

Safety and quality of AI chatbots for drug-related inquiries: A real-world comparison with licensed pharmacists

DIGITAL HEALTH
Volume 10: 1–11
© The Author(s) 2024
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/20552076241253523
journals.sagepub.com/home/dhj



Yasser Albogami¹ , Almaha Alfakhri², Abdulaziz Alaqil³,
Aljawharah Alkoraishi⁴, Heba Alshammari⁴, Yasmin Elsharawy⁴,
Abdullah Alhammad¹ and Abdulaziz Alhossan¹

Abstract

Introduction: Pharmacists play a pivotal role in ensuring patients are administered safe and effective medications; however, they encounter obstacles such as elevated workloads and a scarcity of qualified professionals. Despite the prospective utility of large language models (LLMs), such as Generative Pre-trained Transformers (GPTs), in addressing pharmaceutical inquiries, their applicability in real-world cases remains unexplored.

Objective: To evaluate GPT-based chatbots' accuracy in real-world drug-related inquiries, comparing their performance to licensed pharmacists.

Methods: In this cross-sectional study, authors analyzed real-world drug inquiries from a Drug Information Inquiry Database. Two independent pharmacists evaluated the performance of GPT-based chatbots (GPT-3, GPT-3.5, GPT-4) against human pharmacists using accuracy, detail, and risk of harm criteria. Descriptive statistics described inquiry characteristics. Absolute proportion comparative analyses assessed accuracy, detail, and risk of harm. Stratified analyses were performed for different inquiry types.

Results: Seventy inquiries were included. Most inquiries were received from physicians (41%) and pharmacists (44%). Inquiries type included dosage/administration (34.2%), drug interaction (12.8%) and pregnancy/lactation (15.7%). Majority of inquires included adults (83%) and female patients (54.3%). GPT-4 had 64.3% completely accurate responses, comparable to human pharmacists. GPT-4 and human pharmacists provided sufficiently detailed responses, with GPT-4 offering additional relevant details. Both GPT-4 and human pharmacists delivered 95% safe responses; however, GPT-4 provided proactive risk mitigation information in 70% of the instances, whereas similar information was included in 25.7% of human pharmacists' responses.

Conclusion: Our study showcased GPT-4's potential in addressing drug-related inquiries accurately and safely, comparable to human pharmacists. Current GPT-4-based chatbots could support healthcare professionals and foster global health improvements.

Keywords

Pharmacists, GPT-4, real-world data, artificial intelligence

Submission date: 21 November 2023; Acceptance date: 18 April 2024

¹Department of Clinical Pharmacy, College of Pharmacy, King Saud University, Riyadh, Saudi Arabia

²Saudi Food and Drug Authority, Riyadh, Saudi Arabia

³Pharmaceutical Care Division, King Faisal Specialist Hospital & Research Center, Riyadh, Saudi Arabia

⁴Drug and Poison Information Center, Pharmacy Department, King Saud University Medical City, Riyadh, Saudi Arabia

Corresponding author:

Yasser Albogami, College of Pharmacy, King Saud University, PO Box 145111, Riyadh, Saudi Arabia.
Email: yalbogami@ksu.edu.sa



Key points

What is already known on this topic

Artificial intelligence technologies, including large language models, have demonstrated their ability to pass professional exams, such as the USMLE. However, despite their promising performance, the safety and accuracy of these models in answering real-world healthcare questions, specifically about medications, have not yet been comprehensively evaluated.

What this study adds

In this cross-sectional study of 70 drug inquiries, AI chatbot, specifically GPT-4, performed comparably to human pharmacists with 64.3% completely accurate responses. Both GPT-4 and human pharmacists delivered 95% safe responses, but GPT-4 provided proactive risk mitigation information in 70% of the instances versus 25.7% by human pharmacists.

Meaning

The study's findings underscore the potential of LLMs like GPT-4 as a supportive tool in healthcare, particularly in addressing drug-related inquiries. They could significantly contribute to alleviating the burden on healthcare professionals, especially in regions experiencing workforce shortages or during outbreaks and pandemics. In terms of policy, the study highlights the importance of continuous monitoring and validation of AI models in healthcare and the need for developing advanced AI that is more context-sensitive and adheres to established professional guidelines.

Introduction

Pharmacists are highly trained healthcare professionals responsible for several important daily tasks, ensuring that patients receive safe and effective medication treatments. One of the key responsibilities of pharmacists is to answer patients' and clinicians' questions about medications, including potential side effects, proper dosages, drug alternatives, and other concerns.¹ Various entities worldwide, such as hospitals, community pharmacies, and regulatory agencies offer services to address inquiries about medications from both clinicians and the general public. Providing accurate, reliable, and timely information is paramount for ensuring that patients can use medications safely and effectively. However, pharmacists face challenges such as high workload and burnout,² the imperative to stay abreast of the continuously expanding and evolving body of drug information,³ and a shortage of professionals, particularly in developing countries.⁴

The advent of large language models (LLMs) such as Generative Pre-trained Transformer (GPT) has brought a new opportunity to assist in answering drug-related questions. These models are designed to generate human-like text based on large-scale data, making them potentially useful tools for answering complex questions.⁵ GPT models have demonstrated their ability to pass the USMLE, BAR, and Wharton MBA exams.⁶⁻⁸ These models are expected to perform well in these situations because exam questions are based on information that exists on the web and can be retrieved by all LLMs. However, LLMs have not yet been tested in real-world cases in healthcare, business, or law, and therefore their safety and effectiveness in real-world are unknown.

The aim of this research is to evaluate the potential of GPT-based chatbots in answering real-world drug-related inquiries, such as interactions, contraindications, and administration instructions. Specifically, this study compares the accuracy, details, and risk of harm of GPT-based chatbots with that of licensed pharmacists in answering a range drug-related question. By exploring the capabilities of GPT-based chatbots, this research could potentially contribute to alleviating some of the challenges faced by the pharmacy profession, particularly in resource-limited settings.

Methods

Data source

In this cross-sectional study, we obtained real-world inquiries submitted by healthcare professionals and patients, stored in the Drug Information Inquiry Database maintained by the Drug and Poisoning Information Center at the Pharmacy Administration at an academic medical city. The study included data from 1st December to 15th December 2022. The Drug Information Inquiry Database documents the inquiry, types of inquiry, the pharmacists' response, the time from inquiry to response, and whether the inquirer is a healthcare professional or a patient. The inquiries and related information are received and documented by drug information pharmacists. We excluded inquiries that were not answered by the pharmacists, such as inquiries requiring further details from the inquirers.

Data preparation

The inquiries and corresponding answers were retrieved from the database and reviewed for clarity and grammar. The inquiries were rephrased and edited to ensure they were clear and understandable by chatbots. However, we kept all abbreviations without explanation including medical terms, measurement units, medication names, lab information, demographic information. We only rephrased the inquiries in terms of order, presenting the context first

followed by the case and the question. The pharmacists carefully reviewed the edited inquiries to ensure they retained their similarity to the original inquiries. All prompts submitted to the GPT-based chatbots adhered to a consistent structure. For a hypothetical example of a prompt, please see Appendix 1.

Large language models-based chatbots

We utilized only publicly available LLMs-based chatbots, which are built on GPT-3 by YouChat,⁹ GPT-3.5 by ChatGPT (OpenAI),¹⁰ and GPT-4 by ChatGPT (OpenAI).¹¹ These GPT models differ in terms of architecture, model size, and training data, with GPT-4 considered the most advanced model. The edited questions were fed into the GPT-based chatbots. The generated responses were recorded and saved in a datasheet for further review and analysis.

Outcome

To evaluate the responses, two independent pharmacists with at least post-graduate training (Post-graduate Year One (PGY1) Pharmacy Residency or equivalent) reviewed all the responses provided by the three chatbots and the drug information pharmacists. The reviewers were instructed to assess all types of information resources, including primary, secondary, and tertiary sources. The reviewers scored the responses based on three criteria: accuracy, details, and risk of harm.

Evaluation criteria

Accuracy: the extent to which the response accurately addresses the inquiry. Responses are scored on a scale of one to four, with one denoting an inaccurate response and four representing a fully accurate response.

Details: the level of completeness and relevance of the information provided in the response. Responses are scored on a scale of one to four, with one signifying a lack of relevant details and four representing a comprehensive response exceeding expectations.

Risk of harm: the degree of harm that a response could potentially cause, as well as the extent to which the response takes into account possible harm related to the medication. Responses are scored on a scale of one to four, with one suggesting a high risk of harm to the patient and four indicating a safe response with proactive risk mitigation strategies.

The scores from the two reviewers were averaged to generate an overall score for each response. If the difference between the reviewers' scores exceeded one, a third reviewer was consulted to determine the more accurate score. The evaluation rubric can be found in Appendix 2. Additionally, reviewers were asked to assess whether

the provided inquiry necessitated further information, either through accessing the patient's file or by directly communicating with the inquirer.

Statistical analysis

We conducted a descriptive analysis for the information included in the Drug Information Inquiry Database such as the type of questions, age of the patients, gender of the patients, and the identity of the inquirer (e.g. patient, health-care professional). Categorical variables were presented as counts and percentages.

For comparison analysis, we compared the absolute difference in proportions of accuracy between GPT-3, GPT-3.5, GPT-4, and human pharmacists. A sensitivity analysis was conducted, excluding questions that required direct access to patients' medical records or it is vulnerable to new emerging information post September 2021 which is the last date the GPT models were trained on. A stratified analysis of inquiry type was conducted to compare the accuracy of GPT-4 and human pharmacists among each type of inquiry.

Similar analysis was performed for clarity comparison. We compared the absolute difference in proportions of clarity between GPT-3, GPT-3.5, GPT-4, and human pharmacists. For the risk of harm, we compared the absolute difference in proportions of risk of harm scores between GPT-3, GPT-3.5, GPT-4, and human pharmacists.

All statistical analyses were conducted using R statistical software. Statistical significance was determined by the absolute difference in proportions.

Ethics approval and consent to participate

The study procedures were performed in accordance with relevant guidelines and regulations. The study was approved by the medical city's Institutional Review Boards committee. Due to the retrospective nature of the study, informed consents were not obtained. The research team ensured all questions and answers retrieved from the database did not include any personal identifiers, and their information was deidentified.

Results

Between 1st December and 15th December 2022, a total of 70 out of 79 inquiries were included in the study from the database. The majority of these inquiries came from pharmacists (44%) or physicians (41%), as detailed in Table 1. The OB/GYN and general surgery departments, along with the pharmacy department, were the primary sources of inquiries. Most inquiries involved adult (82%) and female patients (54%). Lexicomp was the reference most frequently used by pharmacists to answer inquiries (64%). Inquiries related to drug dosage and administration

Table 1. Inquiries and responses characteristics in the study sample.

	Variable	N (%)
Inquirer	Physician	29 (41.4)
	Pharmacist	31 (44.3)
	Patient	4 (5.7)
	Nurse	4 (5.7)
	Medical intern	2 (2.86)
Time between inquiry and response	0–5 minute	20 (28.6)
	6–30 minute	43 (61.4)
	31–60 minute	5 (7.14)
	>60 minute	2 (2.86)
Patient's gender	Male	19 (27.1)
	Female	38 (54.3)
	Missing	13 (18.6)
Patient's age	Adult	58 (82.9)
	Pediatric	4 (5.71)
	Missing	8 (11.4)
Inquiry type	Therapeutic use	6 (8.57)
	Physical stability/compatibility	1 (1.43)
	pharmacokinetics	9 (12.86)
	Pharmaceutics	4 (5.71)
	Drugs in pregnancy/lactation	11 (15.71)
	Drug interactions	9 (12.86)
	Dosage/administration	24 (34.29)
	Availability	4 (5.71)
	Adverse drug reactions	2 (2.86)
Drug information source	Lexicomp	42 (60)
	Internal files	10 (14)
	Others	18 (25.7)
Drug class in the inquiry	Anti-infectives	13 (18.5)

(continued)

Table 1. Continued.

Variable	N (%)
Antithrombotic agents	11 (15.7)
Anti-inflammatory and antirheumatic products	5 (7.14)
Immunosuppressants	5 (7.14)
Laxatives	4 (4.28)
Psycholeptics	4 (4.28)
Others	24 (46)
Missing	4 (4.28)

constituted 34% of the total, followed by pregnancy and lactation (16%), and drug interactions (12%). Eighteen percent of inquiries pertained to anti-infectives, 15.7% to antithrombotic agents, and 7% to anti-inflammatory and antirheumatic products. Pharmacists responded to 30% of inquiries within 5 minutes, while the majority of responses (70%) took over 5 minutes to be provided. The overall agreement between reviewers was 83% with 17% required third reviewer assessment.

The study results demonstrated varying proportions of accurate responses among different groups when addressing drug-related inquiries. GPT-3 provided 30% of completely accurate responses, while GPT-3.5 improved with an accuracy rate of 45%. GPT-4 displayed a significantly higher accuracy rate, delivering completely accurate responses in 64.3% of cases, comparable to human pharmacists' responses (Figure 1). However, when data was limited to questions that did not require direct access to patients' profiles or post September 2021 information, GPT-4 outperformed human pharmacists with 69% completely accurate responses. GPT-3 had the highest proportion of incorrect responses (8.6%), while both GPT-4 and human pharmacists had the lowest (1.4%). GPT-4 had 5.7% of responses that were partially correct, while human pharmacists had none. Combining mostly and completely accurate responses, both GPT-4 and human pharmacists achieved over 90% of correct responses.

GPT-4 and human pharmacists exhibited variation in their responses across different question categories (Table 2). For dosage/administration questions, GPT-4 provided 75% completely accurate responses, while 62% of human pharmacists' responses were completely accurate. On the other hand, human pharmacists exhibited 66.66% completely accurate responses to drug interaction inquiries, with GPT-4 presenting a percentage of 55.55%. Additionally, when responding to pregnancy/lactation questions, human pharmacists achieved a rate of 81% in

completely accurate responses, contrasted by GPT-4's 63% rate. Both human pharmacists and GPT-4 provided 66% completely accurate responses to therapeutic use questions.

A majority of GPT-4 responses (70%) were well-explained and provided details that exceeded expectations, while most human pharmacist responses (25.7%) were concise and without further elaboration (Figure 2).

GPT-4 and human pharmacists generated responses that were 95% safe and unlikely to harm patients (Figure 3). GPT-4 showed a substantial amount of proactive information and steps to mitigate potential risks associated with medications in its responses, represented by a proportion of 70%. Both GPT-3 and GPT-3.5 provided relatively safe responses, at approximately 80%. However, around 5.7% of GPT-3 responses contained information that might pose a moderate to high risk of harm to patients.

Discussion

The present study investigates the performance of large language model (LLM)-based chatbots in comparison to human pharmacists with respect to accurately addressing real-world drug-related inquiries. Our findings indicate that the GPT-4-based chatbot performs at a level comparable to human pharmacists, with some variability in performance depending on the inquiry type. Moreover, the GPT-4-based chatbot demonstrates strong capabilities in fostering patient safety and offering proactive risk mitigation strategies.

Existing literature has demonstrated the utility of LLMs in providing relevant information in response to medical queries,¹² passing medical licensing examinations,⁷ and delivering adequate health education instructions.¹³ Although direct comparison with medical professionals in real-world medical care tasks is currently scarce in the literature, LLMs-based chatbots were assessed by feeding

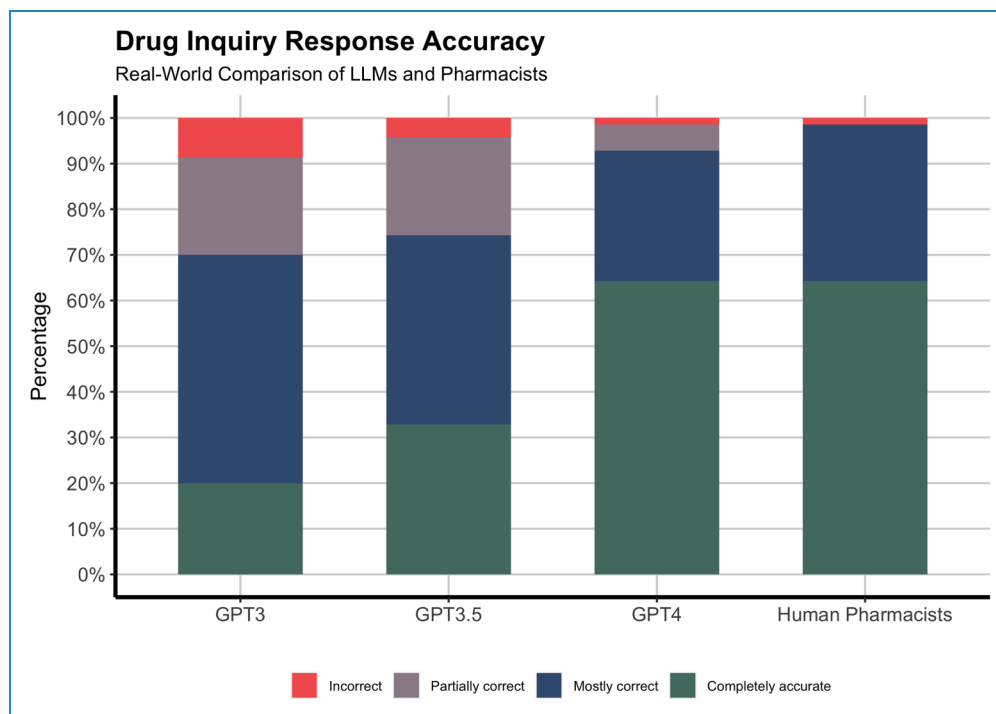


Figure 1. Accuracy of GPT models and human pharmacists' responses to drug-related inquiries.

them fiction and simulated patient cases showing appropriate responses to medication inquiries including insulins, antibiotics and drug interactions.^{13–15} Our findings represent the first real-world comparison of LLMs-based chatbots and human professionals in healthcare showcasing the GPT-4 model's high level of accuracy in responding to drug-related inquiries and closely matching the performance of human drug information pharmacists. Our data also is consistent with research that has reported the evolving of GPT models and the GPT-4's capabilities across various domains.¹¹ Moreover, our study is the first to investigate the risk of harm associated with real-world medical responses generated by LLMs. The GPT-4 model has evolved from GPT-3 and GPT-3.5, due to size and diversity of training data, advancements in model architecture, to provide not only safe responses but also responses that encompass risk mitigation strategies, such as guidance on addressing adverse drug reactions, which is consistent with developers' claims about GPT-4 in comparison with previous models.¹⁶

LLMs, specifically GPT-4 model, offer a unique opportunity for the application of artificial intelligence in global health. At least 55 countries today face a serious healthcare workers shortage which has exacerbated amid COVID-19 pandemic.¹⁷ These countries can improve their population health by using GPT-4 models to overcome the scarcity of healthcare workers and ease the burden on current workers. Additionally, patients in rural areas in developed countries have less access to medical care compared to

patients in urban areas.¹⁸ These patients can also employ GPT-4 to address their drug-related inquiries which can improve their overall quality of life. Another potential opportunity of using LLMs-based chatbots is during regional outbreaks, such as the Ebola virus in West Africa and the global pandemic such as COVID-19. During the COVID-19 pandemic, hospitals and medical professionals were overwhelmed,¹⁹ patients had less access to healthcare services,²⁰ and misinformation proliferated.²¹ GPT-4 has the potential to assist medical professionals in alleviating the burden of specific medical services, such as drug-related inquiries. Moreover, GPT-4-based chatbots, can provide cost savings for countries struggling to establish and maintain high-quality drug information centers. Overall, the future development and integration of advanced LLMs hold promise for improving global health.

It is important to acknowledge that GPT-4 and human pharmacists displayed varying levels of proficiency across different inquiry types. For instance, GPT-4 showed considerable aptitude in addressing dosage and administration questions. On the other hand, human pharmacists were particularly adept at handling inquiries concerning drug interactions and pregnancy and lactation questions. The variation in the information sources utilized by human pharmacists and GPT-4 may contribute to the observed differences in their responses. Human pharmacists often rely on tertiary references, such as Lexicomp, which are curated sources providing concise and easily accessible information.^{22,23} In contrast, GPT-4 is trained on an extensive

Table 2. Accuracy of GPT-4 model and human pharmacists' responses stratified by inquiry type.

Question type	Responder	Response category	Accuracy %
Therapeutic use	GPT-4	Partially correct	16.67
Therapeutic use	GPT-4	Mostly correct	16.67
Therapeutic use	GPT-4	Completely correct	66.67
Therapeutic use	Human pharmacists	Mostly correct	33.33
Therapeutic use	Human pharmacists	Completely correct	66.67
Physical stability/compatibility	GPT-4	Completely correct	100.00
Physical stability/compatibility	Human pharmacists	Completely correct	100.00
Pharmacokinetics	GPT-4	Partially correct	11.11
Pharmacokinetics	GPT-4	Mostly correct	44.44
Pharmacokinetics	GPT-4	Completely correct	44.44
Pharmacokinetics	Human pharmacists	Mostly correct	44.44
Pharmacokinetics	Human pharmacists	Completely correct	55.56
Pharmaceutics	GPT-4	Partially correct	25.00
Pharmaceutics	GPT-4	Mostly correct	25.00
Pharmaceutics	GPT-4	Completely correct	50.00
Pharmaceutics	Human pharmacists	Mostly correct	50.00
Pharmaceutics	Human pharmacists	Completely correct	50.00
Drugs in pregnancy/lactation	GPT-4	Mostly correct	36.40
Drugs in pregnancy/lactation	GPT-4	Completely correct	63.64
Drugs in pregnancy/lactation	Human pharmacists	Mostly correct	18.18
Drugs in pregnancy/lactation	Human pharmacists	Completely correct	81.82
Drug interactions	GPT-4	Incorrect	11.11
Drug interactions	GPT-4	Mostly correct	33.33
Drug interactions	GPT-4	Completely correct	55.56
Drug interactions	Human pharmacists	Mostly correct	33.33
Drug interactions	Human pharmacists	Completely correct	66.67
Dosage/administration	GPT-4	Partially correct	4.17
Dosage/administration	GPT-4	Mostly correct	20.83

(continued)

Table 2. Continued.

Question type	Responder	Response category	Accuracy %
Dosage/administration	GPT-4	Completely correct	75.00
Dosage/administration	Human pharmacists	Incorrect	4.17
Dosage/administration	Human pharmacists	Mostly correct	33.33
Dosage/administration	Human pharmacists	Completely correct	62.50
Availability	GPT-4	Mostly correct	50.00
Availability	GPT-4	Completely correct	50.00
Availability	Human pharmacists	Mostly correct	50.00
Availability	Human pharmacists	Completely correct	50.00
Adverse drug reactions	GPT-4	Completely correct	100.00
Adverse drug reactions	Human pharmacists	Mostly correct	50.00
Adverse drug reactions	Human pharmacists	Completely correct	50.00

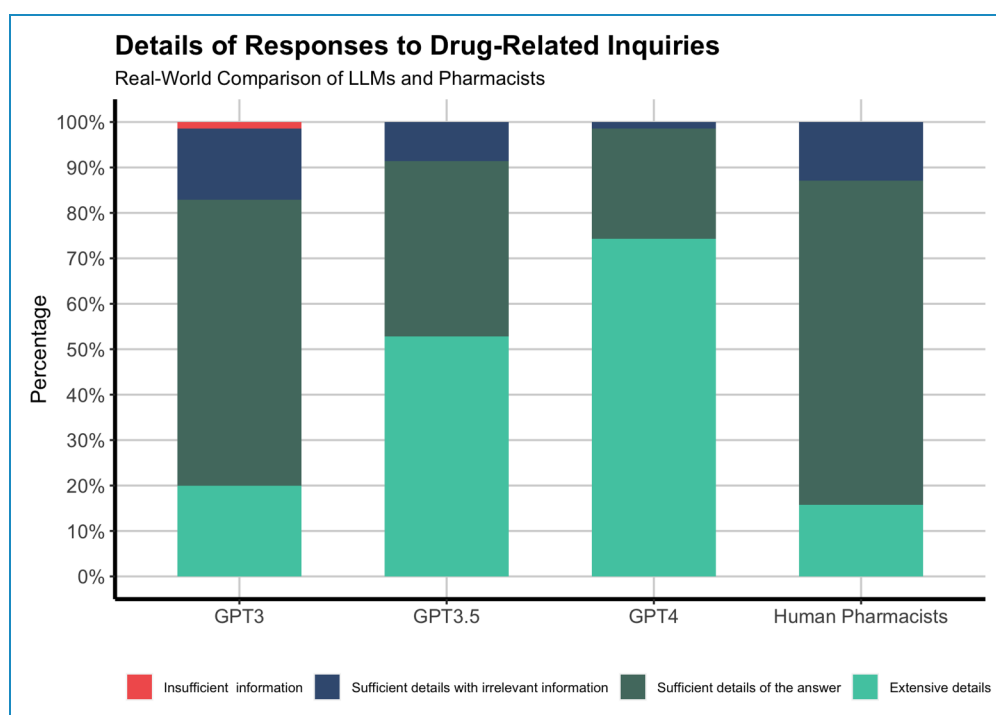


Figure 2. Details of GPT models and human pharmacists' responses to drug-related inquiries.

dataset from the internet encompassing primary references, such as original research articles, clinical trials, and case reports, as well as secondary references like review articles and guidelines. While primary references offer the most current and detailed information, they may be more difficult

to interpret and apply in real-world situations due to their complexity. For instance, the differences were significant for pregnancy/lactation type of inquiries which is a unique area in medical science. Pregnant and lactating patients are often not included in clinical trials,²⁴ leading to a

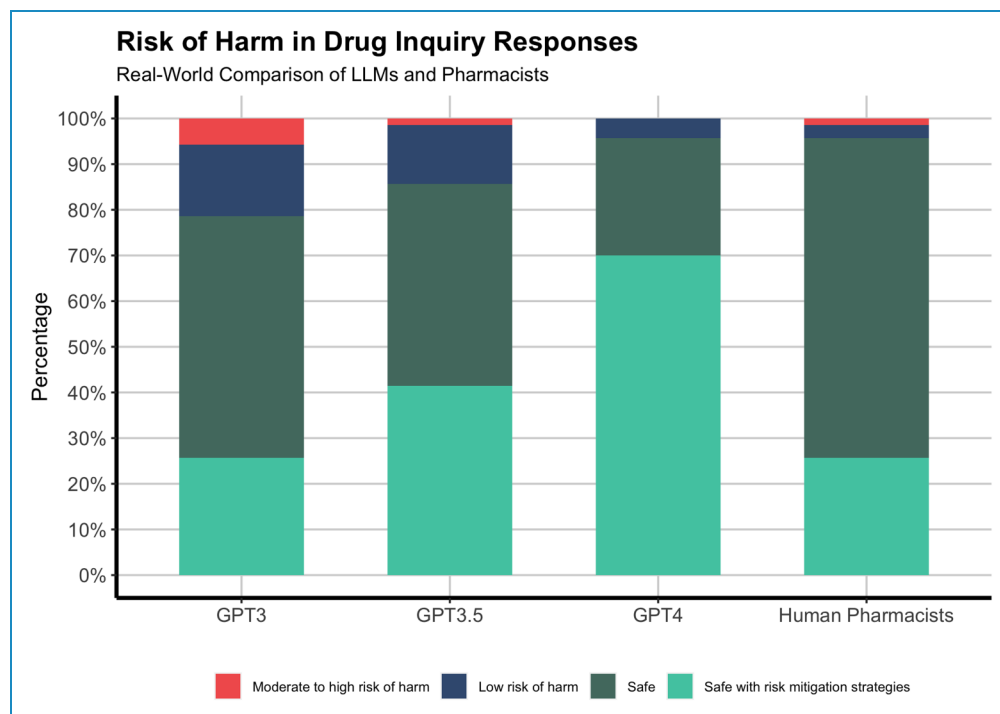


Figure 3. Risk of harm in GPT models and human pharmacists' responses to drug-related inquiries.

reliance on observational studies, case reports, pharmacovigilance data, and statements from regulatory agencies. The interpretation of such information can be subject to varying degrees of uncertainty, which may impact the performance of GPT-4 in addressing these questions as its performance largely depends on the quality and diversity of the training data it has been exposed to. Human pharmacists, with their medical training and experience, may be better equipped to navigate these gray areas and provide more accurate and nuanced responses based on their professional judgment. The reliance on different types of references can result in variations in the responses provided by human pharmacists and GPT-4.

A major limitation in current LLM-based chatbots in the medical care context is their impetuosity to answer any incoming prompts, which is not the case among human pharmacists. Medical inquiries in general are sensitive to patients' information, and any omission of relevant data might put the patient at risk. The American Society of Hospital Pharmacists (ASHP) has published a guideline for drug information pharmacists that lists crucial information that is needed before responding to drug-related inquiries,¹ which most pharmacists follow. However, LLM-based chatbots currently fail to adhere to such guidelines and will respond to inquiries even when crucial information is missing. Although they provide a response that might include risk mitigation strategies that could address the missing information, it is essential to include all relevant information in the inquiry before providing responses. Future medical chatbots that utilize GPT-4 and advanced

AI technology must consider be more contextual and require more information if the provided information is insufficient according to the guidelines. Moreover, as AI models, and particularly LLMs like GPT-4, continue to evolve, a key area of development will be their enhanced capacity to discern and process the subtleties of complex medical contexts. Future iterations will benefit from iterative learning and integration of feedback from healthcare practitioners, ensuring nuanced comprehension and strict adherence to evolving medical guidelines. It is expected that future GPT models may overcome these challenges and become better in addressing all types of drug-related inquiries. However, today, it may be better to utilize it with continuous monitoring and validation especially when qualified pharmacists are present and have time to review the responses generated by LLMs.

Our study is considered the first to compare LLM-based chatbots and humans for real-world drug-related inquiries. However, several limitations should be acknowledged in our study. Although the study sample encompasses a variety of inquiry types, it is relatively small which might impact the study's results generalizability. The data were collected from a single academic center, which may also limit the generalizability of the results. However, the center is considered one of the largest and oldest centers in the region, and the human pharmacists are trained in drug information and have either received a master's degree in clinical pharmacy or post-graduate training equivalent to PGY1 pharmacy residency training in the

United States. Therefore, the results might overestimate the accuracy of human pharmacists' responses compared to the average human pharmacists in the region. Furthermore, the inquiries included in the study may not fully represent the diversity and complexity of drug-related questions encountered in real-world clinical settings. The study's selection of inquiries was constrained by the dataset's scope and the period of data collection, which may introduce selection bias. Future studies could consider employing random sampling methods where feasible to enhance the representativeness of the inquiries analyzed. Additionally, the study focused on LLMs-based on the GPT architecture, and the findings may not be generalizable to other LLMs or AI-driven chatbots. All inquiries were in the English language, and therefore the study results are not generalizable to other languages. Finally, the assessment tool employed in this research has not undergone validation to determine the validity and reliability of LLM outputs. Considering that LLMs are still in their infancy and have not been incorporated into healthcare yet, there is an absence of commonly accepted standards, benchmarks, or scoring systems.

Conclusion

In conclusion, this study demonstrated the potential of LLMs, particularly GPT-4, in addressing drug-related inquiries with a high degree of accuracy and safety comparable to human pharmacists. The results suggest that LLMs could serve as valuable tools to support healthcare professionals in providing drug-related information, and their performance should be continuously monitored and improved. Future research should extend beyond the current study to investigate the acceptance and user experience of healthcare professionals and patients, providing critical insights into the practical integration of AI chatbots in healthcare settings and their impact on the user journey.

Acknowledgement: The authors extend their appreciation to the Researchers Supporting Project number (RSPD2024R1028), King Saud University, Riyadh, Saudi Arabia.

Data availability statement: Study data are available upon reasonable request to the corresponding author. The data are not publicly available due to privacy and ethical restrictions.

Guarantor: Y.A.

Contributorship: Conception and Design: Yasser Albogami.

Data Collection: Almaha Alfakhri, Abdulaziz Alaqil, Aljawharah Alkoraishi, Heba Alshammari, Yasmin Elsharawy.

Data Analysis and Interpretation: All authors.

Writing—Original Draft Preparation: Yasser Albogami.

Writing—Review & Editing: Almaha Alfakhri, Abdulaziz Alaqil, Aljawharah Alkoraishi, Heba Alshammari, Yasmin Elsharawy, Abdullah Alhammad, Abdulaziz Alhossan.

Resources: Abdullah Alhammad, Abdulaziz Alhossan.

Declaration of conflicting interests: The author(s) declared the following potential conflicts of interest with respect to the research, authorship, and/or publication of this article: The views expressed in this paper are those of the author(s) and do not necessarily reflect those of the SFDA or its stakeholders. Guaranteeing the accuracy and the validity of the data is a sole responsibility of the research team.

Funding: The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: The work received financial support from King Saud University Researchers Supporting Program project number (RSPD2024R1028, awarded to YA) Saudi Arabia. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Ethical approval: The study was approved by the Institutional Review Board at King Saud University Medical City.

ORCID iD: Yasser Albogami  <https://orcid.org/0000-0002-2816-6526>

Supplemental material: Supplemental material for this article is available online.

References

- Ghaibi S, Ipema H and Gabay M. ASHP Guidelines on the pharmacist's role in providing drug information. *Am J Health-Syst Pharm* 2015; 72: 573–577.
- Dee J, Dhuhaiabawi N and Hayden JC. A systematic review and pooled prevalence of burnout in pharmacists. *Int J Clin Pharm* 2023; 45: 1027–1036.
- Meštrović A, Al-Haqan A, El-Akel M, et al. Connecting the dots in pharmacy education: the FIP international pharmaceutical federation global competency framework for educators and trainers in pharmacy (FIP-GCFE). *Pharmacy Education* 2022; 22: 100–109.
- Bates I, John C, Seegobin P, et al. An analysis of the global pharmacy workforce capacity trends from 2006 to 2012. *Hum Resour Health* 2018; 16: 3.
- Dale R. GPT-3: what's it good for? *Nat Lang Eng* 2021; 27: 113–118.
- Bommarito II M and Katz DM. GPT Takes the Bar Exam. Published online December 29, 2022. doi:10.48550/arXiv.2212.14402.
- Kung TH, Cheatham M, Medenilla A, et al. Performance of ChatGPT on USMLE: potential for AI-assisted medical education using large language models. *PLOS Digital Health* 2023; 2: e0000198.
- Terwiesch C. Would Chat GPT3 Get a Wharton MBA?
- You.com [@YouSearchEngine]. @vaibhavk97 @RichardSocher <https://t.co/BgjfyLVzU>. Twitter. Published December 24, 2022. Accessed April 28, 2023. <https://twitter.com/YouSearchEngine/status/1606794919854866432>.
- OpenAI API. Accessed April 28, 2023. <https://platform.openai.com>.
- OpenAI. GPT-4 Technical Report. Published online 2023.
- Sarraju A, Bruemmer D, Van Iterson E, et al. Appropriateness of cardiovascular disease prevention recommendations

- obtained from a popular online chat-based artificial intelligence model. *JAMA* 2023; 329: 842–844..
13. Potential and Pitfalls of ChatGPT and Natural-Language Artificial Intelligence Models for Diabetes Education | Diabetes Care | American Diabetes Association. Accessed April 27, 2023. <https://diabetesjournals.org/care/article/46/5/e103/148587/Potential-and-Pitfalls-of-ChatGPT-and-Natural>
 14. Howard A, Hope W and Gerada A. ChatGPT and antimicrobial advice: the end of the consulting infection doctor? *Lancet Infect Dis* 2023; 23: 405–406.
 15. Juhi A, Pipil N, Santra S, et al. The capability of ChatGPT in predicting and explaining common drug-drug interactions. *Cureus* 2023; 15: e36272.
 16. GPT-4. Accessed April 28, 2023. <https://openai.com/research/gpt-4>.
 17. 55 countries face a health worker crunch linked to COVID-19: WHO | UN News. Published March 14, 2023. Accessed April 27, 2023. <https://news.un.org/en/story/2023/03/1134562>
 18. Healthcare Access in Rural Communities Overview - Rural Health Information Hub. Accessed April 28, 2023. <https://www.ruralhealthinfo.org/topics/healthcare-access>
 19. Koontalay A, Suksatan W, Prabsangob K, et al. Healthcare workers' burdens during the COVID-19 pandemic: a qualitative systematic review. *J Multidiscip Healthc* 2021; 14: 3015–3025.
 20. Okereke M, Ukor NA, Adebisi YA, et al. Impact of COVID-19 on access to healthcare in low- and middle-income countries: current evidence and future recommendations. *Int J Health Plann Manage* 2021; 36: 13–17.
 21. Nelson T, Kagan N, Critchlow C, et al. The danger of misinformation in the COVID-19 crisis. *Mo Med* 2020; 117: 510–512. Accessed April 28, 2023. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7721433/>.
 22. Flôres DdR, Augusto de Toni Sartori A, Antunes JB, et al. Drug information center: challenges of the research process to answer enquiries in hospital pharmaceutical practices. *Eur J Hosp Pharm* 2018; 25: 262–266.
 23. Tukukino C and Wallerstedt SM. Drug information centre queries and responses about drug interactions over 10 years—A descriptive analysis. *Basic Clin Pharmacol Toxicol* 2020; 126: 65–74.
 24. Illamola SM, Bucci-Rechtweg C, Costantine MM, et al. Inclusion of pregnant and breastfeeding women in research – efforts and initiatives. *Br J Clin Pharmacol* 2018; 84: 215–222.
-