

## Review

# Integrating Retrieval-Augmented Generation with Large Language Models in Nephrology: Advancing Practical Applications

Jing Miao <sup>1</sup> , Charat Thongprayoon <sup>1</sup>, Supawadee Suppadungsuk <sup>1,2</sup> , Oscar A. Garcia Valencia <sup>1</sup>   
and Wisit Cheungpasitporn <sup>1,\*</sup> 

<sup>1</sup> Division of Nephrology and Hypertension, Department of Medicine, Mayo Clinic, Rochester, MN 55905, USA; miao.jing@mayo.edu (J.M.); charat.thongprayoon@gmail.com (C.T.); s.suppadungsuk@hotmail.com (S.S.); garciavalencia.oscar@mayo.edu (O.A.G.V.)

<sup>2</sup> Chakri Naruebodindra Medical Institute, Faculty of Medicine Ramathibodi Hospital, Mahidol University, Samut Prakan 10540, Thailand

\* Correspondence: wcheungpasitporn@gmail.com

**Abstract:** The integration of large language models (LLMs) into healthcare, particularly in nephrology, represents a significant advancement in applying advanced technology to patient care, medical research, and education. These advanced models have progressed from simple text processors to tools capable of deep language understanding, offering innovative ways to handle health-related data, thus improving medical practice efficiency and effectiveness. A significant challenge in medical applications of LLMs is their imperfect accuracy and/or tendency to produce hallucinations—outputs that are factually incorrect or irrelevant. This issue is particularly critical in healthcare, where precision is essential, as inaccuracies can undermine the reliability of these models in crucial decision-making processes. To overcome these challenges, various strategies have been developed. One such strategy is prompt engineering, like the chain-of-thought approach, which directs LLMs towards more accurate responses by breaking down the problem into intermediate steps or reasoning sequences. Another one is the retrieval-augmented generation (RAG) strategy, which helps address hallucinations by integrating external data, enhancing output accuracy and relevance. Hence, RAG is favored for tasks requiring up-to-date, comprehensive information, such as in clinical decision making or educational applications. In this article, we showcase the creation of a specialized ChatGPT model integrated with a RAG system, tailored to align with the KDIGO 2023 guidelines for chronic kidney disease. This example demonstrates its potential in providing specialized, accurate medical advice, marking a step towards more reliable and efficient nephrology practices.

**Keywords:** large language models (LLMs); nephrology; chronic kidney disease; artificial intelligence; retrieval-augmented generation (RAG)



**Citation:** Miao, J.; Thongprayoon, C.; Suppadungsuk, S.; Garcia Valencia, O.A.; Cheungpasitporn, W. Integrating Retrieval-Augmented Generation with Large Language Models in Nephrology: Advancing Practical Applications. *Medicina* **2024**, *60*, 445. <https://doi.org/10.3390/medicina60030445>

Academic Editor: Davide Bolignano

Received: 3 February 2024

Revised: 1 March 2024

Accepted: 6 March 2024

Published: 8 March 2024



**Copyright:** © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Large language models (LLMs) are a sophisticated type of artificial intelligence (AI), specifically designed for understanding and generating human-like language, thereby earning the alternative designation of chatbots. These models, like the generative pre-trained transformer (GPT) series, are trained on vast datasets of text and can generate text that is often indistinguishable from that written by humans. They can answer questions, write essays, generate creative content, and even write code, based on the patterns they have learned from the data they were trained on. The capabilities of LLMs continue to expand, making them a crucial and transformative technology in the field of AI [1]. ChatGPT, a prominent generative LLM developed by OpenAI, was released towards the close of 2022 [2]. The most recent version, GPT-4, is equipped with advanced capabilities in both text and image analysis and is collectively referred to as GPT-4 Vision [3]. Alongside ChatGPT, the current landscape of widely utilized LLMs encompasses Google's Bard AI [4] and Microsoft's Bing Chat [5]. These technological advancements have enabled

their adoption and implementation across various domains, including business, academic institutions, and healthcare fields [1].

Within the healthcare sector, the evolving influence of AI is reshaping traditional practices [6]. Tools like AI-driven LLMs possess a significant potential to enhance multiple aspects of healthcare, encompassing patient management, medical research, and educational methodologies [7–11]. For example, some investigators have shown that they can offer tailored medical guidance [12], distribute educational resources [7], and improve the quality of medical training [7,13,14]. These tools can also support clinical decision making [15–17], help identify urgent medical situations [18], and respond to patient inquiries with understanding and empathy [19–21]. Extensive research has shown that ChatGPT, particularly its most recent version GPT-4, excels across various standardized tests. This includes the United States Medical Licensing Examination [22–25]; medical licensing tests from different countries [26–30]; and exams related to specific fields such as psychiatry [31], nursing [32], dentistry [33], pathology [34], pharmacy [35], urology [36], gastroenterology [37], parasitology [38], and ophthalmology [39]. Additionally, there is evidence of ChatGPT's ability to create discharge summaries and operative reports [40,41], record patient histories of present illness [42], and enhance the documentation process for informed consent [43], although its effectiveness requires further improvement. Within the specific scope of our research in nephrology, we have explored the use of chatbots in various areas such as innovating personalized patient care, critical care in nephrology, and kidney transplant management [44], as well as dietary guidance for renal patients [45,46] and addressing nephrology-related questions [47].

Despite these advancements, LLMs face notable challenges. A primary concern is their tendency to generate hallucinations—outputs that are either factually incorrect or not relevant to the context [48,49]. For instance, the references or citations generated by these chatbots are unreliable [50–57]. Our analysis of 610 nephrology-related references showed that only 62% of ChatGPT's references existed. Meanwhile, 31% were completely fabricated, and 7% were partial or incomplete [58]. We also compared the relevance of ChatGPT, Bing Chat, and Bard AI in nephrology literature searches, with accuracy rates of only 38%, 30%, and 3%, respectively [59]. The occurrence of hallucinations during the literature searches, combined with the suboptimal accuracy in responding to nephrology inquiries [47] and correctly identifying oxalate, potassium, and phosphorus in diets [45,46], compromises the reliability or dependability of LLM outputs, raising significant concerns about their practical application. In critical areas like healthcare decision making, the impact of such inaccuracies is considerably heightened, highlighting the need for models that are more reliable and precise.

To address these challenges, various strategies have been developed. One such strategy is prompt engineering, like the multiple-shot or chain-of-thought prompting techniques [60–64]. This approach involves structuring the input prompt to encourage the model to break down the problem into intermediate steps or reasoning sequences before arriving at a final answer. By explicitly asking the model to generate a step-by-step explanation or “thought process”, chain-of-thought prompting helps the model tackle multistep reasoning problems more effectively, potentially leading to more accurate and interpretable answers [65–67]. Although this approach has proven beneficial in several contexts, it is not without its limitations. Concerns like scalability and the risk of embedding biases present significant challenges, necessitating meticulous prompt engineering to maintain the model's adaptability while safeguarding its efficiency.

Another strategy to enhance LLMs' ability is the retrieval-augmented generation (RAG) technique [68]. The primary advantage of the RAG approach is that it allows the LLM to access a vast external database of information, effectively extending its knowledge beyond what was available in its training data. This can significantly improve the model's performance, especially in generating responses that require specific factual information or up-to-date knowledge.

This review aims to explore the potential application of LLMs integrated with RAG in nephrology. This review also provides an analysis of the strengths and weaknesses of RAG.

\* 思維鏈

These observations are essential for appraising the potential of sophisticated AI models to drive notable advancements in healthcare sectors, where both precision and contemporary knowledge are of utmost importance, thus redefining the benchmarks for AI deployment in key domains.

## 2. What Is the RAG System?

The RAG approach is a method used in natural language processing and machine learning that combines the strengths of retrieval-based and generative models to improve the quality of generated text [68,69]. This approach is particularly useful in tasks such as question answering, document summarization, and conversational agents. In the dynamic field of medicine, the unique capability of the RAG system to access external medical databases in real time allows the LLM to base its responses on the latest research, clinical guidelines, and drug information [70,71].

To generate more accurate and contextually relevant responses, the RAG approach combines the strengths of two components including the retrieval and generation components. The former component is responsible for fetching relevant information or documents from a large database or knowledge source provided to the LLMs. The retrieval is typically based on the input query or context, aiming to find content that is most likely to contain the information needed to generate an accurate response. The latter component takes the input prompt along with the retrieved documents or information from the retrieval component and generates a response. The generation component uses the context provided by the retrieved documents to inform its responses, making them more accurate, informative, and contextually relevant.

The RAG approach is particularly beneficial in scenarios where the model needs to provide information that may not have been present in its training set or when the information is continually updated. By grounding the responses in factual data, the RAG approach effectively reduces the occurrence of inaccuracy or hallucinations. However, the success of RAG depends on the quality and timeliness of the external data sources, and integrating these sources introduces additional technical complexities. Complementing these approaches is the process of fine-tuning, which involves adapting a pre-trained model to specific tasks or domains. This enhances the model's capacity to process certain types of queries or content, thereby improving its efficiency and specificity for certain domains. While this method improves the model's performance in specific areas, it also poses the risk of over-fitting in certain datasets, potentially limiting its broader applicability and increasing the demands on training resources.

## 3. Current Research Regarding the Application of RAG in Medical Domain

A recent study experimentally developed a liver disease-focused LLM model named **LiVersa**, incorporating the RAG approach with 30 guidelines from the American Association for the Study of Liver Diseases. This integration was intended to enhance LiVersa's functionality. In the study, **LiVersa accurately answered all 10 questions related to hepatitis B virus treatment and hepatocellular carcinoma surveillance**. However, the explanations provided for three of these cases **were not entirely accurate** [72]. Another study introduced **Almanac**, an LLM framework enhanced with RAG functions, which was specifically integrated with medical guidelines and treatment recommendations [73]. This framework's effectiveness **was evaluated using a new dataset comprising 130 clinical scenarios**. In terms of accuracy, Almanac outperformed ChatGPT by an average of 18% across various medical specialties. The most notable improvement was seen in cardiology, where Almanac **achieved 91% accuracy compared to ChatGPT's 69%** [73]. Moreover, they evaluated the performance of Almanac against conventional LLMs (ChatGPT-4 [May 24, 2023 version], BingChat [June 28, 2023], and Bard AI [June 28, 2023]) by testing the LLMs with a new dataset comprising 314 clinical questions across nine medical specialties. Almanac demonstrated **notable enhancements in accuracy, comprehensiveness, user satisfaction, and resilience to adversarial inputs when compared to the standard LLMs** [74]. A recent investigation introduced a **RAG system** named **RECTIFIER** (RAG-Enabled Clinical Trial Infrastructure for Inclusion Exclusion Review), assessing its efficacy against that of expert

clinicians in a clinical trial screening [75]. The comparison revealed a high concordance between the responses from RECTIFIER and those from expert clinicians, with RECTIFIER's accuracy spanning from 98% to 100% and the study staff's accuracy from 92% to 100%. Notably, RECTIFIER outperformed the study staff in identifying the inclusion criterion of "symptomatic heart failure", achieving an accuracy of 98% compared to 92%. In terms of eligibility determination, RECTIFIER exhibited a sensitivity of 92% and a specificity of 94%, whereas the study staff recorded a sensitivity of 90% and a specificity of 84%. These findings indicate that integrating a RAG system into GPT-4-based solutions could significantly enhance the efficiency and cost effectiveness of clinical trial screenings [75].

The RAG's strengths lie in its access to current information and its ability to tailor relevance. By utilizing the most recent data, the likelihood of offering outdated or incorrect information is greatly reduced. However, this approach also presents several challenges. The effectiveness of RAG's responses is heavily dependent on the quality and currency of the data sources it uses. Adding RAG to LLMs also introduces an extra layer of complexity, which can complicate implementation and ongoing management. Moreover, there is a risk of retrieval errors. Should the retrieval system malfunction or fetch incorrect information, it could result in inaccuracies in the output it generates.

#### 4. The Potential Applications of RAG in Nephrology

The RAG integration is also valuable in nephrology, where staying abreast of the latest developments is crucial. This integration of current, validated data from external sources significantly reduces the likelihood of the LLMs providing outdated or incorrect information.

##### 4.1. Integrating Latest Research and Guidelines

The RAG approach has the unique capability to dynamically integrate the most recent findings from nephrology-related sources into the model's outputs. This includes new research from nephrology journals, results from the latest clinical trials, or any updates in treatment guidelines. By doing so, the RAG approach ensures that LLMs are not only up-to-date but also highly relevant and accurate in the field of nephrology. For instance, consider a scenario where a nephrology specialist or an internist is seeking information about the latest management strategies for polycystic kidney disease (PKD). In such cases, the RAG can actively search for, retrieve, and incorporate information from the most recent guidelines and treatment protocols, such as the KDIGO 2023 clinical practice guideline for autosomal dominant polycystic kidney disease (ADPKD), and studies published in the PubMed database. This process involves not just accessing this information but also synthesizing it in a way that is coherent and directly applicable to the query at hand.

By utilizing RAG, the physician is thus provided with information that is not only current but is also directly relevant to their specific inquiry. This approach is especially valuable in a field like nephrology, where advancements in research and changes in treatment protocols can have a significant impact on patient care. The ability of RAG to provide the latest knowledge helps healthcare professionals stay informed and make well-founded decisions in their practice.

##### 4.2. Case-Based Learning and Discussion

Employing RAG in educational settings can significantly enhance the learning process by incorporating detailed and real-life case studies into lectures, discussions, or interactive learning modules. This application of RAG is particularly useful in complex and dynamic fields like medicine. Take, for example, the education of medical students on the topic of complex electrolyte imbalances in chronic kidney disease (CKD). The RAG approach can be utilized to access and reference specific, real-world case reports or clinical scenarios relevant to this topic. By doing so, it can provide students with practical, tangible examples that illustrate the theoretical concepts they are learning. This not only aids in a deeper understanding of the subject matter but also helps students appreciate the real-world implications and applications of their knowledge.



Moreover, RAG's ability to retrieve the latest studies and reports ensures that the educational content is not only rich in practical examples but also current. This is especially vital in medical education, where staying abreast of the latest research and clinical practices is crucial. By integrating up-to-date case studies and scenarios, RAG can help create a more engaging and informative educational experience, preparing students for the challenges they will face in their medical careers. This approach can be extended to other complex medical topics, making learning more interactive, relevant, and evidence-based.

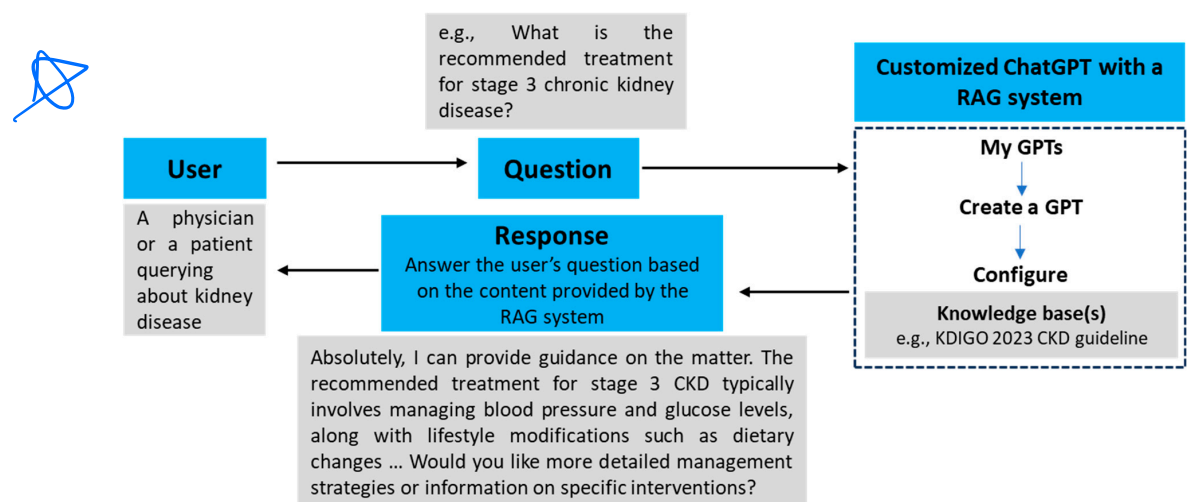
#### 4.3. Multidisciplinary Approach

In situations where a multidisciplinary perspective is essential, RAG proves to be particularly valuable as it can draw upon a wide array of medical disciplines to offer a more comprehensive understanding. This capability is critical in treating conditions that intersect multiple areas of healthcare. Consider the case of a patient suffering from diabetic nephropathy, for instance. This condition, being at the crossroads of diabetes and kidney health, requires a nuanced understanding from several medical specialties. The RAG system can effectively consolidate relevant information from endocrinology, focusing on diabetes management strategies; from cardiology, addressing the cardiovascular risks associated with the condition; and from nephrology, providing insights into preserving renal function.

By integrating this diverse information, the RAG system can greatly assist healthcare professionals in developing a holistic and multifaceted treatment plan. This approach ensures that all aspects of the patient's condition are considered, leading to more effective and comprehensive patient care. Such an integrated approach is beneficial not just in diabetic nephropathy but in any complex medical condition where multiple body systems are affected or where various specialties need to collaborate for optimal patient management. The ability of RAG to seamlessly merge insights from different medical fields into a cohesive whole enhances its utility in planning and implementing effective treatment strategies.

### 5. Creation of a CKD-Specific Knowledge Base for RAG

To illustrate the process of creating a customized ChatGPT model with a RAG strategy, we will use the field of nephrology as a reference, specifically focusing on CKD due to its prevalence in nephrology encounters (Figure 1). This example will serve to demonstrate the steps and considerations involved in tailoring a ChatGPT model to a specific medical specialty, incorporating a specialized knowledge base. The aim is to enhance the model's responses with precise, specialized knowledge, in this case, centered around CKD, guided by insights from the KDIGO 2023 Clinical Practice Guideline [76]. Below is a detailed breakdown of the steps involved in this process.



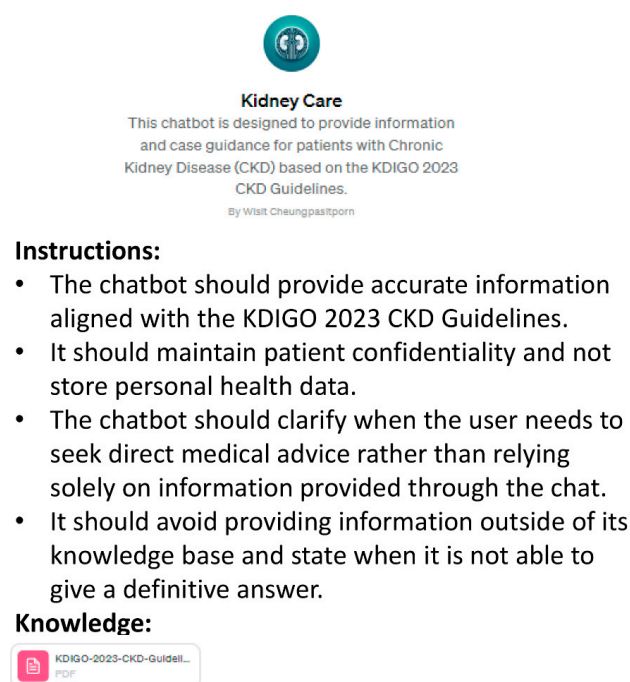
**Figure 1.** The process of creating a customized ChatGPT model with the retrieval-augmented generation (RAG) strategy in nephrology.

### 5.1. Creation of a CKD-Focused Retrieval System

This process involves the careful selection of knowledge sources, integration of guidelines, and regular updates to ensure accuracy and relevancy. The first step is to meticulously select a comprehensive database rich in information about CKD. This database should draw from a range of reliable sources, such as peer-reviewed academic journals, reports from clinical trials, and authoritative nephrology textbooks. A key focus is placed on incorporating the KDIGO 2023 CKD guidelines [76], which are recognized for their currency and authority in the field.

Next, it is vital to directly integrate these KDIGO 2023 guidelines into the chosen database by creating a customized ChatGPT model (Figure 2). This process involves navigating to “My GPTs” and selecting “Create a GPT”. Following this, we have the opportunity to customize/configure our GPT by entering a name, description, and instructions, and by uploading the knowledge bases(s) we wish to embed within the model. We can choose to restrict access to the model by selecting one of the following options: “Only me”, “Anyone with a link”, or “Everyone”. Once customized, the GPT will be accessible under “My GPTs”, where it will produce responses utilizing the incorporated database(s).

This integration covers the detailed aspects of CKD, including diagnosis, staging, management, and treatment protocols. Such incorporation ensures that the model’s responses are in line with the most recent and accepted clinical practices. While ChatGPT operates based on its internal knowledge gained during training, RAG takes this a step further by dynamically incorporating external information into the generation process. The integration of a retrieval component in RAG could theoretically enhance ChatGPT by providing it access to a wider range of current information and specific data not covered during its training.



**Figure 2.** The creation of a CKD-specific knowledge base by customizing GPT-4 with the retrieval-augmented generation (RAG) approach.

### 5.2. Development of a CKD-Focused Retrieval System

The RAG system, specialized for CKD, is specifically configured to identify and respond to CKD-related queries accurately. It is adept at grasping the intricacies of CKD, including its various stages, the comorbid conditions often accompanying it, and the diverse methods of treatment available. Additionally, the system is fine-tuned for both speed and relevance, ensuring rapid and efficient access to relevant information from the

comprehensive CKD database when processing queries. This optimization guarantees prompt and pertinent responses tailored to the specifics of CKD. Moreover, establishing a system for continuous updates to the database is crucial. This involves regularly reviewing and including new research findings, updated medical guidelines, and emerging treatment methods in nephrology. Keeping the database up to date guarantees that the information remains both current and authoritative, making it a reliable foundation for the model's knowledge base.

### *5.3. Integration with the Customized GPT-4 Model*

Integrating the customized GPT-4 model with the CKD retrieval system involves establishing strong and secure API (Application Programming Interface) connections. Firstly, it focuses on creating a robust connection that allows for the seamless flow of data between the customized ChatGPT model and the CKD retrieval system. This connection must be secure to protect sensitive medical information and ensure data integrity. Secondly, the customized ChatGPT model undergoes fine-tuning to harmonize the in-depth CKD information with its innate natural language processing abilities. This fine-tuning is critical to ensure that the model not only provides responses that are accurate and rich in CKD-specific information but also maintains clarity and appropriateness in the context of the user's query.

Through this integration, the model becomes capable of delivering responses that are not just factually correct but also tailored to the specific context of the query, whether it is a patient's inquiry, a healthcare professional's detailed question, or an educational scenario. This ensures that the model's outputs are highly relevant, understandable, and useful for various users, ranging from medical practitioners and students to patients seeking information about CKD.

### *5.4. Customized Response for CKD Inquiries*

The integration of a customized GPT-4 model with a CKD-specialized RAG system brings a significant advancement in handling CKD-related inquiries. This integration leverages sophisticated algorithms to ensure that the ChatGPT model precisely recognizes the context and specific details of queries related to CKD, leading to highly relevant and tailored responses. This process operates on multiple levels, including contextual understanding, relevance of responses, access to updated information, and dynamic information integration.

Through this integrated approach, the ChatGPT model becomes a powerful tool for providing accurate, up-to-date, and highly specific responses to a wide range of CKD-related inquiries. This capability is particularly valuable for healthcare professionals seeking quick and reliable information, patients looking for understandable explanations of their condition, and researchers needing the latest data in the field of nephrology.

### *5.5. Rigorous Testing with CKD Scenarios*

The system undergoes comprehensive testing in a variety of CKD situations. This testing encompasses a spectrum of patient histories, various stages of CKD, and the intricacies involved in treatment plans. Such extensive testing is crucial for confirming the model's reproductivity and its ability to adapt to diverse clinical conditions. The feedback obtained from these rigorous tests is instrumental to the ongoing enhancement of the system. It aids in refining the precision of information retrieval and boosting the effectiveness of how the ChatGPT model works in conjunction with the CKD database. This process of continuous improvement ensures the system remains reliable and effective in addressing the complex needs of CKD management.

### *5.6. Regular System Monitoring and Updating*

The system's performance in providing accurate and relevant CKD information is consistently monitored. This includes assessing the accuracy of responses, the relevance of

information provided, and the speed of retrieval. Moreover, the CKD database is regularly updated with the latest research, guidelines, and treatment protocols, ensuring the model's responses remain current and authoritative.

#### 5.7. Healthcare Professional Engagement and Feedback

Healthcare professionals are trained on how to effectively use the customized ChatGPT model for CKD queries. This includes understanding its capabilities, limitations, and the best ways to phrase queries for optimal results. A feedback loop is established to continuously improve the system based on real-world user experiences and suggestions from healthcare professionals.

### 6. Examples of Responses Generated by GPT-4 with and without RAG System

The effectiveness of the responses generated by GPT-4, both with and without the RAG approach, is evaluated using a straightforward query: “List medication treatment to help slow progression of CKD and end-stage kidney disease (ESKD)”. This test aims to compare the quality and accuracy of the information provided by GPT-4 under both methodologies (Figures 3 and 4).

When using the general GPT-4 to address treatment approaches for slowing the progression of CKD to ESKD, the responses tend to offer a broad overview, lacking in-depth adherence to the latest KDIGO guidelines. However, the customized GPT-4 model enhanced with a RAG system provides responses that are more specific, detailed, and nuanced. Upon verification, these responses are found to be in close alignment with the KDIGO 2023 CKD guidelines, accurately reflecting the current research and clinical practices within nephrology. ChatGPT's recommendations included SGLT-2 inhibitor and GLP-1 receptor agonists for patients with CKD and type 2 diabetes. However, ChatGPT failed to mention some targeted pharmaceutical interventions that may offer a way to slow CKD progression in individuals with specific causes, such as tolvaptan for ADPKD patients. To enhance its precision, it is necessary to incorporate additional resources, such as the ADPKD guidelines, into its reference database. This will enable ChatGPT to access a broader array of documents, facilitating the generation of more precise advice for CKD patients with specific conditions.

Significantly, utilizing a series of prompts or exploring varied prompting techniques in standard ChatGPT, such as the chain-of-thought method and determining a specific CKD guideline for use, could also lead to more consistent responses with the RAG system. This review seeks to present an alternative strategy, the RAG system, for enhancing the effectiveness of LLMs and their applications, including in the context of CKD, to illustrate its utility. This method proves to be advantageous, efficient, and expedient when responses require dependence on specific or particular documents. Therefore, creating a customized ChatGPT model specifically for nephrology, with a focus on CKD and based on the KDIGO 2023 CKD guidelines, is an extensive and meticulous process. It involves building a specialized knowledge base, developing a dedicated retrieval system tailored to nephrology, and integrating this with the ChatGPT model. The process also includes fine-tuning the model to generate precise responses, conducting thorough testing to ensure reliability, continuously updating the system with the latest information, and maintaining engagement with healthcare professionals for feedback and validation. This development results in a model that stands out in offering specialized and accurate medical guidance for managing CKD. As such, it becomes an invaluable resource for healthcare providers, enhancing their ability to deliver informed and up-to-date care to patients with CKD. Notably, the ChatGPT model has merely presented an instance to illustrate the utility of the RAG approach. Additional research is required to confirm its dependability and enhance its efficacy in nephrology applications.





**You**

List medication treatment to help slow progression of CKD to ESKD



**ChatGPT**

1. **Angiotensin-Converting Enzyme (ACE) Inhibitors and Angiotensin I Receptor Blockers (ARBs):** Lower blood pressure and reduce proteinuria to decrease kidney workload. Examples include lisinopril and enalapril (ACE inhibitors), losartan, and valsartan (ARBs).
2. **Diuretics:** Help manage fluid overload, particularly in patients with heart failure and high blood pressure, aiding kidney function.
3. **Statins:** Cholesterol-lowering medications. Important for managing cardiovascular risk factors in CKD patients.
4. **Erythropoiesis-Stimulating Agents (ESAs):** Used in cases of anemia associated with CKD to stimulate red blood cell production.
5. **Phosphate Binders:** Reduce phosphate absorption from the diet to manage high levels due to reduced kidney function.
6. **Vitamin D Supplements and Calcimimetics:** Help manage mineral and bone disorders in CKD.
7. **Antihypertensive Medications:** Essential for controlling blood pressure, may include calcium channel blockers or beta-blockers.
8. **Blood Glucose Control Medications:** For diabetic CKD patients, maintaining optimal blood sugar levels is key, which might involve insulin or oral hypoglycemic agents.
9. **Dietary Modifications:** Limiting protein, potassium, sodium, and phosphorus intake can help reduce kidney workload.
10. **Smoking Cessation Aids:** Crucial for reducing CKD progression, as smoking is a significant risk factor.
11. **Newer Therapies:** Medications like sodium-glucose cotransporter-2 (SGLT2) inhibitors are emerging as beneficial in slowing CKD progression, especially in diabetic patients.

Treatment regimens should be tailored to individual needs, considering health, CKD stage, and comorbid conditions. Regular monitoring and treatment adjustments, along with non-pharmacological interventions like lifestyle changes, are important. Collaboration with a healthcare team, including nephrologists and dietitians, is essential for optimal care.

**Figure 3.** Responses of general GPT-4 to the question regarding medication treatment to help slow the progression of CKD to ESKD.



**You**

List medication treatment to help slow progression of CKD to ESKD



### **Kidney Care**

Based on the KDIGO 2023 CKD Guidelines, several medications are recommended to help slow the progression of chronic kidney disease (CKD) to end-stage kidney disease (ESKD). Here are some key recommendations:

1. **Sodium-Glucose Cotransporter-2 Inhibitors (SGLT2i):** These medications are suggested for use in adults and children with CKD but not diabetes. Examples include canagliflozin, dapagliflozin, empagliflozin, and others. They are compared against placebo, usual care, or an active comparator in terms of their effects on the progression of CKD, complications of CKD, and adverse events.
2. **Nonsteroidal Mineralocorticoid Receptor Antagonists (MRAs):** These are recommended for adults with Type 2 Diabetes (T2D), an eGFR  $>25$  ml/min per  $1.73$  m<sup>2</sup>, normal serum potassium concentration, and albuminuria despite the maximum tolerated dose of RAS inhibitor (RASi). They can be added to a RASi and an SGLT2i for the treatment of T2D and CKD in adults.
3. **Glucagon-Like Peptide-1 Receptor Agonists (GLP-1 RA):** For adults and children with CKD and T2D, GLP-1 RAs like albiglutide, dulaglutide, and liraglutide may be used. These are compared against other glucose-lowering agents, placebo, or usual care, specifically looking at their impact on kidney failure, all-cause hospitalizations, and other critical outcomes.
4. **Lipid Management:** In adults aged  $\geq 50$  years with an eGFR  $<60$  ml/min per  $1.73$  m<sup>2</sup> but not treated with chronic dialysis or kidney transplantation, treatment with a statin or statin/ezetimibe combination is recommended.

These recommendations are aimed at managing the progression of CKD and its complications, and the choice of specific medication should be made based on individual patient factors and in consultation with a healthcare provider. It's also important to note that these medications are part of a broader management strategy that includes lifestyle changes and management of other health conditions.

If you have CKD or are at risk, it's essential to consult with a healthcare professional for personalized advice and treatment.

**Figure 4.** Responses of the customized GPT-4 with the retrieval-augmented generation (RAG) system to the question regarding medication treatment to help slow the progression of CKD to ESKD.

## 7. Future Studies in the Context of LLMs with RAG Systems in Nephrology

Future studies in the context of LLMs with RAG systems in nephrology are suggested to address several promising avenues. These could significantly enhance both the depth and breadth of nephrology research, clinical decision support, patient education, and personalized medicine.

Prospective studies would likely involve deploying RAG-enhanced LLMs in clinical settings as decision-support tools. Their effectiveness in assisting with real-time patient care decisions could be evaluated against traditional decision-making processes. Key metrics could include improvements in treatment time efficiency, accuracy in diagnosis, and patient satisfaction levels. Research could explore the seamless integration of LLMs with RAG systems into electronic health record (EHR) platforms, which is essential for enabling real-time, context-aware decision support for clinicians treating patients with kidney diseases. For instance, by leveraging the latest research findings, current guidelines, and patient-specific data, these models could assist in identifying subtle patterns or rare conditions that are difficult for humans to discern, thus improving the diagnostic accuracy for complex kidney diseases and tailoring treatment plans for individual patients with kidney diseases such as CKD or AKI. Future research might also explore automating the process of conducting systematic reviews and meta-analyses using LLMs with RAG systems. This could significantly speed up the synthesis of new research findings, ensuring that the nephrology practice remains at the cutting edge.

Moreover, the integration of nephrology-focused RAG systems with other medical domains could provide a more comprehensive patient care model. For instance, combining nephrology with cardiovascular data might better predict renal patients' risk of heart disease. Studies could examine the outcomes of such integrations in improving the management of comorbid conditions. Combining insights from genomics, proteomics, and other omics technologies with LLMs and RAG systems also could lead to a more comprehensive understanding of kidney diseases and breakthroughs in precision medicine and novel therapeutic targets.

The development of adaptive learning modules using RAG-enhanced LLMs could offer personalized educational pathways for medical professionals. These modules could use real-time data to simulate patient scenarios, adapting to the learner's responses and providing immediate feedback grounded in the latest clinical guidelines. To mitigate the risk of misinformation, future research might develop advanced fact-checking algorithms tailored to medical data nuances. These algorithms could cross-reference multiple authoritative databases before generating patient advice, ensuring a higher degree of accuracy in the information provided. The LLMs with a RAG system can also be utilized to provide personalized, easy-to-understand educational materials and support for patients with kidney diseases.

Furthermore, studies may explore the establishment of international consortia for the standardization of AI applications in nephrology. These networks could facilitate the sharing of best practices, the creation of diverse and comprehensive datasets, and the development of AI models that are generalizable across different populations and healthcare systems. This includes customizing models to account for genetic, environmental, and socioeconomic factors affecting kidney disease prevalence and treatment outcomes across different populations.

As LLMs with RAG systems rely on extensive data, future studies must address ethical and privacy concerns, ensuring patient data are used responsibly and securely. Therefore, research into the ethical implications of AI in nephrology will need to address consent processes for patient data, biases in AI training, and the transparency of AI decision-making processes. Regulatory studies might focus on developing frameworks for AI accountability and compliance with healthcare regulations like the Health Insurance Portability and Accountability Act (HIPAA).

## 8. Conclusions

Combining LLMs with RAG systems in nephrology is a big step forward. It has the potential to change how we care for and educate patients in this specialized area. However, one of the main challenges is making sure the information they provide is accurate and reliable. To make these models better for their use in nephrology, strategies like using detailed prompting techniques, carefully applying RAG, and fine-tuning the models are important. As we move into this new phase, it is essential to have teams that include AI experts, kidney specialists, and ethicists. The goal is to improve AI so that it not only matches the skills of healthcare professionals but also adds to them. Achieving this is complex and very important. It requires a constant commitment to accuracy, innovation, and ethical practice. Through ongoing research, improvement, and a focus on patient welfare, we are getting closer to a future where AI plays a transformative role in healthcare, leading to better patient outcomes and more effective, knowledgeable healthcare systems.

**Author Contributions:** Conceptualization, J.M., C.T., S.S., O.A.G.V. and W.C.; methodology, J.M., C.T. and W.C.; validation, J.M., C.T. and W.C.; investigation, J.M., C.T. and W.C.; resources, J.M., C.T. and W.C.; data curation, J.M., C.T. and W.C.; writing—original draft preparation, J.M., C.T., S.S., O.A.G.V. and W.C.; writing—review and editing, J.M., C.T., S.S., O.A.G.V. and W.C.; visualization, J.M. and W.C.; supervision, C.T. and W.C.; project administration, C.T. and W.C.; All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Availability Statements are available in the original publication, reports, and preprints that were cited in the reference citation.

**Conflicts of Interest:** The authors declare no conflict of interest.

## Abbreviations

ADPKD: autosomal dominant polycystic kidney disease; AI: artificial intelligence; CKD: chronic kidney disease; EHRs: electronic health records; ESKD: end-stage kidney disease; GPT: generative pre-trained transformer; HIPAA: Health Insurance Portability and Accountability Act; LLMs: large language models; PKD: polycystic kidney disease; RAG: retrieval-augmented generation.

## References

1. Michael Kerner, S. Large Language Models (LLMs). Available online: [https://www.techtarget.com/whatis/definition/large-language-model-LLM#:~:text=A%20large%20language%20model%20\(LLM,generate%20and%20predict%20new%20content](https://www.techtarget.com/whatis/definition/large-language-model-LLM#:~:text=A%20large%20language%20model%20(LLM,generate%20and%20predict%20new%20content) (accessed on 13 September 2023).
2. Introducing ChatGPT. Available online: <https://openai.com/blog/chatgpt> (accessed on 30 November 2022).
3. OpenAI. GPT-4V(ision) System Card. Available online: [https://cdn.openai.com/papers/GPTV\\_System\\_Card.pdf](https://cdn.openai.com/papers/GPTV_System_Card.pdf) (accessed on 25 September 2023).
4. Bard. Available online: <https://bard.google.com/chat> (accessed on 21 March 2023).
5. Bing Chat with GPT-4. Available online: <https://www.microsoft.com/en-us/bing?form=MA13FV> (accessed on 14 October 2023).
6. Majnaric, L.T.; Babic, F.; O'Sullivan, S.; Holzinger, A. AI and Big Data in Healthcare: Towards a More Comprehensive Research Framework for Multimorbidity. *J. Clin. Med.* **2021**, *10*, 766. [CrossRef] [PubMed]
7. Sallam, M. ChatGPT Utility in Healthcare Education, Research, and Practice: Systematic Review on the Promising Perspectives and Valid Concerns. *Healthcare* **2023**, *11*, 887. [CrossRef] [PubMed]
8. Ruksakulpiwat, S.; Kumar, A.; Ajibade, A. Using ChatGPT in Medical Research: Current Status and Future Directions. *J. Multidiscip. Healthc.* **2023**, *16*, 1513–1520. [CrossRef] [PubMed]
9. Jamal, A.; Solaiman, M.; Alhasan, K.; Temsah, M.H.; Sayed, G. Integrating ChatGPT in Medical Education: Adapting Curricula to Cultivate Competent Physicians for the AI Era. *Cureus* **2023**, *15*, e43036. [CrossRef] [PubMed]
10. van Dis, E.A.M.; Bollen, J.; Zuidema, W.; van Rooij, R.; Bockting, C.L. ChatGPT: Five priorities for research. *Nature* **2023**, *614*, 224–226. [CrossRef]



11. Yu, P.; Xu, H.; Hu, X.; Deng, C. Leveraging Generative AI and Large Language Models: A Comprehensive Roadmap for Healthcare Integration. *Healthcare* **2023**, *11*, 2776. [\[CrossRef\]](#) [\[PubMed\]](#)
12. Joshi, G.; Jain, A.; Araveeti, S.R.; Adhikari, S.; Garg, H.; Bhandari, M. FDA Approved Artificial Intelligence and Machine Learning (AI/ML)-Enabled Medical Devices: An Updated Landscape. *Electronics* **2024**, *13*, 498. [\[CrossRef\]](#)
13. Oh, N.; Choi, G.S.; Lee, W.Y. ChatGPT goes to the operating room: Evaluating GPT-4 performance and its potential in surgical education and training in the era of large language models. *Ann. Surg. Treat. Res.* **2023**, *104*, 269–273. [\[CrossRef\]](#)
14. Eysenbach, G. The Role of ChatGPT, Generative Language Models, and Artificial Intelligence in Medical Education: A Conversation with ChatGPT and a Call for Papers. *JMIR Med. Educ.* **2023**, *9*, e46885. [\[CrossRef\]](#)
15. Reese, J.T.; Danis, D.; Caulfield, J.H.; Casiraghi, E.; Valentini, G.; Mungall, C.J.; Robinson, P.N. On the limitations of large language models in clinical diagnosis. *medRxiv* **2023**. [\[CrossRef\]](#)
16. Eriksen, A.V.; Möller, S.; Ryg, J. Use of GPT-4 to Diagnose Complex Clinical Cases. *NEJM AI* **2023**, *1*, A1p2300031. [\[CrossRef\]](#)
17. Kanjee, Z.; Crowe, B.; Rodman, A. Accuracy of a Generative Artificial Intelligence Model in a Complex Diagnostic Challenge. *JAMA* **2023**, *330*, 78–80. [\[CrossRef\]](#) [\[PubMed\]](#)
18. Zuniga Salazar, G.; Zuniga, D.; Vindel, C.L.; Yoong, A.M.; Hincapie, S.; Zuniga, A.B.; Zuniga, P.; Salazar, E.; Zuniga, B. Efficacy of AI Chats to Determine an Emergency: A Comparison between OpenAI's ChatGPT, Google Bard, and Microsoft Bing AI Chat. *Cureus* **2023**, *15*, e45473. [\[CrossRef\]](#)
19. Ayers, J.W.; Poliak, A.; Dredze, M.; Leas, E.C.; Zhu, Z.; Kelley, J.B.; Faix, D.J.; Goodman, A.M.; Longhurst, C.A.; Hogarth, M.; et al. Comparing Physician and Artificial Intelligence Chatbot Responses to Patient Questions Posted to a Public Social Media Forum. *JAMA Intern. Med.* **2023**, *183*, 589–596. [\[CrossRef\]](#) [\[PubMed\]](#)
20. Lee, P.; Bubeck, S.; Petro, J. Benefits, Limits, and Risks of GPT-4 as an AI Chatbot for Medicine. *N. Engl. J. Med.* **2023**, *388*, 1233–1239. [\[CrossRef\]](#)
21. Mello, M.M.; Guha, N. ChatGPT and Physicians' Malpractice Risk. *JAMA Health Forum* **2023**, *4*, e231938. [\[CrossRef\]](#) [\[PubMed\]](#)
22. Mihalache, A.; Huang, R.S.; Popovic, M.M.; Muni, R.H. ChatGPT-4: An assessment of an upgraded artificial intelligence chatbot in the United States Medical Licensing Examination. *Med. Teach.* **2024**, *46*, 366–372. [\[CrossRef\]](#)
23. Mbakwe, A.B.; Lourentzou, I.; Celi, L.A.; Mechanic, O.J.; Dagan, A. ChatGPT passing USMLE shines a spotlight on the flaws of medical education. *PLoS Digit. Health* **2023**, *2*, e0000205. [\[CrossRef\]](#)
24. Gilson, A.; Safranek, C.W.; Huang, T.; Socrates, V.; Chi, L.; Taylor, R.A.; Chartash, D. How Does ChatGPT Perform on the United States Medical Licensing Examination? The Implications of Large Language Models for Medical Education and Knowledge Assessment. *JMIR Med. Educ.* **2023**, *9*, e45312. [\[CrossRef\]](#)
25. Kung, T.H.; Cheatham, M.; Medenilla, A.; Sillos, C.; De Leon, L.; Elepano, C.; Madriaga, M.; Aggabao, R.; Diaz-Candido, G.; Maningo, J.; et al. Performance of ChatGPT on USMLE: Potential for AI-assisted medical education using large language models. *PLoS Digit. Health* **2023**, *2*, e0000198. [\[CrossRef\]](#)
26. Meyer, A.; Riese, J.; Streichert, T. Comparison of the Performance of GPT-3.5 and GPT-4 with That of Medical Students on the Written German Medical Licensing Examination: Observational Study. *JMIR Med. Educ.* **2024**, *10*, e50965. [\[CrossRef\]](#) [\[PubMed\]](#)
27. Tanaka, Y.; Nakata, T.; Aiga, K.; Etani, T.; Muramatsu, R.; Katagiri, S.; Kawai, H.; Higashino, F.; Enomoto, M.; Noda, M.; et al. Performance of Generative Pretrained Transformer on the National Medical Licensing Examination in Japan. *PLoS Digit. Health* **2024**, *3*, e0000433. [\[CrossRef\]](#)
28. Zong, H.; Li, J.; Wu, E.; Wu, R.; Lu, J.; Shen, B. Performance of ChatGPT on Chinese national medical licensing examinations: A five-year examination evaluation study for physicians, pharmacists and nurses. *BMC Med. Educ.* **2024**, *24*, 143. [\[CrossRef\]](#) [\[PubMed\]](#)
29. Wojcik, S.; Rulkiewicz, A.; Pruszczyk, P.; Lisik, W.; Pobozy, M.; Domienik-Karłowicz, J. Reshaping medical education: Performance of ChatGPT on a PES medical examination. *Cardiol. J.* **2023**. [\[CrossRef\]](#) [\[PubMed\]](#)
30. Lai, U.H.; Wu, K.S.; Hsu, T.Y.; Kan, J.K.C. Evaluating the performance of ChatGPT-4 on the United Kingdom Medical Licensing Assessment. *Front. Med.* **2023**, *10*, 1240915. [\[CrossRef\]](#)
31. Li, D.J.; Kao, Y.C.; Tsai, S.J.; Bai, Y.M.; Yeh, T.C.; Chu, C.S.; Hsu, C.W.; Cheng, S.W.; Hsu, T.W.; Liang, C.S.; et al. Comparing the performance of ChatGPT GPT-4, Bard, and Llama-2 in the Taiwan Psychiatric Licensing Examination and in differential diagnosis with multi-center psychiatrists. *Psychiatry Clin. Neurosci.* **2024**. [\[CrossRef\]](#)
32. Su, M.C.; Lin, L.E.; Lin, L.H.; Chen, Y.C. Assessing question characteristic influences on ChatGPT's performance and response-explanation consistency: Insights from Taiwan's Nursing Licensing Exam. *Int. J. Nurs. Stud.* **2024**, *153*, 104717. [\[CrossRef\]](#)
33. Chau, R.C.W.; Thu, K.M.; Yu, O.Y.; Hsung, R.T.; Lo, E.C.M.; Lam, W.Y.H. Performance of Generative Artificial Intelligence in Dental Licensing Examinations. *Int. Dent. J.* **2024**, *in press*. [\[CrossRef\]](#)
34. Wang, A.Y.; Lin, S.; Tran, C.; Homer, R.J.; Wilsdon, D.; Walsh, J.C.; Goebel, E.A.; Sansano, I.; Sonawane, S.; Cockenpot, V.; et al. Assessment of Pathology Domain-Specific Knowledge of ChatGPT and Comparison to Human Performance. *Arch. Pathol. Lab. Med.* **2024**. [\[CrossRef\]](#)
35. Wang, Y.M.; Shen, H.W.; Chen, T.J. Performance of ChatGPT on the pharmacist licensing examination in Taiwan. *J. Chin. Med. Assoc.* **2023**, *86*, 653–658. [\[CrossRef\]](#)
36. Deebel, N.A.; Terlecki, R. ChatGPT Performance on the American Urological Association Self-assessment Study Program and the Potential Influence of Artificial Intelligence in Urologic Training. *Urology* **2023**, *177*, 29–33. [\[CrossRef\]](#)



37. Suchman, K.; Garg, S.; Trindade, A.J. Chat Generative Pretrained Transformer Fails the Multiple-Choice American College of Gastroenterology Self-Assessment Test. *Am. J. Gastroenterol.* **2023**, *118*, 2280–2282. [CrossRef] [PubMed]
38. Huh, S. Are ChatGPT's knowledge and interpretation ability comparable to those of medical students in Korea for taking a parasitology examination?: A descriptive study. *J. Educ. Eval. Health Prof.* **2023**, *20*, 1. [CrossRef] [PubMed]
39. Sakai, D.; Maeda, T.; Ozaki, A.; Kanda, G.N.; Kurimoto, Y.; Takahashi, M. Performance of ChatGPT in Board Examinations for Specialists in the Japanese Ophthalmology Society. *Cureus* **2023**, *15*, e49903. [CrossRef] [PubMed]
40. Abdelhady, A.M.; Davis, C.R. Plastic Surgery and Artificial Intelligence: How ChatGPT Improved Operation Note Accuracy, Time, and Education. *Mayo Clin. Proc. Digit. Health* **2023**, *1*, 299–308. [CrossRef]
41. Singh, S.; Djalilian, A.; Ali, M.J. ChatGPT and Ophthalmology: Exploring Its Potential with Discharge Summaries and Operative Notes. *Semin. Ophthalmol.* **2023**, *38*, 503–507. [CrossRef]
42. Baker, H.P.; Dwyer, E.; Kalidoss, S.; Hynes, K.; Wolf, J.; Strelzow, J.A. ChatGPT's Ability to Assist with Clinical Documentation: A Randomized Controlled Trial. *J. Am. Acad. Orthop. Surg.* **2024**, *32*, 123–129. [CrossRef]
43. Decker, H.; Trang, K.; Ramirez, J.; Colley, A.; Pierce, L.; Coleman, M.; Bongiovanni, T.; Melton, G.B.; Wick, E. Large Language Model-Based Chatbot vs Surgeon-Generated Informed Consent Documentation for Common Procedures. *JAMA Netw. Open* **2023**, *6*, e2336997. [CrossRef]
44. Miao, J.; Thongprayoon, C.; Suppadungsuk, S.; Garcia Valencia, O.A.; Qureshi, F.; Cheungpasitporn, W. Innovating Personalized Nephrology Care: Exploring the Potential Utilization of ChatGPT. *J. Pers. Med.* **2023**, *13*, 1681. [CrossRef]
45. Qarajeh, A.; Tangpanithandee, S.; Thongprayoon, C.; Suppadungsuk, S.; Krisanapan, P.; Aiumtrakul, N.; Garcia Valencia, O.A.; Miao, J.; Qureshi, F.; Cheungpasitporn, W. AI-Powered Renal Diet Support: Performance of ChatGPT, Bard AI, and Bing Chat. *Clin. Pract.* **2023**, *13*, 1160–1172. [CrossRef]
46. Aiumtrakul, N.; Thongprayoon, C.; Arayangkool, C.; Vo, K.B.; Wannaphut, C.; Suppadungsuk, S.; Krisanapan, P.; Garcia Valencia, O.A.; Qureshi, F.; Miao, J.; et al. Personalized Medicine in Urolithiasis: AI Chatbot-Assisted Dietary Management of Oxalate for Kidney Stone Prevention. *J. Pers. Med.* **2024**, *14*, 107. [CrossRef]
47. Miao, J.; Thongprayoon, C.; Garcia Valencia, O.A.; Krisanapan, P.; Sheikh, M.S.; Davis, P.W.; Mekraksakit, P.; Suarez, M.G.; Craici, I.M.; Cheungpasitporn, W. Performance of ChatGPT on Nephrology Test Questions. *Clin. J. Am. Soc. Nephrol.* **2023**, *19*, 35–43. [CrossRef]
48. Shah, D. The Beginner's Guide to Hallucinations in Large Language Models. Available online: <https://www.lakera.ai/blog/guide-to-hallucinations-in-large-language-models#:~:text=A%20significant%20factor%20contributing%20to,and%20factual%20correctness%20is%20challenging> (accessed on 23 August 2023).
49. Metze, K.; Morandin-Reis, R.C.; Lorand-Metze, I.; Florindo, J.B. Bibliographic Research with ChatGPT may be Misleading: The Problem of Hallucination. *J. Pediatr. Surg.* **2024**, *59*, 158. [CrossRef] [PubMed]
50. Temsah, O.; Khan, S.A.; Chaiah, Y.; Senjab, A.; Alhasan, K.; Jamal, A.; Aljamaan, F.; Malki, K.H.; Halwani, R.; Al-Tawfiq, J.A.; et al. Overview of Early ChatGPT's Presence in Medical Literature: Insights From a Hybrid Literature Review by ChatGPT and Human Experts. *Cureus* **2023**, *15*, e37281. [CrossRef] [PubMed]
51. Wagner, M.W.; Ertl-Wagner, B.B. Accuracy of Information and References Using ChatGPT-3 for Retrieval of Clinical Radiological Information. *Can. Assoc. Radiol. J.* **2023**, *75*, 8465371231171125. [CrossRef] [PubMed]
52. King, M.R. Can Bard, Google's Experimental Chatbot Based on the LaMDA Large Language Model, Help to Analyze the Gender and Racial Diversity of Authors in Your Cited Scientific References? *Cell Mol. Bioeng.* **2023**, *16*, 175–179. [CrossRef] [PubMed]
53. Dumitru, M.; Berghi, O.N.; Taciuc, I.A.; Vranceanu, D.; Manole, F.; Costache, A. Could Artificial Intelligence Prevent Intraoperative Anaphylaxis? Reference Review and Proof of Concept. *Medicina* **2022**, *58*, 1530. [CrossRef]
54. Bhattacharyya, M.; Miller, V.M.; Bhattacharyya, D.; Miller, L.E. High Rates of Fabricated and Inaccurate References in ChatGPT-Generated Medical Content. *Cureus* **2023**, *15*, e39238. [CrossRef] [PubMed]
55. Alkaissi, H.; McFarlane, S.I. Artificial Hallucinations in ChatGPT: Implications in Scientific Writing. *Cureus* **2023**, *15*, e35179. [CrossRef]
56. Athaluri, S.A.; Manthena, S.V.; Kesapragada, V.; Yarlagadda, V.; Dave, T.; Duddumpudi, R.T.S. Exploring the Boundaries of Reality: Investigating the Phenomenon of Artificial Intelligence Hallucination in Scientific Writing Through ChatGPT References. *Cureus* **2023**, *15*, e37432. [CrossRef] [PubMed]
57. Masters, K. Medical Teacher's first ChatGPT's referencing hallucinations: Lessons for editors, reviewers, and teachers. *Med. Teach.* **2023**, *45*, 673–675. [CrossRef]
58. Suppadungsuk, S.; Thongprayoon, C.; Krisanapan, P.; Tangpanithandee, S.; Garcia Valencia, O.; Miao, J.; Mekraksakit, P.; Kashani, K.; Cheungpasitporn, W. Examining the Validity of ChatGPT in Identifying Relevant Nephrology Literature: Findings and Implications. *J. Clin. Med.* **2023**, *12*, 5550. [CrossRef]
59. Aiumtrakul, N.; Thongprayoon, C.; Suppadungsuk, S.; Krisanapan, P.; Miao, J.; Qureshi, F.; Cheungpasitporn, W. Navigating the Landscape of Personalized Medicine: The Relevance of ChatGPT, BingChat, and Bard AI in Nephrology Literature Searches. *J. Pers. Med.* **2023**, *13*, 1457. [CrossRef]
60. Mayo, M. Unraveling the Power of Chain-of-Thought Prompting in Large Language Models. Available online: <https://www.kdnuggets.com/2023/07/power-chain-thought-prompting-large-language-models.html> (accessed on 13 November 2023).
61. Ott, S.; Hebenstreit, K.; Lievin, V.; Hother, C.E.; Moradi, M.; Mayrhauser, M.; Praas, R.; Winther, O.; Samwald, M. ThoughtSource: A central hub for large language model reasoning data. *Sci. Data* **2023**, *10*, 528. [CrossRef]

62. Wei, J.; Wang, X.; Schuurmans, D.; DBosma, M.; Ichter, B.; Xia, F.; Chi, E.; Le, Q.; Zhou, D. Chain-of-Thought Prompting Elicits Reasoning in Large Language Models. Available online: <https://arxiv.org/abs/2201.11903> (accessed on 13 November 2023).
63. Ramlochan, S. Master Prompting Concepts: Zero-Shot and Few-Shot Prompting. Available online: <https://promptengineering.org/master-prompting-concepts-zero-shot-and-few-shot-prompting/> (accessed on 25 April 2023).
64. Miao, J.; Thongprayoon, C.; Suppadungsuk, S.; Krisanapan, P.; Radhakrishnan, Y.; Cheungpasitporn, W. Chain of Thought Utilization in Large Language Models and Application in Nephrology. *Medicina* **2024**, *60*, 148. [CrossRef]
65. Wolff, T. How to Craft Prompts for Maximum Effectiveness. Available online: <https://medium.com/mllearning-ai/from-zero-shot-to-chain-of-thought-prompt-engineering-choosing-the-right-prompt-types-88800f242137> (accessed on 14 November 2023).
66. Shin, E.; Ramanathan, M. Evaluation of prompt engineering strategies for pharmacokinetic data analysis with the ChatGPT large language model. *J. Pharmacokinet. Pharmacodyn.* **2023**. [CrossRef]
67. Wadhwa, S.; Amir, S.; Wallace, B.C. Revisiting Relation Extraction in the era of Large Language Models. *Proc. Conf. Assoc. Comput. Linguist. Meet.* **2023**, *2023*, 15566–15589. [CrossRef]
68. Merritt, R. What Is Retrieval-Augmented Generation, Aka RAG? Available online: [https://blogs.nvidia.com/blog/what-is-retrieval-augmented-generation/#:~:text=Generation%20\(RAG\)?,Retrieval-augmented%20generation%20\(RAG\)%20is%20a%20technique%20for%20enhancing,how%20many%20parameters%20they%20contain](https://blogs.nvidia.com/blog/what-is-retrieval-augmented-generation/#:~:text=Generation%20(RAG)?,Retrieval-augmented%20generation%20(RAG)%20is%20a%20technique%20for%20enhancing,how%20many%20parameters%20they%20contain) (accessed on 15 November 2023).
69. Guo, Y.; Qiu, W.; Leroy, G.; Wang, S.; Cohen, T. Retrieval augmentation of large language models for lay language generation. *J. Biomed. Inform.* **2023**, *149*, 104580. [CrossRef]
70. Luu, R.K.; Buehler, M.J. BioinspiredLLM: Conversational Large Language Model for the Mechanics of Biological and Bio-Inspired Materials. *Adv. Sci.* **2023**, e2306724. [CrossRef]
71. Wang, C.; Ong, J.; Wang, C.; Ong, H.; Cheng, R.; Ong, D. Potential for GPT Technology to Optimize Future Clinical Decision-Making Using Retrieval-Augmented Generation. *Ann. Biomed. Eng.* **2023**. [CrossRef]
72. Ge, J.; Sun, S.; Owens, J.; Galvez, V.; Gologorskaya, O.; Lai, J.C.; Pletcher, M.J.; Lai, K. Development of a Liver Disease-Specific Large Language Model Chat Interface using Retrieval Augmented Generation. *medRxiv* **2023**. [CrossRef] [PubMed]
73. Zakka, C.; Chaurasia, A.; Shad, R.; Dalal, A.R.; Kim, J.L.; Moor, M.; Alexander, K.; Ashley, E.; Boyd, J.; Boyd, K.; et al. Almanac: Retrieval-Augmented Language Models for Clinical Medicine. *Res. Sq.* **2023**. [CrossRef] [PubMed]
74. Zakka, C.; Shad, R.; Chaurasia, A.; Dalal, A.R.; Kim, J.L.; Moor, M.; Fong, R.; Phillips, C.; Alexander, K.; Ashley, E.; et al. Almanac—Retrieval-Augmented Language Models for Clinical Medicine. *NEJM AI* **2024**, *1*, AIoa2300068. [CrossRef]
75. Unlu, O.; Shin, J.; Mailly, C.J.; Oates, M.F.; Tucci, M.R.; Varugheese, M.; Waghlikar, K.; Wang, F.; Scirica, B.M.; Blood, A.J.; et al. Retrieval Augmented Generation Enabled Generative Pre-Trained Transformer 4 (GPT-4) Performance for Clinical Trial Screening. *medRxiv* **2024**. [CrossRef]
76. KDIGO 2023 Clinical Practice Guideline for the Evaluation and Management of Chronic Kidney Disease. Available online: <https://kdigo.org/guidelines/ckd-evaluation-and-management/> (accessed on 1 July 2023).

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.