


# Large Language Models in Gastroenterology and Gastrointestinal Surgery: A New Frontier in Patient Communication and Education

Dushyant Singh Dahiya<sup>a,1</sup> , Hassam Ali<sup>b</sup>, Vishali Moond<sup>c</sup>, M. Danial Ali Shah<sup>d</sup>, Christina Santana<sup>e</sup>, Noor Ali<sup>f</sup>, Abu Baker Sheikh<sup>g</sup>, Muhammad Ahmad Nadeem<sup>h</sup>, Aqsa Munir<sup>i</sup>, Mohammed A. Quazi<sup>g</sup>, Hareesha Rishab Bharadwaj<sup>j</sup>, Amir Humza Sohail<sup>k</sup>

## Abstract

When integrated into healthcare, large language models (LLMs) have transformative and revolutionary effects, including significant potential for improving patient care and streamlining clinical processes. However, one specialty that particularly requires data on LLM use is gastroenterology and gastrointestinal surgery, a gap we sought to address in our research. Advanced artificial intelligence (AI) systems like LLMs have demonstrated the ability to mimic human communication, assist in diagnosis, provide patient education, and support medical research simultaneously. Despite these advantages, challenges such as biases, data privacy concerns, and lack of transparency in decision-making remain critical. The role of regulations in mitigating these risks is widely debated, with proponents advocating for

structured oversight to enhance trust and patient safety, while others caution against potential barriers to innovation. Rather than replacing human expertise, AI should be integrated thoughtfully to complement clinical decision-making. Ensuring a balanced approach requires collaboration between medical professionals, AI developers, and policymakers to optimize its responsible implementation in healthcare.

**Keywords:** Artificial intelligence; Gastroenterology; Gastroenterology surgery; Healthcare; Ethical medical care

## Introduction

Large language models (LLMs) are a complex form of artificial intelligence (AI) trained on an extensive corpus of text data, often encompassing terabytes of information. These models are designed to generate texts that closely mimic human language and have the ability to understand natural language queries with a high degree of accuracy [1]. Features that particularly differentiate LLMs from other forms of AI include their ability to identify and grasp context-specific pointers, infer the underlying meaning and implications, and generate coherent and context-appropriate responses. This makes LLMs invaluable across a myriad of applications in business and healthcare, including, but not limited to, customer service chatbots, virtual personal assistants, and advanced language translation systems.

While AI has been a subject of research since the 1950s, interest in its medical applications has fluctuated over time [1]. The recent surge in AI interest, particularly within healthcare, has been driven by advancements in LLMs over the past few years [2]. This renewed focus is largely due to improvements in deep learning, increased computational power, and the development of sophisticated models like Generative Pre-trained Transformer (GPT)-3 and GPT-4, which have demonstrated unprecedented capabilities in clinical decision support, patient education, and research [3]. These models are not just academic possibilities or theoretical models but are transforming the face and landscape of healthcare globally. LLMs possess the computational power to rapidly analyze vast repositories

Manuscript submitted December 20, 2024, accepted March 12, 2025  
Published online March 24, 2025

<sup>a</sup>Division of Gastroenterology, Hepatology and Motility, The University of Kansas School of Medicine, Kansas City, KS, USA

<sup>b</sup>Department of Gastroenterology, Division of Internal Medicine, ECU Health Medical Center/Brody School of Medicine, Greenville, NC, USA

<sup>c</sup>Department of Internal Medicine, Saint Peter's University Hospital/Robert Wood Johnson Medical School, New Brunswick, NJ, USA

<sup>d</sup>Department of Internal Medicine, King Edward Medical University, Lahore, Pakistan

<sup>e</sup>Department of Internal Medicine, Division of Internal Medicine, ECU Health Medical Center/Brody School of Medicine, Greenville, NC, USA

<sup>f</sup>Department of Public Health, East Carolina University, Greenville, NC, USA

<sup>g</sup>Department of Internal Medicine, University of New Mexico Health Sciences, Albuquerque, NM, USA

<sup>h</sup>Department of Hepato-Pancreato-Biliary and Liver Transplant Surgery, Digestive Diseases and Surgery Institute, Cleveland Clinic Foundation, Cleveland, OH, USA

<sup>i</sup>Department of Internal Medicine, Dow Medical College, Karachi, Pakistan

<sup>j</sup>Faculty of Biology, Medicine and Health, The University of Manchester, Manchester, UK

<sup>k</sup>Department of Surgery, University of New Mexico Health Sciences, Albuquerque, NM, USA

<sup>1</sup>Corresponding Author: Dushyant Singh Dahiya, Division of Gastroenterology, Hepatology and Motility, The University of Kansas School of Medicine, Kansas City, KS 66160, USA. Email: dush.dahiya@gmail.com

doi: <https://doi.org/10.14740/gr2011>

of medical literature, scientific publications, clinical studies, patient education material and other online data. This enables them to assist healthcare professionals in diagnosing a wide array of diseases with greater accuracy, suggesting evidence-informed management plans, and predict patient outcomes and prognosis based on historical data and patient symptomatology and workup [4]. Thus, LLMs play a vital role in enhancing the quality of medical decision-making processes, ultimately contributing to superior patient care and thus clinical outcomes.

Interestingly, LLMs have shown considerable promise in the specialized fields of gastrointestinal pathology, offering a wide array of applications. Their proficiency in handling various gastrointestinal issues is demonstrated most effectively by their ability to address common gastrointestinal inquiries [5]. With versatile linguistic capabilities, they are able to comprehend and respond to questions in a host of languages, whether that is answering questions about cirrhosis in Arabic or addressing patient inquiries regarding upcoming procedures in their native language [6, 7]. However, there is significant room for improvement; for instance, while being able to respond accurately to queries related to hepatocellular carcinoma and cirrhosis, LLMs have demonstrated significant gaps in their understanding of current management guidelines [8]. This deficiency could preclude LLMs from providing comprehensive and fully accurate information to patients and their providers. Therefore, further research and development are crucial to strengthening their capabilities, especially in the highly specialized subsets of medicine, such as gastroenterology and gastrointestinal surgery.

While LLMs have demonstrated significant potential across various areas, their application in medicine brings unique challenges and ethical considerations, including the risk of introducing errors, biases, and human prejudices in medical practice. This problem arises since these models are trained on unstructured free text from diverse sources, including scientific literature, online forums, books, and publicly available internet data, which may contain inaccuracies, biases, and inconsistencies [9]. This can result in the dissemination of inaccurate or even harmful medical advice and treatment recommendations. Another domain in need of major improvement is the “black-box” nature of these AI systems, which limits transparency and makes it difficult to fully understand the rationale behind their decision-making processes. However, it is important to recognize that human decision-making is similarly complex; many human decisions are influenced by subconscious biases, heuristics, and prior experiences, which are not always explicitly understood or easily explained. While enhancing explainability in AI is a priority, there are tradeoffs: requiring models to generate detailed explanations can decrease computational efficiency, increase processing time, and, in some cases, reduce predictive performance. The balance between explainability, accuracy, and efficiency is context-dependent, while transparency is critical in high-stakes medical applications, there are scenarios where prioritizing accuracy and speed may be more beneficial [10].

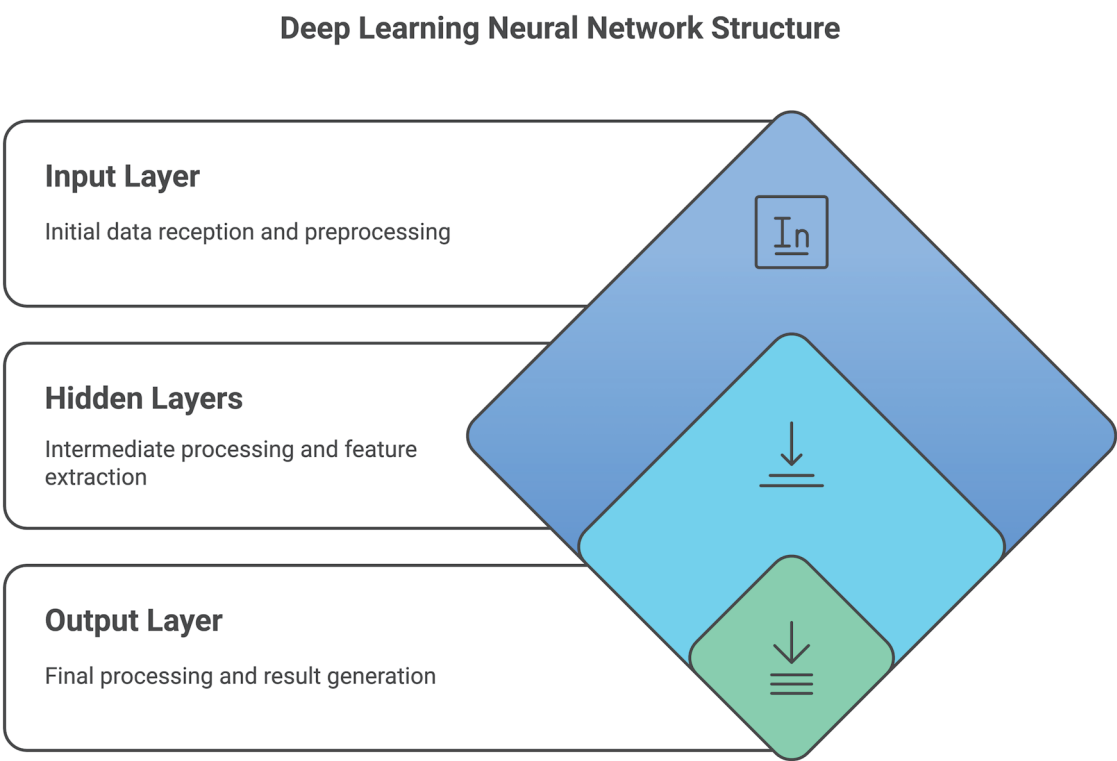
Furthermore, the utilization of LLMs in healthcare settings raises legitimate concerns surrounding patient privacy, confidentiality of sensitive patient information, and data security, since these models often require access to large vol-

umes of sensitive patient data to function effectively [11]. The increasing integration of AI in healthcare is transforming the role of medical professionals, automating certain tasks while enhancing other advancements, such as X-rays and computed tomography (CT) scans, did not diminish medical care but rather improved diagnostic accuracy and efficiency. Similarly, AI may reduce reliance on some traditional clinical practices, such as auscultation, while strengthening physicians’ ability to make data-driven diagnoses, optimize treatment strategies, and improve workflow efficiency. The key challenge lies in balancing AI integration with maintaining the human aspects of patient care, ensuring that automation enhances rather than replaces meaningful physician-patient interactions. In light of these concerns, it is imperative that effective quality-control mechanisms and error mitigation processes be integrated within LLMs. Self-reflective LLMs improve reliability by identifying errors, learning from feedback, and refining responses. However, HIPAA restricts real-time learning from clinical interactions, limiting AI’s ability to adapt dynamically. While this protects patient privacy, it also slows AI improvement, requiring a balance between adaptability and confidentiality in medical AI development [12]. Recent studies have thus explored the potential of LLMs in being able to engage in such reflective processes, aiming to minimize the incidence of misinformation and bias [12, 13]. By incorporating complex self-correction algorithms, these advanced models can limit the propagation of errors in clinical decision support systems [13]. Furthermore, the implementation of transparent AI frameworks can contribute to demystifying the “black-box” issue, allowing healthcare professionals to understand and trust the AI’s reasoning. This transparency, besides building trust, also enables clinicians to provide input that can guide the AI towards more accurate and ethically sound decisions.

## How LLMs Work

LLMs such as GPT-4 are a phenomenon of modern technology designed to understand and generate human-like text. At the core of these models lies a computational framework known as “deep learning,” a specialized form of AI. The human brain’s neural networks inspired the concept of deep learning, aiming to mimic how humans learn from experience and data (Fig. 1) [1].

Inside LLMs, such as GPT-4, an intricate structure, known as a “neural network,” consists of layers of interconnected nodes, or “neurons.” These layers are categorized into three main types: the input layer, which receives the initial data; the hidden layers, which process this data; and the output layer, which produces the final result (Fig. 2). Before LLMs can generate text or answer questions, they must be trained on a comprehensive and detailed datasets [1]. Thus, LLMs are repeatedly exposed to colossal amounts of text data, from academic manuscripts, book chapters, fiction, online data, such as websites and social media posts. This diversity of information allows the model to comprehend the nuances, details, idioms, and structures inherent to human language. Therefore, utilizing sophisticated optimization algorithms, the model learns

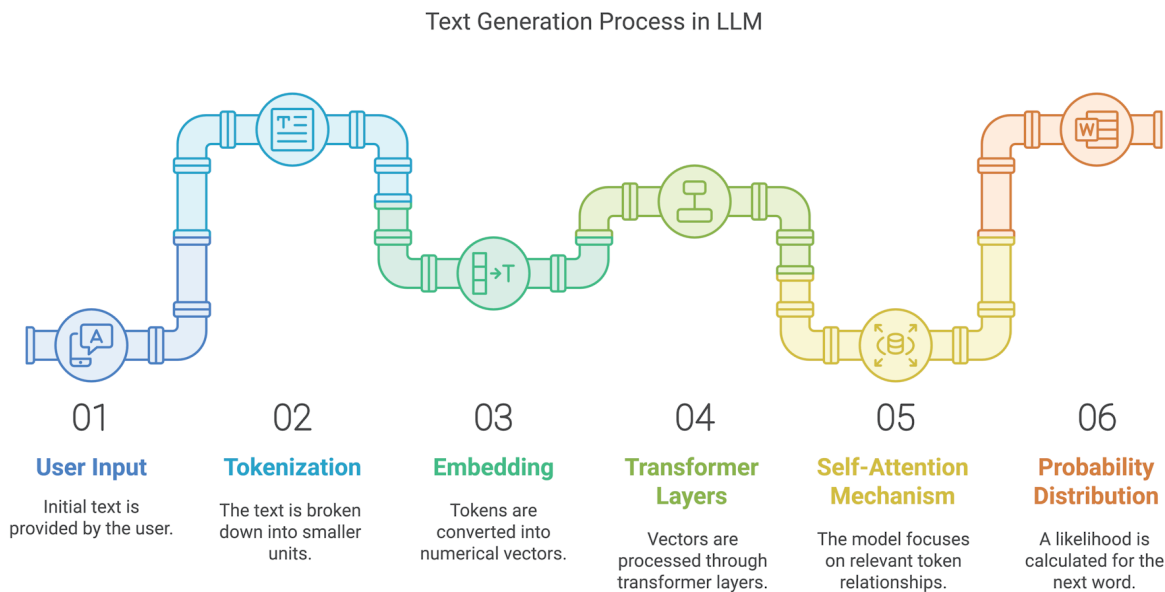


**Figure 1.** A deep learning neural network structure used in large language models (LLMs), including input, hidden, and output layers.

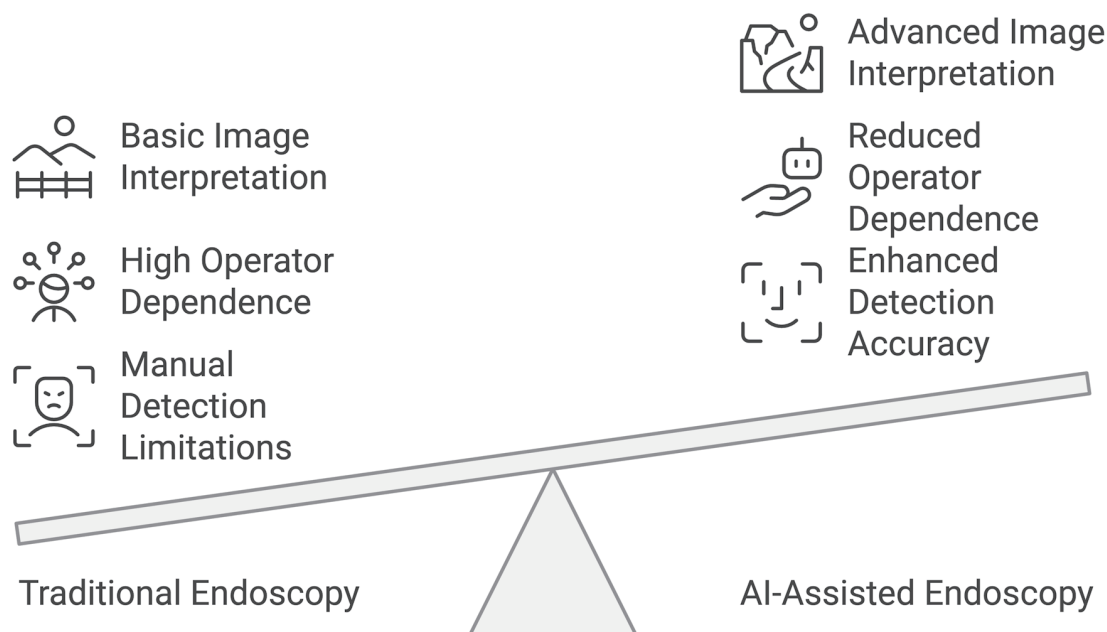
and re-learns to adjust internal parameters (often numbering in billions) to better predict the next word in a given sequence of words. This concept is known as “word prediction.” For instance, if the input is “The cat sat on the,” the model might predict the next word to be “mat,” based on its learning patterns

and previous exposure (Fig. 2).

After initial training, these models can be further specialized for particular tasks or domains, such as healthcare; this is known as “fine-tuning.” During fine-tuning, the model is trained on a narrower dataset that contains domain-specific in-



**Figure 2.** The process of text generation in large language models (LLMs). User input is tokenized, processed through transformer layers, and refined into a probabilistic text output.



### AI enhances endoscopy precision and efficiency.

**Figure 3.** AI-assisted endoscopy compared to traditional endoscopy. AI: artificial intelligence.

formation, in this case: medical textbooks and research papers and abstracts. This allows the LLMs to generate more accurate and relevant responses when queried about topics in that specific field [1]. Another distinguishing feature is their contextual understanding; as opposed to simpler models, LLMs are not focused solely on individual words but can consider the surrounding words to generate coherent and contextually appropriate responses. For example, the word “bank” could refer to a financial institution or the side of a river. The model uses the context of the surrounding words to decide which meaning is appropriate.

### Current Studies and Data on the Use of LLMs in Gastroenterology

Although the research on the application of LLMs in gastroenterology is still in its nascent stages, there is a growing interest in exploring its massive potential [2]. Recent studies have focused on the effectiveness of LLMs in providing information to patients, assisting in symptom assessment, and even aiding healthcare professionals in decision-making [14]. There is significant potential in these models to revolutionize diagnostic and therapeutic strategies in the field, which is evident by the ongoing research and projects.

Machine learning models are being developed to predict disease flares in conditions such as inflammatory bowel disease, aiding in early intervention and customized treatment plans [15]. By analyzing a patient’s medical history, genetic

data, and other relevant factors, LLMs can be vital in devising patient-specific treatment regimens and optimizing therapeutic outcomes. AI models have shown promising results in classifying and differentiating endoscopic diagnosis hence guiding clinical decisions (Fig. 3) [16].

### Current Studies and Data on the Use of LLMs in Gastroenterology Surgery

There is a paucity of data regarding the utilization of LLMs in the realm of gastrointestinal surgery. This gap highlights the need for further research in this domain. Nevertheless, it is noteworthy that AI technology, in general, has made considerable advancements in the field of surgery, finding successful applications across various surgical procedures [17]. The integration of ChatGPT has led to significant breakthroughs in the field. Moreover, leveraging ChatGPT’s capability to analyze extensive medical databases can enhance the accuracy of diagnoses, aiding in the identification of rare conditions and suggesting pertinent investigations. Additionally, ChatGPT can contribute to streamlining and enhancing surgical planning by developing personalized preoperative plans, ensuring efficiency and safety. Postoperative care and rehabilitation can also be fortified by delivering tailored recovery guidance and monitoring patients through ongoing communication and other enhancements [18]. Recently, the utilization of AI-powered ChatGPT has seen a rising trend in various applications within surgery. These include contributing to educational initiatives

for the patients, providing medical care suggestions, and conducting individual case analyses in the context of surgery planning [19]. These can be extrapolated to gastrointestinal surgery as well, but it remains an area of further exploration.

## General Avenues for Incorporation of LLMs Into Healthcare

LLMs could be indispensable in assisting physicians with provision of information on rare diseases, the latest research findings, or treatment guidelines [20]. Less influenced by outside factors, LLMs can also help prevent the cognitive errors sometimes seen in medicine, such as anchoring bias or availability bias. Their ability to generate a broad list of differential diagnoses based on the information gathered by a clinician helps ensure patients are accurately assessed, and rare diagnoses are not missed. LLMs' application in gastroenterology and gastrointestinal surgery, and healthcare in general, can extend to creation of virtual assistants to support patients in healthcare management by generating concise summaries of patient encounters and medical histories optimizing recordkeeping for healthcare providers.

### Documentation and data gathering

Thus, LLMs offer significant potential to enhance the efficiency and accuracy of clinical processes, especially in interpreting clinical notes. With the expansive amount of data in electronic health records (EHRs), sifting through such extensive information to extract pertinent details can be time-consuming for healthcare professionals. LLMs can be trained to process and understand the intricacies of clinical documentation, identify key patterns, and extract vital information. As a result, they can generate concise summaries from voluminous clinical notes, highlighting the most relevant and critical details [21]. This capability ensures that physicians can rapidly access a synthesized view of a patient's medical history, including previous diagnoses, treatments, laboratory and imaging results, and other pertinent data. The ability of LLMs to rapidly review and summarize vast amounts of medical literature facilitates evidence-based clinical decision-making. If integrated with EHRs, these AI-powered systems can alert physicians about potential drug interactions, recommend relevant guidelines, or suggest diagnostic tests based on the patient's symptoms and history (Fig. 4) [22]. As clinical documentation continues to become more onerous for those practicing, LLMs could provide a solution to help lessen the tedious portions. Their ability to manage data contextually would allow them to synthesize a focused review of a patient's chart or listen to a practitioner's observations and summarize the critical aspects of an encounter, such as the details surrounding the symptoms described, working diagnoses, and possible interventions. By automating certain aspects of the documentation process, clinicians would experience a reduced cognitive burden, allowing them to dedicate more time and attention to their patients [23]. Furthermore,

by rapidly reviewing and summarizing EHRs, LLMs can assist in identifying potential risk factors, ensuring that no crucial details are overlooked. This can be particularly beneficial when faced with complex cases where multiple factors are at play or when a comprehensive overview of a patient's history is required to determine the best course of action.

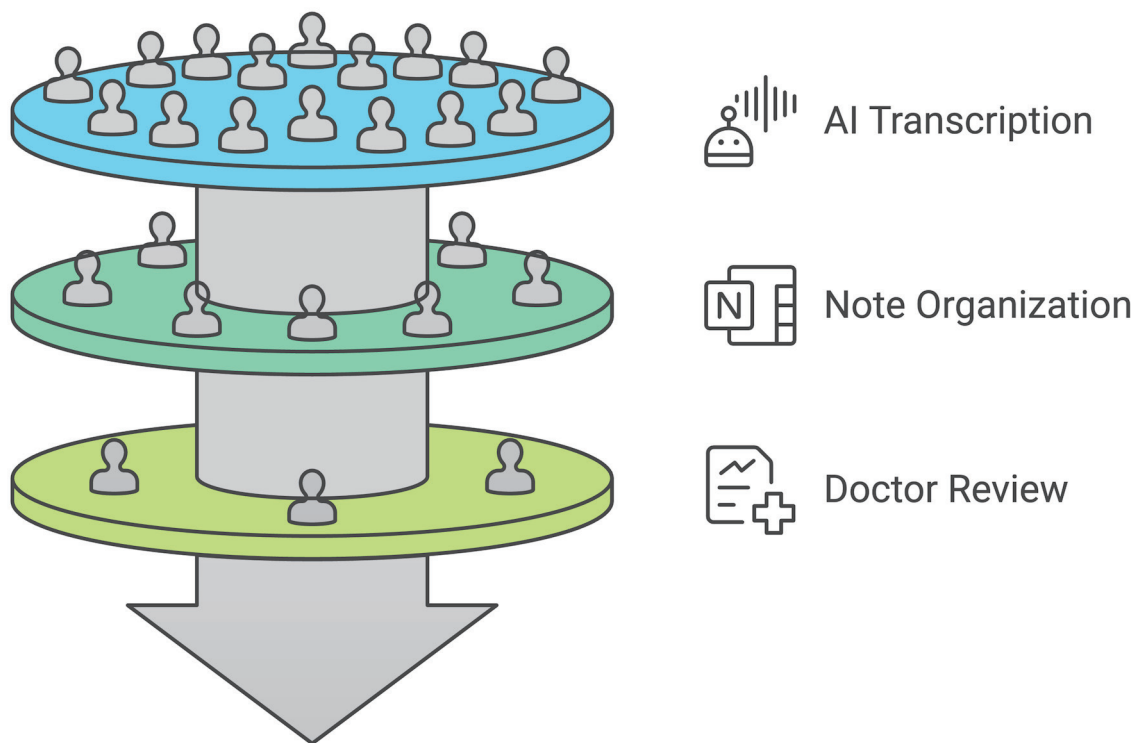
### Patient education and communication

LLMs, with their text-simplifying potential, can enhance communication between gastrointestinal physicians and their patient population [24]. Unlike human experts who are bound by time restrictions, LLMs are available at all hours, making them more accessible and allowing for the facilitation of more relaxed interactions [25]. This feature would be of particular interest to those who perhaps do not seek medical attention for conditions they feel are associated with a degree of social stigma.

LLMs have the potential to enhance patient education and communication in gastroenterology and gastrointestinal surgery by simplifying complex medical concepts into more accessible language. Studies have shown that AI-driven text simplification can improve patient comprehension of medical information [25]. Additionally, advanced techniques such as structured prompt interrogation enable AI to extract and present relevant medical knowledge in a user-friendly format [26]. Furthermore, research comparing AI-generated responses to patient queries with those provided by physicians found that AI responses were often rated higher for completeness and clarity, though human oversight remains essential to ensure accuracy and avoid misinformation [27]. These findings suggest that while LLMs can play a valuable role in patient education, their implementation should be carefully monitored to maintain reliability and trust. This can empower patients to make informed decisions about their health and better understand their conditions and treatment options. Similarly, these models can be utilized by medical students and residents as educational tools to enhance learning in gastroenterology. For example, LLMs can generate clinical case simulations for differential diagnosis, assist in interpreting endoscopic images, provide step-by-step guidance for procedures like colonoscopy and endoscopic retrograde cholangiopancreatography (ERCP), and aid in board exam preparation by summarizing guidelines and generating practice questions. These applications support interactive learning, improve diagnostic reasoning, and reinforce procedural knowledge within the specialty [28].

One of the strengths of LLMs is their potential to offer individualized care by tailoring information and advice to each patient's unique circumstances. LLMs could serve as a preliminary point of contact for patients with gastrointestinal complaints, offering basic information and/or suggesting when they should seek medical attention [28]. AI-powered chatbots and applications can inform patients about their conditions, treatment options, the importance of compliance with medications/dietary restrictions, and even provide reminders on when to take their medication and upcoming appointments. This per-

## AI-Enhanced EHR Documentation



**Figure 4.** AI's role in automating medical documentation. AI listens to physician-patient interactions, transcribes key details, structures note in EHRs, and assists in clinical decision-making. AI: artificial intelligence; EHRs: electronic health records.

sonalization has the potential to enhance patient engagement and adherence to treatment plans.

### Administrative work

For administrative staff in gastroenterology or gastrointestinal surgery practices, LLMs may be of value in scheduling, billing queries, and educating patients with information regarding the pre-procedural preparation required prior to undergoing endoscopic or surgical interventions. LLMs can also serve as an initial screening tool at telemedicine appointments, gathering basic patient information and symptoms before a virtual consultation with a physician. LLMs can provide patients with post-procedural instructions, answer common questions, and guide them on when to seek help for potential complications [27].

### Research

The use of LLMs is not limited to interventional fields, but across various areas of medical research. One promising avenue for future research is integrating LLMs with other natural language processing tools, such as topic modeling, to identify relevant research domains and formulate more concise

and targeted research queries. [29] Further exploration might also delve into LLMs' utility in the sub-disciplines within gastroenterology i.e., hepatology and pancreatic surgery. LLMs can process vast datasets to identify patterns, correlations, or trends in gastrointestinal disease prevalence, incidence, and outcomes, offering insights into disease epidemiology.

LLMs not only have the potential to enhance the health-care field but also to help further it. By analyzing data from existing scientific publications, they are able to provide insight into pertinent issues within specific research sectors, facilitate drug development, and contribute to advancing scientific exploration [30]. LLMs can assist in summarizing vast amounts of research data, providing insights, and suggesting relevant references [13]. LLMs can be a resource to help summarize recent research summaries and clinical trial outcomes and provide updates to existing guidelines in gastroenterology and gastrointestinal surgery. Researchers in clinical medicine can use these models for quick literature searches, to brainstorm research ideas, or to aid in statistical analysis [13]. LLMs have the potential to function as direct-response search engines, synthesizing information and providing immediate answers rather than simply directing users to external sources. This can enhance efficiency in medical decision-making and patient education. However, unlike traditional search engines that present multiple sources for validation, LLMs generate responses

based on pre-existing data, which may lack real-time updates and explicit citations. While this improves accessibility, it also underscores the need for careful verification of AI-generated medical information to ensure accuracy and reliability. This could simplify the research paper writing process by reducing the tedious task of article hunting and selection. Consequently, researchers can dedicate more time to their core research and its methodology. LLM's ability to sift through extensive patient data could extend to identifying suitable participants for clinical trials. In addition to selection, LLMs can use real patient data to generate simulated control patients that reflect attributes of the broader population to make clinical trials more efficient and less costly [13].

## Human Oversight

In terms of research, LLMs draw from expansive textual databases and can synthesize information by identifying patterns and intersections across multiple sources. This ability enhances knowledge accessibility and may generate novel insights through contextualization. However, unlike human researchers, LLMs do not independently form hypotheses or conduct experimental validation, meaning their outputs, while informative, rely entirely on pre-existing data rather than original discovery [31]. Unlike humans, who derive insights from personal experiences, critical thinking, and experimentation, LLMs generate responses by predicting text based on statistical patterns in pre-existing data. While they can identify correlations and synthesize information, they lack independent reasoning, intentionality, and the ability to truly comprehend or innovate beyond their training data. Consequently, their outputs, though informative, are extrapolations rather than original insights.

While LLMs can offer valuable assistance, human oversight remains indispensable in healthcare. Despite the immense potential of LLMs in aiding healthcare professionals, they should be leveraged as instruments to complement, not supplant, human expertise. It is imperative to deeply comprehend the dependability, repeatability, and stability of decisions made by these models, along with their performance indicators, both context-independent and dependent.

Ensuring the accuracy and reliability of information provided by LLMs is paramount. There is a need for ongoing validation and calibration to guarantee that the medical advice and information given aligns with current medical standards by healthcare professionals. In order to maintain reliability, LLMs must be updated regularly to ensure the information provided is based on the latest research and most up-to-date practice guidelines. Additionally, LLMs should be programmed to recognize when a situation requires human intervention, ensuring that patients receive the appropriate level of care.

## Ethical Implications

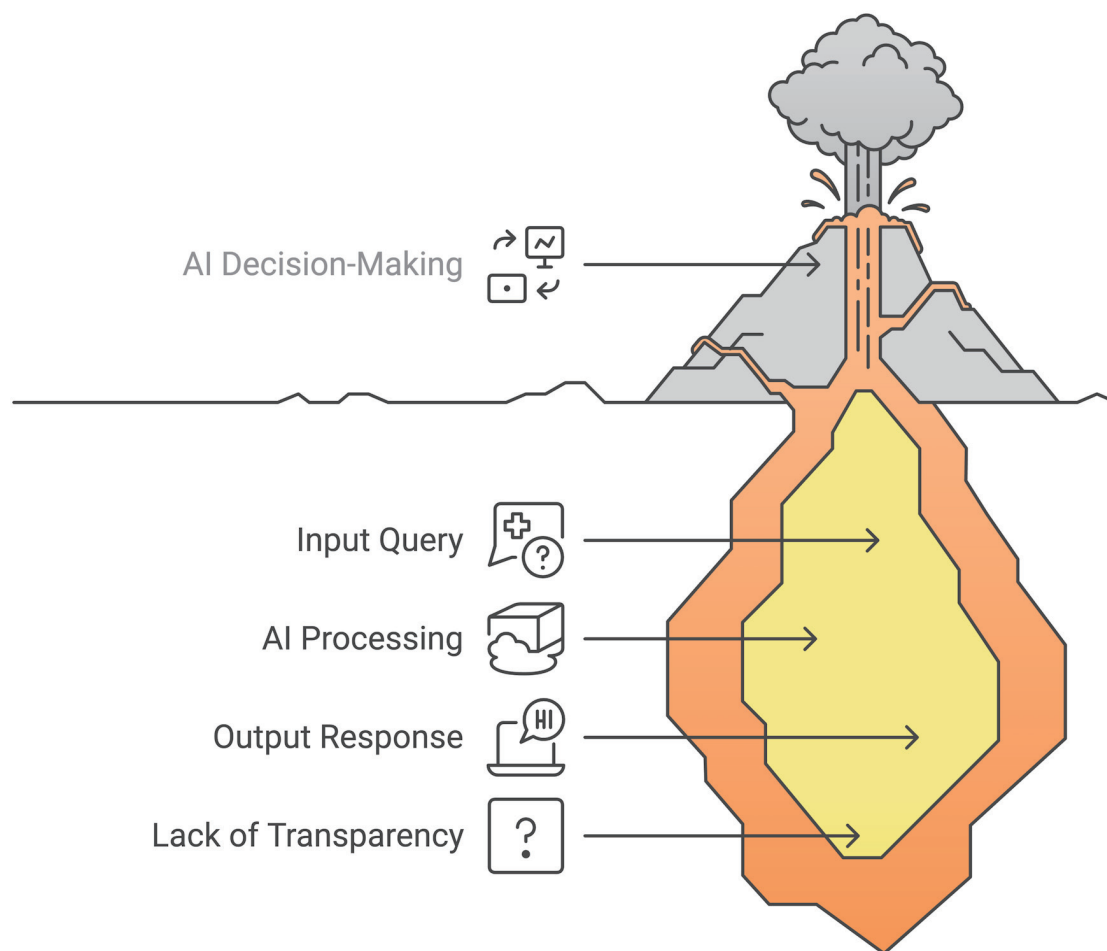
The rapid integration of LLMs into healthcare has been met with enthusiasm and caution. While the transformative potential of LLMs is broad, from aiding in clinical decision-making

to impacting whole fields such as surgical oncology, their use raises a host of ethical concerns [32]. Data privacy and the possibility of re-identification, even after data anonymization, are pressing issues [33]. While this presents challenges for AI learning systems, privacy-preserving techniques such as federated learning, differential privacy, and synthetic data can help mitigate risks. These approaches enable models to improve without directly accessing or storing sensitive patient information, balancing AI advancement with regulatory compliance and ethical considerations. There is growing concern about the amplification of biases, particularly in clinical phenotyping, where state-of-the-art LLMs have been found to underdiagnose vulnerable intersectional subpopulations [34]. For example, AI-driven dermatology models have shown reduced accuracy in diagnosing skin conditions in patients with darker skin tones [35]. Similarly, cardiovascular risk prediction algorithms have underestimated risk in Black patients, and some gastrointestinal (GI) symptom analysis models have exhibited gender bias, leading to potential disparities in care [36, 37]. Addressing these biases is critical to ensuring AI models contribute to equitable healthcare delivery.

Another ethical challenge is the issue of informed consent. While AI is already embedded in many medical technologies, its role in direct clinical decision-making and patient communication raises unique ethical considerations. Patients may not need explicit consent for AI-driven imaging enhancements, but when LLMs influence diagnostic decisions or provide medical guidance, transparency is essential. Informing patients about AI's role in their care helps maintain trust, ensures accountability, and clarifies the boundaries between human and AI-driven decision-making. The role of LLMs in clinical decision-making is also under scrutiny. While AI can assist healthcare professionals by enhancing diagnostics, streamlining workflows, and reducing errors, it is not yet capable of fully replacing human expertise in complex medical decision-making. AI lacks contextual understanding, ethical reasoning, and accountability, which are essential factors in patient care. Rather than replacing physicians, AI should complement human judgment, ensuring that clinical decisions integrate both data-driven insights and patient-centered considerations (Fig. 5). The commercialization of LLMs raises questions about transparency and potential conflicts of interest. The inter-mixing of medical advice and advertising is a growing concern. Ethical education for healthcare professionals is also being advocated to help providers navigate these complex ethical landscapes. The future of LLMs in healthcare necessitates a close partnership between the medical community and AI developers, emphasizing the need for technological advancements adherent to the highest moral and ethical values.

## Legislation and Regulation

It is crucial to mention that while these advancements bring immense potential, as mentioned above, challenges such as data privacy, model interpretability, clinical validation, and ethical considerations are bound to arise. As LLMs are integrated into healthcare, legislation and regulation will play a



**Figure 5.** The “black box dilemma” in AI models. While LLMs generate responses based on learned patterns, their decision-making pathways remain opaque, making interpretability a challenge in clinical applications. AI: artificial intelligence; LLMs: large language models.

key role in balancing patient safety, data privacy, and ethical AI use. While guidelines can help standardize AI implementation and mitigate risks, overly rigid regulations may hinder innovation and delay the adoption of beneficial advancements. The current Food and Drug Administration (FDA) approval process for medical software can take months to years, while AI models evolve rapidly, raising concerns about whether traditional regulatory frameworks can keep pace. A dynamic regulatory approach may be needed to ensure both safety and continuous innovation in AI-driven healthcare [38]. Collaboration between technologists, clinicians, and researchers will be essential to harness the potential of LLMs in gastroenterology and gastrointestinal surgery.

## Future Strategies

The future of LLMs in gastroenterology and gastrointestinal surgery may involve the development of more specialized models fine-tuned for these realms. Collaborative strategies combining the strengths of LLMs with human expertise will

likely become more prevalent. Future strategies should explore multiple approaches to AI integration in healthcare, including developing specialized models, enhancing generalized AI systems, and fostering interdisciplinary collaboration. Integrating AI with human medical professionals remains a key pathway to optimizing patient care, but progress will be multifaceted, requiring a combination of targeted advancements and broad adaptability.

## Conclusions

The integration of large LLMs in gastroenterology and gastrointestinal surgery offers significant potential but also presents challenges. While LLMs can enhance personalized care and medical research, their opaque decision-making, risk of generating inaccurate information, and potential biases limit reliability. Human expertise remains essential for nuanced judgment, accountability, and ethical considerations. To maximize benefits, AI should complement rather than replace clinicians, with efforts focused on improving model transparency, rigor-

ous validation, and responsible regulation. Striking a balance between innovation and oversight is key to ensuring AI contributes to safer, more effective patient care.

## Acknowledgments

None to declare.

## Financial Disclosure

No funding sources to declare.

## Conflict of Interest

All other authors report no conflict of interest.

## Author Contributions

Conceptualization: Dushyant Singh Dahiya, Hassam Ali and Amir Humza Sohail. Data curation and analysis: Dushyant Singh Dahiya, Hassam Ali, Vishali Moond, M. Danial Ali Shah, Christina Santana. Investigation: all authors. Methodology: all authors. Project administration: Dushyant Singh Dahiya and Hassam Ali. Resources and software: Dushyant Singh Dahiya and Hassam Ali. Supervision: Dushyant Singh Dahiya, Hassam Ali and Amir Humza Sohail. Validation: all authors. Visualization: Dushyant Singh Dahiya and Hassam Ali. Writing - original draft: all authors. Writing - review and editing: all authors.

## Data Availability

The authors declare that data supporting the findings of this study are available within the article.

## References

1. Thirunavukarasu AJ, Ting DSJ, Elangovan K, Gutierrez L, Tan TF, Ting DSW. Large language models in medicine. *Nat Med*. 2023;29(8):1930-1940. [doi](#) [pubmed](#)
2. Singhal K, Azizi S, Tu T, Mahdavi SS, Wei J, Chung HW, Scales N, et al. Large language models encode clinical knowledge. *Nature*. 2023;620(7972):172-180. [doi](#) [pubmed](#)
3. LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature*. 2015;521(7553):436-444. [doi](#) [pubmed](#)
4. Hu X, Ran AR, Nguyen TX, Szeto S, Yam JC, Chan CKM, Cheung CY. What can GPT-4 do for diagnosing rare eye diseases? A pilot study. *Ophthalmol Ther*. 2023;12(6):3395-3402. [doi](#) [pubmed](#)
5. Lahat A, Shachar E, Avidan B, Glicksberg B, Klang E. Evaluating the utility of a large language model in answering common patients' gastrointestinal health-related questions: are we there yet? *Diagnostics (Basel)*. 2023;13(11):1950. [doi](#) [pubmed](#)
6. Samaan JS, Yeo YH, Ng WH, Ting PS, Trivedi H, Vipani A, Yang JD, et al. ChatGPT's ability to comprehend and answer cirrhosis related questions in Arabic. *Arab J Gastroenterol*. 2023;24(3):145-148. [doi](#) [pubmed](#)
7. Tariq R, Malik S, Khanna S. Evolving landscape of large language models: an evaluation of ChatGPT and bard in answering patient queries on colonoscopy. *Gastroenterology*. 2024;166(1):220-221. [doi](#) [pubmed](#)
8. Yeo YH, Samaan JS, Ng WH, Ting PS, Trivedi H, Vipani A, Ayoub W, et al. Assessing the performance of ChatGPT in answering questions regarding cirrhosis and hepatocellular carcinoma. *Clin Mol Hepatol*. 2023;29(3):721-732. [doi](#) [pubmed](#)
9. Kitamura FC, Marques O. Trustworthiness of artificial intelligence models in radiology and the role of explainability. *J Am Coll Radiol*. 2021;18(8):1160-1162. [doi](#) [pubmed](#)
10. Varghese J. Reply to: "Can LLMs improve existing scenario of healthcare?". *J Hepatol*. 2024;80(1):e29-e30. [doi](#) [pubmed](#)
11. Xu L, Sanders L, Li K, Chow JCL. Chatbot for health care and oncology applications using artificial intelligence and machine learning: systematic review. *JMIR Cancer*. 2021;7(4):e27850. [doi](#) [pubmed](#)
12. Wang C, Zhang J, Lassi N, Zhang X. Privacy protection in using artificial intelligence for healthcare: Chinese regulation in comparative perspective. *Healthcare (Basel)*. 2022;10(10):1878. [doi](#) [pubmed](#)
13. Shah K, Xu AY, Sharma Y, Daher M, McDonald C, Diebo BG, Daniels AH. Large language model prompting techniques for advancement in clinical medicine. *J Clin Med*. 2024;13(17):5101. [doi](#) [pubmed](#)
14. Alowais SA, Alghamdi SS, Alsuhebany N, Alqahtani T, Alshaya AI, Almohareb SN, Aldairem A, et al. Revolutionizing healthcare: the role of artificial intelligence in clinical practice. *BMC Med Educ*. 2023;23(1):689. [doi](#) [pubmed](#)
15. Javaid A, Shahab O, Adorno W, Fernandes P, May E, Syed S. Machine learning predictive outcomes modeling in inflammatory bowel diseases. *Inflamm Bowel Dis*. 2022;28(6):819-829. [doi](#) [pubmed](#)
16. Pecere S, Milluzzo SM, Esposito G, Dilaghi E, Telese A, Eusebi LH. Applications of artificial intelligence for the diagnosis of gastrointestinal diseases. *Diagnostics (Basel)*. 2021;11(9):1575. [doi](#) [pubmed](#)
17. Zhou XY, Guo Y, Shen M, Yang GZ. Application of artificial intelligence in surgery. *Front Med*. 2020;14(4):417-430. [doi](#) [pubmed](#)
18. Maida M, Celsa C, Lau LHS, Ligresti D, Baraldo S, Ramai D, Di Maria G, et al. The application of large language models in gastroenterology: a review of the literature. *Cancers (Basel)*. 2024;16(19):3328. [doi](#) [pubmed](#)
19. Li S. ChatGPT has made the field of surgery full of opportunities and challenges. *Int J Surg*. 2023;109(8):2537-2538. [doi](#) [pubmed](#)

20. Wojtara M, Rana E, Rahman T, Khanna P, Singh H. Artificial intelligence in rare disease diagnosis and treatment. *Clin Transl Sci.* 2023;16(11):2106-2111. [doi pubmed](#)
21. Ravi A, Neinstein A, Murray SG. Large language models and medical education: preparing for a rapid transformation in how trainees will learn to be doctors. *ATS Sch.* 2023;4(3):282-292. [doi pubmed](#)
22. Khosravi M, Zare Z, Mojtabaeian SM, Izadi R. Artificial intelligence and decision-making in healthcare: a thematic analysis of a systematic review of reviews. *Health Serv Res Manag Epidemiol.* 2024;11:23333928241234863. [doi pubmed](#)
23. Radford A, Kim JW, Xu T, Brockman G, McLeavey C, Sutskever I. Robust speech recognition via large-scale weak supervision. *arXiv:221204356 [cs, eess]*. Published online December 6, 2022.
24. Devaraj A, Wallace BC, Marshall IJ, Li JJ. Paragraph-level simplification of medical texts. *Proc Conf.* 2021;2021:4972-4984. [doi pubmed](#)
25. Becker G, Kempf DE, Xander CJ, Momm F, Olschewski M, Blum HE. Four minutes for a patient, twenty seconds for a relative - an observational study at a university hospital. *BMC Health Serv Res.* 2010;10:94. [doi pubmed](#)
26. Caufield JH, Hegde H, Emonet V, Harris NL, Joachimiak MP, Matentzoglou N, Kim H, et al. Structured Prompt Interrogation and Recursive Extraction of Semantics (SPIRES): a method for populating knowledge bases using zero-shot learning. *Bioinformatics.* 2024;40(3):btac104. [doi pubmed](#)
27. Ayers JW, Poliak A, Dredze M, Leas EC, Zhu Z, Kelley JB, Faix DJ, et al. Comparing physician and artificial intelligence chatbot responses to patient questions posted to a public social media forum. *JAMA Intern Med.* 2023;183(6):589-596. [doi pubmed](#)
28. Abd-Alrazaq A, AlSaad R, Alhuwail D, Ahmed A, Healy PM, Latifi S, Aziz S, et al. Large language models in medical education: opportunities, challenges, and future directions. *JMIR Med Educ.* 2023;9:e48291. [doi pubmed](#)
29. Hutson M. Could AI help you to write your next paper? *Nature.* 2022;611(7934):192-193. [doi pubmed](#)
30. Arora G, Joshi J, Mandal RS, Shrivastava N, Virmani R, Sethi T. Artificial intelligence in surveillance, diagnosis, drug discovery and vaccine development against COVID-19. *Pathogens.* 2021;10(8):1048. [doi pubmed](#)
31. Stokel-Walker C, Van Noorden R. What ChatGPT and generative AI mean for science. *Nature.* 2023;614(7947):214-216. [doi pubmed](#)
32. Ramamurthi A, Are C, Kothari AN. From ChatGPT to treatment: the future of AI and large language models in surgical oncology. *Indian J Surg Oncol.* 2023;14(3):537-539. [doi pubmed](#)
33. Jeyaraman M, Balaji S, Jeyaraman N, Yadav S. Unraveling the ethical enigma: artificial intelligence in healthcare. *Cureus.* 2023;15(8):e43262. [doi pubmed](#)
34. Pal R, Garg H, Patel S, Sethi T. Bias amplification in intersectional subpopulations for clinical phenotyping by large language models. *medRxiv.* Published online March 25, 2023:2023.03.22.23287585.
35. Daneshjou R, Vodrahalli K, Novoa RA, Jenkins M, Liang W, Rotemberg V, Ko J, et al. Disparities in dermatology AI performance on a diverse, curated clinical image set. *Sci Adv.* 2022;8(32):eabq6147. [doi pubmed](#)
36. Vyas DA, Eisenstein LG, Jones DS. Hidden in plain sight - reconsidering the use of race correction in clinical algorithms. *N Engl J Med.* 2020;383(9):874-882. [doi pubmed](#)
37. Zand A, Stokes Z, Sharma A, van Deen WK, Hommes D. Artificial intelligence for inflammatory bowel diseases (IBD); Accurately predicting adverse outcomes using machine learning. *Dig Dis Sci.* 2022;67(10):4874-4885. [doi pubmed](#)
38. Haupt CE, Marks M. AI-generated medical advice-GPT and beyond. *JAMA.* 2023;329(16):1349-1350. [doi pubmed](#)