



(REVIEW ARTICLE)



AI Chatbots for Primary Care Triage: A systematic review

Michael Ajemba ^{1,*} and Mgbeahuru Mgbedikearu ²

¹ *Clinical Associate Professor, ULAN.*

² *Primi Dona Magni Research Lab, Awka, Nigeria.*

World Journal of Advanced Research and Reviews, 2026, 29(03), 845-855

Publication history: Received on 10 January 2026; revised on 26 February 2026; accepted on 28 February 2026

Article DOI: <https://doi.org/10.30574/wjarr.2026.29.3.0138>

Abstract

Background: Demand pressures in primary care—rising patient volumes, workforce shortages, and administrative overload—have intensified interest in artificial intelligence (AI) tools that can support access, triage, and workflow efficiency. AI chatbots, particularly those using natural language processing and machine learning, are increasingly positioned as scalable interfaces for symptom assessment, patient navigation, and documentation support; however, concerns persist regarding clinical accuracy, safety, bias, privacy, and the patient experience.

Objective: This systematic review evaluated the effectiveness of AI-powered chatbots for primary care triage compared with conventional clinician-led triage, focusing on accuracy, efficiency, and patient satisfaction/usability, and summarised key implementation risks and governance considerations.

Methods: A PRISMA 2020-informed systematic search of PubMed, PubMed Central, ResearchGate, and Google Scholar was conducted for studies published from 2015 to the present. Screening and selection were guided by a PICO framework (primary care triage populations; AI chatbot interventions; clinician/standard triage comparators; outcomes of accuracy, efficiency, satisfaction, and utilisation). Narrative synthesis was used to integrate findings, and methodological quality was appraised using the Critical Appraisal Skills Programme (CASP) checklists.

Results: After deduplication and staged screening, eight studies met inclusion criteria, comprising surveys, comparative evaluations, narrative and scoping reviews, and health-services analyses. Across included evidence, chatbots demonstrated potential efficiency gains, particularly in administrative support and rapid generation of clinical documentation, with one comparative study reporting substantially faster discharge summary production than clinicians while achieving broadly similar scoring on quality/accuracy metrics. In symptom-related triage contexts, performance varied by model and clinical complexity; comparative assessment in ophthalmology triage reported higher expert-rated accuracy and clarity for ChatGPT relative to another large language model, while narrative syntheses consistently highlighted limitations in complex reasoning, inconsistent handling of nuanced presentations, and lack of access to non-verbal cues—factors central to safe primary care triage. [2–4] Studies also raised recurring concerns regarding hallucinations, equity-related harms driven by training data limitations, information governance, and patient trust—suggesting that chatbot triage is best deployed as clinician-supervised decision support rather than a replacement for professional assessment.

Conclusion: AI chatbots show promise for improving primary care triage efficiency and supporting administrative workflows, but current evidence indicates variable diagnostic/triage accuracy and unresolved challenges in safety, equity, privacy, and patient experience. Responsible integration should prioritise clinically validated use-cases, transparent governance, human oversight, and continuous evaluation using patient-centred outcomes and real-world safety monitoring. [1–4,6,8]

* Corresponding author: Michael Ajemba

Keywords: Artificial intelligence; Chatbots; Primary care; Triage; Symptom assessment; Natural language processing; Patient satisfaction.

1. Introduction

In recent years, the use of Artificial intelligence in healthcare has been rising. From administrative duties, to record keeping and diagnosis, more clinicians are adopting artificial intelligence to help reduce the burden of the rapidly increasing patient population and workload. The increase in demand for primary care services has put a strain on the health system leading to increased wait time, shortage of practitioners, preventable deaths and patient dissatisfaction. [1]The administrative burden, including time spent documenting in the electronic health record (EHR), is a known contributor to clinician burnout and waste in the healthcare system (Matulis and McCoy, 2023). From image analysis in radiology to natural language processing (NLP) for electronic health records, AI technologies are optimizing healthcare workflows, improving diagnostic accuracy, and enabling the customization of treatment strategies. [2]Machine learning algorithms are aiding in the early detection of diseases and provide healthcare professionals with data-driven insights (David et al., 2024), as well as providing representation to augment the shortage in healthcare practitioners.

Among the many AI technologies being adopted by healthcare practitioners, AI Chatbots have emerged as a promising solution to streamline triage, symptoms assessment and patient guidance. [3]A Chatbot is an automated software application designed to simulate and facilitate conversations with human users, typically through text- or voice-based interactions (Sabaner et al., 2024). [4]They leverage natural language processing (NLP) and machine learning (ML) algorithms to comprehend and respond to user inquiries conversationally and intuitively (Chakraborty et al., 2023). Chatbots can be used in a variety of applications, such as customer service, healthcare, and e-commerce, to provide instant responses and personalized experiences. [4]The essential components or technology used for Chatbots are Pattern Matching, Natural Language Understanding (NLU), Artificial Intelligence Markup Language (AIML), Latent Semantic Analysis (LSA), Natural Language Processing (NLP), Chatscript, entity, RiveScript, contexts (Chakraborty et al., 2023). Alan Turing first proposed the idea of AI and conceived the term in 1950 when he proposed the "imitation game" to understand the machine's capacity, which can be distinguished from humans. The evolution of Chatbots dates back to the 1960s, when Joseph Weizenbaum developed the first Chatbot, ELIZA a rule-based program that uses pattern recognition and substitution to simulate human like dialogue. [5]After ELIZA other Chatbots evolved such as Parry by Colby in the mid-1970s, ALICE in 1995, Smartchild by ActiveBuddy Inc. in 2001, siri by Apple In 2011, IBM Watson, Amazon Alexa, Microsoft Cortana, and then Chatgpt (Wang, 2024). [6]The launch of Chatgpt in November 2022 popularized the use of AI Chatbots among researchers and students. Following its launch, interest in large language model (LLM)-powered Chatbots has surged with increasing focus on the clinical potential of these tools (Blease et al., 2024). [4]Chatbots are trained to carry out these intelligent functions in several ways; the rule-based method which uses a set of rules to guide the chatbot's responses, mostly suitable for user queries; the machine learning method which involves training the Chatbot on a vast range of data using algorithms, suitable for complex user queries; Lastly, there is the hybrid approach which combines the rule-based and machine learning methods, hence, the chatbot follows rules for simple queries and ML for complex queries thus, it is able to handle a wider Range of user queries (Chakraborty et al., 2023).

Although only recently introduced, in the medical context, AI-based Chatbots already have diverse applications. They serve as popular and accessible resources for answering medical questions, offering information on symptoms, treatments, and general health advice. [2]Patients can now inquire about a wide range of medical topics, without the need for immediate medical consultation (David et al., 2024).

[7]Primary care is structured to be the first point of contact for patients making primary care clinicians are the first point of contact to sick people in most communities. Their services include diagnosis and treatment of health conditions and long-term care (Care, 2023), most of which involves interaction. [6]Several AI Chatbots have been developed to carry out some of these functions of primary care physicians. In an online survey by Blease et al. (2024), general practitioners in the UK showed to have used AI Chatbots like Chatgpt, Microsoft Bing AI, Google's Bard, and others. These AI Chatbots were reported to have provided assistance in duties such as generating documentation after appointment, suggesting differential analysis, suggesting treatment options, and writing letters.

The use of AI Chatbots in primary care not only makes the workload easier for primary care clinicians, it provides representation to make up for shortage of primary care professionals. It also ensures accurate record keeping of patient data, thereby providing enough information for accurate diagnosis. In addition, the use of AI Chatbots reduces carbon foot print from patients commuting to the health center while also leaving the centres less cluttered. [4]Chatbots can reduce the carbon footprint of healthcare facilities by decreasing the necessity for patients to go to hospitals for minor issues and uncertainties (Chakraborty et al., 2023). [4]This not only saves the time of the patients but also that of the

clinicians as they would focus on the aspect of their job that requires more human participation. Some examples of chatbots used in medical science as identified by Chakraborty et al. (2023) includes **Smart Chatbot** - that performs follow up and assessment on beta thalassemia patients and monitor their status; **Emoha** - provides cognitive support through CBT-Bot exercises and guided conversations for mental healthcare and support in China; **ChatGPT** - generates treatment advice and serves as decision tools in many specialties including ophthalmology, mental diseases, cancer etc; **EQRbot Chatbot** - used to generate explanations for patients' treatment advice specifically for as a pattern of explanation-question-response interactions; **Ask Rosa** - Used to identify best practices for future patients about hereditary breast and ovarian cancer for future patient-focused chatbots; **Tana Chatbot** - Used as a complement to care by health personnel during the COVID-19 period; **Vickybot** - used to screen, monitor, and reduce anxiety-depressive symptoms and work-related burnout, and in detecting suicide risk in patients from PC and health care workers; **COVID-Bot** - Used to screen, design and develop COVID-19 vaccination and its confirming status; **Otis Chatbot** - mixed methods of cognitive behavioral therapy (CBT)-based Chatbot applied for patient's health anxiety management for adults in New Zealand during the COVID-19 pandemic, **SnehAIChatbot** - was used to provide young people with sexual and reproductive health advice in India, Anti-TB Chatbot, and ChatbotWakamola

AI Chatbots have been shown to improve efficiency in primary care systems, but there are still questions on how well they attend to the need of the patients. Aside providing accurate answers and feedback to the patients, how much satisfaction does the patient feel after interacting with the bots as compared to the human clinicians. It is not just important to know that AI Chatbots are beneficial to primary care triage, it is important to know the level of impact this has. A comprehensive understanding of both the benefits and limitations of AI chatbots is essential for their effective and responsible integration into primary care settings.

This systematic review seeks to examine the current applications of AI chatbots in primary care, evaluating their effectiveness, benefits, and limitations. By analyzing recent studies and real-world implementations, the review aims to provide insights into how these technologies can be integrated responsibly into primary care settings to enhance patient engagement and support healthcare providers.

Research question: How effective are AI-powered chatbots in primary care triage compared to conventional triage methods in terms of accuracy, efficiency, and patient satisfaction

2. Research method

A thorough literature search was done on the subject under review. It was conducted using a comprehensive range of electronic databases which included PubMed, PubMed Central, ResearchGate, and Google Scholar. The search covered publications from 2015 to the present. Medical subject headings and keywords were used relating to 'Artificial intelligence, 'Chatbots' 'primary care'.

The databases were used because they are medical databases. [8]Inclusion and exclusion criteria were utilized in the screening and selection of literature in addition to the PRISMA process (Page et al., 2021). [9]According to Lisy et al. (2016), narrative synthesis was employed to extract data. [10]The CASP program according to CASP (2023) was used to analyze a selection of the literature, and Microsoft Excel was used to retrieve the data for each article.

3. Project research design & methodology

To better understand the roles and impacts of AI Chatbots on primary care delivery, a systematic literature is an important tool. [11]Based on published research, in the medicine education, using the qualitative review process, research issues were evaluated and critical conclusions for a few chosen quantitative publications were reached (Couto et al. 2020). [12]The results and observations from these studies emphasized on the need for control from app developers, family and schools. To locate, access, and summarize pertinent data from this literature, the review questions were methodically addressed in this study (Chen, 2017).

4. Search strategy & inclusion & exclusion criteria

The systematic review in this study employed the PICO (population, intervention, comparison and outcome) framework. For all reviews, it is important to define criteria such as the population, intervention, comparison and outcomes, and to identify potential risks of bias. [13]Reviews of the effect of rehabilitation interventions or reviews of data from observational studies, diagnostic test accuracy, or qualitative data may be more methodologically challenging

than reviews of effectiveness (Pollock A, Berge E, 2018). This PICO format will help guide the systematic review by focusing on AI Chatbots and their role in primary care.

Population: The search words used for this component determine location that will be investigated. The cohorts being investigated in this study are patients receiving care in primary care settings.

Intervention is the phrase used to define the treatments or strategies work best to improve outcomes and making a difference. [14]It was thus claimed that with a PICO-based research question, one can only generate study designs that detect correlations between a treatment and a desired outcome (Andreas Nishikawa-Pacher, 2022). The intervention in this study is the use of AI Chatbots in primary care delivery.

Comparison refers to the alternative interventions or exposures compared with the one being Investigated. The comparison is services delivered by AI chatbots and human clinicians.

The Outcome result that should be visible, measurable, or detectable in the population under observation. In this study, the outcome are patient satisfaction, better hospital administration and efficiency in primary care delivery [13]Based on the research perspective that this systematic review gave the framework for primary articles was used (Pollock and Berge, 2018). In this systematic review, a predetermined database query strategy was created using the PICO framework to find and obtain literature from the chosen databases. The database query procedure was uniform, systematic, and repeatable because to the PICO-driven predetermined search strategy. The criteria for including and excluding the retrieved articles were created using the PICO framework, search keywords, and phrases.

Table 1 Inclusion/Exclusion Criteria

S/N	PICO	Inclusion Criteria	Exclusion Criteria
Population (P):	Patients seeking primary care for symptom assessment and triage.	Adults or children (any age) using AI chatbots for primary care triage, symptom checking, or initial diagnosis or non-emergency primary care guidance.	Patients seeking emergency or specialist care
Intervention (I):	AI-powered chatbots used for primary care triage and symptom assessment.	Studies assessing AI chatbots designed for primary care triage, including symptom checkers and virtual assistants.	Studies focusing on AI in non-healthcare domains (e.g., finance, customer service).
Comparison (C):	Traditional or clinician-based triage.	Studies comparing AI chatbot with human-based triage (e.g., nurse triage, physician consultation, or telephone-based assessment).	Studies that do not include a human comparator group.
Outcome (O):	Accuracy, efficiency, patient experience, and impact on healthcare utilization.	AI chatbot accuracy in symptom assessment and triage recommendations. Patient satisfaction, usability, and trust in AI chatbots.	Outcomes unrelated to patient triage, diagnosis, or primary care service delivery.

As described in the next section of this study, database result filters were also employed to further limit the search results in addition to the inclusions/exclusions.

During the screening process, the titles and abstracts of the articles were reviewed. The screening and selection process was utilized once the record was collected to make sure that only studies that matched the inclusion/exclusion criteria were included. [8]In order to make sure that the studies were pertinent to the research issue, the PICO framework was employed in conjunction with the screening and selection process (Page et al., 2021). [8]This reduced reviewer bias and made it possible to include those papers that had a direct bearing on the research issue (Page et al., 2021). PRISMA, the preferred reporting item for systematic reviews and meta-analyses, served as the basis for the article screening and selection procedure (Page et al., 2021).

Table 2 Summary of the data base query strategy

S/N	PICO	Research Definition	Search Terminology
1	Population	Patients seeking primary care for symptom assessment and triage.	"Patients using AI chatbots for symptom assessment in primary care" OR "Primary care patients interacting with AI chatbots"
2	Intervention	AI-powered chatbots used for primary care triage and symptom assessment.	"AI chatbot for symptom checking" OR "Artificial intelligence in primary care triage" OR "Machine learning for patient triage"
3	Comparison	Traditional clinician-based triage methods.	"Nurse triage" OR "Physician-led triage" OR "Telephone-based triage"
4	Outcome	Accuracy, efficiency, patient experience, and impact on healthcare utilization.	"Diagnostic accuracy of AI chatbots" OR "Patient satisfaction with AI triage" OR "Healthcare access and efficiency"

4.1. Critical Appraisal

[12]Critical evaluation and quality assessment of certain papers were also included in the systematic review (Chen, 2017). [12]This procedure involved evaluating how well the publications adhered to the necessary methodological rigour (Chen, 2017). The CASP instrument, a commonly used technique for assessing the calibre of a research project, was used for this assessment. The rigour applied in the execution of various areas of the original research article is evaluated by the CASP tool using ten questions. CASP tool was used among other tools due to its benefits which includes organization of and management of large volume of literature by organizing studies based on criteria, facilitating a more systematic and transparent approach to data analysis. Additionally, it identifies patterns, trends and gaps in the existing literature, enhancing the synthesis of evidence or a comprehensive review. While the critical appraisal skills programme (CASP) tools are widely used or systematic review or assessing study quality, they have some limitations. One limitation is that CASP tools may oversimplify complex meteorological issues, potentially leading to subjective interpretations. Additionally, the tool may not cover all relevant aspects of study quality and their application can be time-consuming.

4.2. Data Extraction and Analysis

After being chosen, pertinent data was retrieved using Microsoft Excel from the chosen literature. [15]Extracting pre-defined information from chosen articles and structuring it for additional analysis and synthesis is known as data extraction (Aromataris& Pearson, 2014).The categories of data extraction include: Basic information such as author, year of publishing, and title of publication, study features which include the methodology, the aim, and the design. [15]Study population and sampling such as sample traits, selection standards, and sample size, strategy for gathering data and finally research analysis, study findings, limitations and recommendations for more research (Aromataris& Pearson, 2014).

4.3. Database Query and Literature Selection

The databases assessed for literature search include: Cochrane, Research gate, Medline, Web of science, Pubmed, Ebsco, Pubmed Central and Goggle scholar. The search strategy detailed in Chapter 3 was implemented during the database query process, resulting in the retrieval of the following articles:

6962 from PMC, 89 from PubMed, 15, 220 from ResearchGate, 17,200 from Google Scholar, and 0 articles from CINAHL. The retrieved articles then underwent a rigorous selection process involving multiple stages to ensure that only the most relevant studies were included in the review.

A preliminary process was conducted using automated tools and filters in each of the databases for the purpose of eliminating duplicates and sorting the articles that needed to meet the inclusion criteria at generic level. The process eliminated 23,096 articles, with 2,566 articles left. Afterwards, the articles left were screened for title and abstract conformity to the study objectives by reading through each article's title and abstract in order to ensure it is relevant to this review. This process produced 37 articles; eliminating 2529 articles with titles or abstracts did not meet the inclusion criteria.

Thorough text screening was then carried out on the remaining articles. At this stage, the full text of each article was read through in order to determine suitability for inclusion, after which the non-suitable ones were eliminated. At this stage, 29 articles were excluded due to payment restriction (i.e. a journal subscription amount of article fee was required before these articles could be accessed). Hence, they were removed from the selected articles; likewise articles with studies designs that do not conform to the inclusion criteria. Likewise, articles which did not fulfill any of the inclusion criteria regarding period of publication, study design and region of interest were excluded.

Eight articles, which meet the inclusion criteria for this review, were finally selected and included in this review. A graphical illustration of the screening and selection procedure is shown in the PRISMA flow diagram (Figure 4.1). The flow chart shows the selection process, as well as the retrieved number of articles, articles excluded at each stage and the number finally selected.

In addition, the details of extracted data are presented in Table 4.1, which represents a summary from the selected article.

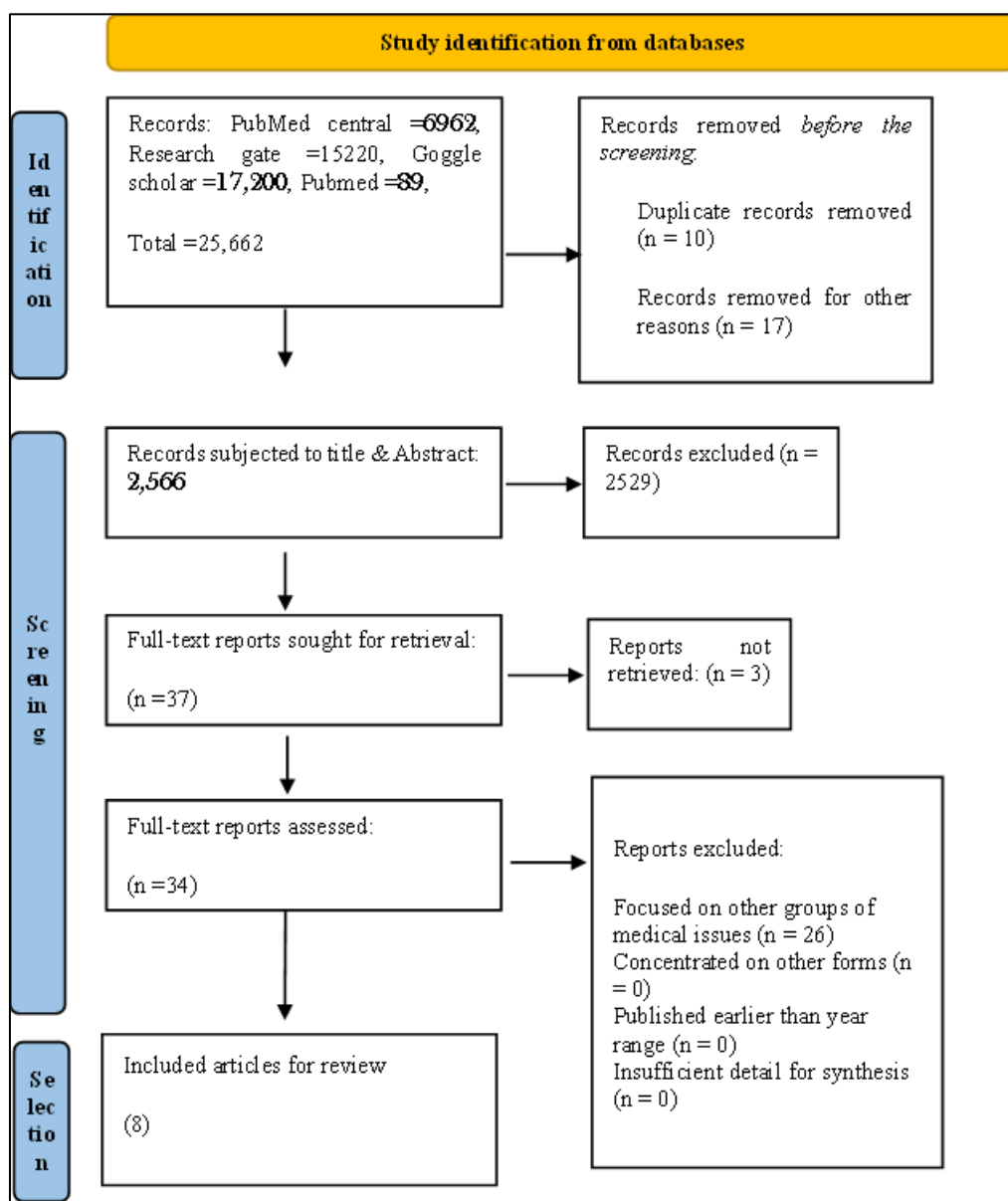


Figure 1 Prisma flow diagram diagram

Table 3 Summary of details from selected articles

S/n	Names	Title	Aim	Type of study	Country	Findings
1	Blease et al.	Generative artificial intelligence in primary care: an online survey of UK general practitioners	Measure GPs' use of generative AI in clinical practice	survey	UK	20% of general practitioners used generative AI, mainly for administrative tasks (29% for documentation, 28% for differential diagnosis). Concerns about errors, biases, and privacy risks.
2	David et al.	The use of artificial intelligence-based chat bots in ophthalmology triage	Evaluate AI chatbots' accuracy in answering ophthalmology patient questions	Comparative analysis	Israel	ChatGPT outperformed Bard in accuracy (median 4.0 vs. 3.0), comprehensiveness (4.5 vs. 3.0), and clarity (5.0 vs. 4.0). Both tools showed promise but had limitations in handling complex cases.
3	Jin et al.	Comparative study of Claude 3.5-Sonnet and human physicians in generating discharge summaries for patients with renal insufficiency	Assess Claude 3.5-Sonnet's performance in generating discharge summaries vs. human physicians	comparative study	China	Claude 3.5 gave results similar to human accuracy (90 vs. 92 points) but was significantly faster (30 seconds vs. 15 minutes).
4	Sabaner et al.	Opportunities and Challenges of Chatbots in Ophthalmology: A Narrative Review	Review chatbots' applications, accuracy, and challenges in ophthalmology	review	Several countries	Chatbots showed promise in education, triage, and patient queries but faced limitations in image analysis, ethical concerns, and variability across languages. ChatGPT generally outperformed other models
5	Chakraborty et al.	Overview of Chatbots with special emphasis on AI-enabled ChatGPT	To explore how AI chatbots, especially ChatGPT, are used in medical science	Review	India	ChatGPT has great potential in healthcare for patient support, education, and diagnostics. However, it also raises concerns about accuracy, privacy, and ethical issues.
6	Matulis & McCoy	Relief in Sight: Chatbots, In-baskets, and the Overwhelmed Clinician	To discuss how chatbots can help reduce primary care workload	Analysis	United States	Chatbots can help reduce workload for clinicians by managing simple tasks and patient communication. Proper

						design and integration are key to success.
7	Wang, Kaicheng	From ELIZA to ChatGPT: History of Chatbots	To review the history and development of chatbots, from early models to ChatGPT	Literature Review	China	Chatbots evolved from rule-based models like ELIZA to modern AI-based models like ChatGPT. Personalization and better conversation skills are key future goals.

The quality assessment carried out on these articles using the CASP appraisal tool produced results that clarifies that these articles identified the specific subjects of interest for this study, as well as discussed those related to the study objectives in details. Hence, the quality assessment of selected articles was deemed satisfactory. A quick glance at the observations from database query up till the selection and quality assessment of articles, it was evident in the selected articles discussed the application of AI chat bots in primary care settings, these articles adjusted for confounders before findings and conclusions were established.

For the purpose of understanding full view of the impact of AI chatbots in primary care triage, this review analyzed the role of AI chatbots based on the available published data.

5. Discussion

How effective are AI-powered chatbots in primary care triage compared to conventional triage methods in terms of accuracy, efficiency, and patient satisfaction?

The conventional primary care triage involves the primary care clinician assessing the patient for symptoms to determine what kind of care is needed or how urgently they require care. Over the years, researchers have sought ways to streamline this triage to improve efficiency and better patient outcome. Delegating some clinician function to an intelligent system like an AI Chatbot is one of the ways. Primary care can be described as preventive medicine. As defined by the WHO, primary care is supposed to offer accessible, equitable, comprehensive, continuous and people centered care to people at all places, but over the years, this role placed on primary care clinicians have become more burdensome. Not just because of an increasing number of people to attend to and mirage of new disease symptoms, but also because of shortage of clinicians. AI Chatbots come in handy to help ease these challenges in Primary care. [4]Being a technology developed using cutting-edge programming languages, including but not limited to C++, J, Python, Java, Lisp, PHP, Ruby, Clojure, or any other programming language with similar capabilities (Chakraborty et al., 2023), to imitate human intelligence, they are able to use machine learning and Natural language processing to assist in carrying out primary care functions such as responding to patients' questions and helping with record keeping, thus bringing efficiency to the primary care system.

AI Chatbots have been effective in attending to non-emergency cases, attending to patients questions and providing feedbacks especially on routine patient interactions. It has been utilized in several aspects of health to provide reliable information to patients and providing guidance towards receiving proper care where necessary. For instance, during the Covid -19 pandemic several chatbots were developed to attend to patients queries. They were used in diagnosis and monitoring of patients for example the COVIDBot, COMPASS, Otis and Vickybot. [4]In India, Chat bots have also been used to provide a private, non-judgmental, and safe space for young people to spur conversations about safe sex and family planning, and offer accurate, relatable, and trustworthy information (Chakraborty et al., 2023). In busy healthcare systems, AI chatbots could fill gaps by answering questions that arise at home that might otherwise go unaddressed. For ophthalmic cancer patients, who may forget key information due to attentional limitations, a common phenomenon in cancer diagnoses, chatbots trained in ophthalmic oncology could provide accessible, tailored support. [3]This would not only improve understanding, but also empower patients to become active participants in their care (Sabaner et al., 2024).

Despite the vast amount of data that these Chatbots have been equipped with, authors argue that compared to the clinicians they may be lacking in some important characteristics required for efficient treatment such as empathy and complex logical reasoning. [3]Chatbots perform well with straightforward cases, but they struggle with complex scenarios, frequently failing to provide comprehensive answers (Sabaner et al., 2024). [4]Modern-day Chatbots rely on

the use of NLU and Natural Language Generation (NLG) to recognize and respond to users by leveraging the ML and DL elements of AI, which provide responses based on user interactions, however, they may lack empathy, have a limited understanding of complex queries, and are limited by technical constraints (Chakraborty et al., 2023). AI Chatbots responses are generated based on the data they have been trained on and may lack the personalized context that a healthcare provider can offer. [2]Therefore, their use is best suited for providing general information and initial guidance, complementing the work of medical professionals (David et al., 2024). In addition, AI Chatbots reviewed in this literature have not shown to recognize non-verbal cues which the primary care clinician other hand can use to make important inferences about the patient's health status whether physical or mental. [6]Blease et al. (2024) also asserts that AI Chatbots are prone to creating erroneous information ('hallucination') and outputs of these models risk perpetuating or potentially worsening racial, gender and disability inequities in healthcare. While these concerns are valid, one could argue that some instances where AI Chatbots could deliver better service to patients. AI Chatbots could help remove biases that can stem from personal prejudices of the clinician. [2]There are also zero chances of AI Chatbots giving biased responses due to any emotional state or connection to patient's condition. Although blinded to the specific AI model, expert's evaluations are inherently biased and effected by their own clinical knowledge and experience (David et al., 2024). AI bias stems from the training data. Due to lack of electronic health records from many rural communities and developing countries, data from these areas are usually excluded from the training datasets so AI Chatbots may not give accurate information on patients from those demographics.

Performing administrative duties is another area where AI chatbots has proven to be very impactful in primary care. [4]Literatures reviewed highlight some administrative duties as setting appointments, gathering patient data, monitoring, and health promotion (Chakraborty et al., 2023), [6]generating patient report and summarizing patient data(Blease et al., 2024), [16]taking notes during a patient's appointment and generating complex medical documents such as surgical records, progress notes, and discharge summaries (Jin et al., 2024). Literature reviewed show that AI Chatbots are efficient in this areas, hence takes its workload off the physicians and allows them attend to more direct patient need.

[16]Jin et al. (2024) evaluated the performance of Claude 3.5-Sonnet in producing discharge summaries for patients with renal insufficiency, comparing its accuracy, efficiency, and quality to those generated by human physicians. These authors found that in terms of accuracy, there was no significant difference between the Chatbot and physician performance; however, the Chatbot provided reports within a significantly shorter period of time. The AI model takes an average of only about 30 s to complete a discharge summary, while human physicians require over 15 min on average. Although statistically similar, the little difference between the Chatbot's and physician's accuracy could have a significant impact in medicine, so while this Chatbot aids in streamlining triage by taking care of some administrative duty, it is important for physicians to supervise/verify its activities.

It is also worthy of note that these AI Chatbots models have different levels of accuracy, clarity, and efficiency and it is the duty of the clinician to conduct proper research and choose one suitable for their needs. [2]In a comparative study on AI chatbot use in ophthalmology, ChatGPT exhibited higher median ratings for Accuracy (4.0 vs. 3.0), Comprehensiveness (4.5 vs. 3.0), and Clarity (5.0 vs. 4.0) in the expert's evaluations compared to Bard showing variance in the models' capabilities to deliver accurate, comprehensive, and lucid responses to ophthalmology queries (David et al., 2024).

Considering the collection and documentation of electronic health records, AI chatbots have proven to be more efficient in maintaining patient records compared to its human counterpart. By interacting with the patients, it collects needed data which is very important in personalized medicine, enabling customized treatment strategies. Having a good patient record can also help in tracing precedence in terms of family history and identifying behavioral patterns. [4]This gives clinicians adequate information with which they can draw accurate inference about a patient's health. AI Chatbots provide healthcare professionals with data-driven insights(David et al., 2024).

Looking at reports presented in reviewed literature, AI Chatbots introduce a lot of ease and efficiency to the primary care triage, making primary care services more readily available, practical, and effective. Despite their immense contribution, when compared to human clinicians, they have been found wanting in some important aspects of primary care which includes demonstrating empathy, lack of consideration for non-verbal cues, and disparity that exist in electronic health records. However, in terms of time of execution, AI Chatbots have shown greater speed (approximately 30s to 30 minutes clinician time) in producing health documents – diagnosis, appointment summary, discharge letters and health history retrieval.

In the future, developers may consider building Chatbots that can recognize nonverbal cues, this could be have very beneficial impacts especially in aspects of healthcare such as therapy. There also needs to be well spelt out and stringent

policies on data privacy. These rules will provide patients with the assurance that their privacy will not be violated and help them trust Chatbots with their personal information, which will provide Chatbots with necessary data to learn from thereby improving reliability and accuracy. [3]Sabaner et al. (2024) suggests that offline chatbots are best suited to collect sensitive health information to avoid bridge of privacy: “..inputting protected health information into chatbots for medical advice risks severe and unpredictable consequences in the event of data breaches or hacking. However, chatbots that can be accessed locally and operate offline offer a potential solution to these privacy issues”

6. Conclusion

AI Chatbots have become an important part of our daily lives, utilized in several sectors to ease workload and improve efficiency. This technology has extended into the primary health sector and has shown to be tremendously beneficial to the system, however, Chatbots cannot replace the clinician neither can it work in solitary. To ensure that AI Chatbots are used accurately to improve efficiency, not only do clinicians and patients need to be trained on how to properly use it to generate useful information, it needs the supervision of health experts to ensure that it doesn't generate erroneous information. So, while the Chatbots help to produce results faster, the clinicians needs to verify them. AI Chatbots would be more beneficial if approached as a virtual assistant rather than a clinician.

Compliance with ethical standards

Disclosure of conflict of interest

The authors declare that there is no conflict of interest in the publication of this article

References

- [1] Matulis, J., McCoy, R., 2023. Relief in Sight? Chatbots, In-baskets, and the Overwhelmed Primary Care Clinician. *J. Gen. Intern. Med.* 38, 2808–2815. <https://doi.org/10.1007/s11606-023-08271-8>
- [2] David, D., Zloto, O., Katz, G., Huna-Baron, R., Vishnevskia-Dai, V., Armarnik, S., Zauberman, N., Barnir, E., Singer, R., Hostovsky, A., Klang, E., 2024. The use of artificial intelligence based Chatbots in ophthalmology triage. *Eye* 39. <https://doi.org/10.1038/s41433-024-03488-1>
- [3] Sabaner, M.C., Anguita, R., Antaki, F., Balas, M., Boberg-Ans, L.C., Ferro Desideri, L., Grauslund, J., Hansen, M.S., Klefter, O.N., Potapenko, I., Rasmussen, M.L.R., Subhi, Y., 2024. Opportunities and Challenges of Chatbots in Ophthalmology: A Narrative Review. *J. Pers. Med.* 14, 1165. <https://doi.org/10.3390/jpm14121165>
- [4] Chakraborty, C., Pal, S., Bhattacharya, M., Dash, S., Lee, S.-S., 2023. Overview of Chatbots with special emphasis on artificial intelligence-enabled ChatGPT in medical science. *Front. Artif. Intell.* 6, 1237704. <https://doi.org/10.3389/frai.2023.1237704>
- [5] Wang, K., 2024. From ELIZA to ChatGPT: A brief history of chatbots and their evolution. *Appl. Comput. Eng.* 39, 57–62. <https://doi.org/10.54254/2755-2721/39/20230579>
- [6] Blease, C.R., Locher, C., Gaab, J., Hägglund, M., Mandl, K.D., 2024. Generative artificial intelligence in primary care: an online survey of UK general practitioners. *BMJ Health Care Inform.* 31, e101102. <https://doi.org/10.1136/bmjhci-2024-101102>
- [7] Care, A.G.D. of H. and A., 2023. About primary care [WWW Document]. URL <https://www.health.gov.au/topics/primary-care/about> (accessed 4.3.25).
- [8] Page, M.J., McKenzie, J.E., Bossuyt, P.M. et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *Syst Rev* 10, 89 (2021). <https://doi.org/10.1186/s13643-021-01626-4>Lisy K. Narrative synthesis as a method for data extraction: Comprehensive support for review tools, strategies, and methods. *Res Synth Methods.* 2016;7(2):123-36.
- [9] Lisy K. Narrative synthesis as a method for data extraction: Comprehensive support for review tools, strategies, and methods. *Res Synth Methods.* 2016;7(2):123-36.
- [10] Critical Appraisal Skills Programme (CASP).(2023)CASP Checklists
- [11] J.G. Couto, S. McFadden, P. McClure, P. Bezzina, C. Hughes,Competencies of therapeutic radiographers working in the linear accelerator across Europe: A systematic search of the literature and thematic analysis, *Radiography*, Volume 26, Issue 1,2020, Pages 82-91, ISSN 1078-8174,<https://doi.org/10.1016/j.radi..2019.06.004>

- [12] Chen, L.-T. & Liu, L. (2017). Content analysis of statistical power in educational technology research: sample size matters. *International journal o technology in teaching and learning*, 15(1), 49-75
- [13] Pollock A, Berge E. How to do a systematic review. *International Journal of Stroke*. 2018;13(2):138-156. doi:10.1177/1747493017743796
- [14] Nishikawa-pacher, A. Research Questions with PICO: A Universal Mnemonic. *Publications* 2022, 10, 21. <https://doi.org/10.3390/publications10030021>
- [15] Aromataris, I., & pearson, B. (2014). Understanding the systematic review process: A structured overview. *JBI Library o systematic reviews*, 12 (7), 111-128
- [16] Jin, H., Guo, J., Lin, Q., Wu, S., Hu, W., Li, X., 2024. Comparative study of Claude 3.5-Sonnet and human physicians in generating discharge summaries for patients with renal insufficiency: assessment of efficiency, accuracy, and quality. *Front. Digit. Health* 6, 1456911. <https://doi.org/10.3389/fdgth.2024.1456911>