.4% of cases as well as designing a management plan for 100% of cases.[15] On the contrary, a research conducted by Al-Ashwal et al. found that when used to identify DDIs, ChatGPT was able to recognize a greater number of possible interactions with lower levels of accuracy than other AI models. It is particularly prone to false positives, in which it falsely labels certain medication combinations as interacting when they do not.[16]

Furthermore Haung et al., compared the performance of ChatGPT to pharmacists dealing with real clinical cases. ChatGPT demonstrated strong performance in drug counseling (8.77 vs. 9.50, $P = 0.0791$), but showed poor performance in prescription review (5.23 vs. 9.90, $P = 0.0089$), patient medication education (6.20 vs. 9.07, $P = 0.0032$), ADR recognition (5.07 vs. 9.70, $P = 0.0483$), and ADR causality assessment (4.03 vs. 9.73, $P = 0.023$) compared to clinical pharmacists.[17] These data corroborated our results, demonstrating the restricted efficacy of ChatGPT in addressing clinical pharmacy issues and queries. Similarly, Fournier et al., examined the ability of ChatGPT to answer questions from their database and they found that in general no ChatGPT answer was better than those of the pharmacists. The accuracy rate of replies ranged from 30 to 57.1% across different question types, with an overall average of 44.9%. Among the 38 responses that were inaccurate, 20 were incorrect and 18 did not provide any answer.[18] Al-Dujaili et al. assessed the answers of ChatGPT to 20 cases of pharmacotherapy at different points of time. ChatGPT had a success rate of 70.83% in week one, 79.2% in week three, and 75% in week five.[19] Overall, the study found that ChatGPT may generate clinically meaningful pharmacological information, albeit with some variance in accuracy and consistency which counteracts our findings in this study. On the Other hand, Morath et al.**,** opposed the idea that ChatGPT could answer questions as effective as pharmacists which came in consistent with our results in this study.[20] Within the study conducted by Morath et al., a precise set of 50 inquiries pertaining to drugs were gathered and systematically inputted into ChatGPT. Out of the 50 replies received, just 13 provided accurate items and enough information to start managing the situation without any danger of harm to the patient. The majority of responses were either incorrect (38%, $n = 19$) or partially accurate (36%, $n = 18$), without any provided references. High risk of patient damage was evaluated in 26% ($n = 13$) of the scenarios, whereas low risk was estimated in 28% ($n = 14$) of the cases.[20]

Some studies showed that adoption of ChatGPT in healthcare settings can help clinicians choose and optimize patients' prescription therapy. ChatGPT can deliver tailored drug recommendations by analyzing massive volumes of medical data alongside professional knowledge customized to each patient's particular needs. ChatGPT can aid healthcare practitioners in minimizing prescription error and reducing the probability of adverse pharmacological events.[21] Moreover, ChatGPT possesses the capacity to perform intricate functions including notifying physicians about drug interactions and mitigating drug adverse effects.[22] Nevertheless, another study

revealed that almost one-third of the studies assessing the effectiveness and possible use of ChatGPT in the healthcare field encountered difficulties related to the possibility of inaccurate information.[23]

The main strengths of this study is that it examined the ability of ChatGPT to answer questions in 10 different areas of knowledge. However a large number of questions in each area need to be used in the upcoming studies to provide generalizable data. Also the questions should be categorized according to the skill it examines in order to provide a detailed image about the points of strength and points of weakness of ChatGPT. Utilizing the points of strength of ChatGPT is highly recommended for improving the skills of the users.

Conclusion

ChatGPT displayed a moderate level of accuracy while responding to English inquiries, but it displayed a slight level of accuracy when responding to Turkish inquiries, contingent upon the question format. Improving the accuracy of ChatGPT in languages other than English requires the incorporation of several components, including acquiring a broader array of diverse and high-quality data, conducting language-specific fine-tuning, optimizing training protocols, and consistently assessing and enhancing the system. The integration of the English version of ChatGPT into clinical practice has the potential to improve the effectiveness, precision, and standard of patient care provision by supplementing personal expertise and professional judgment. Healthcare practitioners can utilize ChatGPT to gain access to current research and scholarly articles. Furthermore, ChatGPT can produce patient-friendly descriptions of medical issues, available treatments, and preventative steps, encouraging patients to be engaged participants in their own healthcare decision-making. However, it is crucial to utilize technology as an adjunct and not a replacement for human decision-making and critical thinking.

Financial disclosure statement

Authors statement role

Nur Ozturk: Conceptualization, Formal analysis, Investigation, Methodology, Project administration, Resources, Validation, Visualization, Roles/Writing - original draft, and Writing - review & editing.

Irem Yakak: Data curation, Formal analysis, Funding acquisition, Investigation, Writing - review & editing.

Melih Buğra Ağ: Data curation, Formal analysis, Funding acquisition, Investigation, Writing - review & editing.

Nilay Aksoy: Conceptualization, Methodology, Project administration, Resources, Supervision, Validation, Visualization, Roles/Writing - original draft, and Writing - review & editing.

Data availability

The datasets generated during and/or analyzed during the current study are available from the corresponding author upon reasonable request.

References

1. Alofi ES. Evaluating Chatgpt in health diagnostic symptoms. *J Namib Stud*. 2023;26:65–89. https://doi.org/10.59670/jns.v35i.4253.
2. Hamid H, Zulkifli K, Naimat F, Che Yaacob NL, Ng KW. Exploratory study on student perception on the use of chat AI in process-driven problem-based learning. *Curr Pharm Teach Learn*. 2023;15:1017–1025. https://doi.org/10.1016/j.cptl.2023.10.001.
3. Johnson D, Goodman R, Patrinely J, et al. *Assessing the accuracy and reliability of AI-generated medical responses: an evaluation of the chat-GPT model*. 2023. https://doi.org/10.21203/rs.3.rs-2566942/v1.
4. Shihab Sk Rezwan, Sultana Nahida, Samad Abdul. Revisiting the use of ChatGPT in business and educational fields: possibilities and challenges. *Bullet [Internet]*. 2023;2:534–545. Available from: https://journal.mediapublikasi.id/index.php/bullet/article/view/2761.
5. Gilson A, Safranek CW, Huang T, et al. How does CHATGPT perform on the United States medical licensing examination? The implications of large language models for medical education and knowledge assessment. *JMIR Med Educ*. 2023;9. https://doi.org/10.2196/45312.
6. Huh S. Are ChatGPT's knowledge and interpretation ability comparable to those of medical students in Korea for taking a parasitology examination?: a descriptive study. *J Educ Eval Health Prof*. 2023;20:1. https://doi.org/10.3352/jeehp.2023.20.1.
7. Nisar S, Aslam MS. Is chatgpt a good tool for t&cm students in studying pharmacology? *SSRN Electron J*. 2023. https://doi.org/10.2139/ssrn.4324310.
8. Wang YM, Shen HW, Chen TJ. Performance of ChatGPT on the pharmacist licensing examination in Taiwan. *J Chin Med Assoc*. 2023;86:653–658. https://doi.org/10.1097/JCMA.0000000000000942.
9. Alowais SA, Alghamdi SS, Alsuhebany N, et al. Revolutionizing healthcare: the role of artificial intelligence in clinical practice. *BMC Med Educ*. 2023;23:689. https://doi.org/10.1186/s12909-023-04698-z.
10. Jamshidi MB, Lalbakhsh A, Talla J, et al. Artificial intelligence and COVID-19: deep learning approaches for diagnosis and treatment. *IEEE Access*. 2020;8:109581–109595. https://doi.org/10.1109/ACCESS.2020.3001973.
11. Ray PP. CHATGPT: a comprehensive review on background, applications, key challenges, bias, ethics, limitations and future scope. *IoT CPS*. 2023;3:121–154. https://doi.org/10.1016/j.iotcps.2023.04.003.
12. Kusunose K, Kashima S, Sata M. Evaluation of the accuracy of ChatGPT in answering clinical questions on the Japanese Society of Hypertension Guidelines. *Circ J*. 2023;87(7):1030–1033. https://doi.org/10.1253/circj.CJ-23-0308.
13. Almagazzachi A, Mustafa A, Eighaei Sedeh A, et al. Generative artificial intelligence in patient education: ChatGPT takes on hypertension questions. *Cureus*. 2024;16(2), e53441. Published 2024 Feb 2 https://doi.org/10.7759/cureus.53441.
14. Meo SA, Al-Khlaiwi T, AbuKhalaf AA, Meo AS, Klonoff DC. The scientific knowledge of bard and ChatGPT in endocrinology, diabetes, and diabetes technology: multiple-choice questions examination-based performance. *J Diabetes Sci Technol*. 2024. https://doi.org/10.1177/19322968231203987.

15. Roosan D, Padua P, Khan R, Khan H, Verzosa C, Wu Y. Effectiveness of chatgpt in clinical pharmacy and the role of artificial intelligence in medication therapy management. *JAPhA*. 2023. https://doi.org/10.1016/j.japh.2023.11.023.

16. Al-Ashwal FY, Zawiah M, Gharaibeh L, Abu-Farha R, Bitar AN. Evaluating the sensitivity, specificity, and accuracy of ChatGPT-3.5, ChatGPT-4, bing AI, and bard against conventional drug-drug interactions Clinical tools. *Drug Healthc Patient Saf*. 2023;15:137–147. Published 2023 Sep 20 https://doi.org/10.2147/DHPS.S425858.

17. Huang X, Estau D, Liu X, Yu Y, Qin J, Li Z. Evaluating the performance of Chatgpt in clinical pharmacy: a comparative study of chatgpt and clinical pharmacists. *Br J Clin Pharmacol*. 2023. https://doi.org/10.1111/bcp.15896.

18. Fournier A, Fallet C, Sadeghipour F, Perrottet N. Assessing the applicability and appropriateness of chatgpt in answering clinical pharmacy questions. *Ann Pharm Fr*. 2023. https://doi.org/10.1016/j.pharma.2023.11.001.

19. Al-Dujaili Z, Omari S, Pillai J, Al Faraj A. Assessing the accuracy and consistency of ChatGPT in clinical pharmacy management: a preliminary analysis with clinical pharmacy experts worldwide. *Res Soc Adm Pharm*. 2023;19:1590–1594. https://doi.org/10.1016/j.sapharm.2023.08.012.

20. Morath B, Chiriac U, Jaszkowski E, et al. Performance and risks of CHATGPT used in drug information: an exploratory real-world analysis. *EJHP*. 2023. https://doi.org/10.1136/ejhpharm-2023-003750.

21. Roosan D, Roosan MR, Kim S, Law AV, Sanine C. *Applying artificial intelligence to create risk stratification visualization for underserved patients to improve population health in a community health setting*. 2022. https://doi.org/10.21203/rs.3.rs-1650806/v1.

22. Donovan T, Abell B, Fernando M, McPhail SM, Carter HE. Implementation costs of hospital-based computerised decision support systems: a systematic review. *Implement Sci*. 2023;18:7. https://doi.org/10.1186/s13012-023-01261-8.

23. Sallam M. ChatGPT utility in healthcare education, research, and practice: systematic review on the promising Perspectives and valid concerns. *Healthcare (Basel)*. 2023;11(6):887. Published 2023 Mar 19 https://doi.org/10.3390/healthcare11060887.