

Emergence of artificial generative intelligence and its potential impact on urology

Mohamed Javid, MS,¹ Madhu Reddiboina, MS,² Mahendra Bhandari, MD³

¹Department of Urology, Chengalpattu Medical College, Chengalpattu, India

²RediMinds, Inc., Southfield, Michigan, USA

³Vattikuti Urology Institute, Henry Ford Hospital, Detroit, Michigan, USA

JAVID M, REDDIBOINA M, BHANDARI M. Emergence of artificial generative intelligence and its potential impact on urology. *Can J Urol* 2023;30(4):11588-11598.

Introduction: Artificial generative intelligence (AGI) and large language models (LLMs) have gained significant attention in healthcare and hold enormous promise for transforming every aspect of our life and urology is no exception.

Materials and methods: We conducted a comprehensive literature search of electronic databases and included articles discussing AGI and LLMs in healthcare. Additionally, we have incorporated our experiences interacting with the ChatGPT and GPT-4 in different situations with real case reports and case constructs.

Results: Our review highlights the potential applications and likely impact of these technologies in urology, for differential diagnosis, prioritizing treatment options, and facilitating research, surgeon, and patient education. At their current developmental stage, we have recognized the need for concurrent validation and continuous

human interaction necessary to induce inverse reinforced learning with human feedback to mature them to authenticity. We need to consciously adjust to the hallucinations and guard patients' confidentiality before their extensive implementations in clinical practice. We propose possible remedies for these shortcomings and emphasize the critical role of human interaction in their evolution.

Conclusion: The integration of these tools has the potential to revolutionize urology, but it also presents several challenges needing attention. To harness the full potential of these models, urologists must consistently engage in training these tools with their clinical sense and experience. We urge the urology community to actively participate in AGI and LLM development to address potential challenges. These models could help us in unleashing our full potential and help us achieve a better work-life balance.

Key Words: artificial intelligence, artificial generative intelligence, large-language models, urology, ChatGPT, Med-PaLM 2, BARD

Introduction

Artificial intelligence (AI) represents an interdisciplinary fusion of computer science and linguistics, with the overarching goal of developing systems capable of executing tasks that are conventionally reliant on human cognitive aptitude and human intelligence.¹ Artificial Generative Intelligence (AGI) refers to a

recent subset of AI models and algorithms that focus on creating new content or data based on the patterns and structures they have learned from their training data.² Large language models (LLMs) are a subset of AI that encompass machine learning frameworks, which are trained on a large corpus of text datasets, enabling them to proficiently perform a diverse array of natural language processing operations, Figure 1.^{5,6} These AI tools employ multilayer transformer-based neural networks and are coupled with extensive data training, to produce text from the internet comprising many unique words and phrases that closely resemble human written communication.⁷ Notable examples of this technology that utilize AGI include Open AI's conversational agent ChatGPT (Chat-based Generative

Accepted for publication July 2023

Address correspondence to Dr. Mohamed Javid, Department of Urology, Chengalpattu Medical College, Chengalpattu 603001 India

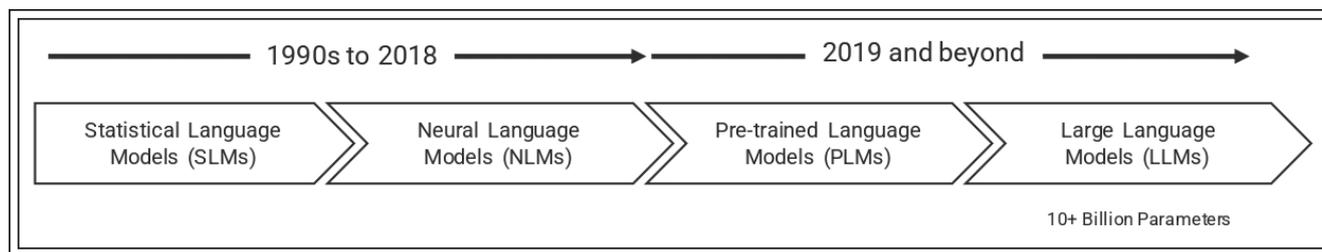


Figure 1. Evolution of language models from 1990s.^{3,4}

Pre-trained Transformer), as well as Google’s dialogue systems LaMDA (Language Model for Dialogue Applications) and Bard.⁸ Bard is the newly unveiled experimental conversational AGI chatbot, that is currently predicated on Google’s LLM, the PaLM - 2, and has recently been made available to over 180 countries and territories. Notably, PaLM - 2 is the enhanced incarnation of its previous version, the PaLM (Pathways Language Model).⁹

The field of healthcare is facing unprecedented challenges due to the rapid progression in the domain of AI and its integration into the healthcare sector. The rise and penetration of AGI and LLMs in urological practice signal the onset of a ground-breaking and exciting era that is expected to reconfigure day-to-day practice. On the horizon is an imminent reality where human intellectual capabilities become augmented with amplified instant recall assisting in complex decisions at the peak of uncertainty. This automation simplifies academic and research methodologies. These represent only a fraction of the extensive applications of AGI and LLMs and the extent of their utility in urology is restricted only by the limits of human imagination. However, to fully harness the maximum capability of these tools, it is prudent to critically examine their implications in the practice of urology.

This comprehensive review aims to explore the integration of AGI and LLMs in healthcare and the possible profound implications of these technologies in the field of urology. It will encompass the potential applications of AGI and LLM in daily clinical practice, in the realm of academia, and in research, highlighting the conceivable benefits and constraints of these tools. As AI technology continues to evolve at a tremendous pace, it is imperative for urologists to update their knowledge and enhance their skill set with a profound understanding of the scope and impact of AGI and LLM on urology. This comprehensive insight empowers urologists to seamlessly integrate these technologies into their practice, leading to optimized healthcare delivery and improved patient outcomes.

Literature review

We conducted a comprehensive literature review to investigate the plausible applications of AGI and LLMs in the healthcare domain, with a particular focus on the field of urology. We used a comprehensive search strategy in the PubMed database, that included the following search terms: “Artificial General Intelligence AND medicine”, “Artificial General Intelligence AND urology”, “Large Language Model AND Urology”, “Large Language Model AND Medicine”, “ChatGPT AND Urology”, “ChatGPT AND Medicine”, “Generative artificial intelligence AND Medicine”, and “Generative artificial intelligence AND Urology”. We independently screened and included articles

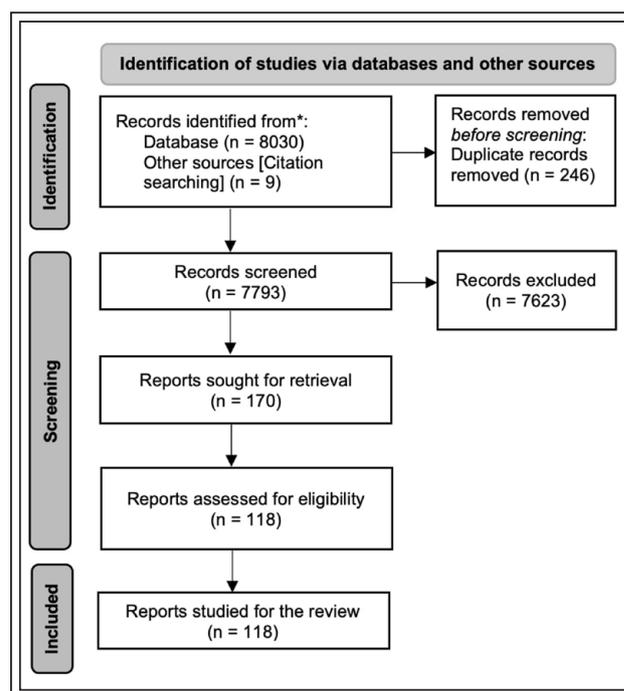


Figure 2. A flow diagram of the comprehensive search strategy for our literature review.

that discussed the applications of AGI and LLMs in healthcare, with a particular focus on the field of urology, in our literature review. The articles selected for review described the possible uses of AGI and LLMs in medicine and urology, with a view to exploring the potential impact of these technologies on clinical practice and other possible avenues. After removing duplicates from the initial identification of 8030 records, we screened 7793 records, assessed the eligibility of 170 articles, and ultimately studied and referred 118 articles in writing our literature review, Figure 2. However, considering the space constraints, not all articles have been listed in the reference section of this review and only the most pertinent articles have been included.

Our experiences with conversational AI

Interactive activity

To uphold conciseness, we have consolidated the prompts and responses, and excluded AI-generated disclaimers, while maintaining the integrity of the original content and its context.

Methods

We interacted with ChatGPT during different phases of clinical assessment and treatment across a diverse set of six clinical scenarios that encompassed localized renal cell carcinoma (RCC) with severe chemotherapy complications, incidentally detected localized chromophobe RCC, metastatic sarcomatoid RCC demonstrating treatment improvement, a complicated pregnancy involving RCC with a level IV IVC thrombus, recurrent high-grade pT1 urothelial bladder cancer following BCG therapy in a patient with a urological cancer history, and an unusual case of testicular torsion involving a supernumerary testicle.

We also leveraged the capabilities of ChatGPT to offer live surgical assistance during the following surgical case construct: A complex surgery for right-sided ureteral cancer, in which the patient experienced massive bleeding, graft thrombosis, and iliac vessel injury, with the complications exacerbated by an undiagnosed Factor VIII deficiency (Hemophilia A).

Prompt (P): A 75-year-old female patient with no known comorbidities underwent a right radical nephrectomy, resulting in a histology report of clear cell RCC (pT3 N0) and an uneventful postoperative recovery. After receiving T. sunitinib 50 mg for 3 weeks, she developed severe throat pain, abdominal pain, and hematemesis. Can you list the likely cause for this?

Response (R): *The severe throat pain, abdominal pain, and hematemesis in the patient who underwent radical nephrectomy and was subsequently treated with sunitinib*

may be due to several causes, including 1) Sunitinib-induced gastrointestinal (GI) toxicity. 2) Gastric ulcer. 3) Gastroesophageal reflux disease (GERD). 4) Acute pancreatitis. 5) Pulmonary embolism. 6) Cardiac toxicity.

Validation

Incidentally, this patient had succumbed to death due to gastric perforation and a literature search did confirm this to be a rare but reported complication of sunitinib.¹⁰

P: During the surgery for right-sided ureteral cancer, the ureter is adherent to iliac vessels, and I am afraid I will injure the vein and cause torrential bleeding. Please, advise me on what should I do as the next surgical step.

R: *It's important to consider the following: 1) Safety first: Consult with a colleague or expert in the field. Consider converting to an open surgery if you're not already performing one. 2) Careful dissection: Work slowly and meticulously to avoid injuring the vein. Use a combination of blunt and sharp dissection to carefully separate the ureter from the iliac vessels. You might also want to use vessel loops or a similar device. 3) Consider alternative approaches: If you're unable to safely separate the ureter from the iliac vessels, it might be necessary to consider a partial ureterectomy, bypassing the affected section, or even autotransplantation. 4) Be prepared for complications: Make sure you have all the necessary equipment and staff on hand to manage any potential complications, such as rapid hemorrhage control and transfusion. 5) Document and communicate.*

P: As feared there is an injury to a major vessel and the vascular surgeon could put clamps proximal and distal to the injury. It is a massive loss of the internal iliac vein. Please tell me what are the surgical options to reconstruct the internal iliac vein.

R: *In a crisis like this, you may want to consider the following: 1) Direct repair. 2) Patch angioplasty. 3) End-to-end anastomosis. 4) Interposition graft. 5) Bypass graft. 6) Ligation.*

P: This patient is under operation for 6 hours and is bleeding from all the sites including the wound, the site where the IV drip is going on. What is happening?

R: *Some possible causes for bleeding from multiple sites during surgery may include: 1) Coagulopathy. 2) Disseminated intravascular coagulation (DIC). 3) Hypothermia. 4) Platelet dysfunction. 5) Vessel injury.*

P: It was completely missed that this patient had a deficiency of Factor VIII. What should we do now?

R: *I'm not a doctor, but I can try to help you think through some general steps to manage a patient with a deficiency of Factor VIII, which is consistent with*

hemophilia A. 1) Administer Factor VIII replacement. 2) Monitor Factor VIII levels. 3) Consider Antifibrinolytic agents. 4) Adequate pain management. 5) Monitor for complications. 6) Consult with a hematologist. 7) Postoperative care: Appropriate follow-up care with a hematologist.

Validation

All the listed items in all of the above responses were elaborated more in the AI's actual response. Our interactions were related to learning about the AI's scope in clinical and surgical scenarios for:

- 1) Potentially predicting complications and identifying overlooked findings that may explain the possible complications.
- 2) The Usage of AI to provide logical and reliable information to support and guide clinical decision-making.
- 3) Providing general information on rare cases with proper references.
- 4) Offering alternative management options or suggestions for alternative surgical steps and approaches in the surgical scenario and explaining the rationale behind the chosen approach.
- 5) Outlining short term and long term consequences of surgical complications and providing suggestions for preoperative and postoperative management.
- 6) Suggesting potential causes of complications during surgery and offering suggestions for further assessment and management.
- 7) Providing a summary of the clinical or surgical scenario and complications discussed.

Patient education blogs

We utilized ChatGPT to prepare patient educational blogs and then conversed and customized those educational blogs.

P: I need to write patient educational blogs for our websites around the following keywords: Robotic surgery AND cancers AND prevention AND Robotic Urological cancers OR Robotic Colorectal surgery OR Robotic Liver Surgery OR Trans oral surgery. Could you please propose some relevant topics?

R: Here are some relevant topics for patient education blogs related to the keywords provided: Title: "The Advantages of Robotic Surgery for Urological Cancers: A Patient's Guide" Content: Discuss the benefits of robotic-assisted surgery in the treatment of urological cancers, including prostate, bladder, and kidney cancers. Explain how robotic surgery can lead to faster recovery times, reduced pain, and improved surgical precision.

Validation

A total of 7 such topics were suggested by ChatGPT. Coherent, well-structured, and lucid blogs on each topic were subsequently generated upon further prompts. Our conversations were to identify:

- 1) The provision of relevant, accurate, and simplified information.
- 2) The provision of specific and customized information with proper references.

Decoding complex journal articles and scientific texts

We interacted with AI by feeding it journal articles that may be difficult to understand, including the statistical analytical aspects of the study. One such example cited here is during our interaction based on the article by Schwen et al.¹¹

P: Can you quote the statistical methods used in this study and simplify them and explain why these tests have probably been used in this study design?

R: The study used several statistical methods to analyze the data, including 1) Descriptive statistics. 2) Univariate analysis. 3) Multivariate logistic regression analysis. 4) Area under the curve (AUC) analysis. 5) Negative predictive value (NPV) calculations.

Validation

The AI's response was elaborated on each of the above sections and on further discussion in each section, more details were provided such as: "Chi-squared test: This test is used to determine if there is a significant association between two categorical variables. In the context of the study, the researchers could have used the chi-squared test to analyze the relationship between categorical biomarker levels (e.g., above or below a certain cut off value) and the presence of GR." Our exercises were to study the AI's ability in the following: 1) Simplifying and explaining the content of the article. 2) Providing guidance on the advantages and limitations of various aspects of research. 3) Offering insight and ideas for future research topics. 4) Analyzing journal articles as a review with recent literature.

Preparing systematic review

We used ChatGPT to provide guidance and assistance in writing a systematic review.

P: Please help me to write a systemic review on low PSA and cancer prostate

R: Title: A Systematic Review of Low Prostate-Specific Antigen (PSA) Levels and Prostate Cancer Detection.

(A) Introduction (B) Methods: 1) Search strategy. 2) Inclusion and exclusion criteria. 3) Study selection. 4) Data extraction. 5) Quality assessment. 6) Data synthesis and analysis. (C) Results. (D) Discussion. (E) Conclusion.

Validation

Each section was actually elaborated by the AI, and it provided insights into each step when further interacted. Our assessments were to find the AI's capabilities to aid in the development of a systematic review on a given topic during various stages of the review.

Literature search

We employed ChatGPT, particularly GPT – 4, to assist us in bibliographic searches.

P: I am planning to write an article on the recent advances in PCNL. Can you provide me the most relevant references for “mini PCNL” AND “ultra mini PCNL” till August 2021?

R: 1) Guohua Z, Wen Z, Xun L et al. Comparative study of the efficacy and safety of the miniaturised percutaneous nephrolithotomy and standard percutaneous nephrolithotomy for the treatment of 2-3 cm kidney stones. *Urol Int.* 2019;102(2):153-60. doi: 10.1159/000494290.
2) Kılıç Ö, Akand M, Van Cleynenbreugel B. Mini, ultra-mini and micro percutaneous nephrolithotomy (PCNL) vs. standard PCNL: A systematic review and meta-analysis of randomised controlled trials. *BJU Int.* 2020;126(1):13-24. doi: 10.1111/bju.15006.

Validation

GPT – 4 had provided various references that were very relevant. However, it also generated responses like the references above, that do not actually exist. This is an example of a phenomenon that has been called “Hallucination”, in which the model produces text that is either factually incorrect or nonsensical.⁶ Our discussions were to explore the AI's aptitude in identifying the most relevant article by feeding in topics or keywords and verifying the actual correctness in the provision of the articles that might be needed.

ChatGPT versus BARD

The emergence of Bard, a novel conversational AI, naturally precipitates comparisons with pre-existing AIs such as ChatGPT. Bard's distinct advantage resides in its capability to access real-time internet-based data, forming a striking contrast with ChatGPT, which relies

on a pre-existing dataset extending only up to late 2021. Also, while Bard is freely available, access to the recent and enhanced ChatGPT – 4 model is behind a paywall, with only the 3.5 version available for free public use. From our preliminary and limited interactions with Bard, we observed that its risk awareness pertaining to the responses generated appeared somewhat limited. Moreover, the responses often tend to be more generic rather than specific often necessitating supplementary prompts to extract the desired answer. Further, the flexibility and the iterative learning from its own errors in Bard appear to be somewhat restricted. Bard predominantly seemed to gravitate towards the provision of information, whereas GPT – 4 seemed to demonstrate an additional proficient execution of complex textual functions. As the conversation AI sphere expands and as more objective comparisons emerge, we will gain more vivid insights into the strengths and weaknesses of each system, and determine the most efficacious application areas for each of the respective AI.

Potential benefits of AGI and LLMs

Clinical practice

The possibility of utilizing ChatGPT as a “virtual medical assistant” or “digital healthcare companion” in healthcare presents a promising avenue for exploration. During our interactions with ChatGPT, we noted that it demonstrated a remarkable capacity to provide insightful and clinically relevant responses across various stages of clinical evaluation and management. Its ability to suggest alternative treatment options, predict outcomes and complications, and identify alternative explanations or overlooked possibilities was particularly noteworthy. What is particularly striking is ChatGPT's adaptive nature, evidenced by its ability to learn from errors and refine its responses based on better reasoning or alternative explanations. This feature is particularly valuable in clinical decision-making processes, as it allows ChatGPT to continually improve its accuracy and usefulness. Its predictive capabilities, as well as its ability to identify potential alternative explanations, can aid healthcare providers in making informed decisions and in improving patient outcomes. The multifaceted abilities of ChatGPT hold the potential to augment human intelligence and amplify instantaneous recall. Additionally, these capabilities can facilitate support for decision-making processes in intricate and time-critical scenarios. In addition, ChatGPT demonstrated its capability in predicting patient prognosis and recurrence, utilizing either general statements or

specific tools such as nomograms when provided with sufficient data. Its ability to effectively summarize key findings at the conclusion of each scenario further highlights its potential as a valuable clinical assistant. ChatGPT can analyze multiple treatment options and provide evidence-based reasoning behind selecting a particular approach. This will help urologists make well-informed decisions that are based on the most current research and take into account individual patient factors.

A recent exploratory study examined the potential of ChatGPT-3.5 and ChatGPT – 4 as clinical assistants in urology decision-making.¹² The study posed a total of 30 urology questions to the two AI models and the responses were then reviewed by three experts who evaluated the consistency of the responses with the best available evidence. The findings indicated that ChatGPT had a high degree of capability in providing responses consistent with strong recommendations based on high-quality evidence in most cases. While there were some minor discrepancies from the recommendations, the advice provided by ChatGPT was generally aligned with the latest scientific evidence and expert guidelines and protocols.¹²

Although ChatGPT serves as a valuable adjunct for clinical decision-making, constraints exist as its training data is limited to information available only till late 2021. This limitation has significant implications on the accuracy and quality of responses generated by such models, particularly in relation to recent guidelines and emerging clinical evidence.

The AI is capable of providing not only real-time support on surgical steps, anatomy, and decision-making but also assisting surgeons by offering logical recommendations for subsequent steps during intricate operations. By taking into account the current surgical status, specific patient factors, and individual anatomy, the AI can supply tailored suggestions that enhance the overall effectiveness and safety of the surgery. Moreover, it can also offer insight into any unforeseen complications that may arise and ultimately lead to improved surgical outcomes and a reduction in complications.

As an extension to its potential role as a “virtual medical assistant”, ChatGPT can help urologists stay organized by managing secretarial tasks, correspondence, phone services, and message filtering. By prioritizing daily schedules, it can ensure that critical tasks are addressed first, allowing urologists to focus their time and energy on the most pressing matters. During consultations, these AIs can provide urologists with instant access to relevant information and guidelines, ensuring that patients receive accurate,

up-to-date care. By acting as a readily available source of knowledge, ChatGPT can help urologists address patient concerns more efficiently, leading to higher patient satisfaction and better outcomes. The AI can generate patient education materials customized to individual needs, considering various factors such as age, medical history, and desired information. By providing accurate and easy-to-understand information, AI can help improve patient adherence to treatment plans and facilitate better communication between patients and healthcare providers. Additionally, it can supply relevant references for the details mentioned in the blog, further enhancing the credibility of the information provided. However, it is always advisable to verify the accuracy of these references once to ensure the utmost reliability of the patient education materials.

Large Language Models (LLMs) hold the potential for aiding in the creation of medical reports or other documentation works through the utilization of specific keywords. Documentation accounts for approximately 25% to 50% of a physician’s time and 20% of a nurse’s time and it has been demonstrated that LLMs can significantly reduce the time spent by clinicians and other healthcare professionals on generating documentation content.⁸ ChatGPT is also currently being explored as a potential instrument for assisting doctors in generating discharge summaries with remarkable speed and efficiency. By inputting brief and specific information, concepts for elaboration, and guidance, ChatGPT can generate formal discharge summaries in mere seconds.¹³ Looking to the future, AGI and LLMs hold great promise for automated data extraction from electronic health records, removing the need for time-consuming manual inputting.¹³ This technology would enable hospitals to access a comprehensive patient database containing detailed information on patient complaints, examination findings, investigations, and management, which can be utilized to produce discharge summaries.¹³ Consequently, LLMs offer the possibility of transforming clinical practice by enabling medical practitioners to allocate more time for patient interaction.¹³ ChatGPT is an example of how AGI and LLM-based models can be helpful in the process of clinical decision support systems in healthcare.

Further exploration of the capabilities of ChatGPT and similar models in the context of healthcare is essential to fully understand their potential and limitations. Specifically, the development of AI models trained on more up-to-date medical databases has the potential to significantly enhance the accuracy and quality of their responses. In alignment with this

strategic direction, Google has recently announced the imminent availability of Med-PaLM 2, an advanced iteration of its initial LLM, the Pathways Language Model (PaLM), which has been specifically adapted for medical applications.¹⁴⁻¹⁶ Med-PaLM 2 is built upon Google's LLM, PaLM-2, and has been meticulously tuned for the medical domain.¹⁴⁻¹⁶ In a bid to explore and probe its diverse applications and assess its effectiveness, Google intends to provide a select group of Google Cloud customers with testing access to this advanced model with the overarching aim of ensuring that this innovative technology is deployed in a safe, responsible, and meaningful manner.¹⁴⁻¹⁶ Med-PaLM has been reported to have made a significant impact as the first-ever AI system to successfully pass the United States Medical Licensing Examination (USMLE) with a score of 67.4% while its successor, Med-PaLM 2, has been claimed to have made further strides, achieving a score exceeding 85.4%.¹⁴⁻¹⁶ These specialized LLMs, such as Med-PaLM 2, serve as pivotal elements in

the swiftly advancing field of AGI and harbor the capability to fundamentally transform the healthcare sector.¹⁴⁻¹⁶ Further research endeavors to imbue these models with multimodal capabilities, with the intent of not only developing a system that is proficient in text comprehension but also in the interpretation of complex medical data that encompasses, but is not limited to, medical records, various imaging modalities, and even genomic data.¹⁴⁻¹⁶ The advent of these game-changing technologies is poised to revamp the framework of present-day medical practice. However, it is crucial to ensure that their deployment strictly complies with ethical guidelines and is directed toward the improvement of patient outcomes.¹⁴⁻¹⁶ The promise held by AGI technologies such as Med-PaLM 2 is substantial, and it is incumbent upon us to explore their full capabilities in an ethically rigorous manner, Figure 3. Thus, the possibilities for enhanced healthcare practices and patient care are truly limitless with the implementation of AGI and LLMs in healthcare settings.

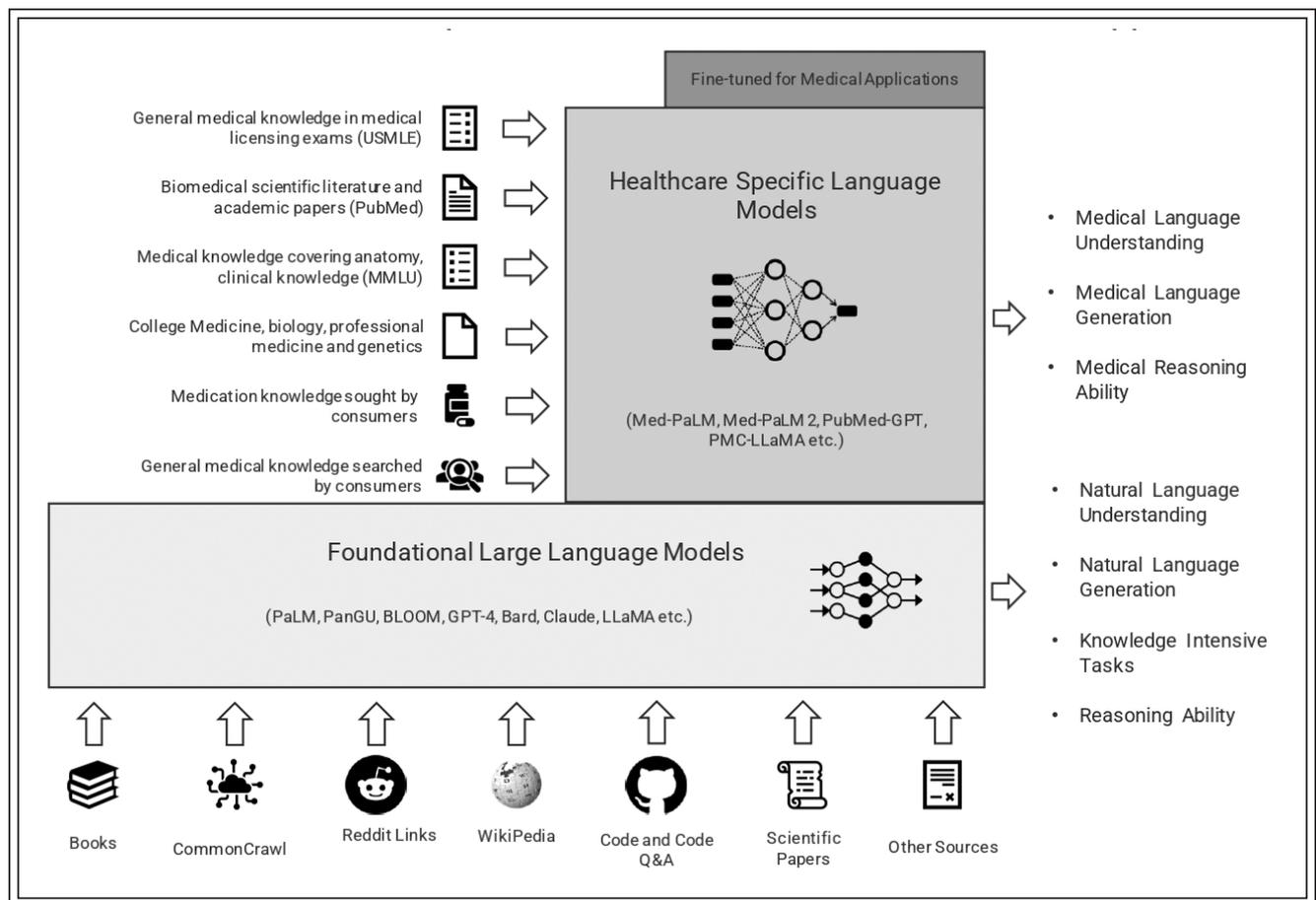


Figure 3. Data sources and capabilities of AGI and LLMs for medical applications.^{3,17-19}

Academics and research

Given that LLMs fundamentally operate on textual data, their applications and implications have been extensively investigated within the realms of academia, encompassing scholarly and research environments. Large language models (LLMs), such as ChatGPT, offer potential assistance in drafting manuscripts. Nevertheless, there are mixed responses concerning their efficacy in producing scientific texts. There exists concern regarding ChatGPT's insufficient domain-specific knowledge and expertise, which are indispensable for generating high-quality scientific manuscripts.²⁰ Moreover, ChatGPT cannot ensure the accuracy, validity, and reliability of scientific claims and findings—factors critical for upholding research integrity.²⁰ In light of these concerns, it is suggested that while ChatGPT might be helpful in generating initial ideas and basic drafts, it should not be utilized for completely authoring scientific manuscripts intended for publication. Furthermore, ChatGPT can assist in outlining research articles, ensuring that the content is organized logically and coherently. By recommending relevant resources, the AI enables urologists to construct quality content and well-substantiated arguments while upholding a high standard in their writing.

Some researchers propose that utilizing ChatGPT solely for language editing purposes poses no ethical concerns in the context of preparing scientific manuscripts.²¹ Nonetheless, it is imperative to subject any novel concepts generated by ChatGPT to rigorous experimental validation, with subsequent verification of the outcomes by human experts.²¹ Some editors and authors assert that using ChatGPT or similar language models for creating biomedical papers should not be entirely restricted.^{22,23} On the contrary, it may prove beneficial in situations where researchers from non-English-speaking countries encounter difficulties when submitting their work to English-language academic journals.^{22,23}

ChatGPT can aid in comprehending intricate texts and extracting conclusions and summaries from a journal article paper or textual content. Upon evaluating ChatGPT's capacity to comprehend information, as well as to derive conclusions from the text within the Background, Methods, and Results sections of an abstract, it was observed that GPT could accurately discern the context and provide a concise summation of the primary outcome of the study.²⁴ ChatGPT has been proposed as a tool for conducting extensive literature reviews. Nonetheless, when employing ChatGPT for referencing scientific

articles, its suggestions may exhibit partial correctness with the possibility of inaccuracies. Consequently, it is essential for researchers to exercise vigilance and conduct thorough verification when relying on such information.^{21,25,26} In light of our experiences, we have noted that the GPT – 4 model exhibits a higher level of accuracy in generating references as opposed to its predecessor, thereby offering a more dependable resource in the academic domain.

ChatGPT possesses the capability to assist in the development of research protocols, drafting a research article, and the generation of innovative ideas.²⁷ Based on our observations, we solicited ChatGPT's guidance in composing a systematic review. Impressively, it delineated the requisite steps for crafting a systematic review and supplied pertinent examples at each stage. Moreover, the references offered during its suggestions were highly relevant to the topic under investigation. Consequently, it may be advantageous to employ these LLMs as a navigational tool that directs the trajectory of a study or research, rather than entrusting them with the comprehensive execution of the task.

ChatGPT can potentially analyze the curriculum vitae (CVs) of candidates to identify unique skills and experiences, helping interviewers create personalized questions that delve deeper into each candidate's background. This will enable a more thorough evaluation, allowing interviewers to identify the most suitable candidates for residency, fellowship, and recruitment positions. ChatGPT can also be a valuable resource for interviewees preparing for the interviews. By analyzing their own CVs, the AI can help predict possible questions that might be asked during the interview, allowing candidates to better prepare their responses. Additionally, ChatGPT can be used to conduct mock interviews, simulating the real interview experience, and can provide constructive feedback on their performance. This interactive practice can help candidates gain confidence and hone their communication skills, ultimately leading to improved performance during the actual interview.

Challenges and concerns related to AGI in urology and possible solutions

The challenges and concerns associated with LLMs raise important questions regarding medical ethics. These issues include data privacy and security, as well as the potential for incorrect or misleading information that could potentially harm patients.

LLMs can generate significant challenges concerning accuracy and impartiality. For instance, if an LLM is trained on biased data, it may produce biased results

when generating outputs. In other words, the LLM's output may be skewed or influenced by the data that it was trained on. Moreover, LLMs rely on specific datasets to learn and generate outputs, and this may introduce biases into their outputs based on the training data.^{6,8,28-30} LLMs trained on vast amounts of internet data with minimal filtering may absorb a mixture of accurate information, misinformation, biased content, and harmful as well as harmless materials. Unfortunately, since LLMs lack the means to evaluate these criteria before generating a response, they face the risk of reproducing, amplifying, and disseminating problematic content and misinformation. Therefore, proper training and validation of the LLMs algorithm with evidence-based knowledge bases to train the model for applications related to diagnosis and therapy support can decrease such flaws.^{6,8,28-30}

Another potential issue with utilizing ChatGPT is the potential risk of reducing critical thinking and creativity among researchers. By relying heavily on these technological models to perform tasks such as data analysis and paper writing, researchers may gradually lose the ability to think creatively and critically.²⁸ It is vital to acknowledge that ChatGPT lacks the capacity to carry out statistical analyses. However, it is noteworthy that ChatGPT can offer helpful guidance for statistical analyses and this feature may prove useful to readers who are not familiar with performing statistical analyses but this should be taken with caution.²⁴

LLMs operate using probabilistic algorithms, meaning that when presented with the same task or question multiple times, the model may generate varying responses. These responses could either be different versions of previous incorrect or problematic answers, improved or correct answers, different variations of prior correct responses, or a combination of these options. This feature raises concerns about reliability and reproducibility, necessitating continuous human supervision of the model's operation.⁸ A major problem with LLMs is referred to as 'neural hallucinations,' where the model produces text that is either factually incorrect or nonsensical, or it makes up things despite displaying a high degree of confidence in its output.^{6,31} Moreover, the varying accuracy of some of the ChatGPT responses may pose a risk to education if the model is used without proper authorization and guidance. This could lead to medical personnel in training, students, or even patients being swayed by ChatGPT and misunderstanding medical information. Consequently, they may make vital clinical decisions without undertaking critical evaluation and validation.³⁰

The use of LLMs in patient management necessitates a cautious approach due to potential ethical and legal implications. Contemporary versions of GPT do not comply with the Health Insurance Portability and Accountability Act (HIPAA), which safeguards patient privacy and so, it has been advised that clinicians refrain from inputting such information until professional-grade versions with appropriate safeguards are available.³² As astutely highlighted in the article by Haupt et al, the crux of comprehending ethical and legal concerns eclipsing novel technologies lies not solely in the technologies per se, but rather in their impact on the social dynamics between users.³² While the patient-physician relationship is heavily regulated and governed by trust, rules, and established norms, GPT currently operates outside these boundaries.³² Also when the information generated by AI is filtered and disseminated through professionals' judgment, it has been suggested that existing legal frameworks can assign liability to professionals for their advice and so it is imperative that urologists scrutinize and verify their outputs when employing AI tools.³² Given the existing dearth of professional oversight and apprehensions surrounding accuracy, reliability, and possible legal ramifications, it is strongly recommended to abstain from utilizing GPT in direct patient-facing AI applications within healthcare delivery that aim to supplant clinician judgment, at least for the foreseeable future.³² When emotional support or medical opinion is directly sought from AI by patients, they operate beyond the patient-physician relationship, with limited regulatory safeguards in place and so, irrespective of regulatory measures, it is crucial for urologists to enlighten patients about the cautious use of LLMs outside the spectrum of a patient-physician relationship.³² They should emphasize that such tools are largely unregulated, potentially misleading, and lacks the legal and ethical responsibilities that healthcare practitioners uphold.³² Therefore, it is imperative to establish clear guidelines and regulations for LLM application in clinical settings to minimize potential risks and ensure responsible use.⁶ It is crucial to acknowledge that, despite their remarkable capabilities in natural language processing, LLMs are not without limitations. As such, it is imperative to utilize these models with caution and awareness of their potential drawbacks. Also, it is highly essential to thoroughly evaluate and ensure the accuracy and validity of the output of an LLM before integrating it into any work.

The role of human interaction in AGI development

Human interaction plays a critical role in the development of AGI and LLM models, ensuring their robustness, predictability, and accuracy. By engaging with these models and providing input based on their expertise, urologists contribute to the continuous refinement of AGI and LLM systems. This iterative process helps to fine-tune the models' performance, resulting in more reliable and precise tools for urological practice. Finding and establishing a balance between human input and AI capabilities is essential for optimizing the performance of these models. While AI systems offer unparalleled computational power and pattern recognition, human insight and experience, and the distinctive quality of a compassionate "human touch" remain invaluable for contextualizing and interpreting the generated data. Consequently, developing a collaborative relationship between urologists and AGI and LLM models will enable the development of more effective and sophisticated tools tailored to the specific needs of the field of urology. To fully harness the potential of AGI in urology, it is crucial to encourage urologists to actively engage with these models and participate in the process of human feedback learning. This can be achieved by providing training opportunities, promoting interdisciplinary collaboration, and highlighting the benefits of AGI integration in clinical practice. By reinforcing the importance of human feedback learning, urologists will be better equipped to contribute to the development of AGI models, ultimately resulting in enhanced patient care and improved clinical outcomes.

Preparing for the future of AGI and LLM in urology

Given the rapid advancement of AGI and the fact that AGI represents the future direction of the field, it is imperative that urologists adapt to these changes and learn to employ AGI tools efficiently and ethically. As the AGI models become more accurate and appropriate with increased application and usage, the efficacy of these models highly relies on their consistent utilization and interaction. So, it is thus highly recommended that urologists actively engage with these models for various applications, as enumerated earlier, to refine and enhance their effectiveness in urological practice. The integration of AGI and LLM models into urology is bound to present various challenges, some of which may emerge as the models evolve and are applied in diverse contexts. Proactively anticipating these potential obstacles and diligently

seeking appropriate solutions will facilitate their timely resolution, ultimately leading to more efficient and optimized models. The earlier urologists begin incorporating these models into their practice, the more rapidly these models can improve and contribute to the advancement of the field of urology.

Conclusion

In conclusion, the advent of the AGI and LLMs as exemplified by models like ChatGPT envisions a plausible monumental paradigm shift in the field of urology. These cutting-edge technologies withhold a transformative potential to revolutionize clinical decision-making, surgical practice, research methodologies, and educational approaches, thereby amplifying the expertise of urologists across various domains. Consequently, these models are to be seen as a means to augment and supplement human capabilities rather than to replace them. This vision can ensure that a synergistic relationship can be formed that maximizes the combined output while alleviating any concerns about the potential impact from the AI's expeditious growth and proficient aptitudes. However, it is vital to acknowledge and navigate through the multifaceted challenges that might accompany these developments, such as the risk of generating biased or inaccurate information, the potential depletion of creativity and critical thinking skills, and the intricate ethical and legal considerations in patient management. To profoundly unlock the underlying monumental abilities of these AIs, active engagement from urologists is indispensable. By promptly participating in the ongoing refinement and optimization of these models through valuable feedback and human input, urologists can guide the direction of the growth of these models. Through proactive collaboration and meticulous scrutiny of the possible shortcomings and challenges, urologists can overcome these and can eventually tap into the unparalleled capabilities of these technologies. This can ensure proper and successful integration into the urological practice and ultimately improve patient care and outcomes. It is our ardent call to the urology community to fervently partake in the development and advancement of the AGI and LLMs. □

References

1. Sarker IH. AI-based modeling: techniques, applications and research issues towards automation, intelligent and Smart Systems. *SN Comput Sci* 2022;3(2):158.

- Baidoo-Anu, David and Owusu Ansah, Leticia, Education in the Era of Generative Artificial Intelligence (AI): Understanding the Potential Benefits of ChatGPT in Promoting Teaching and Learning (January 25, 2023). Available at SSRN: <https://ssrn.com/abstract=4337484> or <http://dx.doi.org/10.2139/ssrn.4337484>.
- A Survey of Large Language Models; arXiv:2303.18223 [cs.CL] <https://doi.org/10.48550/arXiv.2303.18223>.
- Harnessing the Power of LLMs in Practice: A Survey on ChatGPT and Beyond; arXiv:2304.13712 [cs.CL] <https://doi.org/10.48550/arXiv.2304.13712>.
- Ahn S. The impending impacts of large language models on medical education. *Korean J Med Educ* 2023;35(1):103-107.
- Janssen BV, Kazemier G, Besselink MG. The use of CHATGPT and other large language models in surgical science. *BJS Open* 2023;7(2):zrad032.
- Alberts IL, Mercolli L, Pyka T et al. Large language models (LLM) and chatgpt: What will the impact on nuclear medicine be? *Eur J Nucl Med Mol Imaging* 2023;50(6):1549-1552.
- Harrer S. Attention is not all you need: the complicated case of ethically using large language models in healthcare and medicine. *EBioMedicine* 2023;90:104512.
- Hsiao S. What's ahead for bard: More global, more visual, more integrated [Internet]. Google; 2023 [cited 2023 May 15]. Available from: <https://blog.google/technology/ai/google-bard-updates-io-2023/>.
- Porta C, Paglino C, Imarisio I, Bonomi L. Uncovering pandora's vase: The growing problem of new toxicities from novel anticancer agents. The case of sorafenib and sunitinib. *Clin Exp Med* 2007;7(4):127-134.
- Schwen ZR, Mamawala M, Tosoian JJ et al. Prostate health index and multiparametric magnetic resonance imaging to predict prostate cancer grade reclassification in active surveillance. *BJU Int* 2020;126(3):373-378.
- Zhou Z, Wang X, Li X, Liao L. Is ChatGPT an evidence-based doctor? *Eur Urol*. 2023;84:355-356.
- Patel SB, Lam K. ChatGPT: the future of discharge summaries? *Lancet Digit Health* 2023;5(3):e107-e108.
- Med-PaLM [Internet]. [cited 2023 May 15]. Available from: <https://sites.research.google/med-palm/> Available from: <https://sites.research.google/med-palm/>.
- Gupta A, Waldron A. A responsible path to generative AI in healthcare | Google Cloud Blog [Internet]. Google; 2023 [cited 2023 May 15]. Available from: <https://cloud.google.com/blog/topics/healthcare-life-sciences/sharing-google-med-palm-2-medical-large-language-model>.
- Matias Y, Corrado G. Our latest health AI research updates [Internet]. Google; 2023 [cited 2023 May 15]. Available from: <https://blog.google/technology/health/ai-llm-medpalm-research-thecheckup/>.
- Large Language Models Encode Clinical Knowledge; arXiv:2212.13138 [cs.CL] <https://doi.org/10.48550/arXiv.2212.13138>.
- ChatDoctor: A Medical Chat Model Fine-Tuned on a Large Language Model Meta-AI (LLaMA) Using Medical Domain Knowledge; arXiv:2303.14070 [cs.CL] <https://doi.org/10.48550/arXiv.2303.14070>.
- PMC-LLaMA: Further Finetuning LLaMA on Medical Papers; arXiv:2304.14454 [cs.CL] <https://doi.org/10.48550/arXiv.2304.14454>.
- Why CHATGPT should not be used to write academic scientific manuscripts for publication. *The Ann Fam Med* 2023;29:58.
- Kim S-G. Using CHATGPT for language editing in scientific articles. *Maxillofac Plast Reconstr Surg* 2023;45(1):13.
- Mishra D. Indian pediatrics' policy regarding artificial intelligence (AI) — enabled large language models. *Indian Pediatr* 2023;60(3):171.
- The Lancet Digital Health. ChatGPT: Friend or foe? *Lancet Digit Health* 2023;5(3):e102.
- Cascella M, Montomoli J, Bellini V, Bignami E. Evaluating the feasibility of CHATGPT in healthcare: An analysis of multiple clinical and research scenarios. *J Med Syst* 2023; 47(1):33.
- Biswas S. Chatgpt and the future of medical writing. *Radiology* 2023;307(2):e223312.
- Alkaissi H, McFarlane SI. Artificial hallucinations in CHATGPT: implications in scientific writing. *Cureus* 2023;15(2):e35179.
- Macdonald C, Adeloye D, Sheikh A, Rudan I. Can chatgpt draft a research article? an example of population-level vaccine effectiveness analysis. *J Glob Health* 2023;13:01003.
- Marchandot B, Matsushita K, Carmona A, Trimaille A, Morel O. ChatGPT: the next frontier in academic writing for cardiologists or a Pandora's box of ethical dilemmas. *Eur Heart J Open* 2023; 3(2):oead007.
- Baumgartner C. The opportunities and pitfalls of ChatGPT in clinical and translational medicine. *Clin Transl Med* 2023;13(3):1206.
- Xue VW, Lei P, Cho WC. The potential impact of CHATGPT in clinical and translational medicine. *Clin Transl Med* 2023; 13(3):e1216.
- King MR. The future of AI in medicine: a perspective from a chatbot. *Ann Biomed Eng* 2022;51(2):291-295.
- Haupt CE, Marks M. Ai-generated medical advice—GPT and beyond. *JAMA* 2023;329(16):1349.