Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon

A systematic review and meta-analysis on ChatGPT and its utilization in medical and dental research

Hiroj Bagde[a,1,**], Ashwini Dhopte[b], Mohammad Khursheed Alam[c,d,e,1,*], Rehana Basri[f]

[a] *Department of Periodontology, Chhattisgarh Dental College and Research Institute, Rajnandgaon, Chhattisgarh, India*
[b] *Department of Oral Medicine and Radiology, Chhattisgarh Dental College and Research Institute, Rajnandgaon, Chhattisgarh, India*
[c] *Preventive Dentistry Department, College of Dentistry, Jouf University, Sakaka, 72345, Saudi Arabia*
[d] *Department of Dental Research Cell, Saveetha Dental College and Hospitals, Saveetha Institute of Medical and Technical Sciences, Chennai, India*
[e] *Department of Public Health, Faculty of Allied Health Sciences, Daffodil International University, Dhaka, Bangladesh*
[f] *Department of Internal Medicine, College of Medicine, Jouf University, Sakaka, 72345, Saudi Arabia*

ARTICLE INFO

Keywords:
Artificial intelligence
Machine learning
ChatGPT
Medicine
Dentistry

ABSTRACT

Background: Since its release, ChatGPT has taken the world by storm with its utilization in various fields of life. This review's main goal was to offer a thorough and fact-based evaluation of ChatGPT's potential as a tool for medical and dental research, which could direct subsequent research and influence clinical practices. Methods: Different online databases were scoured for relevant articles that were in accordance with the study objectives. A team of reviewers was assembled to devise a proper methodological framework for inclusion of articles and meta-analysis. Results: 11 descriptive studies were considered for this review that evaluated the accuracy of ChatGPT in answering medical queries related to different domains such as systematic reviews, cancer, liver diseases, diagnostic imaging, education, and COVID-19 vaccination. The studies reported different accuracy ranges, from 18.3 % to 100 %, across various datasets and specialties. The meta-analysis showed an odds ratio (OR) of 2.25 and a relative risk (RR) of 1.47 with a 95 % confidence interval (CI), indicating that the accuracy of ChatGPT in providing correct responses was significantly higher compared to the total responses for queries. However, significant heterogeneity was present among the studies, suggesting considerable variability in the effect sizes across the included studies. Conclusion: The observations indicate that ChatGPT has the ability to provide appropriate solutions to questions in the medical and dentistry areas, but researchers and doctors should cautiously assess its responses because they might not always be dependable. Overall, the importance of this study rests in shedding light on ChatGPT's accuracy in the medical and dentistry fields and emphasizing the need for additional investigation to enhance its performance.

* Corresponding author. Preventive Dentistry Department, College of Dentistry, Jouf University, Sakaka 72345, Saudi Arabia.
** Corresponding author.
E-mail addresses: hirojbagde8@gmail.com (H. Bagde), dralam@gmail.com, mkalam@ju.edu.sa (M.K. Alam).

[1] Joined 1st author.

https://doi.org/10.1016/j.heliyon.2023.e23050
Received 7 May 2023; Received in revised form 24 October 2023; Accepted 24 November 2023
Available online 29 November 2023

1. Introduction

Artificial Intelligence (AI) and Machine Learning (ML) are two related technologies that have gained immense popularity in recent years [1]. AI refers to the creation of intelligent machines that can perform tasks that typically require human intelligence, such as visual perception, speech recognition, decision-making, and language translation [2]. ML, on the other hand, is a subset of AI that focuses on developing algorithms and statistical models that enable machines to learn from data and improve their performance over time without being explicitly programmed [3].

One of the key benefits of AI and ML is their ability to process and analyze vast amounts of data, providing insights and predictions that would be difficult or impossible for humans to uncover [4]. In the medical field, AI and ML are being used to improve patient care, support medical diagnosis, and assist with medical research [5]. For example, ML algorithms can be trained to recognize patterns in medical images, assisting radiologists in detecting early signs of cancer or other diseases [6]. AI-powered chatbots and virtual assistants can provide patients with personalized medical advice and support, while also freeing up healthcare providers to focus on more complex cases [7]. However, there are also challenges associated with the use of AI and ML in healthcare, such as ensuring the accuracy and reliability of algorithms, addressing concerns around data privacy and security, and ensuring that these technologies are used in an ethical and responsible manner [8]. Ongoing research and development are needed to address these challenges and ensure that AI and ML technologies are effectively integrated into healthcare systems for the benefit of patients and healthcare providers alike [9,10].

> ChatGPT is a language model based on the GPT (Generative Pre-trained Transformer) architecture, which is part of the family of deep learning models used in natural language processing (NLP) [11]. It was developed by OpenAI and is designed to generate human-like text and engage in conversations with users through chat interfaces [12]. Since its launch, ChatGPT has gained tremendous popularity and has been the inspiration for the development of several mobile applications that incorporate the keywords "chatbot" and "ChatGPT". In fact, in the first ten days of January 2023, ChatGPT - GPT 3 was downloaded 3,771 times by global users, while the app Lia ChatGPT was downloaded 3,560 times [13,14].

ChatGPT is trained on a massive amount of text data and uses deep neural networks to learn the patterns and structures of language [15]. This enables it to generate coherent and contextually relevant responses to input text prompts [16]. ChatGPT has been widely used in various fields, including healthcare, education, customer service, and entertainment [17]. This chatbot has demonstrated the potential to impact a wide range of fields, including but not limited to healthcare, finance, customer service, education, and entertainment [18]. In healthcare, ChatGPT can assist medical professionals with diagnoses, treatment plans, and patient education [19]. In finance, ChatGPT can be used for fraud detection, risk assessment, and investment advice [20]. In customer service, ChatGPT can provide 24/7 support, handle routine inquiries, and escalate complex issues to human representatives [21]. In education, ChatGPT can support students with personalized learning and tutoring [22]. In entertainment, ChatGPT can generate natural language responses in virtual assistants and chatbots for gaming or online assistants [17]. Overall, ChatGPT has the potential to significantly impact

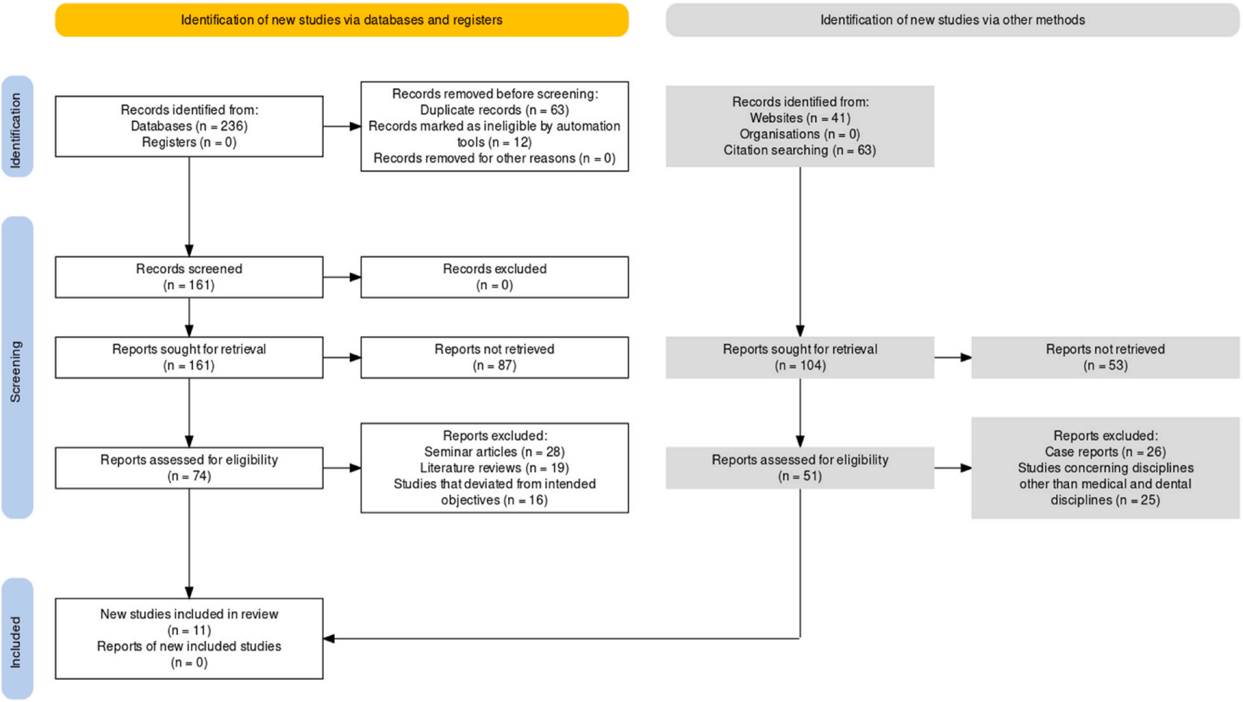

Identification of new studies via databases and registers
Identification of new studies via other methods
Identification
Screening
Included
Records identified from:
Databases (n = 236)
Registers (n = 0)
Records removed before screening:
Duplicate records (n = 63)
Records marked as ineligible by automation tools (n = 12)
Records removed for other reasons (n = 0)
Records identified from:
Websites (n = 41)
Organisations (n = 0)
Citation searching (n = 63)
Records screened
(n = 161)
Records excluded
(n = 0)
Reports sought for retrieval
(n = 161)
Reports not retrieved
(n = 87)
Reports sought for retrieval
(n = 104)
Reports not retrieved
(n = 53)
Reports assessed for eligibility
(n = 74)
Reports excluded:
Seminar articles (n = 28)
Literature reviews (n = 19)
Studies that deviated from intended objectives (n = 16)
Reports assessed for eligibility
(n = 51)
Reports excluded:
Case reports (n = 26)
Studies concerning disciplines other than medical and dental disciplines (n = 25)
New studies included in review
(n = 11)
Reports of new included studies
(n = 0)

Fig. 1. Article selection framework for the studies included in the review.

numerous fields by providing efficient and accurate responses to a range of inquiries and tasks.

Both AI and ML are essential to ChatGPT because they enable the model to learn and improve its responses over time [13]. ChatGPT was trained on a massive amount of data using machine learning algorithms to recognize patterns and generate responses based on context [14]. As a result, ChatGPT can generate increasingly sophisticated and nuanced responses to queries. Furthermore, AI and ML are essential to other applications related to ChatGPT, such as medical diagnosis, drug discovery, and data analysis [15]. These applications rely on machine learning algorithms to analyze large datasets and generate insights that can help medical professionals make informed decisions [17].

The evident literature gap when the usage of ChatGPT is concerned with medical and dental fields lies in the glaring lack of comprehensive systematic reviews and meta-analyses that have been conducted in this area. While there have been several studies conducted on the use of ChatGPT in various medical and dental settings, these studies have been limited in their scope and have not provided a comprehensive overview of the potential benefits and limitations of using ChatGPT in these fields, being either literature reviews or editorials [23–25]. Additionally, there is a lack of consensus regarding the best practices for using ChatGPT in medical and dental settings, as well as a need for further research into the ethical implications of using these technologies in clinical practice [23–25]. Therefore, there is a definite need for a comprehensive review of the available literature on the use of ChatGPT in the aforementioned settings, which can help identify gaps in knowledge and inform future research in this area. The objectives of this review were to comprehensively assess the accuracy, reliability, and usefulness of ChatGPT in medical and dental research, considering studies from different specialties and subjects, published in the year 2023. The review aimed to identify and analyze the existing literature that evaluated the efficacy of ChatGPT in medical and dental research, including its ability to accurately and reliably answer clinical and non-clinical questions and provide support for clinical decisions. Additionally, the review aimed to assess the ability of ChatGPT to interpret medical imaging, detect and diagnose diseases, and produce information in response to educational commands. The overall objective of this review was to provide a comprehensive and evidence-based assessment of ChatGPT's potential as a tool for medical and dental research, which could guide future research and inform healthcare practices.

2. Materials and methods

2.1. Protocol and PICO strategy for review

The current systematic review was conducted as per the PRISMA guidelines [26] which are used for guidance of studies like these (Fig. 1). To utilize the PICO search strategy for this study, the authors first identified the research question they want to answer and then used the PICO framework to develop a search strategy. This involved selecting relevant keywords and synonyms for each element of the PICO framework such as "ChatGPT," "accuracy," "medical," "dental," "specialties," and "healthcare providers", and combining them using Boolean operators (e.g., AND, OR, NOT) to create a comprehensive search strategy, given as follows-

- Population: Patients or healthcare providers who use ChatGPT for medical or dental queries
- Intervention: Use of ChatGPT for answering queries related to medical or dental fields and specialties
- Comparison: Alternative methods of answering queries (e.g., human experts, other AI systems)

3. Outcome: accuracy of ChatGPT in answering queries related to medical and dental fields and specialties

It is pertinent to mention here that the specific version of ChatGPT employed throughout the study was GPT-3.5. This choice of the ChatGPT version was selected due to the universal availability of the said variant without any restrictions pertaining to paywalls/ exclusive access. GPT-3.5 was utilized as the foundation for evaluating the performance, accuracy, and effectiveness of ChatGPT in addressing queries related to medical and dental fields and specialties in this investigation. This version provided the linguistic and contextual understanding necessary for assessing ChatGPT's suitability and potential applications within the medical and dental research domains.

3.1. Database search strategy

For searching across PubMed, the MeSH terms used were "ChatGPT" and "Medical Informatics" which were combined using Boolean operators "AND" and "OR" with other keywords such as "dental", "specialty", "accuracy", "query", "response" and "meta-analysis". The search was limited to articles published between January 1, 2023 and December 31, 2023. The search was conducted in English language only. Similar MeSH terms and keywords were used for searching across Google Scholar, Scopus, EMBASE, Cochrane Library, and UpToDate. Boolean operators "AND" and "OR" were used for combining the search terms to retrieve relevant articles. In addition, reference lists of relevant articles and systematic reviews were screened to identify additional studies. The search process was carried out independently by two reviewers and any disagreement was resolved by discussion. As for other databases, the search strategy used the following keywords and MeSH terms: ("ChatGPT" [MeSH Terms] OR "ChatGPT" [Title/Abstract]) AND ("Medical Informatics" [MeSH Terms] OR "Medical Informatics" [Title/Abstract]) AND ("dental" OR "specialty" OR "accuracy" OR "query" OR "response" OR "meta-analysis") AND ("2023/01/01"[Date - Publication]: "2023/12/31"[Date - Publication]) AND English[lang]. The same strategy with appropriate syntax was used in other databases such as Google Scholar, Scopus, EMBASE, Cochrane Library and UpToDate.

The search strategy aimed to identify relevant studies, with the use of Boolean operators and MeSH terms allowing for a comprehensive and systematic search across different databases, while limiting the search to a specific time frame ensured that the results were up-to-date.

3.1.1. *Selection criterion*

For this investigation, the inclusion and exclusion criteria were established to ensure that relevant studies were included while minimizing the risk of bias. The inclusion criteria were as follows: 1) studies that assessed the accuracy of ChatGPT in answering medical and dental queries; 2) studies that were conducted in the year 2023 only; 3) studies that included different medical and dental specialties and subjects; 4) studies that used any type of study design; and 5) studies that were written in the English language.

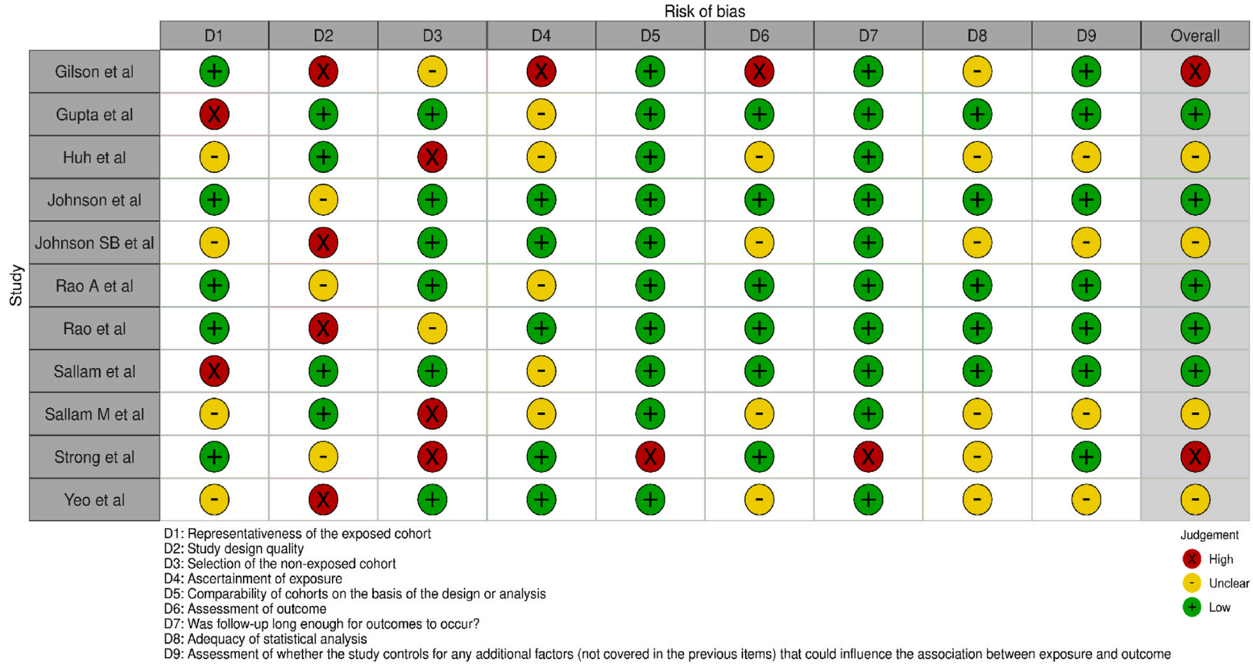
The exclusion criteria were as follows: 1) studies that did not assess the accuracy of ChatGPT in answering medical and dental queries; 2) studies that were conducted before the year 2023; 3) studies that were deviated from their intended objectives; 4) studies that were not written in the English language; and 5) studies that were duplicates.

The inclusion and exclusion criteria were established by the research team to ensure that the selected studies met the research question of this review while minimizing the risk of bias. These criteria were used to screen the titles, abstracts, and full texts of the studies retrieved from the databases. The studies that met the inclusion criteria were included in the review while those that did not meet the criteria were excluded. By establishing clear inclusion and exclusion criteria, the research team was able to select studies that were relevant to the research question while minimizing the risk of bias.

4. Reviewer protocol

The reviewer strategy for this systematic review involved the collaboration of four different experts who were specialized in the field of machine learning and GPT-3. The first step of the reviewer strategy was to provide the reviewers with a clear understanding of the research question, inclusion and exclusion criteria, and the search protocol for identifying relevant studies. The reviewers were also given access to a spreadsheet where they could record their findings and comments on the studies that met the inclusion criteria. Each reviewer was assigned a set of studies to evaluate independently using a predetermined checklist that was based on the Newcastle-Ottawa Scale (NOS) tool [27,28]. The checklist was designed to assess the methodological quality and risk of bias of each study in areas such as study design, sample size, blinding, and reporting of outcomes. The reviewers were required to record their assessment of each study in the spreadsheet provided and provide a rationale for their score.

Once the reviewers had completed their independent assessments, they convened to discuss their findings and resolve any discrepancies in their assessments. Any disagreements were resolved through a consensus-building process, which involved a thorough discussion of the study design, quality, and risk of bias. Overall, the reviewer strategy for this systematic review involved a collaborative effort among four specialized reviewers who worked independently to evaluate the methodological quality and risk of bias of each study. This approach ensured that the systematic review was thorough, objective, and rigorous, and that the conclusions drawn from the evidence were robust and reliable.



Risk of bias

Study	D1	D2	D3	D4	D5	D6	D7	D8	D9	Overall
Gilson et al	+	X	-	X	+	X	+	-	+	X
Gupta et al	X	+	+	-	+	+	+	+	+	+
Huh et al	-	+	X	-	+	-	+	-	-	-
Johnson et al	+	-	+	+	+	+	+	+	+	+
Johnson SB et al	-	X	+	+	+	-	+	-	-	-
Rao A et al	+	-	+	-	+	+	+	+	+	+
Rao et al	+	X	-	+	+	+	+	+	+	+
Sallam et al	X	+	+	-	+	+	+	+	+	+
Sallam M et al	-	+	X	-	+	-	+	-	-	-
Strong et al	+	-	X	+	X	+	X	-	+	X
Yeo et al	-	X	+	+	+	-	+	-	-	-

D1: Representativeness of the exposed cohort
D2: Study design quality
D3: Selection of the non-exposed cohort
D4: Ascertainment of exposure
D5: Comparability of cohorts on the basis of the design or analysis
D6: Assessment of outcome
D7: Was follow-up long enough for outcomes to occur?
D8: Adequacy of statistical analysis
D9: Assessment of whether the study controls for any additional factors (not covered in the previous items) that could influence the association between exposure and outcome

Judgement
X High
- Unclear
+ Low

Fig. 2. Bias assessment of the selected papers using the NOS tool.

4.1. Assessment of bias

NOS was used to assess the risk of bias for all the studies selected in this systematic review and meta-analysis (Fig. 2). The reviewers evaluated each study according to the NOS criteria for case-control and cohort studies. The NOS assesses the risk of bias in three main domains: selection of study groups, comparability of groups, and ascertainment of outcomes. In the selection domain, studies were evaluated for representativeness of the exposed cohort, selection of the non-exposed cohort, and ascertainment of exposure. In the comparability domain, studies were evaluated for comparability of cohorts on the basis of the design or analysis, and adjustment for confounders. In the outcome domain, studies were evaluated for ascertainment of the outcome of interest, follow-up time, and adequacy of follow-up. Each criterion was given a score, and the scores were summed to produce a total score for each study. Any discrepancies between the two reviewers were resolved by a third reviewer. Studies were classified as having a low, moderate, or high risk of bias based on their NOS scores. A high-quality study was considered to have a low risk of bias in all three domains, while a low-quality study was considered to have a high risk of bias in at least one domain. The results of the bias assessment were taken into account in the interpretation of the findings of this systematic review and meta-analysis.

4.2. Meta-analysis strategy

The meta-analysis protocol for this review used the RevMan 5 software, which is a standard software for preparing and maintaining Cochrane systematic reviews. The meta-analysis aimed to generate forest plots of OR and RR using a random effects model and 95 % CI. The random effects model was chosen due to the anticipated heterogeneity in the included studies. The meta-analysis was performed using the data from the primary studies selected for inclusion in the systematic review. The forest plots were used to visually represent the effect sizes and the level of variability in the results across the included studies. The software was used to perform a statistical analysis of the data to calculate the pooled effect size, along with its confidence interval. The 95 % CI was used to determine the level of statistical significance of the pooled effect size. The meta-analysis protocol followed the standard guidelines for conducting a meta-analysis and was reviewed and approved by the review team. The forest plots generated using RevMan 5 were used to present the findings of the meta-analysis in a clear and concise manner.

5. Results

After the completion of the search protocol as devised by the reviewers, we were left with 11 studies [29–39] that were in accordance of our review objectives. Table 1 presents a list of studies along with their respective study ID, country where the study was conducted, sample size, and protocol used in the study. The sample sizes in the studies vary from 4 data sets to 240 novel concepts to 77 medical students to 33 physicians to 36 clinical vignettes to 14 multiple-part cases, and some studies have undefined sample sizes. All the studies used a descriptive protocol. Some of the studies have undefined locations or sample sizes. Overall, this table provides a brief overview of the characteristics of each study, such as the country where it was conducted, the sample size, and the protocol used.

Table 2 on the other hand represents the results of different studies that evaluated the performance of ChatGPT in various domains. The first study by Gilson et al. [29] assessed the accuracy of ChatGPT in relation to medical examination and reported an accuracy of 44 %, 42 %, 64.4 %, and 57.8 % across four different datasets of the United States Medical Licensing Exam. Gupta et al. [30] evaluated ChatGPT's ability to generate novel concepts related to systematic reviews and reported an overall accuracy of 55 %, with 35 % accuracy for generalised concepts and 75 % accuracy for specific concepts. Huh et al. [31] compared the knowledge and interpretation-based queries of medical students and ChatGPT and reported a ChatGPT accuracy of 60.8 % compared to an average of 89.6 % for the students. Johnson et al. [32] evaluated ChatGPT's accuracy in answering medical queries across 17 different specialties and reported a 39.4 % accuracy for perfect answers and 18.3 % accuracy for near-perfect answers. The study by Johnson SB [33] et al. assessed ChatGPT's accuracy in answering queries about cancer and reported an accuracy of 96.9 % for the obtained responses and 100 % for completeness. Rao A et al. [34] evaluated ChatGPT's performance in relation to diagnostic imaging in a clinical setting and reported an accuracy of 88.9 % for breast cancer evaluation and 58.3 % for prompts about breast pain. Rao et al. [35] assessed ChatGPT's efficacy in supporting clinical decisions on clinical vignettes and reported an overall accuracy of 71.7 %. Strong et al. [38]

Table 1
Demographic characteristics as observed in the selected papers.

Study ID	Country	Sample size	Protocol
Gilson et al. [29]	USA	4 data sets	Descriptive
Gupta et al. [30]	Undefined	240 novel concepts	Descriptive
Huh et al. [31]	South Korea	77 medical students	Descriptive
Johnson et al. [32]	USA	33 physicians	Descriptive
Johnson SB et al. [33]	Undefined	Undefined	Descriptive
Rao A et al. [34]	Undefined	Undefined	Descriptive
Rao et al. [35]	Undefined	36 vignettes (clinical)	Descriptive
Sallam et al. [36]	Jordan	Undefined	Descriptive
Sallam M et al. [37]	Undefined	Undefined	Descriptive
Strong et al. [38]	USA	14 multiple-part cases	Descriptive
Yeo et al. [39]	USA	Undefined	Descriptive

Table 2
Accuracy measurements, objectives and inferences pertaining to the selected studies.

Study ID	Objectives related to ChatGPT	Domain of assessment	Accuracy	Advantages	Disadvantages
Gilson et al. [29]	Assessment of ChatGPT accuracy in relation to medical examination	United States Medical Licensing Exam comprising of 4 different data sets	44 %, 42 %, 64.4 %, and 57.8 % across 4 datasets	100 % of a speicific data set outputs contained a logical reason for the answer given by ChatGPT. In 96.8 % of the queries, there was internal information related to the topic.	On two different data sets, the presence of information not directly related to the question was 44.5 % and 27 % lower for erroneous than for correct responses, respectively.
Gupta et al. [30]	Assessment of ChatGPT to demonstrate production of novel concepts related systematic reviews	20 specific concepts for systematic reviews across 12 varied topics of cosmetic surgery	55 % overall (35 % for generalised concepts and 75 % for specific)	The generalised accuracy rate was 55 % overall for the 240 specific concepts generated by ChatGPT and 75 % within the examined disciplines	General concepts generated had only 35 % accuracy
Huh et al. [31]	Comparison between medical students and ChatGPT in knowledge and interpretation-based queries	Parasitology examination consisting of 79 different items	60.8 % of ChatGPT compared to 89.6 % average of students	ChatGPT generated respectable answers which were in relation to the query that was being answered and the number of correct answers given by it were almost comparable to the students' answers	The general level of acceptability of the responses were not so good
Johnson et al. [32]	Assessment of ChatGPT accuracy in relation to medical queries	284 medical questions across 17 different specialities	39.4 % of perfect answers and 18.3 % of near-perfect answers	18.3 % responses were assessed as nearly all correct, while 39.4 % received the highest accuracy rating. Moreover, 26.1 % were satisfactory and 53.3 % were rated as comprehensive in terms of completeness.	8.3 % of the responses were wholly inaccurate, while 8.3 % responses were very poor from a completeness point of view.
Johnson SB et al. [33]	Assessment of ChatGPT accuracy in relation to queries about cancer	Myths surrounding cancer and related misconceptions	96.9 % for the obtained responses and 100 % for completeness	Clinicians considered the information sources to be trustworthy and informative, with clear and concise responses which debunked the illogical myths and associated misconceptions surrounding cancer.	Slightly less accuracy was obtained in terms of the comprehensiveness of obtained responses.
Rao A et al. [34]	Assessment of ChatGPT in relation to diagnostic imaging in a clinical setting	Breast cancer and breast pain evaluation	88.9 % for breast cancer evaluation and 58.3 % for prompts about breast pain	ChatGPT demonstrated significant accuracy in relation to radiological diagnostic protocol in a clinical setting	Unspecified
Rao et al. [35]	Assessment of ChatGPT efficacy in supporting clinical decisions on clinical vignettes	Merck Sharpe & Dohme (MSD) Clinical Manual's 36 clinical vignettes	71.7 % overall	ChatGPT exhibited above significant accuracy with respect to decision making in clinical settings	Poor performance with respect to differential diagnosis and clinical management
Sallam et al. [36]	Evaluation of the information produced in response to ChatGPT commands in connection to education across different domains (both medical and dental)	Pharmaceutical sciences, medical science, dentistry and public health	Undefined	In medical education, the potential to enhance individualised learning, clinical reasoning, and comprehension of difficult medical ideas. Improved abilities were mentioned in relation to dentistry education along with interactive information, step-by-step directions, and immediate feedback on student technique.	Issues related to plagiarism, copyright violations, academic dishonesty, and the lack of interpersonal and emotional interactions
Sallam M et al. [37]	Assessment of ChatGPT responses in relation to COVID-19 vaccination (both the oral and systemic aspect of the vaccine)	Conspiracy theories about COVID-19 vaccination and its associated aspects	85.3 % for completeness and 92.3 % for bias assessment	Clinicians considered the information sources to be trustworthy and informative, with clear and concise responses which debunked the illogical theories surrounding COVID-19 vaccination.	Clinicians relied on a subjective review of ChatGPT, which could produce somewhat different outcomes depending on the situation.

(continued on next page)

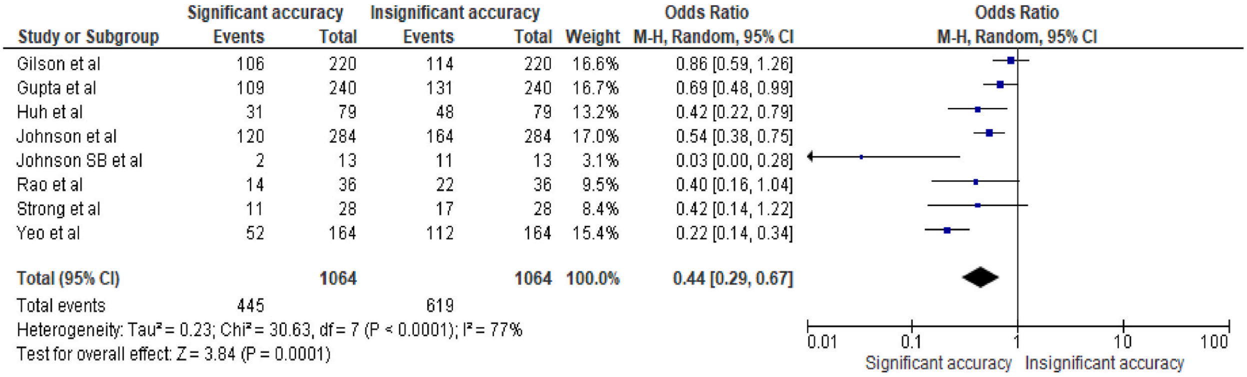
Table 2 (*continued*)

Study ID	Objectives related to ChatGPT	Domain of assessment	Accuracy	Advantages	Disadvantages
Strong et al. [38]	Assessment of ChatGPT in answering clinical reasoning questions based on both medical and dental domains	14 multiple-part cases of clinical reasoning examination for medical students	43 %–81 % across 20 simulations	In approximately half of the number of simulations performed, ChatGPT exhibited performance above the passing mark	Initial accuracy at the start was around 43 %
Yeo et al. [39]	Assessment of ChatGPT accuracy in relation to queries about liver diseases	164 queries about hepatocellular carcinoma and liver cirrhosis	79.1 % for liver cirrhosis, 74 % for carcinoma and 76.9 % for quality measures	Researchers noted that in comparison to the areas of diagnosis and preventive medicine, ChatGPT performed better in terms of basic knowledge, lifestyle, and therapy.	The accuracy was 47.3 % and 41.1 % for cirrhosis and carcinoma respectively.

assessed ChatGPT's accuracy in answering clinical reasoning questions and reported an accuracy range of 43 %–81 % across 20 simulations. Yeo et al. [39] evaluated ChatGPT's accuracy in answering queries related to liver diseases and reported an accuracy of 79.1 % for liver cirrhosis, 74 % for hepatocellular carcinoma, and 76.9 % for quality measures. Sallam et al. [36] evaluated the information produced in response to ChatGPT commands related to education across different domains, while Sallam M et al. [37] assessed ChatGPT's responses related to COVID-19 vaccination and reported an accuracy of 85.3 % for completeness and 92.3 % for bias assessment.

The forest plot shown in Fig. 3 presented an OR of 2.25, with a 95 % CI ranging from 1.49 to 3.40, indicating that the accuracy of ChatGPT in providing correct responses was significantly higher compared to the total responses for queries. The forest plot also showed significant heterogeneity among the studies, with a Tau^2 of 0.23, a Chi^2 of 30.63 with 7 degrees of freedom (df) ($P < 0.0001$), and an I^2 of 77 %, suggesting that there was considerable variability in the effect sizes across the included studies. Additionally, the test for overall effect yielded a Z score of 3.84 ($P = 0.0001$), indicating that there was a statistically significant difference between the accuracy of ChatGPT in providing correct responses compared to the total responses for queries (assuming a random effects model). The results of this meta-analysis suggest that the true effect size of ChatGPT's accuracy in providing correct responses may vary across studies due to both sampling error and genuine differences in the population, intervention, and outcome measures. Therefore, the estimated effect size of 2.25 may not be a constant true effect size but rather a distribution of true effect sizes, with a range from 1.49 to 3.40. This result implies that ChatGPT may have potential in accurately answering queries in the medical and dental fields, and further research is needed to explore the sources of heterogeneity and to identify ways to maximize the accuracy of ChatGPT in these contexts.

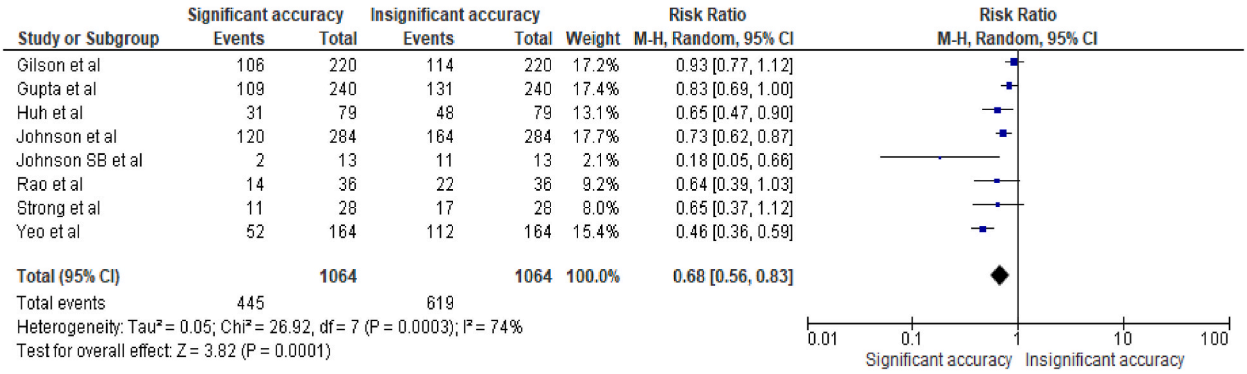
Fig. 4's forest plot presented a RR of 1.47, with a 95 % CI ranging from 1.21 to 1.80, indicating that the accuracy of ChatGPT in providing correct responses was significantly higher compared to the total responses for queries. The forest plot also showed significant heterogeneity among the studies, with a Tau^2 of 0.05, a Chi^2 of 26.92 with 7 degrees of freedom (df) ($P = 0.0003$), and an I^2 of 74 % assuming a random effects model, suggesting that there was considerable variability in the effect sizes across the included studies. Additionally, the test for overall effect yielded a Z score of 3.82 ($P = 0.0001$), indicating that there was a statistically significant difference between the accuracy of ChatGPT in providing correct responses compared to the total responses for queries. The results of this meta-analysis suggest that the true effect size of ChatGPT's accuracy in providing correct responses may vary across studies due to both sampling error and genuine differences in the population, intervention, and outcome measures. Therefore, the estimated effect size of 1.47 may not be a constant true effect size but rather a distribution of true effect sizes, with a range from 1.21 to 1.80. This result implies that ChatGPT may have potential in accurately answering queries in the medical and dental fields, and further research is needed to explore the sources of heterogeneity and to identify ways to maximize the accuracy of ChatGPT in these contexts. The



	Significant accuracy		Insignificant accuracy			Odds Ratio
Study or Subgroup	Events	Total	Events	Total	Weight	M-H, Random, 95% CI
Gilson et al	106	220	114	220	16.6%	0.86 [0.59, 1.26]
Gupta et al	109	240	131	240	16.7%	0.69 [0.48, 0.99]
Huh et al	31	79	48	79	13.2%	0.42 [0.22, 0.79]
Johnson et al	120	284	164	284	17.0%	0.54 [0.38, 0.75]
Johnson SB et al	2	13	11	13	3.1%	0.03 [0.00, 0.28]
Rao et al	14	36	22	36	9.5%	0.40 [0.16, 1.04]
Strong et al	11	28	17	28	8.4%	0.42 [0.14, 1.22]
Yeo et al	52	164	112	164	15.4%	0.22 [0.14, 0.34]
Total (95% CI)		**1064**		**1064**	**100.0%**	**0.44 [0.29, 0.67]**
Total events	445		619			

Heterogeneity: Tau² = 0.23; Chi² = 30.63, df = 7 (P < 0.0001); I² = 77%
Test for overall effect: Z = 3.84 (P = 0.0001)

Fig. 3. Overall accuracy of ChatGPT in providing correct responses as compared to total responses for queries represented in terms of OR on a forest plot.



Study or Subgroup	Significant accuracy Events	Significant accuracy Total	Insignificant accuracy Events	Insignificant accuracy Total	Weight	Risk Ratio M-H, Random, 95% CI
Gilson et al	106	220	114	220	17.2%	0.93 [0.77, 1.12]
Gupta et al	109	240	131	240	17.4%	0.83 [0.69, 1.00]
Huh et al	31	79	48	79	13.1%	0.65 [0.47, 0.90]
Johnson et al	120	284	164	284	17.7%	0.73 [0.62, 0.87]
Johnson SB et al	2	13	11	13	2.1%	0.18 [0.05, 0.66]
Rao et al	14	36	22	36	9.2%	0.64 [0.39, 1.03]
Strong et al	11	28	17	28	8.0%	0.65 [0.37, 1.12]
Yeo et al	52	164	112	164	15.4%	0.46 [0.36, 0.59]
Total (95% CI)		**1064**		**1064**	**100.0%**	**0.68 [0.56, 0.83]**
Total events	445		619			

Heterogeneity: Tau² = 0.05; Chi² = 26.92, df = 7 (P = 0.0003); I² = 74%
Test for overall effect: Z = 3.82 (P = 0.0001)

Fig. 4. Overall accuracy of ChatGPT in providing correct responses as compared to total responses for queries represented in terms of RR on a forest plot.

findings suggest that the accuracy of ChatGPT in providing correct responses to queries related to medical and dental specialties and subjects is promising. However, clinicians and researchers should be cautious and critically evaluate the accuracy of ChatGPT's responses as they may not be consistently reliable. Further studies are needed to investigate the factors that influence the accuracy of ChatGPT in these domains and to identify strategies for optimizing its performance.

6. Discussion

This study provides a comprehensive overview of various studies that have evaluated ChatGPT's accuracy in medical and dental domains. The findings of this study can have future implications for the development and application of natural language processing technology in the field of medicine as well as dentistry. The study demonstrates that ChatGPT can provide accurate responses to queries related to medical examination, systematic reviews, clinical reasoning, diagnostic imaging, liver diseases, and COVID-19 vaccination. These findings suggest that ChatGPT has the potential to serve as a valuable tool for healthcare providers and medical researchers in facilitating the diagnosis, treatment, and prevention of various diseases. Although ChatGPT has been released only around six months ago, the findings of this study are still relevant as they provide initial insights into ChatGPT's accuracy in different medical contexts. As natural language processing technology continues to advance, ChatGPT's accuracy is likely to improve, and its potential applications in medicine may expand. Therefore, this study can serve as a foundation for future research to build upon and further explore the efficacy of ChatGPT and other natural language processing technologies in medicine. The findings of this study can also inspire researchers to develop new protocols that can enhance the accuracy and reliability of ChatGPT's responses to medical queries. Overall, the study's significance lies in its potential to drive innovation and improve healthcare outcomes through the development of natural language processing technology. ChatGPT's potential to assist in optimizing clinical workflow appears promising, with the potential for cost savings and improved healthcare delivery efficiency [21,40–42]. This was recently demonstrated by Patel and Lam, who highlighted ChatGPT's capacity to generate effective discharge summaries, which can be useful to lessen the load of documentation in the healthcare industry [43]. Additionally, ChatGPT has the potential to revolutionise the way healthcare is delivered by improving diagnostics, predicting illness risk and outcome, and discovering new drugs, among other translational research fields [44–46]. Additionally, ChatGPT demonstrated moderate accuracy in identifying the imaging processes required for breast cancer screening and in the assessment of breast discomfort, suggesting that it has potential for use in radiology decision-making [34]. By making accessible and understandable health information available to the general population, ChatGPT in health care settings also has the potential to increase health literacy and advance personalized treatment [20,22,39,47,48]. Responses from ChatGPT illustrated its usefulness by emphasizing the importance of consulting healthcare professionals and other trustworthy sources in particular circumstances [37,49].

On the other hand, a number of issues with ChatGPT's application in healthcare settings were brought up. Transparency difficulties and other ethical concerns, such as the possibility of prejudice, are frequently raised [34,42,44,46]. Furthermore, the creation of erroneous content might have serious negative effects on healthcare; as a result, this legitimate issue should be carefully taken into account in clinical practise [20,22,43,50]. This worry also extends to ChatGPT's capacity to offer explanations for erroneous judgements [34]. Other ChatGPT drawbacks include the problems with interpretability, repeatability, and the treatment of uncertainty [46,48,51], all of which can have negative effects in healthcare settings and health care research. Given the variation in multiple health-related features found across various populations, the lack of openness and ambiguous information on the sources of data used for ChatGPT training are significant problems in health care settings [34]. Reproducibility between ChatGPT prompt runs, which can be a significant drawback in clinical use [44], is a crucial concern.

There are some limitations of this article that should be taken into account. First off, the study sample sizes, which ranged from 4 data sets to 240 unique concepts, were extremely tiny. Small sample sizes may affect statistical power, making it challenging to extrapolate the results to larger populations. Second, while all studies employed a descriptive methodology, several had ambiguous sample sizes or locations, which could have an effect on the reliability of the findings. Thirdly, there was a large amount of heterogeneity across the studies included in the meta-analysis, indicating a wide range of effect sizes. As a result, it is possible that the estimated effect sizes are not constant, and more research is required to investigate the causes of heterogeneity and find strategies for enhancing ChatGPT's accuracy in these situations. Last but not least, the research examined ChatGPT's accuracy in certain fields including medical inquiries, education, and COVID-19 immunisation; hence, the results may not generalise to other fields. The study's limitations must be taken into account when evaluating the findings, even though they shed insight on ChatGPT's accuracy in many sectors. The potential of ChatGPT in diverse circumstances needs to be further investigated, while also taking into account the limitations this study found.

7. Conclusion

Summarily speaking, the findings from this study indicate that ChatGPT has shown promise in providing accurate responses to medical queries and indicated a significantly higher accuracy compared to the total responses for queries. However, noticeable heterogeneity was observed among the included studies, suggesting variability in the effect sizes across different medical contexts. Therefore, further research is needed to explore the sources of heterogeneity and to identify ways to maximize the accuracy of ChatGPT in these contexts. Overall, this study highlights the potential of ChatGPT in the field of medicine and dentistry, and its future implications could range from providing clinical decision support to facilitating medical education and research.

Data availability statement

No data associated with our study been deposited into a publicly available repository. Data included in article/supp. material/referenced in article. All data are available within the manuscript in the form of results/tables/figures in article.

CRediT authorship contribution statement

Hiroj Bagde: Writing – review & editing, Writing – original draft, Software, Resources, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Ashwini Dhopte:** Writing – review & editing, Writing – original draft, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Mohammad Khursheed Alam:** Writing – review & editing, Writing – original draft, Supervision, Software, Project administration, Investigation, Data curation, Conceptualization. **Rehana Basri:** Writing – review & editing, Writing – original draft, Methodology, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.heliyon.2023.e23050.

References

[1] K.H. Yu, A.L. Beam, I.S. Kohane, Artificial intelligence in healthcare, Nat. Biomed. Eng. 2 (10) (2018) 719–731, https://doi.org/10.1038/s41551-018-0305-z.

[2] L. Xu, L. Sanders, K. Li, J.C.L. Chow, Chatbot for health care and oncology applications using artificial intelligence and machine learning: systematic review, JMIR Cancer 7 (4) (2021), e27850, https://doi.org/10.2196/27850.

[3] D.B. Chonde, A. Pourvaziri, J. Williams, et al., Rad translate: an artificial intelligence– powered intervention for urgent imaging to enhance care equity for patients with limited English proficiency during the COVID-19 pandemic, J. Am. Coll. Radiol. 18 (7) (2021) 1000–1008, https://doi.org/10.1016/j.jacr.2021.01.013.

[4] J. Chung, D. Kim, J. Choi, et al., Prediction of oxygen requirement in patients with COVID19 using a pre-trained chest radiograph xAI model: efficient development of auditable risk prediction models via a fine-tuning approach, Sci. Rep. 12 (1) (2022), 21164, https://doi.org/10.1038/s41598-022-24721-5.

[5] M.D. Li, N.T. Arun, M. Aggarwal, et al., Multi-population generalizability of a deep learning based chest radiograph severity score for COVID-19, Medicine (Baltim.) 101 (29) (2022), e29587, https://doi.org/10.1097/MD.0000000000029587.

[6] D. Kim, J. Chung, J. Choi, et al., Accurate auto-labeling of chest X-ray images based on quantitative similarity to an explainable AI model, Nat. Commun. 13 (1) (2022) 1867, https://doi.org/10.1038/s41467-022-29437-8.

[7] A. O'Shea, M.D. Li, N.D. Mercaldo, et al., Intubation and mortality prediction in hospitalized COVID-19 patients using a combination of convolutional neural network-based scoring of chest radiographs and clinical data, BJR|Open. 4 (1) (2022), 20210062, https://doi.org/10.1259/bjro.20210062.

[8] J. Witowski, J. Choi, S. Jeon, et al., MarkIt: a collaborative artificial intelligence annotation platform leveraging blockchain for medical imaging research, Blockchain Healthc Today (2021), https://doi.org/10.30953/bhty.v4.176. Published online May 5.

[9] ChatGPT, Optimizing Language models for dialogue, OpenAI. Published November 30 (2022). https://openai.com/blog/chatgpt/. (Accessed 15 April 2023).

[10] T.H. Kung, M. Cheatham, G.P.T. Chat, et al., Performance of ChatGPT on USMLE: potential for AI-assisted medical education using large language models, Published online December 21 (2022) (2022), https://doi.org/10.1101/2022.12.19.22283643.

[11] M. Bommarito II, D.M. Katz, GPT takes the bar exam, Published online (2022), https://doi.org/10.48550/arXiv.2212.14402. December 29.

[12] J.H. Choi, K.E. Hickman, A. Monahan, D. Schwarcz, ChatGPT Goes to Law School 23 (2023), https://doi.org/10.2139/ssrn.4335905. Published online January.

[13] T.J. Chen, ChatGPT and other artificial intelligence applications speed up scientific writing, J. Chin. Med. Assoc. 86 (4) (2023) 351–353, https://doi.org/10.1097/JCMA.0000000000000900. PMID 36791246.

[14] H.H. Thorp, ChatGPT is fun, but not an author, Science 379 (2023) 313, https://doi.org/10.1126/science.adg7879.

[15] F.C. Kitamura, ChatGPT is shaping the future of medical writing but still requires human judgment, Radiology (2023), 230171, https://doi.org/10.1148/radiol.230171.

[16] J. Lubowitz, ChatGPT, An artificial intelligence chatbot, is impacting medical literature, Arthroscopy 39 (5) (2023) 1121–1122, https://doi.org/10.1016/j.arthro.2023.01.015. PMID 36797148.

[17] Nature editorial Tools such as ChatGPT threaten transparent science; here are our ground rules for their use, Nature 613 (2023) 612, https://doi.org/10.1038/d41586-023-00191-1.

[18] P. Moons, L. Van Bulck, ChatGPT: Can artificial intelligence language models be of value for cardiovascular nurses and allied health professionals, Eur. J. Cardiovasc. Nurs. 22 (7) (2023) e55–e59, https://doi.org/10.1093/eurjcn/zvad022. PMID 36752788.

[19] P. Cahan, B. Treutlein, A conversation with ChatGPT on the role of computational systems biology in stem cell research, Stem Cell Rep. 18 (2023) 1–2, https://doi.org/10.1016/j.stemcr.2022.12.009.

[20] C. Ahn, Exploring ChatGPT for information of cardiopulmonary resuscitation, Resuscitation 185 (2023), 109729, https://doi.org/10.1016/j.resuscitation.2023.109729.

[21] J. Gunawan, Exploring the future of nursing: insights from the ChatGPT model, Belitung Nurs. J. 9 (2023) 1–5, https://doi.org/10.33546/bnj.2551.

[22] R.S. D'Amico, T.G. White, H.A. Shah, D.J. Langer, I asked a ChatGPT to write an editorial about how we can incorporate chatbots into neurosurgical research and patient care, Neurosurgery 92 (2023) 993–994, https://doi.org/10.1227/neu.0000000000002414.

[23] N. Fijačko, L. Gosak, G. Štiglic, C.T. Picard, M. John Douma, Can ChatGPT pass the life support exams without entering the American heart association course? Resuscitation 185 (2023), 109732 https://doi.org/10.1016/j.resuscitation.2023.109732.

[24] A.B. Mbakwe, I. Lourentzou, L.A. Celi, O.J. Mechanic, A. Dagan, ChatGPT passing USMLE shines a spotlight on the flaws of medical education, PLoS Digit. Health. 2 (2023), e0000205, https://doi.org/10.1371/journal.pdig.0000205.

[25] S. Huh, Issues in the 3rd year of the COVID-19 pandemic, including computer-based testing, study design, ChatGPT, journal metrics, and appreciation to reviewers, J. Educ. Eval. Health Prof. 20 (2023) 5, https://doi.org/10.3352/jeehp.2023.20.5.

[26] A. Liberati, The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration, Ann. Intern. Med. 151 (4) (2009), https://doi.org/10.7326/0003-4819-151-4-200908180-00136.

[27] L.A. McGuinness, J.P. Higgins, Risk-of-bias visualization (robvis): an R package and shiny web app for visualizing risk-of-bias assessments, Res. Synth. Methods 12 (1) (2020) 55–61, https://doi.org/10.1002/jrsm.1411.

[28] C.K.L. Lo, D. Mertz, M. Loeb, Newcastle-Ottawa Scale: comparing reviewers' to authors' assessments, BMC Med. Res. Methodol. 14 (2014) 45, https://doi.org/10.1186/1471-2288-14-45.

[29] A. Gilson, C.W. Safranek, T. Huang, V. Socrates, L. Chi, R.A. Taylor, D. Chartash, 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 9 (2023 Feb 8), e45312, https://doi.org/10.2196/45312. PMID: 36753318; PMCID: PMC9947764.

[30] R. Gupta, J.B. Park, C. Bisht, I. Herzog, J. Weisberger, J. Chao, K. Chaiyasate, E.S. Lee, Expanding cosmetic plastic surgery research using ChatGPT, Aesthetic Surg. J. (2023 Mar 21) sjad069, https://doi.org/10.1093/asj/sjad069. Epub ahead of print. PMID: 36943815.

[31] S. Huh, 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 20 (2023) 1, https://doi.org/10.3352/jeehp.2023.20.1. Epub 2023 Jan 11. PMID: 36627845; PMCID: PMC9905868.

[32] D. Johnson, R. Goodman, J. Patrinely, C. Stone, E. Zimmerman, R. Donald, S. Chang, S. Berkowitz, A. Finn, E. Jahangir, E. Scoville, T. Reese, D. Friedman, J. Bastarache, Y. van der Heijden, J. Wright, N. Carter, M. Alexander, J. Choe, C. Chastain, J. Zic, S. Horst, I. Turker, R. Agarwal, E. Osmundson, K. Idrees, C. Kiernan, C. Padmanabhan, C. Bailey, C. Schlegel, L. Chambless, M. Gibson, T. Osterman, L. Wheless, Assessing the accuracy and reliability of AI-generated medical responses: an evaluation of the chat-GPT model, Res. Sq. [Preprint] (2023 Feb 28), rs.3.rs-2566942, https://doi.org/10.21203/rs.3.rs-2566942/v1. PMID: 36909565; PMCID: PMC10002821.

[33] S.B. Johnson, A.J. King, E.L. Warner, S. Aneja, B.H. Kann, C.L. Bylund, Using ChatGPT to evaluate cancer myths and misconceptions: artificial intelligence and cancer information, Mar 1, JNCI Cancer Spectr. 7 (2) (2023) pkad015, https://doi.org/10.1093/jncics/pkad015. PMID: 36929393; PMCID: PMC10020140.

[34] A. Rao, J. Kim, M. Kamineni, M. Pang, W. Lie, M.D. Succi, Evaluating ChatGPT as an adjunct for radiologic decision-making, medRxiv [Preprint] (2023 Feb 7: 2023), https://doi.org/10.1101/2023.02.02.23285399. PMID: 36798292; PMCID: PMC9934725.

[35] A. Rao, M. Pang, J. Kim, M. Kamineni, W. Lie, A.K. Prasad, A. Landman, K.J. Dreyer, M.D. Succi, Assessing the Utility of ChatGPT throughout the Entire Clinical Workflow, 2023.02.21.23285886, medRxiv [Preprint] (2023), https://doi.org/10.1101/2023.02.21.23285886. PMID: 36865204; PMCID: PMC9980239.

[36] Malik Sallam, Nesreen Salim, Muna Barakat, Al-Tammemi, Alaa, ChatGPT applications in medical, dental, pharmacy, and public health education: a descriptive study highlighting the advantages and limitations, Narrative J 3 (2023) e103, https://doi.org/10.52225/narra.v3i1.103.

[37] M. Sallam, N.A. Salim, A.B. Al-Tammemi, M. Barakat, D. Fayyad, S. Hallit, H. Harapan, R. Hallit, A. Mahafzah, ChatGPT output regarding compulsory vaccination and COVID-19 vaccine conspiracy: a descriptive study at the outset of a paradigm shift in online search for information, Feb 15, Cureus 15 (2) (2023), e35029, https://doi.org/10.7759/cureus.35029. PMID: 36819954; PMCID: PMC9931398.

[38] E. Strong, A. DiGiammarino, Y. Weng, P. Basaviah, P. Hosamani, A. Kumar, A. Nevins, J. Kugler, J. Hom, J.H. Chen, Performance of ChatGPT on free-response, clinical reasoning exams, Mar 29:2023.03.24.23287731, medRxiv [Preprint] (2023), https://doi.org/10.1101/2023.03.24.23287731. PMID: 37034742; PMCID: PMC10081420.

[39] Y.H. Yeo, J.S. Samaan, W.H. Ng, P.S. Ting, H. Trivedi, A. Vipani, W. Ayoub, J.D. Yang, O. Liran, B. Spiegel, A. Kuo, Assessing the performance of ChatGPT in answering questions regarding cirrhosis and hepatocellular carcinoma, Clin. Mol. Hepatol. (2023 Mar 22), https://doi.org/10.3350/cmh.2023.0089. Epub ahead of print. PMID: 36946005.

[40] Y. Shen, L. Heacock, J. Elias, K.D. Hentel, B. Reig, G. Shih, L. Moy, ChatGPT and other large language models are double-edged swords, Radiology (2023), 230163, https://doi.org/10.1148/radiol.230163.

[41] M. Mijwil, M. Aljanabi, A. Ali, ChatGPT: exploring the role of cybersecurity in the protection of medical information, Mesop. J. CyberSecurity. (2023) 18–21, https://doi.org/10.58496/MJCS/2023/004.

[42] K. Jeblick, B. Schachtner, J. Dexl, A. Mittermeier, A.T. Stüber, J. Topalis, T. Weber, P. Wesp, B. Sabel, J. Ricke, et al., ChatGPT Makes Medicine Easy to Swallow: an Exploratory Case Study on Simplified Radiology Reports, 2022, https://doi.org/10.48550/arxiv.2212.14882.2212.14882 arXiv.

[43] S.B. Patel, Lam K. ChatGPT, The future of discharge summaries? Lancet Digit. Health. 5 (2023) e107–e108, https://doi.org/10.1016/S2589-7500(23)00021-3.

[44] A. Holzinger, K. Keiblinger, P. Holub, K. Zatloukal, H. Müller, AI for life: trends in artificial intelligence for biotechnology, N. Biotechnol. 74 (2023) 16–24, https://doi.org/10.1016/j.nbt.2023.02.001.

[45] D. Mann, Artificial intelligence discusses the role of artificial intelligence in translational medicine: a jacc: basic to translational science interview with ChatGPT, J. Am. Coll. Cardiol. Basic Trans. Sci. 8 (2023) 221–223, https://doi.org/10.1016/j.jacbts.2023.01.001.

[46] G. Sharma, A. Thakur, ChatGPT in Drug Discovery. ChemRxiv, Preprint, 2023, https://doi.org/10.26434/chemrxiv-2023-qgs3k.

[47] T.H. Kung, M. Cheatham, A. Medenilla, C. Sillos, L. De Leon, C. Elepaño, M. Madriaga, R. Aggabao, G. Diaz-Candido, J. Maningo, et al., Performance of ChatGPT on USMLE: potential for AI-assisted medical education using large language models, PLOS Digit. Health. 2 (2023), e0000198, https://doi.org/10.1371/journal.pdig.0000198.

[48] D. Duong, B.D. Solomon, Analysis of large-language model versus human performance for genetics questions, medRxiv (2023), https://doi.org/10.1101/2023.01.27.23285115. Preprint.

[49] A. Zhavoronkov, Rapamycin in the context of Pascal's Wager: generative pre-trained transformer perspective, Oncoscience 9 (2022) 82–84, https://doi.org/10.18632/oncoscience.571.

[50] S.R. Ali, T.D. Dobbs, H.A. Hutchings, I.S. Whitaker, Using ChatGPT to write patient clinic letters, Lancet Digit. Health. (2023), https://doi.org/10.1016/S2589-7500(23)00048-1. Online ahead of print.

[51] F. Sanmarchi, A. Bucci, D. Golinelli, A step-by-step Researcher's Guide to the use of an AI-based transformer in epidemiology: an exploratory analysis of ChatGPT using the STROBE checklist for observational studies, medRxiv (2023), https://doi.org/10.1007/s10389-023-01936-y. Preprint.