FUTURE OF THE LANGUAGE MODELS IN HEALTHCARE: THE ROLE OF CHATGPT

Francisco Tustumi; Nelson Adami Andreollo; José Eduardo de Aguilar-Nascimento

FUTURE OF THE LANGUAGE MODELS IN HEALTHCARE: THE ROLE OF CHATGPT.

Futuro dos modelos de linguagem nos cuidados em saúde: o papel do ChatGPT

Francisco **TUSTUMI**[1], Nelson Adami **ANDREOLLO**[2], José Eduardo de **AGUILAR-NASCIMENTO**[3],

From [1]Department of Gastroenterology, Faculty of Medicine, Universidade de São Paulo (USP), São Paulo (SP), Brazil ; [2]Department of Surgery, Digestive Disease Surgical Unit, Faculty of Medical Sciences, Universidade Estadual de Campinas (Unicamp), Campinas (SP), Brazil ; [3]Full Professor, Department of Surgery, Medical School, Federal University of Mato Grosso (UFMT), Director of Medical School of the Várzea Grande University Center (UNIVAG), Cuiabá (MT), Brazil.

Francisco TUSTUMI: https://orcid.org/0000-0001-6695-0496

Nelson Adami ANDREOLLO: https://orcid.org/0000-0001-7452-1165

José Eduardo de AGUILAR-NASCIMENTO: https://orcid.org/0000-0002-3583-6612

Correspondence: Francisco Tustumi. Email: franciscotustumi@gmail.com.

ABSTRACT

The field of medicine has always been at the forefront of technological innovation, constantly seeking new strategies to diagnose, treat, and prevent diseases. Guidelines for clinical practice to orientate medical teams regarding diagnosis, treatment, and prevention measures have increased over the years. The purpose is to gather the most medical knowledge to construct an orientation for practice. Evidence-based guidelines follow several of the main characteristics of a systematic review, including systematic and unbiased search, selection, and extraction of the source of evidence. In recent years, the rapid advancement of artificial intelligence (AI) has provided clinicians and patients with access to personalized, data-driven insights, support and new opportunities for healthcare professionals to improve patient outcomes, increase efficiency, and reduce costs. One of the most exciting developments in AI has been the emergence of chatbots. A chatbot is a computer program to simulate conversation with human users. Recently, OpenAI, a research organization focused on machine learning, developed ChatGPT, a large language model that generates human-like text. ChatGPT uses a type of AI known as a deep learning model. ChatGPT can quickly search and select pieces of evidence through numerous databases to

provide answers to complex questions, reducing the time and effort required to research a particular topic manually. Consequently, language models can accelerate the creation of clinical practice guidelines. While there is no doubt that ChatGPT has the potential to revolutionize the way healthcare is delivered, it is essential to note that it should not be used as a substitute for human healthcare professionals. Instead, ChatGPT should be seen as a tool that can be used to augment and support the work of healthcare professionals, helping them to provide better care to their patients.

Keywords: Guidelines as Topic. Artificial Intelligency. Diagnosis. Costs and Cost Analysis. Delivery of Health Care.

RESUMO

A área da medicina sempre esteve na vanguarda da inovação tecnológica, buscando constantemente novas estratégias para diagnosticar, tratar e prevenir doenças. As diretrizes para a prática clínica são para orientar as equipes médicas quanto ao diagnóstico, tratamento e medidas de prevenção aumentaram ao longo dos anos. O objetivo é reunir o máximo de conhecimento médico para construir uma orientação para a prática. As diretrizes baseadas em evidências seguem várias das principais características de uma revisão sistemática, incluindo busca sistemática e imparcial, seleção e extração da fonte de evidência. Nos últimos anos, o rápido avanço da inteligência artificial (IA) forneceu aos médicos e pacientes acesso a informações personalizadas e baseadas em dados, suporte e novas oportunidades para os profissionais de saúde melhorarem os resultados dos pacientes, aumentarem a eficiência e reduzirem custos. Um dos desenvolvimentos mais empolgantes da IA foi o surgimento dos *chatbots*. Um *chatbot* é um programa de computador para simular conversas com usuários humanos. Recentemente, a OpenAI, uma organização de pesquisa focada em aprendizado de máquina, desenvolveu o ChatGPT, um grande modelo de linguagem que gera texto semelhante ao humano. O ChatGPT usa um tipo de IA conhecido como modelo de aprendizado profundo. O ChatGPT pode pesquisar e selecionar rapidamente evidências em vários bancos de dados para fornecer respostas a perguntas complexas, reduzindo o tempo e o esforço necessários para pesquisar um tópico específico

manualmente. Consequentemente, os modelos de linguagem podem acelerar a criação de diretrizes de prática clínica. Embora não haja dúvida de que o ChatGPT tem potencial para revolucionar a forma como os cuidados de saúde são prestados, é essencial observar que não deve ser usado como substituto de profissionais de saúde humanos. Em vez disso, o ChatGPT deve ser visto como uma ferramenta que pode ser usada para aumentar e apoiar o trabalho dos profissionais de saúde, ajudando-os a prestar melhores cuidados aos seus pacientes.

DESCRITORES: Guias como Assunto. Inteligência Artificial. Diagnóstico. Custos e Análise de Custos. Atenção à Saúde.
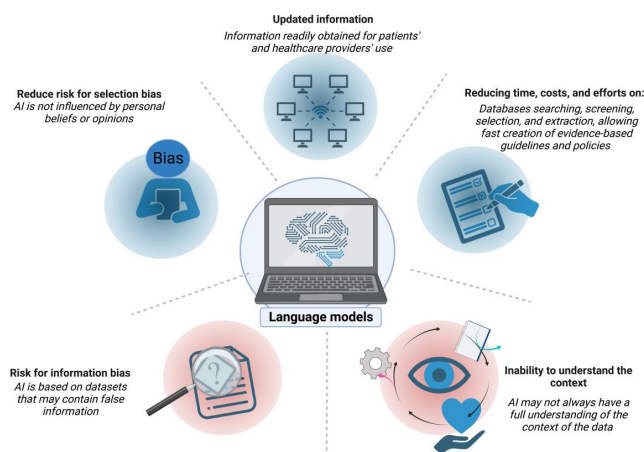
Image

Figure 1. Language models in medicine are promising and have the potential to significantly improve patient outcomes and increase the efficiency of healthcare delivery. However, it is crucial to consider the limitations of this technology carefully. AI: Artificial intelligence.

Central Message

In recent years, the rapid advancement of artificial intelligence (AI) has provided clinicians and patients with access to personalized, data-driven insights, support and new opportunities for healthcare professionals to improve patient outcomes, increase efficiency, and reduce costs. One of the most exciting developments in

AI has been the emergence of chatbots. A chatbot is a computer program to simulate conversation with human users.

Perspectives

The use of ChatGPT in medicine is a promising development that has the potential to significantly improve patient outcomes and increase the efficiency of healthcare delivery. However, it is important to consider this technology's limitations carefully and ensure that it is used responsibly and in conjunction with human healthcare professionals

INTRODUCTION

The field of medicine has always been at the forefront of technological innovation, constantly seeking new strategies to diagnose, treat, and prevent diseases. The number of publications in the main medical databases, such as PubMed and Embase, has been increasing steadily over the years due to the contributions from authors worldwide, the expansion of research areas, the rise of open-access publishing, and advancements in technology[2]. This growth in biomedical literature provides a wealth of information that can be used to advance research and improve patient care[7]. However, this fast science production in the medical field has boosted a new problem: how can clinical practice follow the constantly updated improvement of science?

Guidelines for clinical practice to orientate medical teams regarding diagnosis, treatment, and prevention measures have increased over the years[6]. The purpose is to gather the most medical knowledge to construct an orientation for practice. Guidelines help standardize care, improve patient outcomes, promote efficient use of resources, and reduce the risk of adverse events. In this sense, evidence-based reviews and guidelines are located at the top of the pyramid in the level of evidence (Figure 1)[14] and in the last stages of translational medicine (Figure 2)[18].

Evidence-based guidelines follow several of the main characteristics of a systematic review, including systematic and unbiased search, selection, and extraction of the source of evidence[21, 24]. However, there are numerous obstacles

to constructing evidence-based guidelines. Search, selection, and extraction take many efforts and a long time.[5] Consequently, evidence-based guidelines usually are published with a long-time delay. An evidenced-based guideline can take anywhere from several months to more than a year to complete, depending on the scope and complexity of the review. Conducting a comprehensive search for relevant studies can take several months, depending on the literature database's size and the search terms' complexity. Screening studies for inclusion based on pre-specified criteria can take another couple of months, depending on the number of studies identified and the number of reviewers involved. Updates of these guidelines may take years or even never be performed, and healthcare professionals may take their diagnosis, treatment, and prevention measures based on guidelines out of step with the current technological and scientific level.

Another relevant issue in evidence-based guidelines and systematic reviews refers to the risk of selection bias[10, 22, 23]. Some authors may make a poor selection of the pieces of evidence for supporting guidelines for some reasons. It can be unconscious due to quite restricted eligibility criteria, such as period, language, or databases searched, or even due to human failure during the process of search, selection, and extraction of the source of evidence[17]. Besides, poor selection can also be conscious of improper manipulation of outcomes, influenced by personal beliefs or opinions[8].

In this context, in recent years, the rapid advancement of artificial intelligence (AI) has provided new opportunities for healthcare professionals to improve patient outcomes, increase efficiency, and reduce costs[25]. In the medical field, AI has the potential to provide clinicians and patients with access to personalized, data-driven insights and support.

One of the most exciting developments in AI has been the emergence of chatbots, which have the potential to revolutionize the way healthcare is delivered[1]. A chatbot is a computer program to simulate conversation with human users. Recently, OpenAI, a research organization focused on machine learning, developed ChatGPT, a large language model that generates human-like text. The initial version of GPT (Generative Pre-trained Transformer) was first introduced in 2018. Since then, several improved versions of GPT have been released.

ChatGPT is a variant of the GPT models that have been specifically fine-tuned and optimized for chat-based applications[13].

ChatGPT uses a type of AI known as a deep learning model (Figure 3). Specifically, it is a type of transformer-based language model that uses a neural network architecture known as the Transformer. This architecture was first introduced in 2017 and has since become a popular choice for natural languages processing tasks, such as language translation, language generation, and text classification. The Transformer architecture uses multiple layers of neurons, or "transformer blocks," that allow it to process and extract features from input data hierarchically. This architecture has proven to be very effective for natural language processing tasks, and it forms the basis of many of the most advanced language models, including ChatGPT. ChatGPT is designed to understand and respond to text-based inputs from users to provide helpful and informative responses to their questions and comments[4, 11, 12].

There are different categories of chatbots and a chatbot can belong to more than one category: Knowledge Domain (generic, open damain and close domain); Service Provided (interpersonal, intrapersonal, inter-agent); Goals (informative, chat base / conversational, task based); Response Generation Method (rule based, retrieval based, generative); Human-Aid (human-mediated, autonomus); Permissions (open-source, commercial); Communication Channel (text, voice, image) [1]. Therefore, there are many applications of chatbots, like education environments, customer service, medicine and health, robotics, industrial and others [1].

The chatbots, AI and telemedicine are increasingly being used in healthcare services with good acceptance, such as education, diagnostic imaging and genetic diagnosis, as well as clinical laboratory, screening and health communications [9,15,25].

One of the key benefits of using ChatGPT in medicine is the ability to provide fast, accurate, and up-to-date information to healthcare professionals and patients. ChatGPT can quickly search and select pieces of evidence through numerous databases to provide answers to complex questions, reducing the time

and effort required to research a particular topic manually. Consequently, language models can accelerate the creation of clinical practice guidelines. AI may help screen numerous databases quickly, saving time and accelerating the finishing of the guidelines[11,20].

In addition to providing information, ChatGPT can also assist healthcare professionals with making diagnoses by analyzing symptoms and recommending tests or treatments. This can help to improve the accuracy of diagnoses and reduce the number of misdiagnoses. ChatGPT can also be used to help manage patients with chronic conditions by providing information about medications, lifestyle changes, and treatment options[3, 15].

Besides, the use of AI might reduce the risk of selection bias. The dataset used to train ChatGPT is typically chosen to be as inclusive as possible, including a broad range of sources and perspectives. ChatGPT is not influenced by individual cognitive biases, which can affect human decision-making. As an AI, ChatGPT uses a purely algorithmic approach to generate responses without being influenced by personal beliefs or opinions. This can reduce the risk of selection bias, as the responses are based solely on the data used to train the algorithm[16].

However, ChatGPT is trained on large amounts of data, which may contain biases or inaccuracies that the system could inadvertently propagate. This information bias could result in inaccurate or discriminatory recommendations or treatment options[15,16]. ChatGPT is trained based on web data, and several web information sources may be easily wrong. Besides, ChatGPT may not always have a full understanding of the context of the data in which a question or prompt is being asked. This could lead to inaccurate or inappropriate responses, particularly in complex medical situations or situations comprehending patients' preferences or feelings.

For evidence-based guidelines with quantitative synthesis (meta-analysis), despite AI potentially detecting statistical heterogeneity, clinical heterogeneity depends on the critical analysis of the included articles[19]. Without the human critical analysis capacity, language models may gather information in

a review based on articles whose data cannot be pooled due to their methodological differences!

While there is no doubt that ChatGPT has the potential to revolutionize the way healthcare is delivered, it is essential to note that it should not be used as a substitute for human healthcare professionals. Instead, ChatGPT should be seen as a tool that can be used to augment and support the work of healthcare professionals, helping them to provide better care to their patients (Figure 4).

CONCLUSIONS

The use of ChatGPT in medicine is a promising development that has the potential to significantly improve patient outcomes and increase the efficiency of healthcare delivery. However, it is important to consider this technology's limitations carefully and ensure that it is used responsibly and in conjunction with human healthcare professionals

REFERENCES

1. Adamopoulou E, Moussiades L. Chatbots: History, technology, and applications. Machine Learning with Applications. 2020;2:100006. https://doi.org/10.1016/j.mlwa.2020.100006.

2. AlRyalat SAS, Malkawi LW, Momani SM. Comparing Bibliometric Analysis Using PubMed, Scopus, and Web of Science Databases. J Vis Exp. 2019;(152). doi: 10.3791/58494.

3. Athota L, Shukla VK, Pandey N, Rana A. Chatbot for healthcare system using artificial intelligence. In 2020 8th International Conference on Reliability, Infocom Technologies and Optimization (trends and future directions)(ICRITO), 2020 Jun 4 (pp. 619-622). doi: 10.1109/ICRITO48877.2020.9197833.

4. Aydın, Ö., Karaarslan, E. (2022). OpenAI ChatGPT Generated Literature Review: Digital Twin in Healthcare . In Ö. Aydın (Ed.), Emerging Computer Technologies 2 (pp. 22-31). Available at: https://ssrn.com/abstract=4308687.

5. Boland A, Dickson R, Cherry G. Doing a systematic review: A student's guide. Doing a Systematic Review. 2017:1-304. Available at: http://digital.casalini.it/9781526416582.

6. Brouwers MC, Florez ID, McNair SA, Vella ET, Yao X. Clinical practice guidelines: tools to support high quality patient care Semin Nucl Med.2019; 49:145-152. https://doi.org/10.1053/j.semnuclmed.2018.11.001.

7. Engle RL, Mohr DC, Holmes SK, Seibert MN, Afable M, Leyson J, Meterko M. Evidence-based practice and patient-centered care: Doing both well. Health Care Manage Rev. 2021;46(3):174-184. doi: 10.1097/HMR.0000000000000254.

8. Hernán MA, Hernández-Díaz S, Robins JM. A structural approach to selection bias. Epidemiology. 2004;15(5):615-25. doi: 10.1097/01.ede.0000135174.63482.43.

9. Isolan G, Malafaia O. How does telemedicine fit into healthcare today? Arq Bras Cir Dig. 2022;34(3):e1584. doi: 10.1590/0102-672020210003e1584.

10. Kho ME, Duffett M, Willison DJ, Cook DJ, Brouwers MC. Written informed consent and selection bias in observational studies using medical records: systematic review. BMJ. 2009;338:b866. doi: 10.1136/bmj.b866.

11. Kung TH, Cheatham M, Medenilla A, Sillos C, De Leon L, Elepaño C, et al. Performance of ChatGPT on USMLE: Potential for AI-assisted medical education using large language models. PLOS Digital Health. 2023;2(2):e0000198. https://doi.org/10.1371/journal.pdig.0000198.

12. Li N, Liu S, Liu Y, Zhao S, Liu M. Neural speech synthesis with transformer network. In: Proceedings of the AAAI conference on artificial intelligence, 2019;33(1);6706-6713. https://doi.org/10.1609/aaai.v33i01.33016706.

13. Lund BD, Wang T. Chatting about ChatGPT: how may AI and GPT impact academia and libraries?. Library Hi Tech News. 2023. https://doi.org/10.1108/LHTN-01-2023-0009.

14. Murad MH, Asi N, Alsawas M, Alahdab F. New evidence pyramid. Evid Based Med. 2016;21(4):125-7. doi: 10.1136/ebmed-2016-110401.

15. Nadarzynski T, Miles O, Cowie A, Ridge D. Acceptability of artificial intelligence (AI)-led chatbot services in healthcare: A mixed-methods

study. Digit Health. 2019;5:2055207619871808. doi: 10.1177/2055207619871808.

16. Ntoutsi E, Fafalios P, Gadiraju U, Iosifidis V, Nejdl W, Vidal ME, et al. Bias in data-driven artificial intelligence systems—An introductory survey. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery. 2020;10(3):e1356. https://doi.org/10.1002/widm.1356.

17. Rother ET. Systematic literature review X narrative review. Acta Paul Enferm. 2007;20:v-i. Available from: https://doi.org/10.1590/S0103-21002007000200001.

18. Seyhan AA. Lost in translation: the valley of death across preclinical and clinical divide–identification of problems and overcoming obstacles. Transl Med Commun.2019; **4**:18. https://doi.org/10.1186/s41231-019-0050-7

19. Thompson SG. Why sources of heterogeneity in meta-analysis should be investigated. BMJ. 1994;309(6965):1351-5. doi: 10.1136/bmj.309.6965.1351.

20. Thorp HH. ChatGPT is fun, but not an author. Science. 2023 ;379(6630):313. doi: 10.1126/science.adg7879.

21. Urra Medina E, Barría Pailaquilén RM. Systematic review and its relationship with evidence-based practice in health. Rev Latino-am Enfermagem. 2010;18:824-31. https://doi.org/10.1590/S0104-11692010000400023.

22. Williamson PR, Gamble C, Altman DG, Hutton JL. Outcome selection bias in meta-analysis. Stat Methods Med Res. 2005;14(5):515-24. doi: 10.1191/0962280205sm415oa.

23. Williamson PR, Gamble C. Identification and impact of outcome selection bias in meta-analysis. Stat Med. 2005;24(10):1547-61. doi: 10.1002/sim.2025.

24. Yao X, Vella ET, Sussman J. More Thoughts than Answers: What Distinguishes Evidence-Based Clinical Practice Guidelines from Non-evidence-Based Clinical Practice Guidelines? J Gen Intern Med. 2021;36(1):207-208. doi: 10.1007/s11606-020-05825-y.

25. Yu KH, Beam AL, Kohane IS. Artificial intelligence in healthcare. Nat Biomed Eng. 2018;2(10):719-731. doi: 10.1038/s41551-018-0305-z.
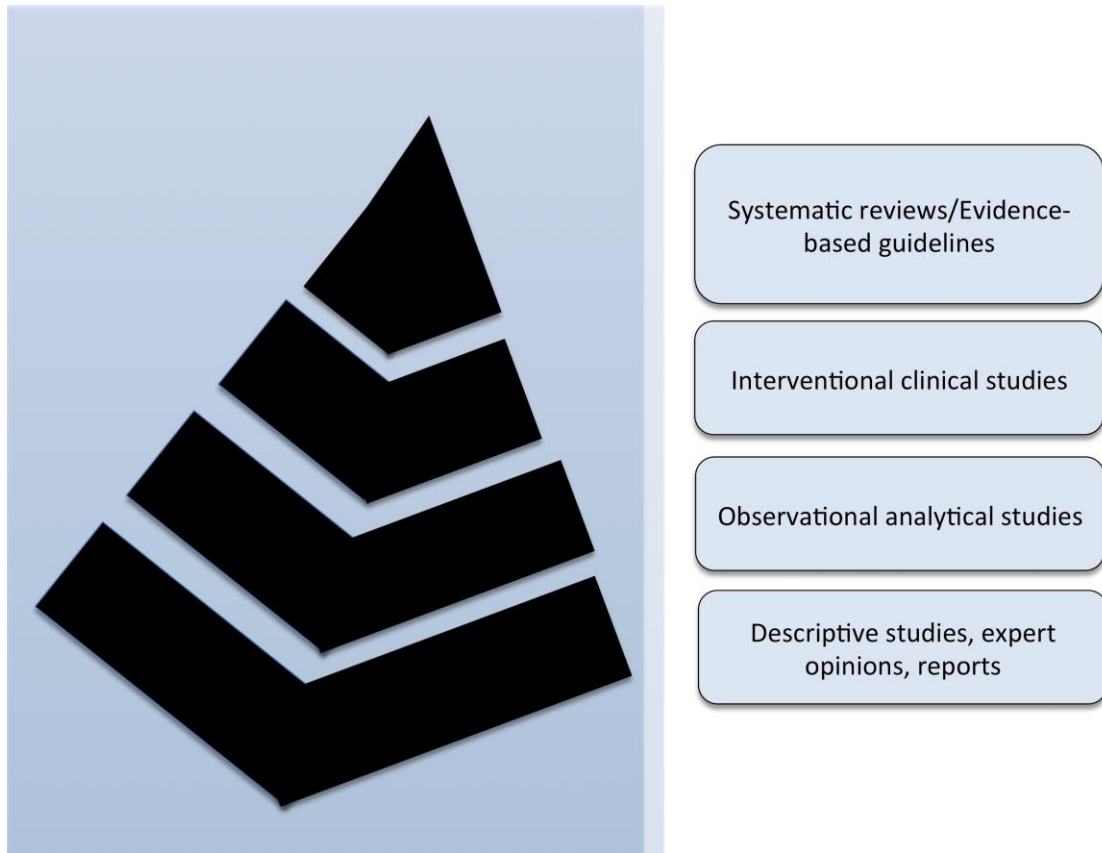
Figure 2. Pyramid of the level of evidence. Systematic reviews and evidence-based guidelines are usually considered the top evidence, gathering all the current evidence for clinical practice[14].
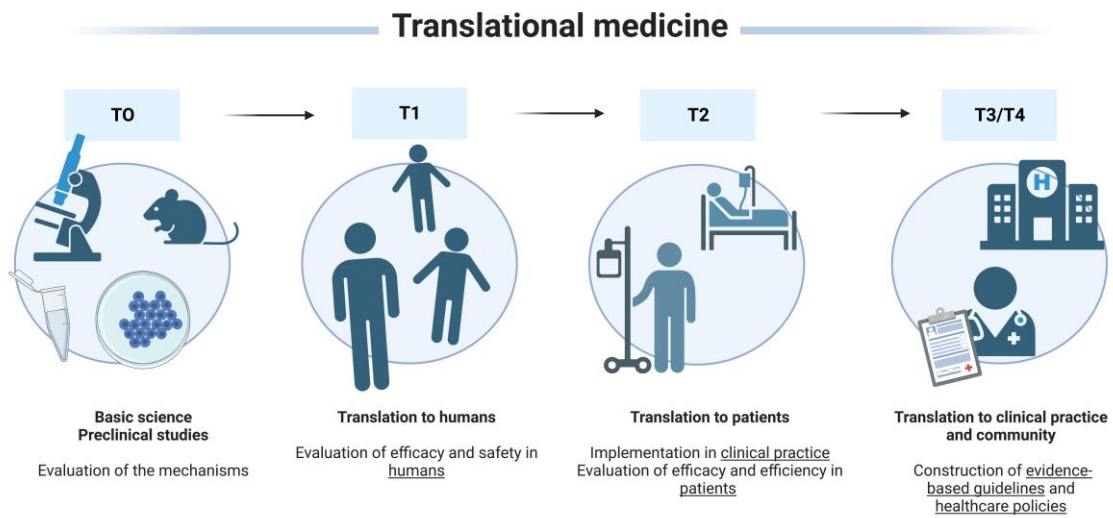
Figure 3. Translational medicine stages. Evidence-based guidelines and healthcare policies are the endpoints of any research line. These guidelines and policies orientate healthcare professionals for patient management[18].
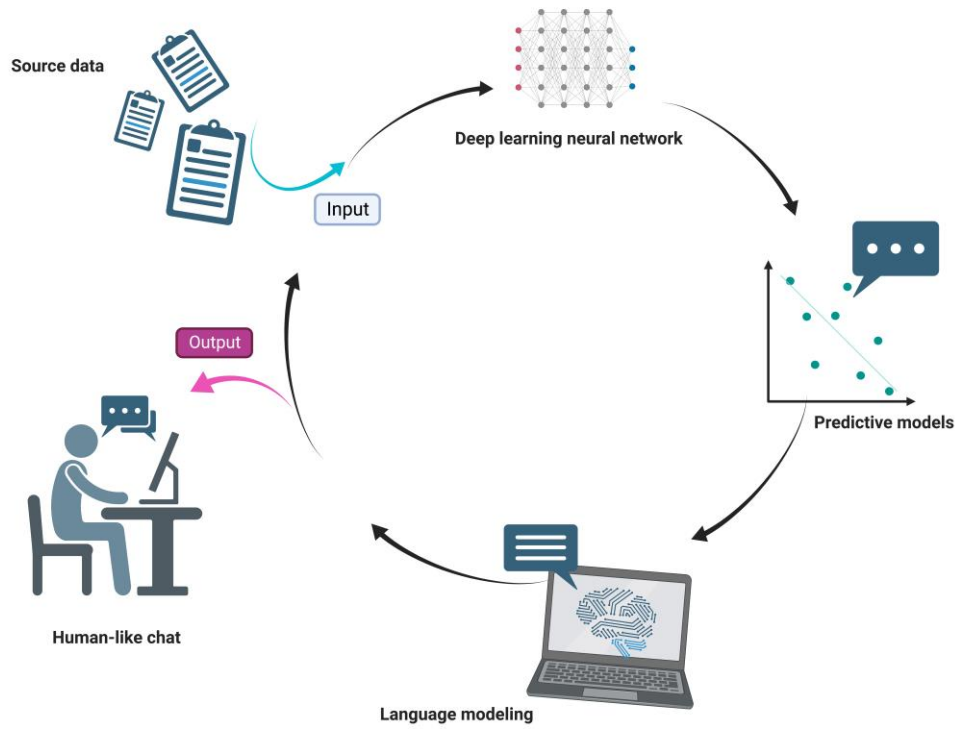
Figure 4. Modern language models use artificial intelligence to create a human-like text.
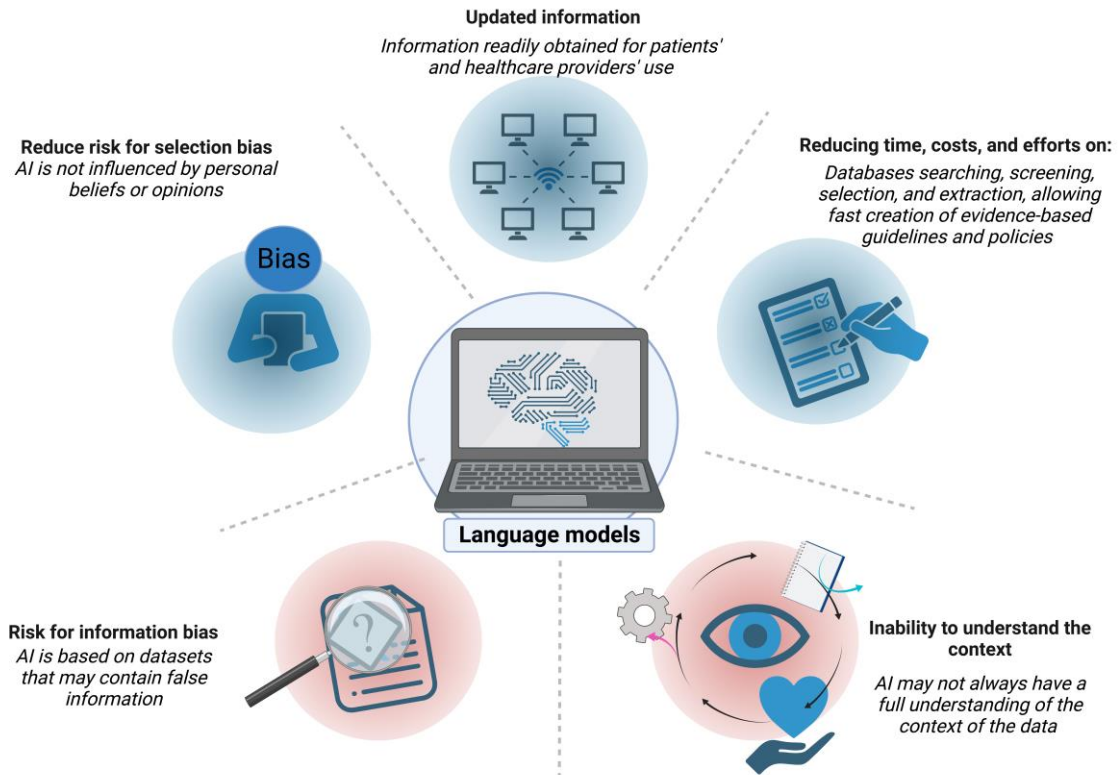
Figure 5. Language models in medicine are promising and have the potential to significantly improve patient outcomes and increase the efficiency of healthcare delivery. However, it is crucial to consider the limitations of this technology carefully. AI: Artificial intelligence.

This preprint was submitted under the following conditions: