The Role of Generative AI in Advancing Precision Oncology: Opportunities and Challenges

Qinglong Xu[1,†], Anqi Lin[2,†,*]

[1] Department of Oncology, Zhujiang Hospital, The Second School of Clinical Medicine, Southern Medical University, Guangzhou 510282, China

[2] Department of Biomedical Sciences, Faculty of Medicine, University of Macau, Macao SAR, China.

[†] Qinglong Xu and Anqi Lin contributed equally to this work and share first authorship.

Correspondence: Anqi Lin (smulinanqi0206@i.smu.edu.cn)

Article ID: iCancer. 2026; Early View: ICA260002. | ISSN 3136-733X

Received: 2026-08-08 | **Revised:** 2026-09-09 | **Accepted:** 2026-10-22

Keywords: Large Language Models | Generative AI | Precision Oncology | Diagnosis | Prognosis

Published: 2026-11-04

ABSTRACT

Large Language Models (LLMs) have recently emerged as a potentially transformative technology in clinical oncology, offering promising opportunities for advancing precision medicine. This comprehensive review critically examines the multifaceted applications of LLMs across the cancer care continuum, encompassing diagnosis, treatment planning, and supportive care. In cancer diagnosis, LLMs exhibit promising potential for enhancing the accuracy and efficiency of pathological and radiological assessments, as well as facilitating the interpretation of complex genomic data. In the context of treatment decision-making, these models demonstrate the capability to integrate vast amounts of clinical data, research findings, and patient characteristics to propose personalized treatment strategies. LLMs also demonstrate potential for enhancing patient education, symptom management, and psychological support. Notwithstanding their potential, significant challenges persist, including the requirement for high-quality annotated datasets, limitations in model interpretability, and ethical considerations regarding medical practice and patient privacy. This review provides a critical analysis of the current state of LLM applications in oncology, elucidating both their transformative potential and the obstacles impeding their widespread clinical implementation. The review examines emerging trends, including the integration of multimodal data analysis and advancements in explainable artificial intelligence (AI), which have the potential to address some of these challenges. Through providing a comprehensive overview of the opportunities and challenges presented by LLMs in oncology, this review seeks to inform clinicians, researchers, and policymakers, thereby promoting further innovation and responsible integration of AI technologies in cancer care with the ultimate goal of improving patient outcomes.

One Sentence Summary: This review summarizes the potential of LLMs in precision oncology, highlighting opportunities and challenges across diagnosis, treatment, and care.

Search strategy and selection criteria

This narrative review integrates studies indexed in the PubMed, Web of Science (WOS), and Google Scholar databases. We included original articles and reviews published in English between 2023 and 2025 that were relevant to our review topic, thereby ensuring coverage of fundamental concepts and recent advances in this rapidly evolving domain. Priority was given to peer-reviewed articles featuring experimental comparisons, clinical relevance, or high citation impact. The content frame-

work for various applications was organized around imaging analysis, staging prediction, report generation, and treatment decision-making, thereby ensuring representation of both pioneering work and recent advances. Additionally, reference lists from retrieved articles and related reviews were manually screened to identify other relevant studies.

1. Introduction

In recent years, the application of artificial intelligence technologies in the medical field has become increasingly widespread, with Large Language Models (LLMs) emerging as a pivotal breakthrough in natural language processing, presenting new opportunities and challenges for clinical oncology[1,2,3,4,5]. LLMs are deep learning-based natural language processing models trained on massive text datasets. These models demonstrate powerful capabilities in understanding and generating human language. These models are capable of answering questions, generating text, and performing complex reasoning and task-solving. In the medical field, particularly in oncology, LLMs hold considerable promise for clinical translation.

Throughout this review, we adopt the following terminology conventions. Artificial Intelligence (AI) refers broadly to computational systems that perform tasks requiring human judgment. Machine Learning (ML) is a subset of AI in which models learn patterns from data. Generative AI encompasses AI systems capable of generating new content, including text, images, and other media. LLMs are a specific class of generative AI models trained on massive text datasets to understand and generate human language. While this review focuses specifically on LLMs and their text-based applications in oncology, we use “generative AI” when referring to the broader category of content generation and “LLMs” for text-based language model applications specifically. Multimodal models capable of processing both text and images are specified explicitly where discussed. All LLMs discussed are assistive tools designed to operate under human oversight.

Cancer remains one of the foremost public health challenges confronting the world today, with its complexity and heterogeneity posing significant challenges to clinical diagnosis, treatment, and management[6,7,8]. The emergence of LLMs presents novel approaches and methodologies to address these issues. By integrating and analyzing vast amounts of multidimensional data, including medical literature, clinical data, and genomic information, LLMs have the potential to provide clinicians with more comprehensive and accurate decision support, thereby enhancing diagnostic accuracy, treatment precision, and patient prognosis management.

This review discusses radiological and pathological applications of LLMs and multimodal models. Furthermore, in genomic data interpretation, LLMs have demonstrated significant potential for rapid analysis of sequencing data and identification of putative carcinogenic mutations and therapeutic targets. In cancer treatment decision-making and planning, LLMs have the capacity to suggest optimal treatment regimens based on individual patient characteristics, while integrating the most recent clinical guidelines and research findings. This process encompasses determining the optimal combination of various treatment modalities, including but not limited to surgery, radiotherapy, chemotherapy, and immunotherapy. In complex cancer cases, LLMs can simulate potential outcomes of various treatment strategies, thereby assisting physicians and patients in making more informed decisions. This personalized and precise treatment planning approach has the potential to significantly improve the efficacy of cancer treatment and enhance patients' quality of life and prognosis. LLMs also have the potential to play a pivotal role in supportive care for cancer patients. These models can function as intelligent information assistants, providing patients with reliable and comprehensible information about their disease and treatment plans, thereby facilitating a better understanding of their condition and treatment process. Moreover, LLMs have the capability to offer personalized symptom management advice and lifestyle guidance, potentially contributing to an improvement in patients' quality of life. Regarding psychological support, LLMs can facilitate emotional support and psychological counseling through natural language interactions, potentially aiding patients in better managing the psychological stress associated with their illness. In surgical oncology, the emergence of AI-guided laparoscopic instrument tracking technology has established the foundation for subsequent artificial intelligence advancement in this field, enabling the application of AI technologies in surgical procedures and enhancing both surgical efficacy and efficiency [9,10]. (Figure 1)

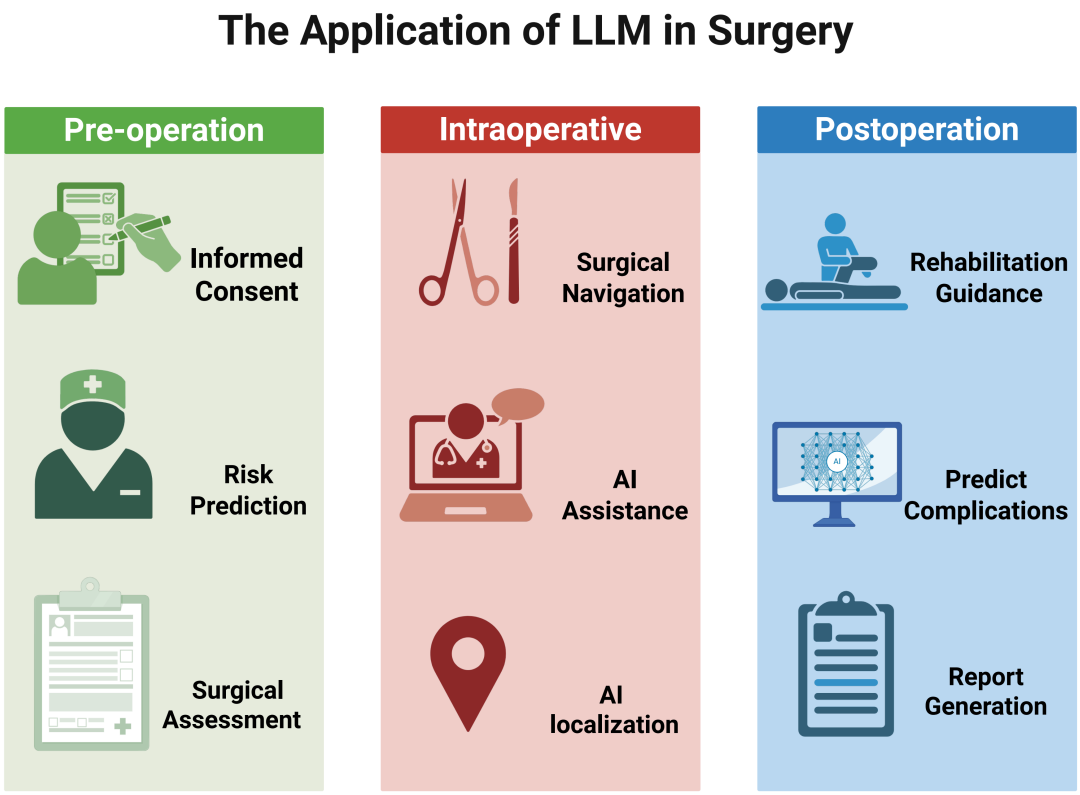


Figure 1. The Application of LLMs in Surgery

The applications of LLMs in surgical procedures can be divided into three distinct phases: preoperative, intraoperative, and postoperative. In the preoperative phase, LLMs can provide patients with disease classification assessments and surgical risk evaluations, thereby enhancing patient understanding of the surgical process. Simultaneously, LLMs can assist physicians in completing the informed consent process and in optimizing preoperative physician-patient communication. During the intraoperative phase, LLMs can provide real-time navigation and precise positioning, assisting surgeons in performing surgical procedures and enhancing surgical precision and efficiency. In the postoperative phase, LLMs can guide patients through rehabilitation training and predict potential complications, thereby facilitating efficient patient recovery. Furthermore, LLMs can assist physicians in generating surgical reports and in optimizing clinical workflows.

Despite the substantial potential that LLMs demonstrate in oncology, their implementation faces several challenges and limitations. First, the training of these models depends on large-scale, high-quality annotated datasets, which are relatively scarce in oncology, potentially limiting the models' performance and generalization capabilities. Second, the interpretability of LLMs remains limited, with their internal decision-making processes being opaque, potentially raising trust issues in clinical applications. Furthermore, the handling of complex medical information and reasoning requires improvement, as the accuracy of these models may diminish when confronted with rare cases or when comprehensive judgment based on multiple information sources is necessary. Finally, issues pertaining to medical ethics and patient privacy protection in the application of LLMs necessitate further research and regulatory considerations. Notwithstanding these challenges, the prospects for applying LLMs in oncology remain highly promising. As the volume of oncological data expands and its quality improves, the performance and generalization capabilities of these models are anticipated to enhance significantly. Concurrently, advancements in explainable artificial intelligence (AI) technologies are likely to enhance the transparency and credibility of LLMs. Moreover, by integrating multimodal data analysis, deep learning, and other advanced technologies, LLMs may be able to process complex medical information more effectively, potentially providing more reliable decision support for precise diagnosis and treatment of cancer.

Beyond the technical and interpretability challenges discussed above, the integration of LLMs into clinical oncology raises fundamental questions regarding the boundaries of clinical responsibility and accountability. Unlike traditional medical devices that perform predefined functions, LLMs generate probabilistic outputs that may vary across prompts, sessions, and model versions, complicating the assignment of liability when AI-generated recommendations contribute to adverse out-

comes. A recent commentary on AI-generated cancer information argues that accountability should not rest with a single stakeholder but rather be viewed as a shared responsibility across the cancer education ecosystem, encompassing patients, healthcare professionals, healthcare organizations, and AI developers[11]. In the clinical decision-making context, human operators retain ultimate accountability and may be legally liable for acting on AI-generated recommendations, especially when those recommendations deviate from established standard procedures[12].

To provide structured guidance, the European Society for Medical Oncology (ESMO) recently released the ESMO Guidance on the Use of Large Language Models in Clinical Practice (ELCAP), the first oncology-specific framework for LLM deployment [13]. ELCAP categorizes LLM applications into three broad types—patient-facing, clinician-facing, and background institutional systems—each with differing accountability and validation requirements, and explicitly states that LLMs are designed to enhance—not replace—clinical workflows, with clinical responsibility remaining with human clinicians. The framework also notes that autonomous ("agentic") AI models pose distinct safety and regulatory challenges requiring future guidance. These frameworks collectively establish that clinical responsibility remains firmly with human clinicians while delineating conditions under which LLMs may safely assist in oncological practice.

Given the increasing significance and potential transformative impact of LLMs in oncology, this review aims to comprehensively examine and analyze the current applications, emerging challenges, and future directions of LLMs in clinical oncology. This paper will focus on the following key areas: First, we will investigate the applications of LLMs in cancer diagnosis, encompassing radiological diagnosis, pathological diagnosis, and genomic analysis. Second, we will critically analyze the role of LLMs in cancer treatment decision-making and planning, elucidating how they support oncologists in developing personalized and precise treatment strategies. Third, we will examine the applications of LLMs in supportive care for cancer patients, focusing on patient education, symptom management, and psychological support. Finally, we will investigate the potential applications of LLMs in cancer prognosis prediction and outcome assessment. Through this comprehensive review, we aim to provide a systematic and evidence-based reference framework for clinicians, researchers, and policymakers, facilitating their understanding and utilization of LLMs in the field of oncology. Concurrently, we aim to stimulate further research and innovation in this field across academia and industry, fostering the deep integration of artificial intelligence technologies with clinical oncology. This integration has the potential to enhance diagnostic accuracy, optimize treatment strategies, and ultimately improve patient outcomes (Figure 2). In accordance with the TITAN 2025 guidelines, we declare that no generative AI tools were used in the core intellectual processes of this review, including the analysis, synthesis, and interpretation of the literature data, as well as the formulation of original insights and conclusions[14].

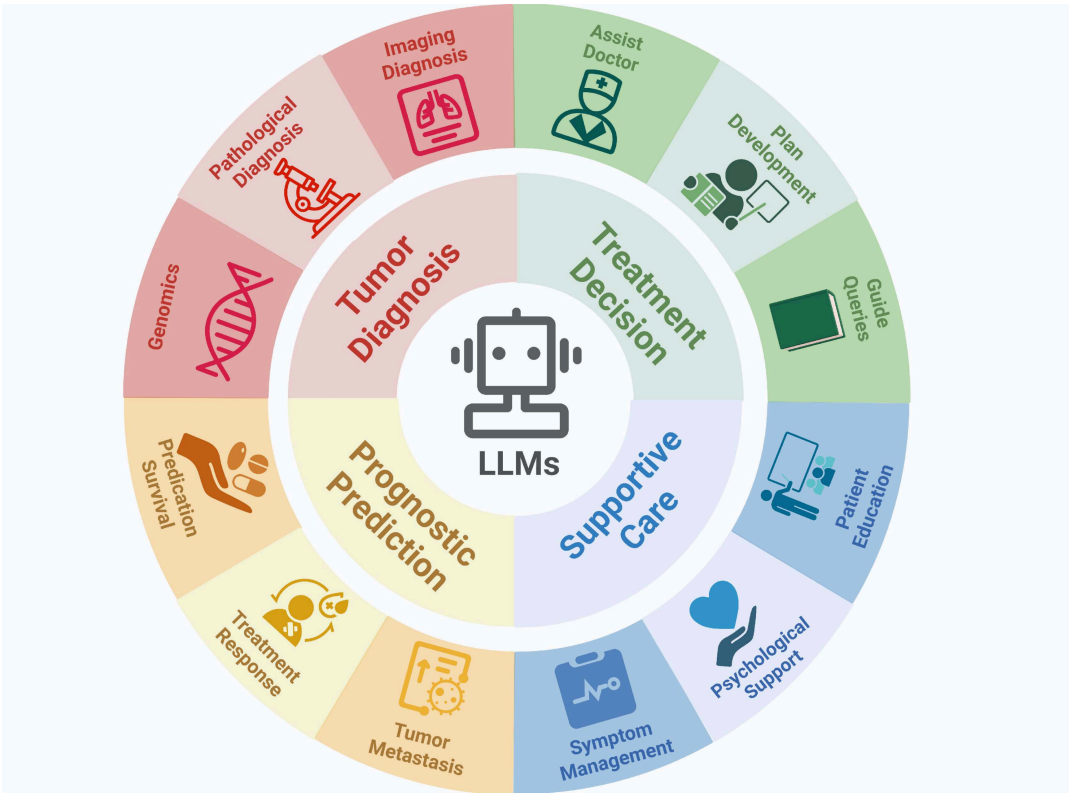


Figure 2. Applications of LLMs in Clinical Oncology

The applications of LLMs in clinical oncology can be categorized into four primary domains: tumor diagnosis, treatment planning, supportive care, and prognostic prediction. In the realm of tumor diagnosis, LLMs can evaluate tumor characteristics, including size and nature, by integrating knowledge from multiple disciplines such as medical imaging, histopathology, and genomic analysis. For treatment planning, LLMs can support oncologists in formulating comprehensive treatment strategies by analyzing and interpreting clinical practice guidelines and current research evidence. In the context of supportive care, LLMs can facilitate symptom management, offer psychological support, and provide tailored educational resources to patients, enhancing their overall quality of life during cancer treatment. Regarding prognostic prediction, LLMs can project potential metastatic patterns, assess treatment efficacy, and estimate patient survival probabilities, thereby aiding in the development of personalized follow-up strategies.

Given the growing importance and potentially transformative impact of LLMs in oncology, this review aims to comprehensively examine current applications, emerging challenges, and future directions of LLMs in precision oncology. However, this field remains in its infancy. The applications discussed herein are primarily at the proof-of-concept, development, or early validation stage. To date, robust evidence from large-scale clinical trials demonstrating the direct impact of LLMs on hard clinical endpoints in surgical oncology remains limited.

2. Applications of LLMs in Cancer Diagnosis

In the context of cancer management, accurate diagnosis is a critical initial step. A precise cancer diagnosis not only informs subsequent treatment strategies but also serves as a foundation for prognostic assessment. Contemporary cancer diagnosis primarily relies on radiological imaging for tumor localization and size determination, complemented by histopathological analysis to assess tumor type and characteristics, culminating in a definitive diagnosis. LLMs, an emerging technology in natural language processing, have demonstrated significant potential in the domain of pathological diagnosis. These models not only assist pathologists in analyzing digital pathology slides and identifying subtle morphological features but also integrate patients' clinical data to generate comprehensive and precise diagnostic recommendations. In the realm of diagnostic imaging, the integration of LLMs with computer vision technologies has revolutionized radiological diagnosis. This synergistic approach not only enhances the accuracy of tumor lesion detection and localization but also facilitates the automatic generation of structured descriptive reports, thereby significantly improving radiologists' efficiency and the quality of diagnostic reporting. Moreover, LLMs have exhibited remarkable proficiency in interpreting genomic data and elucidating its clinical implications. These models efficiently process vast quantities of sequencing data, swiftly identifying potential carcinogenic muta-

tions and putative targeted therapy targets, thereby providing critical evidence for precise cancer diagnosis and the formulation of personalized treatment strategies. LLMs also demonstrate the capacity to integrate multi-omics data, encompassing genomics, transcriptomics, proteomics, and metabolomics, among others. This integration facilitates a profound elucidation of tumor molecular mechanisms, providing robust support for subsequent targeted drug development and the formulation of personalized treatment strategies. As illustrated in Figure 2, LLMs demonstrate significant potential across various aspects of oncology diagnosis by integrating multidisciplinary knowledge such as medical imaging, histopathological analysis, and genomics. As illustrated in Figure 2, we clearly delineate the diverse applications of LLMs within clinical oncology, effectively highlighting the extensive scope and feasibility of their practical implementation.

2.1 Enhancing Diagnostic Accuracy with LLMs

Image-related generative AI research includes histopathology image synthesis using latent diffusion models and bioinformatics plot interpretation using image-enabled ChatGPT [15,16]. These are distinct applications. The plot-interpretation study is a preprint and reports limitations in quantitative interpretation.

2.1.1 Diagnosis of Pathological Images

The traditional gold standard for cancer diagnosis involves examining pathological sections, analyzing tissue structural features under a microscope, and subsequently rendering a diagnosis [17]. This diagnostic process is customarily conducted by experienced pathologists. The rapid development of LLMs has stimulated considerable research interest in pathological image diagnosis, though the clinical translation of these advances remains in early stages. Laohawetwanit et al. assessed the efficacy of ChatGPT-4 in diagnosing hepatobiliary tumors from histopathological images. The study revealed that models specifically trained for this task exhibited higher accuracy and consistency in replicating pathologists' diagnoses [18]. This finding strongly suggests that LLMs require continuous optimization and substantial data support to achieve higher diagnostic accuracy in specific medical fields. In the context of data requirement optimization, Sushil et al. explored the potential of LLMs in mitigating the need for large-scale data annotation. Through comparative analysis with other models, they determined that LLMs exhibited significant advantages in this domain [19]. However, it is important to emphasize that this study predominantly focused on the classification of breast pathology reports. In the realm of image diagnosis, no efficacious methodologies have yet been identified to significantly reduce data dependency. Regarding the processing of pathology report data, LLMs have exhibited the capability to perform without relying on large-scale task-specific training, a significant advancement that offers new avenues for enhancing data processing efficiency [20].

The ongoing development of LLMs has progressively expanded their capabilities, evolving from analyzing microscopic section images to integrating sectional information and extracting structured data. When combined with sophisticated algorithms and computer-aided diagnostic technologies, these advancements substantially enhance the efficiency and accuracy of pathological diagnosis. Nevertheless, LLMs continue to face myriad challenges in clinical applications, including surmounting public bias against algorithms and resolving associated legal issues. Effectively mitigating these challenges will facilitate the further development of LLMs in the domain of cancer diagnosis [21,22].

2.1.2 Diagnosis in Radiation Oncology

Analogous to advancements in pathological image analysis, the advent of LLMs has ushered in novel breakthroughs and developmental opportunities in diagnostic imaging within the field of radiation oncology. In computed tomography (CT) diagnosis, research has demonstrated that ChatGPT-4V exhibits considerable accuracy in non-small cell lung cancer diagnosis, with an overall diagnostic accuracy rate of 56.76%. For non-small cell lung cancer detection, the sensitivity and specificity were 27.27% and 60.47%, respectively. The positive predictive value and negative predictive value were 26.09% and 61.90%, respectively, yielding an overall accuracy of 49.23%. Chi-square statistical analysis revealed significant differences in sensitivity across different case categories, correlating with lesion location and lobar involvement patterns. This demonstrates the promising development potential of ChatGPT-4V in chest CT interpretation [23]. Studies suggest that LLMs demonstrate proficiency in extracting pertinent information from CT reports and can substantially enhance diagnostic efficiency following ad-

ditional training. Nevertheless, when faced with intricate diagnostic tasks, the performance of these models may still require further refinement [24]. Comparative analyses of similar tasks have revealed that GPT-4 exhibits superior performance among various LLMs [25].

For magnetic resonance imaging (MRI) diagnostics, investigators have evaluated the capacity of LLMs to categorize brain MRI reports. Researchers categorized reports into three groups: non-tumor cases (Group 1), post-treatment tumor cases (Group 2), and pre-treatment tumor cases (Group 3), then compared the model's performance on the test dataset with that of two radiologists. The results demonstrated that the fine-tuned LLM achieved an overall accuracy of 0.970 (95% CI: 0.930-0.990). The model demonstrated sensitivities of 1.000, 0.864, and 0.978 for the three groups, respectively. The model exhibited specificities for the three groups of 0.991, 0.993, and 0.958, respectively. No significant statistical differences were found between the LLM and radiologists in terms of accuracy, sensitivity, and specificity ($p \geq 0.371$); however, the LLM completed tasks approximately 20-26 times faster than radiologists. This indicates that trained and fine-tuned models can achieve classification accuracy comparable to that of radiologists while significantly reducing the required time[26]. In the context of liver MRI, GPT-4 was employed to extract reporting and data system features, as well as category information from unstructured MRI reports. This application revealed GPT-4's considerable potential, albeit with the necessity for further refinement of prompts and enhancement of neural network architectures [27]. Furthermore, a novel application of ChatGPT was identified in the generation of MRI reports for prostate cancer patients, enhancing the overall comprehensibility of these reports. Nevertheless, the overall quality and accuracy of these AI-generated reports still necessitate further enhancement [28]. These investigations demonstrate that the current applications of LLMs in MRI are predominantly confined to information extraction and report analysis. The future trajectory of LLMs' development in this field lies in their potential to assist clinicians in interpreting MRI images, thereby enhancing the efficiency and accuracy of tumor diagnosis.

In the domain of ultrasound imaging diagnostics, Wang et al. employed GPT-4 to analyze thyroid ultrasound reports and generate corresponding treatment recommendations. Their findings revealed that GPT-4 exhibited promise in generating thyroid cancer ultrasound reports; however, it demonstrated limitations in diagnostic accuracy. The researchers also proposed an online platform, "ThyroAIGuide," and implemented a chain-of-thought approach to augment the interpretability of AI-driven decisions and facilitate physician-patient communication [29]. Presently, in the realm of ultrasound imaging diagnostics, LLMs predominantly function in the capacity of report summarization and analysis.

2.1.3 Visual Imaging Diagnosis

The advent of advanced visual language models, exemplified by GPT-4 Vision, has ushered in novel technological approaches for the diagnosis of benign and malignant cutaneous neoplasms. Preliminary investigations suggest that ChatGPT Vision's efficacy in melanoma diagnosis using dermoscopic imagery is substantially inferior to established professional AI algorithms currently available in the market. Premature implementation in clinical settings may result in false-negative diagnoses of melanoma or false-positive diagnoses of benign lesions, potentially leading to substantial medical risks [30]. Nevertheless, the rapid progression of LLM technology has enabled the latest iterations, such as GPT-4 Vision and Large Language and Vision Assistant, to demonstrate proficiency in accurately distinguishing between benign lesions and melanoma. This advancement has unveiled potential applications of LLMs technology across dermatological clinical practice, medical research, and educational paradigms [31].

2.1.4 Cancer and Tumor Screening and Monitoring

Beyond image-based diagnostics, LLMs demonstrate the capability to extrapolate tumor and cancer progression trends from diverse textual and imaging data that encapsulate patient information, facilitating real-time monitoring of cancer progression. Zhang et al. assessed the efficacy of diverse LLMs in detecting patients with metastatic cancer through the analysis of discharge summaries. The findings revealed that prompts integrating explicit and succinct reasoning processes substantially enhanced model performance. This discovery not only elucidated potential areas for refinement in existing models but also delineated trajectories for augmenting their clinical applicability [32]. Complementing discharge summaries, radiological reports constitute another crucial repository of clinical information. Meticulously calibrated LLMs possess the capability to de-

duce tumor metastasis sites from unprocessed radiological reports. Leveraging this capability, researchers engineered a natural language processing (NLP) information extraction system that surpassed LLMs in the task of inferring cancer metastasis [33].

LLMs demonstrate proficiency not only in elucidating cancer metastasis from clinical reports but also in evaluating tumor resectability, substantially broadening their utility in clinical decision support systems. Empirical studies have demonstrated that LLMs possess the capability to synthesize comprehensive analytical reports for pancreatic ductal adenocarcinoma (PDAC) from unprocessed radiological reports and evaluate its resectability. This methodology exhibits high precision, furnishing surgeons with more accurate diagnostic evidence to refine treatment strategies [34]. Nevertheless, LLMs continue to exhibit constraints in certain specialized diagnostic tasks. Son et al. explored the viability of employing ChatGPT for the diagnosis of osseous metastases in skeletal scintigraphy. The investigation revealed that ChatGPT exhibited notable limitations in differentiating osseous metastases from benign spinal lesions, with its diagnostic accuracy falling short of clinical application standards [35]. These findings underscore the imperative for judicious application of LLMs in specialized medical diagnostic tasks.

As LLMs continue to advance their natural language processing capabilities, they demonstrate potential to effectively assist clinicians in performing examinations and risk screening as adjunctive tools. Recent studies suggest that appropriately trained LLMs can comprehensively analyze patient histories, colonoscopy reports, and pathology findings to provide evidence-based recommendations for scheduling subsequent colonoscopies, indicating potential for integration into clinical practice [36,37]. However, LLMs necessitate further refinement and validation in the context of colorectal cancer screening. At present, the application of these models in this field is primarily limited to providing basic educational information to patients, with the reliability and provenance of the information provided requiring further verification [38,39,40,41].

In the realm of prostate cancer screening and biomarker-related issues, the performance of LLMs parallels that observed in colorectal cancer screening, with current accuracy levels warranting further enhancement. Comprehensive and rigorous training is imperative for these models to effectively contribute to clinical decision support [42,43,44]. It is noteworthy that different versions of LLMs exhibit performance heterogeneity. Empirical evidence indicates that GPT-4 demonstrates superior accuracy in addressing pertinent queries compared to GPT-3.5 [45]. This finding not only elucidates generational disparities in LLMs' performance but also underscores the burgeoning potential for LLMs in the medical field through iterative refinement and optimization.

Studies have evaluated LLMs in relation to head and neck cancer [46], colorectal cancer [38], and basal cell carcinoma [47]. Results should be interpreted as study-specific rather than as a common precision metric. In the colorectal cancer study, ChatGPT-3.5 averaged 4.5 correct answers out of 10 across four sessions (45% overall performance). Further evaluations have addressed central nervous system tumors [48], ophthalmic neoplasms [49], urological malignancies [50], hepatocellular carcinoma [51,52], oral cancer [53], and breast cancer [54]. These study-specific findings should not be used to rank performance across cancer types. Of particular note, LLMs demonstrate distinctive potential in the diagnosis of oral squamous cell carcinoma. Preliminary research suggests that even with limited datasets, LLMs can execute effective diagnostic reasoning on confocal laser endomicroscopy (CLE) images. However, the capacity of LLMs to attain reliable diagnostic accuracy in real-world clinical settings, particularly when confronted with raw, unprocessed datasets, necessitates further comprehensive investigation and rigorous validation [55].

Through a comprehensive analysis of existing studies, we have found that LLMs exhibit significant performance variations across different oncological domains [56,57]. These variations likely stem from multiple factors, including but not limited to the developmental stage of LLMs, the frequency of training database updates, and differences in model training methodologies. This finding indicates a direction for future research: while LLMs demonstrate immense potential for application in cancer screening, realizing this potential requires continuous refinement of relevant data sources, optimization of neural network architectures, and implementation of targeted model training. Through these improvements, LLMs can more effectively fulfill their role as clinical assistants, potentially bringing revolutionary changes to the field of cancer diagnosis and treatment.

2.2 Applications of LLMs in Tumor Classification and Staging

In the process of tumor diagnosis, in addition to determining the tumor type, clinicians are particularly concerned with tumor characteristics and growth patterns, as these factors significantly influence the development of treatment strategies for different types of tumors. LLMs offer a novel approach to tumor identification and classification that warrants further investigation. Yang et al. utilized LLMs for the differential diagnosis of benign and malignant bone tumors. The results demonstrated that the initial model achieved a diagnostic accuracy of 73%, whereas after a few-shot fine-tuning, the LLMs exhibited enhanced adaptability, increasing diagnostic accuracy to 87%. This research outcome underscores the substantial potential of LLMs in the differential diagnosis of benign and malignant bone tumors, demonstrating improved diagnostic efficiency and reduced misdiagnosis rates, thus enhancing clinical decision-making processes [58].

In addition to demonstrating advantages in the differential diagnosis of benign and malignant tumors, LLMs also exhibit significant application potential in clinical tumor staging. Studies suggest that in TNM staging, LLMs are particularly effective as auxiliary tools, assisting physicians in tumor classification and staging, thus enhancing work efficiency. However, it is crucial to emphasize that the application of LLMs must be conducted under the supervision of professional medical personnel to ensure diagnostic accuracy and patient safety [59]. Furthermore, LLMs have demonstrated the ability to accurately identify clinical phenotypes from Electronic Health Records (EHRs). This capability not only facilitates more precise clinical staging but also enables medical teams to gain a comprehensive understanding of patients' health status, providing a critical foundation for developing personalized treatment plans [60].

2.3 Utilizing LLMs to Predict Cancer Progression

From a molecular biology perspective, the development of tumors and cancers is primarily attributed to cellular gene mutations. The identification of critical gene mutations is fundamental to advancing research in targeted cancer therapies. LLMs, through the integration of multi-modal tumor data, can perform comprehensive analyses of extensive sequencing datasets, facilitating the prediction of potential oncogenic mutations. Researchers have developed innovative artificial intelligence systems utilizing LLMs embedding techniques, which enable simultaneous processing of multiple public databases and adaptation to novel database categories [61]. Furthermore, researchers have investigated the potential applications of LLMs in data extraction from open-source databases. Findings suggest that LLMs-assisted data extraction, when compared to manual methods, not only significantly reduces time and cost but also maintains comparable accuracy [62]. LLMs demonstrate proficiency not only in processing public databases but also exhibit promising performance and high feasibility in clinical settings and data analysis. Nevertheless, ongoing enhancements in accuracy are imperative to meet the rigorous standards of clinical applications [63]. Future research, building upon these applications and harnessing LLMs' efficient processing capabilities for genetic datasets, has the potential to significantly expedite the development of clinical drugs [64].

In the realm of clinical drug development, the integration of LLMs and machine learning techniques has yielded an innovative framework for the research and development of bispecific antibodies (BsAbs). This novel approach, which enables simultaneous targeting of two antigens, has not only significantly improved prediction accuracy but also substantially enhanced the innovative potential in drug design [65]. In addition to data processing and analysis, LLMs have demonstrated the capability to perform in-depth analysis and precise interpretation of cancer sequencing data [66]. Research conducted by Uprety et al. has demonstrated that ChatGPT can effectively review and extract key information from cancer patient records, interpret next-generation sequencing reports, and generate comprehensive protocol lists for potential clinical trials [67]. A separate study explored the potential application of ChatGPT in predicting pathogenic cancer predisposition gene (CPG) mutations through the analysis of clinical notes. The findings suggest that ChatGPT demonstrates promising potential in analyzing clinical notes, particularly genetics-related clinical records, for the prediction of CPGs. However, the researchers emphasized that further comprehensive studies are necessary to validate and expand the clinical application of this technology [68].

3. The Application of LLMs in Cancer Treatment Decision-Making and Planning

LLMs have generated considerable interest in cancer treatment decision-making, with preliminary studies exploring their potential utility. LLMs can integrate and analyze vast quantities of multidimensional data, encompassing clinical records, imaging information, and genomic profiles. By leveraging their robust data processing capabilities and sophisticated deep learning

algorithms, these models provide clinicians with more comprehensive and precise decision support recommendations. LLMs exhibit remarkable potential in the formulation of personalized treatment strategies. These models can propose optimal treatment regimens based on patients' individual characteristics (including age, sex, medical history, and genetic profile), while incorporating the most recent clinical guidelines and research findings. These treatment strategies may encompass multiple therapeutic modalities, including surgical interventions, radiation therapy, systemic chemotherapy, immunotherapeutic approaches, and molecularly targeted therapies. LLMs can employ sophisticated algorithmic analyses to determine the optimal combination of these therapeutic modalities, thereby facilitating the realization of precision medicine. In complex oncological scenarios, the advantages of LLMs become increasingly evident. These models can leverage advanced simulation algorithms to forecast potential outcomes and anticipate possible complications associated with various treatment strategies, thereby providing comprehensive risk-benefit analyses for both clinicians and patients. This approach not only facilitates more informed treatment decisions but also enhances the process of obtaining informed consent and improves clinician-patient communication. As artificial intelligence technologies continue to advance rapidly and clinical applications expand, the potential for LLMs in oncological treatment decision-making and planning continues to grow exponentially. These models have the potential to significantly enhance the precision and efficacy of cancer treatments, not only improving patients' quality of life and clinical outcomes but also potentially optimizing medical resource allocation and reducing healthcare expenditures. However, it is imperative to address the ethical considerations and data security challenges associated with the implementation of LLMs in healthcare, ensuring that this innovative technology can be safely and effectively utilized to benefit patients.

3.1 Assisting in Treatment Plan Formulation

In the complex landscape of cancer management, formulating personalized treatment plans represents a crucial step, typically necessitating interdisciplinary collaboration among oncologists and experts from various specialties through multidisciplinary consultations. As summarized in Figure 2, LLMs, as advanced natural language processing tools, exhibit powerful capabilities in information integration and analysis, holding promise for revolutionizing decision support in future diagnostic and treatment processes. This approach facilitates a comprehensive evaluation of the patient's specific circumstances, enabling the development of optimal diagnostic and therapeutic strategies. LLMs, as state-of-the-art natural language processing tools, demonstrate robust information integration and analytical capabilities, potentially revolutionizing decision support in future diagnostic and treatment processes. Researchers have conducted extensive investigations into the application of LLMs for clinical decision support across diverse cancer types. Notably, studies focusing on breast cancer and head and neck cancer have been particularly prolific, underscoring the clinical significance and intricate nature of these malignancies in oncological practice.

Initial investigations into the application of LLMs for breast cancer treatment decision-making revealed that early iterations of these models exhibited substantial disparities in treatment recommendation accuracy when benchmarked against established clinical guidelines [69,70]. Subsequent technological advancements, however, have facilitated notable progress in LLMs' capabilities for information retrieval and treatment plan formulation, particularly through training based on clinical guidelines. These improvements have led to markedly enhanced accuracy in treatment recommendations [71,72,73]. Comparative analyses of various LLMs' proficiency in formulating breast cancer treatment plans revealed that the most recent iteration of the GPT-4 model demonstrated superior performance in terms of accuracy [74]. This finding underscores the significant potential of advanced AI models in augmenting medical decision-making processes. Despite these notable advancements in breast cancer treatment decision support, LLMs continue to exhibit suboptimal performance when confronted with complex clinical presentations or atypical patient scenarios. This observation underscores the imperative for continued optimization and refinement of LLMs in medical applications, with a particular emphasis on enhancing their capacity to comprehend and navigate intricate clinical contexts [75].

In the field of head and neck cancer treatment planning, the performance of LLMs demonstrates both similarities and distinctions when compared to their application in breast cancer management. Preliminary studies suggest that the accuracy of LLMs in head and neck cancer treatment decision-making, akin to observations in breast cancer research, has not yet attained optimal levels. Nevertheless, LLMs have exhibited exceptional proficiency in interpreting tumor cTNM staging and have significantly surpassed human physicians in recommending additional examinations[76], underscoring the potential superiority

of AI over conventional approaches in specific tasks. Moreover, in evaluating surgical indications for head and neck cancer, LLMs' recommendations garnered approval from 48% of clinical experts[77]. While this result indicates room for improvement, it also highlights the potential value of AI in complex medical decision-making processes. With ongoing technological advancements, LLMs have markedly enhanced their accuracy in supporting head and neck cancer treatment decisions, generally demonstrating consistency with clinical guidelines in the majority of cases. Nonetheless, several critical issues persist, including inappropriate surgical plan selection in specific scenarios and inadequate explainability in clinical applications. These challenges underscore the necessity for continued optimization and stringent clinical validation of LLMs in medical applications [46,78,79]. Comparative analyses of various LLMs in formulating head and neck cancer treatment strategies have revealed that, while models generally exhibit satisfactory overall performance, discrepancies persist in specific areas, such as the application of induction chemotherapy and the selection of surgical approaches [80]. This finding underscores the necessity for careful consideration of inter-model disparities when implementing AI-assisted decision-making tools in clinical practice, as well as the critical importance of integrating clinical expertise for comprehensive assessment.

In addition to the two extensively studied cancer types mentioned above, LLMs have been applied to a wide range of cancer and tumor research areas. In the field of prostate cancer research, LLMs have been employed for preliminary screening and triage of patients with metastatic prostate cancer. Research findings indicate that LLMs demonstrate high sensitivity in identifying emergency admissions for patients with metastatic prostate cancer, suggesting the potential for significant improvements in care efficiency and treatment outcomes in the future [81]. Furthermore, during prostate cancer treatment, LLMs have demonstrated effectiveness in monitoring radiation toxicity, offering not only high accuracy but also substantial time savings [82]. In gynecologic oncology research, various LLMs exhibit comparable performance, offering advantages in rapid response; however, further evaluation of their output quality and clinical applicability remains imperative [83]. A separate research team specifically evaluated the performance of GPT-4 in the field of ovarian cancer. Results indicate that, at a single time point, GPT-4 can provide accurate and comprehensive answers to a limited range of ovarian cancer-related questions, demonstrating particular proficiency in risk factor identification, surgical management strategies, and disease monitoring. However, GPT-4 occasionally produces inaccuracies, which significantly constrain the unsupervised use of LLMs in clinical settings [84]. In the field of gastrointestinal tumors, the majority of overall decisions made by LLMs demonstrate general consistency with existing clinical guidelines. However, a decline in consistency is observed regarding specific treatment plans, particularly in the number and accuracy of recommended chemotherapy regimens. Consequently, prudence is warranted when directly applying LLMs to clinical decision-making in gastrointestinal tumors [85]. In pancreatic cancer research, LLMs exhibit promising developmental potential, providing an optimistic outlook for their future applications in this field [86]. In the diagnosis and treatment of colorectal cancer, LLMs show substantial alignment with recommendations from existing clinical guidelines, indicating their potential clinical value in this field [87]. However, in the research on applications for non-melanoma skin cancer, LLMs exhibit significant limitations. LLMs demonstrate poor concordance with dermatology specialists in diagnosis and treatment decisions, suggesting potential risks in their direct application to clinical practice [88]. In the field of renal cell carcinoma research, the performance of LLMs parallels that observed in pancreatic cancer studies. Despite certain limitations in initial applications, subsequent model fine-tuning and iterative training have substantially enhanced the accuracy of LLMs in answering relevant questions [89]. In the field of immuno-oncology, the performance of LLMs in answering relevant questions has undergone evaluation. Research findings indicate that while LLMs demonstrate promise, the accuracy of their responses necessitates further refinement. Therefore, before applying LLMs to clinical practice in immuno-oncology, rigorous validation and evaluation by experts remain imperative [90].

3.2 Writing Case Reports and Clinical Case Studies

After formulating appropriate treatment plans, physicians must meticulously document and refine patient cases, thereby providing comprehensive and accurate information for subsequent treatment and follow-up. Furthermore, physicians can compile representative or educationally significant cases into case reports, which serve as valuable learning resources for future practitioners and provide critical references for research and clinical practice in relevant fields. The rapid advancement of LLMs offers the potential to assist with these time-consuming documentation tasks, potentially significantly enhancing work efficiency. Zamarud et al. utilized ChatGPT to compose and compare two similar case reports, demonstrating both the powerful text processing capabilities of LLMs and their significant potential in drafting case reports and clinical case studies [91]. Le et

al. employed ChatGPT to compose a case report on a rare disease, providing an in-depth exploration of the potential applications of generative AI in medical research literature writing. The study revealed that LLMs can provide researchers with a clear structural framework for literature and suggest potential future research directions in relevant fields. However, the study also highlighted that LLMs currently have limitations in accurately proofreading reference formats [92].

3.3 Providing Personalized Treatment Recommendations

Contemporary cancer treatments predominantly adhere to established therapeutic guidelines, which are typically based on the majority of common cases. However, due to complex factors such as comorbidities and genetic predispositions in certain patients, standardized guidelines may not always yield optimal treatment outcomes, necessitating the development of personalized treatment plans. In this context, LLMs can serve as powerful decision support tools to assist physicians in formulating personalized treatment plans.

A study conducted by Benary et al. evaluated the capability of four distinct LLMs in formulating treatment plans for 10 patients with advanced cancer harboring genetic mutations. The findings revealed that for each patient, at least one LLM successfully generated a treatment plan considered valuable by human experts. This finding not only underscores the potential of LLMs in precision oncology but also highlights that the application of LLMs in this field has not yet attained a level of quality and credibility comparable to that of human experts, necessitating further refinement and in-depth research [93]. Additionally, LLMs have the potential to propose innovative approaches in the development of personalized treatment plans, which may not yet be acknowledged or validated by human experts. This characteristic emphasizes the innovative potential of LLMs in personalized medicine applications while simultaneously revealing their limitations in clinical practice [57].

Nonetheless, some researchers have demonstrated that through applications in decision support, patient education, and care coordination, LLMs can facilitate more personalized, proactive, and collaborative treatment approaches, potentially leading to significant improvements in patient outcomes [94]. These findings indicate that in the future, physicians may be able to utilize LLMs to devise personalized treatment plans for each patient based on their specific disease characteristics, personal financial considerations, and individual needs. This approach, integrating artificial intelligence and clinical expertise, has the potential to accelerate the development of precision medicine, providing patients with more optimized treatment options. (Figure 3)

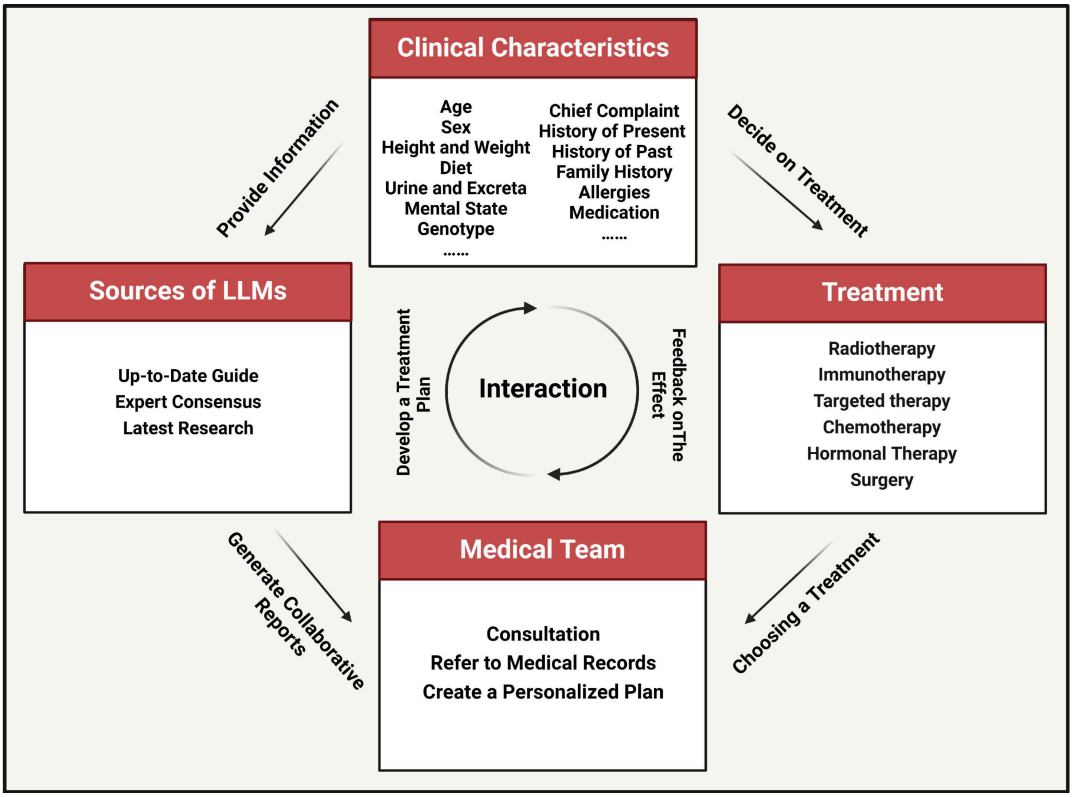


Figure 3. Applications of LLMs in Personalized Medicine

In the context of personalized treatment, LLMs aggregate patient-specific data and integrate it with state-of-the-art information, including the most recent clinical guidelines, emerging research findings, and expert consensus. This integration enables LLMs to generate comprehensive reports for the treatment team's reference, encompassing recommended treatment modalities, risk stratification, and prognostic predictions. Based on these LLM-generated reports, the multidisciplinary treatment team collaboratively selects the most appropriate therapeutic strategies, ultimately making decisions in consultation with the patient. Throughout the treatment course, patients provide real-time feedback on treatment efficacy and adverse effects, enabling the treatment team to promptly formulate and adjust therapeutic plans as necessary.

3.4 Assisting in Surgical Procedures

For cancer patients, surgical intervention remains the most common method for addressing underlying lesions. As depicted in Figure 1, LLMs have the potential to assist surgeons throughout the entire surgical continuum - preoperatively, intraoperatively, and postoperatively - leveraging their advanced natural language processing and visual imaging capabilities to perform surgical procedures with enhanced precision and efficiency. LLMs, due to their advanced natural language processing and visual imaging capabilities, have the potential to assist surgeons in performing procedures with greater precision and efficiency. In the context of specimen processing after breast cancer surgery, researchers have employed ChatGPT to predict surgical margin status using intraoperative mammography images. In a cohort of 100 patients, ChatGPT achieved an accuracy of 84.0% in predicting margin status, with a sensitivity of 60.0% for identifying incomplete excision, a specificity of 86.7% for identifying complete excision, a positive predictive value of 33.3%, and a negative predictive value of 95.1%. These findings indicate that ChatGPT demonstrates moderate accuracy in confirming complete excision but exhibits limited reliability in detecting positive margins[95]. Other investigators have employed AI recognition models to facilitate pancreatic identification during robot-assisted gastrectomy. On 80 test images, the HRNet model achieved precision, recall, intersection over union (IoU), and Dice coefficients of 0.70, 0.59, 0.46, and 0.61, respectively [96]. Luo et al. utilized AI-based models to measure tumor infiltration depth, achieving high sensitivity, specificity, and accuracy [97]. These models can assist clinicians in selecting minimally invasive surgical approaches, thereby optimizing therapeutic decision-making for addressing tumor invasion.

Although the aforementioned applications demonstrate the potential of LLMs in surgical assistance, their seamless real-time integration into dynamic surgical workflows remains hindered by a series of formidable technical and practical challenges. First, system latency constitutes a critical issue. Surgical decisions typically require completion within seconds or even subseconds, and any significant delays introduced by data transmission, model inference, or result output could disrupt the surgi-

cal rhythm and potentially lead to misjudgment. Second, system interoperability represents a substantial barrier. Modern operating rooms encompass diverse hardware from different vendors (such as endoscopes, robotic systems, and vital sign monitors) and software systems (such as picture archiving and communication systems and electronic medical records). LLMs must achieve seamless and stable data integration and communication with these heterogeneous systems, which poses substantial challenges at both technical and standardization levels. Finally, validating real-time decision support systems in dynamic and heterogeneous surgical environments is extraordinarily complex. The anatomical variations, tissue deformations, and unexpected complications encountered during surgery are inherently unpredictable, and ensuring that LLMs maintain high reliability, robustness, and safety across various edge cases and high-pressure scenarios requires extraordinarily rigorous and large-scale clinical validation processes that surpass traditional software verification. Addressing these challenges of latency, interoperability, and dynamic validation represents the critical gap that LLMs must bridge to transition from proof-of-concept studies to routine clinical applications.

4. Applications of LLMs in Supportive Care for Oncology Patients

Oncology patients face multifaceted challenges during treatment, including not only the direct effects of the disease but also significant psychological distress, social burdens, and the necessity to adapt to substantial lifestyle modifications. As outlined in Figure 2, the primary applications of LLMs in supportive care include symptom management, providing psychological support, and delivering personalized educational resources for patients. Supportive care is a comprehensive approach aimed at enhancing patients' quality of life, mitigating suffering, and addressing multidimensional patient needs across physical, psychological, social, and spiritual domains. LLMs, leveraging their advanced natural language processing capabilities, have the potential to deliver personalized, timely, and comprehensive supportive care for oncology patients across multiple domains, potentially leading to significant improvements in treatment experiences and clinical outcomes.

Primarily, LLMs can function as intelligent information assistants for patients, offering accurate, reliable, and comprehensible information regarding their diagnoses and therapeutic options. Oncology patients and their caregivers often face the challenge of assimilating vast amounts of medical information, encompassing but not limited to disease diagnosis, treatment modalities, potential adverse effects, and rehabilitative care strategies. Confronted with an array of technical terminology and intricate medical concepts, patients frequently experience confusion and anxiety. LLMs have the capacity to elucidate complex medical information in layman's terms, tailored to individual patient circumstances. This capability allows for the clarification of patient inquiries, enhancement of their comprehension regarding their condition and treatment protocols, potentially leading to improved treatment adherence and patient engagement.

Furthermore, LLMs have the potential to augment healthcare professionals in patient management, potentially leading to significant improvements in clinical efficiency. LLMs can potentially aid healthcare professionals in the systematic collection and organization of patient medical records, real-time monitoring of treatment progress, and delivery of foundational medical education to patients. This application has the potential to not only alleviate the workload of healthcare professionals but also enhance clinical efficiency, potentially facilitating more timely and effective patient care. Lastly, given the constraints on physicians' ability to address all patient inquiries in real-time, LLMs could potentially serve as adjunct medical assistants, providing prompt responses to patients' general questions. This application may help alleviate communication burdens between healthcare providers and patients, potentially facilitating a more streamlined treatment process.

4.1 Providing Symptom Management and Survivorship Care Guidance

Cancer treatment is typically a prolonged process, wherein symptom control and patient prognosis management are crucial components that not only affect patients' quality of life but also may directly impact treatment outcomes. LLMs have demonstrated significant potential in this area, assisting medical teams in developing personalized treatment plans and providing patients with prognostic management advice and guidance [98]. LLMs have demonstrated considerable potential for improving the quality of prognostic care for cancer patients. These models can provide personalized health education for patients and predict disease progression trends based on specific patient data, thus facilitating the development of more precise treatment and care strategies. The integrated application of these functions has the potential to significantly improve patient treatment outcomes and quality of life [94].

However, when providing care recommendations, LLMs' responses may occasionally deviate from National Comprehensive Cancer Network (NCCN) guidelines, potentially recommending pharmaceutical interventions not endorsed by these guidelines. Nevertheless, LLMs can still offer concise and readily comprehensible supportive care recommendations. Future development should focus on integrating LLMs with authoritative guidelines to generate more accurate and reliable recommendations [99]. In radiation oncology patient care, studies have demonstrated that LLMs can employ language comparable to human experts to deliver accurate, comprehensive, and concise answers at a high professional level, providing patients with more intelligible and implementable care guidance [100].

Notably, LLMs have exhibited consistently high performance across diverse types of cancer and tumor prognostic care planning, suggesting their broad applicability. In the context of prognostic care support for gynecological tumors, researchers found that LLMs can identify post-care scenarios, propose relevant preventive measures and care recommendations, and generate personalized advice based on patient-specific factors. However, it is important to note that they cannot yet supplant expert panels [101]. In the realm of supportive care for head and neck cancer patients, researchers utilized ChatGPT-4 to provide prognostic support and determined it to be highly accurate in delivering evidence-based information and guidance for symptom management and survivorship care [102]. In the context of colorectal cancer, LLMs demonstrated proficiency in addressing questions pertaining to patient prognostic care [103]. In the field of lung cancer, LLMs' care plan recommendations for researcher-provided cases aligned with established guidelines in terms of scope and nature, accurately prioritizing oxygenation and ventilation needs. This demonstrates their potential value as decision support tools for optimizing cancer care documentation [104].

Ultimately, discussing prognosis-related privacy issues with patients presents a significant challenge for physicians. Hanai et al. assessed LLMs' capabilities in addressing cancer survivors' sexual health concerns and observed that the models' content was policy-restricted, with responses biased towards non-pharmacological interventions, potentially fostering patients' misplaced expectations regarding these approaches [105].

4.2 Patient Education and Empowerment

LLMs were originally developed as question-answering systems engineered to tackle a wide range of complex problems [106]. In the realm of clinical oncology, LLMs can be utilized for patient education, facilitating informed consent processes for treatments, and aiding patients in overcoming reading comprehension barriers, thus offering comprehensive support for cancer patients throughout their treatment trajectory. Patients frequently encounter difficulties in comprehending information on cancer-related websites owing to the prevalence of technical terminology and the complexity of the content. LLMs have the capability to transform this specialized information into patient-friendly formats while preserving content accuracy, thereby significantly enhancing the readability of relevant materials [107]. Moreover, LLMs can facilitate patients in surmounting language barriers by not merely translating various medical guidelines into different languages but also by simplifying technical terms and medical jargon, thus enhancing patients' comprehension and acceptance of this information [108]. Emerging research suggests that LLMs, such as ChatGPT, exhibit considerable promise in promoting health behaviors among cancer patients. Nevertheless, these applications continue to confront challenges in terms of regulation, reliability, and privacy protection, underscoring the necessity for further refinement and enhancement of the underlying technologies [109].

Regarding patient empowerment, while LLMs exhibit accuracy comparable to physicians in their responses, inaccuracies persist, and patients still prefer interacting with doctors while voicing concerns about information provided by LLMs. This finding underscores the critical role of human interaction in patient empowerment [110]. Moreover, LLMs can be employed in patient education across various cancer types. In the context of prostate cancer, LLMs can offer patients essential information, albeit with scope for optimization [111]. To enhance patient education outcomes, further refinements in readability and diversification of information sources are imperative [112]. LLMs can function as "reviewers" in evaluating the quality of online medical videos pertaining to urological cancers. While LLMs demonstrate efficacy in evaluating the medical quality of video content, their assessment results warrant cautious interpretation due to low reproducibility [113].

In addressing questions related to urethral malignancies, LLMs provide high-quality information, potentially enhancing healthcare system efficiency and reducing future medical costs. However, LLMs presently lack the capacity to replace urologists [114,115,116]. In the context of laryngeal cancer, LLMs can function as a reliable information source for patients seeking knowledge. A comparative study revealed that among ChatGPT 3.5, ChatGPT4, and Bard, ChatGPT 3.5 yielded the most reliable and safe responses of the evaluated models [117]. In the realm of breast cancer, LLMs can offer screening recommendations to patients. Despite inherent limitations, this process can be implemented under physician supervision, thereby supporting doctors in breast cancer screening [118].

In the management of thyroid nodules, LLMs can aid physicians in patient care. While they are not suitable as the primary source of medical information, they can function as supplementary tools [119]. However, regarding oral cancer, LLMs' responses may be insufficient to educate patients and potentially lead to misinformation. Moreover, the readability of LLM-generated content exceeds the complexity of recommended patient materials [120]. Concerning the early detection of oral cancer, although LLMs show promise, certain content may be potentially misleading [121]. In addressing rare complications, such as breast implant-associated anaplastic large cell lymphoma, a comparative analysis of two LLMs (ChatGPT and Google) revealed that ChatGPT provided more comprehensive and informed responses, underscoring its significant potential in patient education [122].

Beyond providing educational content, LLMs hold potential for promoting patient compliance and self-management in cancer care by delivering personalized guidance on medication adherence, symptom monitoring, and lifestyle modifications. Emerging evidence suggests that LLMs can support health literacy improvement and provide individualized recommendations across diet, physical activity, and psychological support domains. However, these potential benefits require rigorous validation in real-world clinical settings. Current evidence remains largely limited to proof-of-concept studies, with concerns regarding response accuracy, patient over-reliance without clinical oversight, and the absence of data on long-term impact on hard clinical endpoints such as treatment completion rates and survival outcomes. Prospective, workflow-integrated studies with robust safety monitoring are therefore urgently needed before these tools can be widely deployed[94,107,109].

4.3 Responses to Medical-Related Questions

The utilization of LLMs in addressing cancer-related medical inquiries not only enhances patient health education but also establishes a robust foundation for the development and implementation of subsequent treatment strategies. In response to queries regarding immune-related adverse events (irAEs), LLMs typically offer accurate and reliable information, demonstrating a low incidence of erroneous content[123]. Nevertheless, pending further improvements in LLMs’ accuracy, clinicians must rigorously adhere to established clinical guidelines and standards to ensure the provision of accurate and reliable information to patients [123]. Regarding inquiries related to genetic testing and counseling for gynecological malignancies, LLMs exhibit high accuracy, particularly in domains such as genetic syndromes, genetic testing methodologies, and genetic counseling protocols. These findings indicate that LLMs possess significant potential for application in the field of genetic counseling. However, to further augment their reliability and clinical utility, LLMs’ performance requires optimization through training with more comprehensive, high-quality datasets [124]. In addressing inquiries concerning the application of complementary and integrative medicine in oncological treatment, LLMs generate responses of moderate quality, exhibiting relatively objective and unbiased characteristics with minimal misleading content. A significant limitation, however, is the current inability of LLMs to directly cite or reference specific scientific literature, which presents a substantial obstacle to their further development and application in the medical domain. Consequently, pending more comprehensive and systematic validation of LLMs, oncology patients should refrain from relying on their output as a primary basis for clinical decision-making [125].

5. Applications of Large Language Models in Tumor Prognosis Prediction

5.1 Predicting Survival Rates

In contemporary clinical oncology, particularly in the management of advanced malignancies, the limitations of current medical technologies have prompted a paradigm shift in treatment objectives[126,127,128,129]. Medical teams now focus on maximizing patient survival duration and enhancing quality of life, rather than pursuing a complete cure[130,131,132]. Within this evolving landscape, LLMs, as sophisticated artificial intelligence tools, have emerged as pivotal instruments in predicting patient survival rates. As elucidated in Figure 2, the applications of LLMs in prognostic prediction primarily encompass forecasting potential metastatic patterns, evaluating treatment efficacy, and estimating patient survival probabilities. Researchers have pioneered the integration of LLMs with multimodal prediction techniques, harnessing the capacity of LLMs to extract critical data from patient clinical information. This approach employs multimodal prediction models to forecast the five-year survival rate of bladder cancer patients. The results are promising, demonstrating that the performance of this method is comparable to, and in some aspects surpasses, traditional manual information extraction models. This outcome underscores the potential of LLMs in the field of tumor prognosis prediction [133]. In a seminal study, researchers leveraged the advantages of LLMs to develop a novel method for predicting metastatic breast cancer. This method not only exhibited superior predictive accuracy but also yielded new insights for the development of personalized cancer treatment strategies [134]. These state-of-the-art studies provide compelling evidence that LLMs offer significant advantages in predicting patient survival rates, particularly when integrated with various advanced artificial intelligence technologies, including machine learning, deep learning, and multimodal data analysis. This integration not only enhances the accuracy of predictions but also augments the interpretability and adaptability of the models. This multidisciplinary integrated approach substantially expands the scope of LLM applications in clinical oncology, offering robust technical support for precision medicine and the development of personalized treatment plans.

Despite the significant potential demonstrated by LLMs in tumor prognosis prediction, it is crucial to acknowledge that LLMs currently face critical limitations when addressing questions related to cancer prognosis. Janopaul-Naylor et al. conducted a comprehensive assessment of LLMs' performance in addressing prognosis-related questions for breast, colon, lung, and prostate cancers. The study findings indicated that the quality of LLMs' responses necessitates continuous improvement. Notably, in their current state, LLMs' responses have the potential to mislead patients and potentially induce unwarranted emotional distress [135]. This finding underscores the necessity for heightened caution when implementing LLMs in clinical practice and accentuates the imperative for further optimization and refinement of this technology. In the realm of predicting disease progression and survival rates, LLMs have demonstrated potential as a novel class of biomedical simulators. This innovative application empowers LLMs to perform comprehensive analyses of the genetic characteristics and progression of cancer, thereby substantially enhancing the accuracy of prediction tasks [136]. This advancement not only paves the way for novel approaches in cancer research but also offers robust support for the development of more precise treatment strategies.

5.2 Treatment Response Prediction

Beyond survival rate prediction, LLMs demonstrate the capability to forecast drug synergies in patients, evaluate the effects of pharmaceuticals on tumor tissues, and furnish critical experimental data to support subsequent drug development. Moreover, LLMs can be utilized to prognosticate the therapeutic efficacy of radiotherapy and chemotherapy in patients, offering valuable insights for informed clinical decision-making. Li et al. developed a few-shot learning-based methodology leveraging LLMs to predict drug pair synergies in rare tissues characterized by a paucity of structured data and features. The research findings demonstrate that LLM-based prediction models can achieve substantial accuracy even with extremely limited sample sizes. This discovery not only facilitates the prediction of drug synergies in rare tissues under data-scarce conditions but also broadens the potential applications of LLMs in biomedical reasoning tasks [137]. Researchers have also discovered that LLMs exhibit the capacity to accurately deduce cancer disease responses from radiology reports. These models can be employed to evaluate progression-free survival rates of tumors and, through the implementation of specific prompts, can substantially reduce the volume of required training data. It is noteworthy that the models' high accuracy is intricately linked to the quality of the training data [138].

6. Future Prospects

Despite the numerous challenges encountered in applying LLMs to oncology, their potential is immense, and the prospects for development are extensive. To address current limitations, this review explores future directions for LLMs in oncology from the following perspectives (Figure 4).

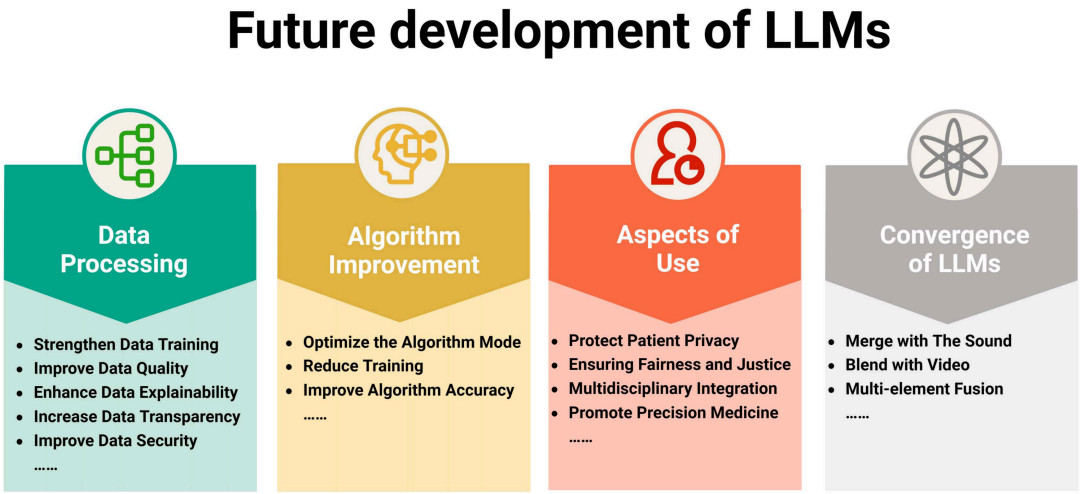


Figure 4. Future Development of LLMs

The development prospects of LLMs are extensive, primarily encompassing four aspects: data processing, algorithm enhancement, user experience, and multimedia integration. Regarding data processing, LLMs can enhance data training, improve output quality, and ensure data interpretability and security. Algorithm enhancement, including optimization, reduction of data-dependent training, and improvement of accuracy, represents a major direction for future LLM development. Concerning user experience, ensuring patient privacy and fairness is essential, while multidisciplinary integration and precision medicine represent additional development directions for LLMs. In terms of multimedia integration, future LLMs may extend beyond current image and text fusion to incorporate audio and video, facilitating more seamless interactive experiences.

6.1 Enhancement of Data Quality and Model Performance

As the volume of oncology-related big data continues to expand and its quality improves, the performance and generalization capabilities of LLMs in oncology are expected to significantly enhance. Future research should focus on establishing more comprehensive and accurate oncology databases while concurrently developing more effective data annotation methods to enhance data quality and usability. These advancements are anticipated to not only enhance LLMs' performance in managing rare cancers and complex cases but also contribute to more precise diagnoses and the formulation of personalized treatment plans. Furthermore, with ongoing advancements in computational power and continuous algorithm optimization, LLMs have the potential to enable real-time medical decision support systems, thereby assisting clinicians in making more rapid and accurate diagnostic and treatment decisions in complex medical scenarios. Most significantly, by integrating multimodal data (including genomics, proteomics, imaging, and clinical information), LLMs will be capable of generating more personalized diagnostic and treatment plans for individual patients, thus not only advancing the field of precision medicine but also potentially improving treatment efficacy and patient prognosis substantially.

6.2 Enhancement of Interpretability, Transparency, and Security

The ongoing advancements in Explainable AI (XAI) technology are poised to significantly enhance the transparency and credibility of LLMs in the medical domain. Future research endeavors should focus on developing more transparent and interpretable AI models while concurrently exploring methodologies to seamlessly integrate XAI techniques with LLMs to enhance model interpretability and credibility in clinical applications [139]. These studies are expected to not only enhance the transparency of medical decision-making but also facilitate clinicians in comprehending and validating the model's reasoning process, thereby augmenting the credibility and acceptance of LLMs in clinical applications. Concurrently, enhancing the security of LLMs in medical applications and developing more robust defense mechanisms against potential cybersecurity threats, data tampering, and privacy breaches represents a crucial direction for future research [140]. These security measures are anticipated to not only safeguard the rights and safety of medical institutions, healthcare professionals, and patients but also establish a robust foundation of trust for the widespread implementation of LLMs in clinical practice. Furthermore, future research should prioritize the development of LLMs with automated citation generation and information source verification capabilities, and establish robust mechanisms to ensure these models can dynamically update with the latest medical knowledge and ethical guidelines in real time. The advancement of explainable intraoperative guidance technologies facilitates the development of a transparent intelligent surgical ecosystem, in which future LLMs will be seamlessly integrated into real-time surgical decision-making processes, thereby enabling dynamic risk warning and operative guidance through multimodal perception (endoscopic imaging, instrument motion trajectories, and vital sign data). The core breakthrough lies in overcoming the limitations of "black-box decision-making": for instance, in thyroid radiofrequency ablation procedures, the system can analyze neurovascular bundle distribution in ultrasound images in real-time, overlay millimeter-level safety margin prompts through augmented reality interfaces, and simultaneously generate decision rationale reports. This transparent guidance not only constrains intraoperative navigation errors to millimeter-level precision but also reshapes the foundation of physician-patient trust through the visualized interpretation of decision processes, thus providing an intelligent collaborative paradigm that combines precision and safety for complex surgical procedures. Additionally, automated surgical report generation represents a significant developmental trajectory for future LLMs. The system integrates real-time surgical field visualization results, monitors intraoperative patient vital signs, comprehensively records surgical procedures, and generates personalized postoperative recovery plans based on individual patient conditions, thus establishing a closed-loop knowledge accumulation cycle from surgical documentation to procedural innovation and ultimately providing a continuously evolving data infrastructure for surgical precision medicine. These advancements are projected to significantly enhance the reliability, timeliness, and traceability of information provided by LLMs in the medical domain, thereby fortifying their role in clinical decision support.

6.3 Strengthening Medical Ethics and Privacy Protection

As the implementation of LLMs in the medical domain becomes increasingly pervasive, associated ethical and privacy concerns are gaining prominence [141]. Future research should concentrate on formulating comprehensive ethical guidelines for medical artificial intelligence, concurrent with the development of more secure AI technologies that can effectively safeguard privacy (Table 1). As summarized in Table 1, multiple regulatory and ethical frameworks have been proposed to address these challenges, including FDA regulations governing AI/ML-based medical devices, the European Union's Artificial Intelligence Act, evaluation standards grounded in the STAGER and QUEST frameworks, and performance-based licensing models, thereby establishing a foundation for future governance. These measures are expected to facilitate the effective utilization of data for model training and application while comprehensively safeguarding patient privacy and ensuring consistency and equity in the output results obtained by diverse users (e.g., clinicians, patients, and researchers) when employing LLMs. Moreover, it is imperative to conduct in-depth investigations into maximizing the value of data sharing and utilization while ensuring robust protection of personal privacy, thereby facilitating advancements in medical research. This endeavor may encompass the development of advanced data anonymization techniques, refinement of federated learning methodologies, and the establishment of highly secure cross-institutional and cross-regional data sharing platforms. These initiatives are anticipated to provide crucial safeguards for the sustainable and ethical development of LLMs in the medical domain, while concomitantly fostering public confidence and acceptance of AI technologies in medical applications. Beyond enhanced patient privacy protection requirements, the proprietary nature of surgical videos constrains the application of traditional centralized

model training approaches, requiring healthcare institutions to retain original videos locally while sharing only model parameter updates (such as gradient information) through encrypted transmission for global model training participation. This paradigm maintains hospital ownership of proprietary video data while enhancing model performance through collaborative efforts, thereby facilitating the establishment of cross-institutional surgical AI infrastructure.

Table 1. Regulatory/ethical framework for LLMs

Author	Regulatory/ethical frameworks	Reference
Iván Miguel García-López	The IEEE's Ethically Aligned Design Bostrom's AI alignment theory	[142]
Geeta Joshi	FDA Regulation of AI/ML-Based Medical Devices	[143]
Aryan Jadon	Health Insurance Portability and Accountability Act	[144]
Armin ZadZiabari	General Data Protection Regulation	[145]
WHO and IMDRF	WHO AI Ethics Guidelines and IMDRF Regulatory Frameworks for AI/ML	[145]
Chen J	STAGER checklist	[146]
Tam TYC	QUEST framework	[147]
Advait Deshpande	Responsible AI Systems	[148]
Danish Contractor	Behavioral use licensing	[149]
The Lancet Regional Health-Europe	high standards of patient care and data security	[150]
The European Union	EU's AI Act	[150]

LLMs: Large Language Models; IEEE: Institute of Electrical and Electronics Engineers; FDA: Food and Drug Administration; AI: Artificial Intelligence; ML: Machine Learning; WHO: World Health Organization; IMDRF: International Medical Device Regulators Forum; EU: European Union

6.4 Expansion and Innovation in Interdisciplinary Applications

LLMs show considerable promise as key tools for integrating knowledge from multiple disciplines, including but not limited to biology, medicine, and pharmacology. The future applications of LLMs in oncology are anticipated to be more comprehensive, sophisticated, and innovative. In the realm of intelligent clinical trial design, LLMs are poised to play pivotal roles in optimizing trial protocols, predicting outcomes, identifying eligible participants, and assessing feasibility. These advancements have the potential to significantly expedite the process of new drug development and treatment method validation. This approach has the potential to not only substantially enhance the efficiency of clinical trials but also effectively reduce associated costs, accelerate the introduction of innovative therapies, and ultimately benefit a broader patient population. Regarding patient education and engagement enhancement, the development of more sophisticated, user-friendly, and personalized patient interaction systems utilizing LLMs has the potential to significantly improve patients' medical literacy, treatment adherence, and self-management capabilities. These advanced interaction systems can offer highly tailored health recommendations, disease management strategies, and psychological support based on individual patient profiles and needs. Such personalized approaches have the potential to comprehensively enhance patients' quality of life, improve treatment adherence, and potentially lead to superior clinical outcomes. Moreover, LLMs are expected to play crucial roles in the systematic analysis of medical literature, formulation of drug repositioning strategies, prediction and validation of molecular targets, and various other domains. These applications have the potential to significantly advance interdisciplinary innovation in cancer research and provide new impetus for the development of precision medicine. Furthermore, within surgical oncology, variability in surgical annotation quality represents a critical challenge for LLMs, as researchers with varying levels of expertise and experience may interpret identical surgical videos differently, resulting in suboptimal model training outcomes. To address this challenge, implementing enhanced model training protocols with timely expert intervention from domain specialists when models exhibit discordance regarding surgical components can correct identified issues and establish a foundation for future advancements in surgical oncology.

6.5 Addressing Real-World Implementation Challenges

Despite the promising prospects of LLMs in both technical and ethical dimensions, seamless integration into real-world surgical oncology practice necessitates overcoming significant practical implementation barriers. This review identifies three major categories of challenges that necessitate collaborative stakeholder efforts for resolution.

First, the heterogeneity of hospital information infrastructure constitutes a fundamental barrier. LLM performance is highly dependent on the quality, structure, and interoperability of electronic health record systems. Disparities in data digitization levels between high-resource academic centers and community or rural hospitals may create a "digital divide," thus exacerbating existing healthcare inequalities. Successful implementation necessitates substantial investment in data standardization (e.g., adoption of Fast Healthcare Interoperability Resources standards) and the development of middleware capable of interfacing with diverse and often legacy hospital IT systems[151].

Second, the regulatory and approval landscape for AI-based clinical decision support systems remains complex and continuously evolving[152]. Regulatory agencies such as the FDA are continually refining their frameworks for medical device software. The "black box" nature of certain LLM decisions complicates the fulfillment of validation and interpretability requirements necessary for regulatory approval. Furthermore, establishing an efficient yet rigorous process for the continual evaluation and approval of iteratively updated LLM versions without stifling innovation represents a critical challenge that regulatory agencies and developers must address collaboratively.

Third, substantial training costs and workflow integration requirements represent another layer of barriers[153]. Beyond the computational overhead of training and fine-tuning models, the hidden costs associated with deploying and maintaining these systems in clinical environments are considerable. These costs include the integration of LLMs into clinical workflows and, most importantly, the training of clinical personnel to use these tools effectively and safely. Failure to budget for and effectively execute comprehensive change management and training programs may result in poor adoption rates and implementation failure, regardless of the technical sophistication of the LLM.

Addressing these barriers necessitates a multi-stakeholder collaborative approach involving hospital administrators, medical informaticists, regulatory experts, and clinicians. Future pilot implementations should specifically document and report these non-technical challenges to provide more realistic roadmaps for large-scale deployment.

6.6 Standardized Evaluation Systems for Clinical LLM Deployment

Finally, the widespread clinical adoption of LLMs in oncology will require the establishment of standardized, multidimensional evaluation systems that systematically assess safety, interpretability, work efficiency, patient outcomes, cost-effectiveness, and fairness. Several frameworks have recently been proposed to address different aspects of this challenge, including the TRIPOD-LLM reporting guideline for LLM studies[154], the MI-CLEAR-LLM checklist for accuracy reporting[155], the DECIDE-AI guideline for early-stage clinical evaluation of AI decision support systems[156], the STAGER checklist for generative AI reliability assessment[146], and the Clinical Safety-Effectiveness Dual-Track Benchmark encompassing 30 metrics across safety and effectiveness domains[157]. However, no single framework currently addresses all six dimensions comprehensively, and future efforts should focus on developing integrated, oncology-specific evaluation pathways to ensure that LLMs can be deployed safely, effectively, and equitably in clinical practice.

7. Abbreviation

AI, Artificial intelligence

BsAbs, Bispecific antibodies

CLE, Confocal laser endomicroscopy

CPG, Cancer predisposition gene

ESMO, European Society for Medical Oncology

ELCAP, ESMO Guidance on the Use of Large Language Models in Clinical Practice

EHRs, Electronic Health Records

irAEs, immune-related adverse events

LLMs, Large Language Models

ML, machine learning

MRI, magnetic resonance imaging

NCCN, National Comprehensive Cancer Network

NLP, natural language processing

PDAC, pancreatic ductal adenocarcinoma

WOS, Web of Science

XAI, Explainable AI

8. Acknowledgements

Not applicable

9. Funding

Not applicable

10. Authors' contributions

Writing-original draft, Q.L.X., A.Q.L.

Conceptualization, A.Q.L.

Investigation, Q.L.X., A.Q.L.

Writing-review and editing, Q.L.X., A.Q.L.

Visualization, Q.L.X., A.Q.L.

All authors have read and agreed to the published version of the manuscript.

11. Competing interests

Anqi Lin is a member of the iCancer Editorial Board and the corresponding author of this article. Anqi Lin was not involved in the peer-review process or editorial decision-making for this manuscript.

12. Availability of supporting data

Not applicable.

13. Ethical Approval and Consent to participate

Not Applicable

14. Consent for publication

Not applicable.

15. Provenance and peer review

Not commissioned, externally peer-reviewed

16. References

[1] Hao Y, Qiu Z, Holmes J, Löckenhoff CE, Liu W, Ghassemi M, et al. Large language model integrations in cancer decision-making: a systematic review and meta-analysis. NPJ Digit Med. 2025;8:450. https://doi.org/10.1038/s41746-025-01824-7

[2] Rengers TA, Thiels CA, Salehinejad H. Academic Surgery in the Era of Large Language Models: A Review. JAMA Surg. 2024;159:445–50. https://doi.org/10.1001/jamasurg.2023.6496

[3] Zhou J, Cheng Y, He S, Chen Y, Chen H. Large Language Models for Transforming Healthcare: A Perspective on DeepSeek-R1. MedComm – Future Medicine. 2025;4:e70021. https://doi.org/10.1002/mef2.70021

[4] Murmu A, Győrffy B. Artificial intelligence methods available for cancer research. Front Med. 2024;18:778–97. https://doi.org/10.1007/s11684-024-1085-3

[5] Gao Y, Baptista-Hon DT, Zhang K. The inevitable transformation of medicine and research by large language models: The possibilities and pitfalls. MedComm – Future Medicine. 2023;2:e49. https://doi.org/10.1002/mef2.49

[6] Lai S, Liu H, Chu T, Cao C. HPV Integration: Moving From Carcinogenesis Mechanisms to Clinical Applications. Med Research. 2025;mdr2.70026. https://doi.org/10.1002/mdr2.70026

[7] Fonseca VC, Sidiropoulou Z. Geriatric Breast Cancer: Staging, Molecular Surrogates, and Treatment. A Review & Meta-analysis. Aging Dis. 2024;15:1602–18. https://doi.org/10.14336/AD.2023.1002

[8] Yan J, Xiao G, Yang C, Liu Q, Lv C, Yu X, et al. Cancer-Associated Fibroblasts Promote Lymphatic Metastasis in Cholangiocarcinoma via the PDGF-BB/PDGFR-β Mediated Paracrine Signaling Network. Aging Dis. 2024;15:369–89. https://doi.org/10.14336/AD.2023.0420

[9] Nema S, Vachhani L. Surgical instrument detection and tracking technologies: Automating dataset labeling for surgical skill assessment. Front Robot AI. 2022;9:1030846. https://doi.org/10.3389/frobt.2022.1030846

[10] Noël J, Moschovas MC, Patel E, Rogers T, Marquinez J, Rocco B, et al. Step-by-step optimisation of robotic-assisted radical prostatectomy using augmented reality. Int Braz J Urol. 2022;48:600–1. https://doi.org/10.1590/S1677-5538.IBJU.2022.9910

[11] Suksatan W. Who Is Responsible When AI Gets Cancer Information Wrong? Implications for Patient Education. J Canc Educ [Internet]. 2026 [cited 2026 Sep 5]; https://doi.org/10.1007/s13187-026-02952-8

[12] Mathes S, Ferber D, Dreyer T, Borm KJ, Modersohn L, Willem T, et al. Collaborative framework on responsible AI in LLM-driven CDSS for precision oncology leveraging real-world patient data. npj Precis Onc. Nature Publishing Group; 2025;10:15. https://doi.org/10.1038/s41698-025-01180-5

[13] Wong EYT, Verlingue L, Aldea M, Franzoi MA, Umeton R, Halabi S, et al. ESMO guidance on the use of Large Language Models in Clinical Practice (ELCAP). Ann Oncol. 2025;36(12):1447-1457. https://doi.org/10.1016/j.annonc.2025.09.001

[14] Agha RA, Mathew G, Rashid R, Kerwan A, Al-Jabir A, Sohrabi C, Franchi T, Nicola M, Agha M; TITAN Group. Transparency in the Reporting of Artificial Intelligence - The TITAN Guideline. Premier Journal of Science. 2025;10:100082. https://doi.org/10.70389/PJS.100082

[15] Osorio P, Jimenez-Perez G, Montalt-Tordera J, Hooge J, Duran-Ballester G, Singh S, et al. Latent Diffusion Models with Image-Derived Annotations for Enhanced AI-Assisted Cancer Diagnosis in Histopathology. Diagnostics (Basel). 2024;14:1442. https://doi.org/10.3390/diagnostics14131442

[16] Wang J, Ye Q, Liu L, Lan Guo N, Hu G. Bioinformatics Illustrations Decoded by ChatGPT: The Good, The Bad, and The Ugly. bioRxiv. 2023;2023.10.15.562423. https://doi.org/10.1101/2023.10.15.562423

[17] Rastogi A, Ashwini NS, Rath I, Bihari C, Sasturkar SV, Pamecha V. Utility and diagnostic accuracy of intraoperative frozen sections in hepato-pancreato-biliary surgical pathology. Langenbecks Arch Surg. 2023;408:390. https://doi.org/10.1007/s00423-023-03124-8

[18] Laohawetwanit T, Apornvirat S, Namboonlue C. Thinking like a pathologist: Morphologic approach to hepatobiliary tumors by ChatGPT. Am J Clin Pathol. 2024;aqae087. https://doi.org/10.1093/ajcp/aqae087

[19] Sushil M, Zack T, Mandair D, Zheng Z, Wali A, Yu Y-N, et al. A comparative study of large language model-based zero-shot inference and task-specific supervised classification of breast cancer pathology reports. J Am Med Inform Assoc. 2024;ocae146. https://doi.org/10.1093/jamia/ocae146

[20] Huang J, Yang DM, Rong R, Nezafati K, Treager C, Chi Z, et al. A critical assessment of using ChatGPT for extracting structured data from clinical notes. NPJ Digit Med. 2024;7:106. https://doi.org/10.1038/s41746-024-01079-8

[21] Truhn D, Loeffler CM, Müller-Franzes G, Nebelung S, Hewitt KJ, Brandner S, et al. Extracting structured information from unstructured histopathology reports using generative pre-trained transformer 4 (GPT-4). J Pathol. 2024;262:310–9. https://doi.org/10.1002/path.6232

[22] Malik S, Zaheer S. ChatGPT as an aid for pathological diagnosis of cancer. Pathol Res Pract. 2024;253:154989. https://doi.org/10.1016/j.prp.2023.154989

[23] Dehdab R, Brendlin A, Werner S, Almansour H, Gassenmaier S, Brendel JM, et al. Evaluating ChatGPT-4V in chest CT diagnostics: a critical image interpretation assessment. Jpn J Radiol. 2024; https://doi.org/10.1007/s11604-024-01606-3

[24] Hu D, Liu B, Zhu X, Lu X, Wu N. Zero-shot information extraction from radiological reports using ChatGPT. Int J Med Inform. 2024;183:105321. https://doi.org/10.1016/j.ijmedinf.2023.105321

[25] Fink MA, Bischoff A, Fink CA, Moll M, Kroschke J, Dulz L, et al. Potential of ChatGPT and GPT-4 for Data Mining of Free-Text CT Reports on Lung Cancer. Radiology. 2023;308:e231362. https://doi.org/10.1148/radiol.231362

[26] Kanzawa J, Yasaka K, Fujita N, Fujiwara S, Abe O. Automated classification of brain MRI reports using fine-tuned large language models. Neuroradiology. 2024; https://doi.org/10.1007/s00234-024-03427-7

[27] Gu K, Lee JH, Shin J, Hwang JA, Min JH, Jeong WK, et al. Using GPT-4 for LI-RADS feature extraction and categorization with multilingual free-text reports. Liver Int. 2024;44:1578–87. https://doi.org/10.1111/liv.15891

[28] Chung EM, Zhang SC, Nguyen AT, Atkins KM, Sandler HM, Kamrava M. Feasibility and acceptability of ChatGPT generated radiology report summaries for cancer patients. Digit Health. 2023;9:20552076231221620. https://doi.org/10.1177/20552076231221620

[29] Wang Z, Zhang Z, Traverso A, Dekker A, Qian L, Sun P. Assessing the role of GPT-4 in thyroid ultrasound diagnosis and treatment recommendations: enhancing interpretability with a chain of thought approach. Quant Imaging Med Surg. 2024;14:1602–15. https://doi.org/10.21037/qims-23-1180

[30] Shifai N, van Doorn R, Malvehy J, Sangers TE. Can ChatGPT vision diagnose melanoma? An exploratory diagnostic accuracy study. J Am Acad Dermatol. 2024;90:1057–9. https://doi.org/10.1016/j.jaad.2023.12.062

[31] Cirone K, Akrout M, Abid L, Oakley A. Assessing the Utility of Multimodal Large Language Models (GPT-4 Vision and Large Language and Vision Assistant) in Identifying Melanoma Across Different Skin Tones. JMIR Dermatol. 2024;7:e55508. https://doi.org/10.2196/55508

[32] Zhang X, Talukdar N, Vemulapalli S, Ahn S, Wang J, Meng H, et al. Comparison of Prompt Engineering and Fine-Tuning Strategies in Large Language Models in the Classification of Clinical Notes. AMIA Jt Summits Transl Sci Proc. 2024;2024:478–87.

[33] Tay SB, Low GH, Wong GJE, Tey HJ, Leong FL, Li C, et al. Use of Natural Language Processing to Infer Sites of Metastatic Disease From Radiology Reports at Scale. JCO Clin Cancer Inform. 2024;8:e2300122. https://doi.org/10.1200/CCI.23.00122

[34] Bhayana R, Nanda B, Dehkharghanian T, Deng Y, Bhambra N, Elias G, et al. Large Language Models for Automated Synoptic Reports and Resectability Categorization in Pancreatic Cancer. Radiology. 2024;311:e233117. https://doi.org/10.1148/radiol.233117

[35] Son HJ, Kim S-J, Pak S, Lee SH. ChatGPT-assisted deep learning for diagnosing bone metastasis in bone scans: Bridging the AI Gap for Clinicians. Heliyon. 2023;9:e22409. https://doi.org/10.1016/j.heliyon.2023.e22409

[36] Chang PW, Amini MM, Davis RO, Nguyen DD, Dodge JL, Lee H, et al. ChatGPT4 Outperforms Endoscopists for Determination of Postcolonoscopy Rescreening and Surveillance Recommendations. Clin Gastroenterol Hepatol. 2024;S1542-3565(24)00429-4. https://doi.org/10.1016/j.cgh.2024.04.022

[37] Lim DYZ, Tan YB, Koh JTE, Tung JYM, Sng GGR, Tan DMY, et al. ChatGPT on guidelines: Providing contextual knowledge to GPT allows it to provide advice on appropriate colonoscopy intervals. J Gastroenterol Hepatol. 2024;39:81–106. https://doi.org/10.1111/jgh.16375

[38] Pereyra L, Schlottmann F, Steinberg L, Lasa J. Colorectal Cancer Prevention: Is Chat Generative Pretrained Transformer (Chat GPT) ready to Assist Physicians in Determining Appropriate Screening and Surveillance Recommendations? J Clin Gastroenterol. 2024; https://doi.org/10.1097/MCG.0000000000001979

[39] Atarere J, Naqvi H, Haas C, Adewunmi C, Bandaru S, Allamneni R, et al. Applicability of Online Chat-Based Artificial Intelligence Models to Colorectal Cancer Screening. Dig Dis Sci. 2024;69:791–7. https://doi.org/10.1007/s10620-024-08274-3

[40] Kerbage A, Kassab J, El Dahdah J, Burke CA, Achkar J-P, Rouphael C. Accuracy of ChatGPT in Common Gastrointestinal Diseases: Impact for Patients and Providers. Clin Gastroenterol Hepatol. 2024;22:1323-1325.e3. https://doi.org/10.1016/j.cgh.2023.11.008

[41] Mukherjee S, Durkin C, Pebenito AM, Ferrante ND, Umana IC, Kochman ML. Assessing ChatGPT's Ability to Reply to Queries Regarding Colon Cancer Screening Based on Multisociety Guidelines. Gastro Hep Adv. 2023;2:1040–3. https://doi.org/10.1016/j.gastha.2023.07.008

[42] Lombardo R, Gallo G, Stira J, Turchi B, Santoro G, Riolo S, et al. Quality of information and appropriateness of Open AI outputs for prostate cancer. Prostate Cancer Prostatic Dis. 2024; https://doi.org/10.1038/s41391-024-00789-0

[43] Lombardo R, Cicione A, Santoro G, De Nunzio C. ChatGPT in prostate cancer: myth or reality? Prostate Cancer Prostatic Dis. 2024;27:9–10. https://doi.org/10.1038/s41391-023-00750-7

[44] Lamy P-J. Conversation between a clinical biologist and an artificial intelligence on prostate cancer biomarkers: a critical reading. Ann Biol Clin (Paris). 2023;81:395–402. https://doi.org/10.1684/abc.2023.1829

[45] Chiarelli G, Stephens A, Finati M, Cirulli GO, Beatrici E, Filipas DK, et al. Adequacy of prostate cancer prevention and screening recommendations provided by an artificial intelligence-powered large language model. Int Urol Nephrol. 2024; https://doi.org/10.1007/s11255-024-04009-5

[46] Kuşcu O, Pamuk AE, Sütay Süslü N, Hosal S. Is ChatGPT accurate and reliable in answering questions regarding head and neck cancer? Front Oncol. 2023;13:1256459. https://doi.org/10.3389/fonc.2023.1256459

[47] Trager MH, Queen D, Bordone LA, Geskin LJ, Samie FH. Assessing ChatGPT responses to common patient queries regarding basal cell carcinoma. Arch Dermatol Res. 2023;315:2979–81. https://doi.org/10.1007/s00403-023-02705-3

[48] Horiuchi D, Tatekawa H, Shimono T, Walston SL, Takita H, Matsushita S, et al. Accuracy of ChatGPT generated diagnosis from patient's medical history and imaging findings in neuroradiology cases. Neuroradiology. 2024;66:73–9. https://doi.org/10.1007/s00234-023-03252-4

[49] Sensoy E, Citirik M. A comparative study on the knowledge levels of artificial intelligence programs in diagnosing ophthalmic pathologies and intraocular tumors evaluated their superiority and potential utility. Int Ophthalmol. 2023;43:4905–9. https://doi.org/10.1007/s10792-023-02893-x

[50] Szczesniewski JJ, Tellez Fouz C, Ramos Alba A, Diaz Goizueta FJ, García Tello A, Llanes González L. ChatGPT and most frequent urological diseases: analysing the quality of information and potential risks for patients. World J Urol. 2023;41:3149–53. https://doi.org/10.1007/s00345-023-04563-0

[51] Yeo YH, Samaan JS, Ng WH, Ting P-S, Trivedi H, Vipani A, et al. Assessing the performance of ChatGPT in answering questions regarding cirrhosis and hepatocellular carcinoma. Clin Mol Hepatol. 2023;29:721–32. https://doi.org/10.3350/cmh.2023.0089

[52] Cao JJ, Kwon DH, Ghaziani TT, Kwo P, Tse G, Kesselman A, et al. Large language models' responses to liver cancer surveillance, diagnosis, and management questions: accuracy, reliability, readability. Abdom Radiol (NY). 2024; https://doi.org/10.1007/s00261-024-04501-7

[53] Bellamkonda N, Farlow JL, Haring CT, Sim MW, Seim NB, Cannon RB, et al. Evaluating the Accuracy of ChatGPT in Common Patient Questions Regarding HPV+ Oropharyngeal Carcinoma. Ann Otol Rhinol Laryngol. 2024;34894241259137. https://doi.org/10.1177/00034894241259137

[54] Haver HL, Gupta AK, Ambinder EB, Bahl M, Oluyemi ET, Jeudy J, et al. Evaluating the Use of ChatGPT to Accurately Simplify Patient-centered Information about Breast Cancer Prevention and Screening. Radiol Imaging Cancer. 2024;6:e230086. https://doi.org/10.1148/rycan.230086

[55] Sievert M, Aubreville M, Mueller SK, Eckstein M, Breininger K, Iro H, et al. Diagnosis of malignancy in oropharyngeal confocal laser endomicroscopy using GPT 4.0 with vision. Eur Arch Otorhinolaryngol. 2024;281:2115–22. https://doi.org/10.1007/s00405-024-08476-5

[56] Odabashian R, Bastin D, Jones G, Manzoor M, Tangestaniapour S, Assad M, et al. Assessment of ChatGPT-3.5's Knowledge in Oncology: Comparative Study with ASCO-SEP Benchmarks. JMIR AI. 2024;3:e50442. https://doi.org/10.2196/50442

[57] Huang Y, Gomaa A, Semrau S, Haderlein M, Lettmaier S, Weissmann T, et al. Benchmarking ChatGPT-4 on a radiation oncology in-training exam and Red Journal Gray Zone cases: potentials and challenges for ai-assisted medical education and decision making in radiation oncology. Front Oncol. 2023;13:1265024. https://doi.org/10.3389/fonc.2023.1265024

[58] Yang F, Yan D, Wang Z. Large-Scale assessment of ChatGPT's performance in benign and malignant bone tumors imaging report diagnosis and its potential for clinical applications. J Bone Oncol. 2024;44:100525. https://doi.org/10.1016/j.jbo.2024.100525

[59] Frank J, Merseburger AS, Landmesser J, Brozat-Essen S, Schramm P, Freimann L, et al. [Large Language Models for Rapid Simplification of Quality Assurance Data Input: Field Trial with Real Data in the Context of Tumour Documentation in Urology]. Aktuelle Urol. 2024; https://doi.org/10.1055/a-2281-8015

[60] Bhattarai K, Oh IY, Sierra JM, Tang J, Payne PRO, Abrams Z, et al. Leveraging GPT-4 for identifying cancer phenotypes in electronic health records: a performance comparison between GPT-4, GPT-3.5-turbo, Flan-T5, Llama-3-8B, and spaCy's rule-based and machine learning-based methods. JAMIA Open. 2024;7:ooae060. https://doi.org/10.1093/jamiaopen/ooae060

[61] Liu J, Zhang Y, Wang K, Yavuz MC, Chen X, Yuan Y, et al. Universal and extensible language-vision models for organ segmentation and tumor detection from abdominal computed tomography. Med Image Anal. 2024;97:103226. https://doi.org/10.1016/j.media.2024.103226

[62] Choi HS, Song JY, Shin KH, Chang JH, Jang B-S. Developing prompts from large language model for extracting clinical information from pathology and ultrasound reports in breast cancer. Radiat Oncol J. 2023;41:209–16. https://doi.org/10.3857/roj.2023.00633

[63] Ge J, Li M, Delk MB, Lai JC. A Comparison of a Large Language Model vs Manual Chart Review for the Extraction of Data Elements From the Electronic Health Record. Gastroenterology. 2024;166:707-709.e3. https://doi.org/10.1053/j.gastro.2023.12.019

[64] Bryant AK, Zamora-Resendiz R, Dai X, Morrow D, Lin Y, Jungles KM, et al. Artificial intelligence to unlock real-world evidence in clinical oncology: A primer on recent advances. Cancer Med. 2024;13:e7253. https://doi.org/10.1002/cam4.7253

[65] Zhang X, Wang H, Sun C. BiSpec Pairwise AI: guiding the selection of bispecific antibody target combinations with pairwise learning and GPT augmentation. J Cancer Res Clin Oncol. 2024;150:237. https://doi.org/10.1007/s00432-024-05740-3

[66] Wang J, Ye Q, Liu L, Guo NL, Hu G. Scientific figures interpreted by ChatGPT: strengths in plot recognition and limits in color perception. NPJ Precis Oncol. 2024;8:84. https://doi.org/10.1038/s41698-024-00576-z

[67] Uprety D, Zhu D, West HJ. ChatGPT-A promising generative AI tool and its implications for cancer care. Cancer. 2023;129:2284–9. https://doi.org/10.1002/cncr.34827

[68] Sultan I, Al-Abdallat H, Alnajjar Z, Ismail L, Abukhashabeh R, Bitar L, et al. Using ChatGPT to Predict Cancer Predisposition Genes: A Promising Tool for Pediatric Oncologists. Cureus. 2023;15:e47594. https://doi.org/10.7759/cureus.47594

[69] Griewing S, Gremke N, Wagner U, Lingenfelder M, Kuhn S, Boekhoff J. Challenging ChatGPT 3.5 in Senology-An Assessment of Concordance with Breast Cancer Tumor Board Decision Making. J Pers Med. 2023;13:1502. https://doi.org/10.3390/jpm13101502

[70] Lukac S, Dayan D, Fink V, Leinert E, Hartkopf A, Veselinovic K, et al. Evaluating ChatGPT as an adjunct for the multidisciplinary tumor board decision-making in primary breast cancer cases. Arch Gynecol Obstet. 2023;308:1831–44. https://doi.org/10.1007/s00404-023-07130-5

[71] Griewing S, Knitza J, Boekhoff J, Hillen C, Lechner F, Wagner U, et al. Evolution of publicly available large language models for complex decision-making in breast cancer care. Arch Gynecol Obstet. 2024;310:537–50. https://doi.org/10.1007/s00404-024-07565-4

[72] Sorin V, Glicksberg BS, Artsi Y, Barash Y, Konen E, Nadkarni GN, et al. Utilizing large language models in breast cancer management: systematic review. J Cancer Res Clin Oncol. 2024;150:140. https://doi.org/10.1007/s00432-024-05678-6

[73] Sorin V, Klang E, Sklair-Levy M, Cohen I, Zippel DB, Balint Lahat N, et al. Large language model (ChatGPT) as a support tool for breast tumor board. NPJ Breast Cancer. 2023;9:44. https://doi.org/10.1038/s41523-023-00557-8

[74] Deng L, Wang T, Yangzhang, Zhai Z, Tao W, Li J, et al. Evaluation of large language models in breast cancer clinical scenarios: a comparative analysis based on ChatGPT-3.5, ChatGPT-4.0, and Claude2. Int J Surg. 2024;110:1941–50. https://doi.org/10.1097/JS9.0000000000001066

[75] Stalp JL, Denecke A, Jentschke M, Hillemanns P, Klapdor R. Quality of ChatGPT-Generated Therapy Recommendations for Breast Cancer Treatment in Gynecology. Curr Oncol. 2024;31:3845–54. https://doi.org/10.3390/curroncol31070284

[76] Lechien JR, Chiesa-Estomba C-M, Baudouin R, Hans S. Accuracy of ChatGPT in head and neck oncological board decisions: preliminary findings. Eur Arch Otorhinolaryngol. 2024;281:2105–14. https://doi.org/10.1007/s00405-023-08326-w

[77] Lee JC, Hamill CS, Shnayder Y, Buczek E, Kakarala K, Bur AM. Exploring the Role of Artificial Intelligence Chatbots in Preoperative Counseling for Head and Neck Cancer Surgery. Laryngoscope. 2024;134:2757–61. https://doi.org/10.1002/lary.31243

[78] Washington CJ, Abouyared M, Karanth S, Braithwaite D, Birkeland A, Silverman DA, et al. The Use of Chatbots in Head and Neck Mucosal Malignancy Treatment Recommendations. Otolaryngol Head Neck Surg. 2024; https://doi.org/10.1002/ohn.818

[79] Marchi F, Bellini E, Iandelli A, Sampieri C, Peretti G. Exploring the landscape of AI-assisted decision-making in head and neck cancer treatment: a comparative analysis of NCCN guidelines and ChatGPT responses. Eur Arch Otorhinolaryngol. 2024;281:2123–36. https://doi.org/10.1007/s00405-024-08525-z

[80] Lorenzi A, Pugliese G, Maniaci A, Lechien JR, Allevi F, Boscolo-Rizzo P, et al. Reliability of large language models for advanced head and neck malignancies management: a comparison between ChatGPT 4 and Gemini Advanced. Eur Arch Otorhinolaryngol. 2024; https://doi.org/10.1007/s00405-024-08746-2

[81] Gebrael G, Sahu KK, Chigarira B, Tripathi N, Mathew Thomas V, Sayegh N, et al. Enhancing Triage Efficiency and Accuracy in Emergency Rooms for Patients with Metastatic Prostate Cancer: A Retrospective Analysis of Artificial Intelligence-Assisted Triage Using ChatGPT 4.0. Cancers (Basel). 2023;15:3717. https://doi.org/10.3390/cancers15143717

[82] Wu DJ, Bibault J-E. Pilot applications of GPT-4 in radiation oncology: Summarizing patient symptom intake and targeted chatbot applications. Radiother Oncol. 2024;190:109978. https://doi.org/10.1016/j.radonc.2023.109978

[83] Piazza D, Martorana F, Curaba A, Sambataro D, Valerio MR, Firenze A, et al. The Consistency and Quality of ChatGPT Responses Compared to Clinical Guidelines for Ovarian Cancer: A Delphi Approach. Curr Oncol. 2024;31:2796–804. https://doi.org/10.3390/curroncol31050212

[84] Finch L, Broach V, Feinberg J, Al-Niaimi A, Abu-Rustum NR, Zhou Q, et al. ChatGPT compared to national guidelines for management of ovarian cancer: Did ChatGPT get it right? - A Memorial Sloan Kettering Cancer Center Team Ovary study. Gynecol Oncol. 2024;189:75–9. https://doi.org/10.1016/j.ygyno.2024.07.007

[85] Aghamaliyev U, Karimbayli J, Giessen-Jung C, Matthias I, Unger K, Andrade D, et al. ChatGPT's Gastrointestinal Tumor Board Tango: A limping dance partner? Eur J Cancer. 2024;205:114100. https://doi.org/10.1016/j.ejca.2024.114100

[86] Bresler TE, Pandya S, Meyer R, Htway Z, Fujita M. From Bytes to Best Practices: Tracing ChatGPT-3.5's Evolution and Alignment With the National Comprehensive Cancer Network® Guidelines in Pancreatic Adenocarcinoma Management. Am Surg. 2024;31348241248801. https://doi.org/10.1177/00031348241248801

[87] Choo JM, Ryu HS, Kim JS, Cheong JY, Baek S-J, Kwak JM, et al. Conversational artificial intelligence (chatGPT^{TM}) in the management of complex colorectal cancer patients: early experience. ANZ J Surg. 2024;94:356–61. https://doi.org/10.1111/ans.18749

[88] Cuellar-Barboza A, Brussolo-Marroquín E, Cordero-Martinez FC, Aguilar-Calderon PE, Vazquez-Martinez O, Ocampo-Candiani J. An evaluation of "ChatGPT" Compared to Dermatological Surgeons' Choice of Reconstruction of Mohs Surgical Defects. Clin Exp Dermatol. 2024;llae184. https://doi.org/10.1093/ced/llae184

[89] Liang R, Zhao A, Peng L, Xu X, Zhong J, Wu F, et al. Enhanced Artificial Intelligence Strategies in Renal Oncology: Iterative Optimization and Comparative Analysis of GPT 3.5 Versus 4.0. Ann Surg Oncol. 2024;31:3887–93. https://doi.org/10.1245/s10434-024-15107-0

[90] Iannantuono GM, Bracken-Clarke D, Karzai F, Choo-Wosoba H, Gulley JL, Floudas CS. Comparison of Large Language Models in Answering Immuno-Oncology Questions: A Cross-Sectional Study. Oncologist. 2024;29:407–14. https://doi.org/10.1093/oncolo/oyae009

[91] Zamarud A, Marianayagam N, Sekar V, Chang SD, Meola A. Treatment Outcomes of Leiomyosarcoma Metastasis Affecting the Brachial Plexus: A Comparative Case Report Using Chat Generative Pre-trained Transformer (ChatGPT). Cureus. 2023;15:e36715. https://doi.org/10.7759/cureus.36715

[92] Le DP, Hall SC. Medical Literature Writing With ChatGPT: A Rare Case of Choriocarcinoma Syndrome With Hemorrhagic Brain Metastases Due to Burned Out Metastatic Mixed Testicular Cancer. Cureus. 2023;15:e36655. https://doi.org/10.7759/cureus.36655

[93] Benary M, Wang XD, Schmidt M, Soll D, Hilfenhaus G, Nassir M, et al. Leveraging Large Language Models for Decision Support in Personalized Oncology. JAMA Netw Open. 2023;6:e2343689. https://doi.org/10.1001/jamanetworkopen.2023.43689

[94] Nashwan AJ, Bani Hani S. Enhancing oncology nursing care planning for patients with cancer through Harnessing large language models. Asia Pac J Oncol Nurs. 2023;10:100277. https://doi.org/10.1016/j.apjon.2023.100277

[95] Andras D, Ilies RA, Esanu V, Agoston S, Marginean Jumate TF, Dindelegan GC. Artificial Intelligence as a Potential Tool for Predicting Surgical Margin Status in Early Breast Cancer Using Mammographic Specimen Images. Diagnostics (Basel). 2025;15:1276. https://doi.org/10.3390/diagnostics15101276

[96] Nakamura T, Kobayashi N, Kumazu Y, Fukata K, Murakami M, Kohno S, et al. Precise highlighting of the pancreas by semantic segmentation during robot-assisted gastrectomy: visual assistance with artificial intelligence for surgeons. Gastric Cancer. 2024;27:869–75. https://doi.org/10.1007/s10120-024-01495-5

[97] Luo X, Wang J, Han Z, Yu Y, Chen Z, Huang F, et al. Artificial intelligence-enhanced white-light colonoscopy with attention guidance predicts colorectal cancer invasion depth. Gastrointest Endosc. 2021;94:627-638.e1. https://doi.org/10.1016/j.gie.2021.03.936

[98] Zhang C, Xu J, Tang R, Yang J, Wang W, Yu X, et al. Novel research and future prospects of artificial intelligence in cancer diagnosis and treatment. J Hematol Oncol. 2023;16:114. https://doi.org/10.1186/s13045-023-01514-5

[99] Lazris D, Schenker Y, Thomas TH. AI-Generated Content in Cancer Symptom Management: A Comparative Analysis Between ChatGPT and NCCN. J Pain Symptom Manage. 2024;S0885-3924(24)00839-X. https://doi.org/10.1016/j.jpainsymman.2024.06.019

[100] Yalamanchili A, Sengupta B, Song J, Lim S, Thomas TO, Mittal BB, et al. Quality of Large Language Model Responses to Radiation Oncology Patient Care Questions. JAMA Netw Open. 2024;7:e244630. https://doi.org/10.1001/jamanetworkopen.2024.4630

[101] Krückel A, Brückner L, Psilopatis I, Fasching PA, Beckmann MW, Emons J. Evaluation of ChatGPT's Potential in Tailoring Gynecological Cancer Therapies. In Vivo. 2024;38:1649–59. https://doi.org/10.21873/invivo.13614

[102] Zhu L, Rong Y, McGee LA, Rwigema J-CM, Patel SH. Testing and Validation of a Custom Retrained Large Language Model for the Supportive Care of HN Patients with External Knowledge Base. Cancers (Basel). 2024;16:2311. https://doi.org/10.3390/cancers16132311

[103] Peng W, Feng Y, Yao C, Zhang S, Zhuo H, Qiu T, et al. Evaluating AI in medicine: a comparative analysis of expert and ChatGPT responses to colorectal cancer questions. Sci Rep. 2024;14:2840. https://doi.org/10.1038/s41598-024-52853-3

[104] Dos Santos FC, Johnson LG, Madandola OO, Priola KJB, Yao Y, Macieira TGR, et al. An example of leveraging AI for documentation: ChatGPT-generated nursing care plan for an older adult with lung cancer. J Am Med Inform Assoc. 2024;ocae116. https://doi.org/10.1093/jamia/ocae116

[105] Hanai A, Ishikawa T, Kawauchi S, Iida Y, Kawakami E. Generative artificial intelligence and non-pharmacological bias: an experimental study on cancer patient sexual health communications. BMJ Health Care Inform. 2024;31:e100924. https://doi.org/10.1136/bmjhci-2023-100924

[106] Hopkins AM, Logan JM, Kichenadasse G, Sorich MJ. Artificial intelligence chatbots will revolutionize how cancer patients access information: ChatGPT represents a paradigm-shift. JNCI Cancer Spectr. 2023;7:pkad010. https://doi.org/10.1093/jncics/pkad010

[107] Abreu AA, Murimwa GZ, Farah E, Stewart JW, Zhang L, Rodriguez J, et al. Enhancing Readability of Online Patient-Facing Content: The Role of AI Chatbots in Improving Cancer Information Accessibility. J Natl Compr Canc Netw. 2024;22:e237334. https://doi.org/10.6004/jnccn.2023.7334

[108] Teixeira da Silva JA. Can ChatGPT rescue or assist with language barriers in healthcare communication? Patient Educ Couns. 2023;115:107940. https://doi.org/10.1016/j.pec.2023.107940

[109] Alanezi F. Examining the role of ChatGPT in promoting health behaviors and lifestyle changes among cancer patients. Nutr Health. 2025;31(2):739-748. https://doi.org/10.1177/02601060241244563

[110] Moll M, Heilemann G, Georg D, Kauer-Dorner D, Kuess P. The role of artificial intelligence in informed patient consent for radiotherapy treatments-a case report. Strahlenther Onkol. 2024;200:544–8. https://doi.org/10.1007/s00066-023-02190-7

[111] Collin H, Keogh K, Basto M, Loeb S, Roberts MJ. ChatGPT can help guide and empower patients after prostate cancer diagnosis. Prostate Cancer Prostatic Dis. 2024; https://doi.org/10.1038/s41391-024-00864-6

[112] Hershenhouse JS, Mokhtar D, Eppler MB, Rodler S, Storino Ramacciotti L, Ganjavi C, et al. Accuracy, readability, and understandability of large language models for prostate cancer information to the public. Prostate Cancer Prostatic Dis. 2024; https://doi.org/10.1038/s41391-024-00826-y

[113] Demirci A. A Comparison of ChatGPT and Human Questionnaire Evaluations of the Urological Cancer Videos Most Watched on YouTube. Clin Genitourin Cancer. 2024;22:102145. https://doi.org/10.1016/j.clgc.2024.102145

[114] Thia I, Saluja M. ChatGPT: Is This Patient Education Tool for Urological Malignancies Readable for the General Population? Res Rep Urol. 2024;16:31–7. https://doi.org/10.2147/RRU.S440633

[115] Ozgor F, Caglar U, Halis A, Cakir H, Aksu UC, Ayranci A, et al. Urological Cancers and ChatGPT: Assessing the Quality of Information and Possible Risks for Patients. Clin Genitourin Cancer. 2024;22:454-457.e4. https://doi.org/10.1016/j.clgc.2023.12.017

[116] Choi J, Kim JW, Lee YS, Tae JH, Choi SY, Chang IH, et al. Availability of ChatGPT to provide medical information for patients with kidney cancer. Sci Rep. 2024;14:1542. https://doi.org/10.1038/s41598-024-51531-8

[117] Ostrowska M, Kacała P, Onolememen D, Vaughan-Lane K, Sisily Joseph A, Ostrowski A, et al. To trust or not to trust: evaluating the reliability and safety of AI responses to laryngeal cancer queries. Eur Arch Otorhinolaryngol. 2024; https://doi.org/10.1007/s00405-024-08643-8

[118] Braithwaite D, Karanth SD, Divaker J, Shoenborn N, Lin K, Richman I, et al. Evaluating ChatGPT's accuracy in providing screening mammography recommendations among older women: Artificial Intelligence and cancer communication. J Am Geriatr Soc. 2024;72:2237–40. https://doi.org/10.1111/jgs.18854

[119] Köroğlu EY, Fakı S, Beştepe N, Tam AA, Çuhacı Seyrek N, Topaloglu O, et al. A Novel Approach: Evaluating ChatGPT's Utility for the Management of Thyroid Nodules. Cureus. 2023;15:e47576. https://doi.org/10.7759/cureus.47576

[120] Davis RJ, Ayo-Ajibola O, Lin ME, Swanson MS, Chambers TN, Kwon DI, et al. Evaluation of Oropharyngeal Cancer Information from Revolutionary Artificial Intelligence Chatbot. Laryngoscope. 2024;134:2252–7. https://doi.org/10.1002/lary.31191

[121] Hassona Y, Alqaisi D, Al-Haddad A, Georgakopoulou EA, Malamos D, Alrashdan MS, et al. How good is ChatGPT at answering patients' questions related to early detection of oral (mouth) cancer? Oral Surg Oral Med Oral Pathol Oral Radiol. 2024;138:269–78. https://doi.org/10.1016/j.oooo.2024.04.010

[122] Liu HY, Alessandri Bonetti M, De Lorenzi F, Gimbel ML, Nguyen VT, Egro FM. Consulting the Digital Doctor: Google Versus ChatGPT as Sources of Information on Breast Implant-Associated Anaplastic Large Cell Lymphoma and Breast Implant Illness. Aesthetic Plast Surg. 2024;48:590–607. https://doi.org/10.1007/s00266-023-03713-4

[123] Burnette H, Pabani A, von Itzstein MS, Switzer B, Fan R, Ye F, et al. Use of artificial intelligence chatbots in clinical management of immune-related adverse events. J Immunother Cancer. 2024;12:e008599. https://doi.org/10.1136/jitc-2023-008599

[124] Patel JM, Hermann CE, Growdon WB, Aviki E, Stasenko M. ChatGPT accurately performs genetic counseling for gynecologic cancers. Gynecol Oncol. 2024;183:115–9. https://doi.org/10.1016/j.ygyno.2024.04.006

[125] Lam CS, Hua R, Koon HK, Zhou KR, Lam TTN, Lee CP, et al. Can ChatGPT Provide Quality Information on Integrative Oncology? A Brief Report. J Integr Complement Med. 2024;30:196–205. https://doi.org/10.1089/jicm.2023.0290

[126] Zhang N, Zhang H, Liu Z, Dai Z, Wu W, Zhou R, et al. An artificial intelligence network-guided signature for predicting outcome and immunotherapy response in lung adenocarcinoma patients based on 26 machine learning algorithms. Cell Prolif. 2023;56:e13409. https://doi.org/10.1111/cpr.13409

[127] Yu L, Huang K, Liao Y, Wang L, Sethi G, Ma Z. Targeting novel regulated cell death: Ferroptosis, pyroptosis and necroptosis in anti-PD-1/PD-L1 cancer immunotherapy. Cell Prolif. 2024;57:e13644. https://doi.org/10.1111/cpr.13644

[128] Liu H, Wang C, Wang R, Cao H, Cao Y, Huang T, et al. New insights into mechanisms and interventions of locoregional therapies for hepatocellular carcinoma. Chin J Cancer Res. 2024;36:167–94. https://doi.org/10.21147/j.issn.1000-9604.2024.02.06

[129] Guo F, Zhang Y, Cui J. Manufacturing CAR-NK against tumors: Who is the ideal supplier? Chin J Cancer Res. 2024;36:1–16. https://doi.org/10.21147/j.issn.1000-9604.2024.01.01

[130] Ji Q, Xu C, Wang Q, Wang W, Guo Z, Li Z, et al. Expert Consensus on the Diagnosis and Treatment of Malignant Pericardial Mesothelioma. Med Research. 2025;1:122–30. https://doi.org/10.1002/mdr2.70003

[131] Liu C, Chen Y, Xu X, Yin M, Zhang H, Su W. Utilizing Macrophages Missile for Sulfate-Based Nanomedicine Delivery in Lung Cancer Therapy. Research (Wash D C). 2024;7:0448. https://doi.org/10.34133/research.0448

[132] Tan R, Zhao J, Shubhra QTH. Eclipsing Toxicity, Enhancing Efficacy: Clofazimine's Vanguard in the Realm of Anti-PD-1 and CTLA-4 Immunotherapy. Research (Wash D C). 2024;7:0419. https://doi.org/10.34133/research.0419

[133] Sun D, Hadjiiski L, Gormley J, Chan H-P, Caoili E, Cohan R, et al. Outcome Prediction Using Multi-Modal Information: Integrating Large Language Model-Extracted Clinical Information and Image Analysis. Cancers (Basel). 2024;16:2402. https://doi.org/10.3390/cancers16132402

[134] Basaad A, Basurra S, Vakaj E, Eldaly AK, Abdelsamea MM. A BERT-GNN Approach for Metastatic Breast Cancer Prediction Using Histopathology Reports. Diagnostics (Basel). 2024;14:1365. https://doi.org/10.3390/diagnostics14131365

[135] Janopaul-Naylor JR, Koo A, Qian DC, McCall NS, Liu Y, Patel SA. Physician Assessment of ChatGPT and Bing Answers to American Cancer Society's Questions to Ask About Your Cancer. Am J Clin Oncol. 2024;47:17–21. https://doi.org/10.1097/COC.0000000000001050

[136] Schaefer M, Reichl S, Ter Horst R, Nicolas AM, Krausgruber T, Piras F, et al. GPT-4 as a biomedical simulator. Comput Biol Med. 2024;178:108796. https://doi.org/10.1016/j.compbiomed.2024.108796

[137] Li T, Shetty S, Kamath A, Jaiswal A, Jiang X, Ding Y, et al. CancerGPT for few shot drug pair synergy prediction using large pretrained language models. NPJ Digit Med. 2024;7:40. https://doi.org/10.1038/s41746-024-01024-9

[138] Tan RSYC, Lin Q, Low GH, Lin R, Goh TC, Chang CCE, et al. Inferring cancer disease response from radiology reports using large language models with data augmentation and prompting. J Am Med Inform Assoc. 2023;30:1657–64. https://doi.org/10.1093/jamia/ocad133

[139] Musheyev D, Pan A, Gross P, Kamyab D, Kaplinsky P, Spivak M, et al. Readability and Information Quality in Cancer Information From a Free vs Paid Chatbot. JAMA Netw Open. 2024;7:e2422275. https://doi.org/10.1001/jamanetworkopen.2024.22275

[140] Das A, Tariq A, Batalini F, Dhara B, Banerjee I. Exposing Vulnerabilities in Clinical LLMs Through Data Poisoning Attacks: Case Study in Breast Cancer. medRxiv. 2024;2024.03.20.24304627. https://doi.org/10.1101/2024.03.20.24304627

[141] Haltaufderheide J, Ranisch R. The ethics of ChatGPT in medicine and healthcare: a systematic review on Large Language Models (LLMs). NPJ Digit Med. 2024;7:183. https://doi.org/10.1038/s41746-024-01157-x

[142] García-López IM, Trujillo-Liñán L. Ethical and regulatory challenges of Generative AI in education: a systematic review. Frontiers in Education [Internet]. Frontiers; 2025 [cited 2025 Oct 18]. p. 1681252. https://www.frontiersin.org/journals/education/articles/10.3389/feduc.2025.1681252/abstract. Accessed 18 Oct 2025

[143] Joshi G, Jain A, Araveeti SR, Adhikari S, Garg H, Bhandari M. FDA-Approved Artificial Intelligence and Machine Learning (AI/ML)-Enabled Medical Devices: An Updated Landscape. Electronics. Multidisciplinary Digital Publishing Institute; 2024;13:498. https://doi.org/10.3390/electronics13030498

[144] Jadon A, Kumar S. Leveraging Generative AI Models for Synthetic Data Generation in Healthcare: Balancing Research and Privacy. 2023 International Conference on Smart Applications, Communications and Networking (SmartNets) [Internet]. 2023 [cited 2025 Oct 18]. p. 1–4. https://doi.org/10.1109/SmartNets58706.2023.10215825

[145] ZadZiabari A, Tabatabaei A. Ethics and Regulations in Generative AI. In: Zamanifar A, Faezipour M, editors. Application of Generative AI in Healthcare Systems [Internet]. Cham: Springer Nature Switzerland; 2025 [cited 2025 Jul 18]. p. 197–211. https://doi.org/10.1007/978-3-031-82963-5_8

[146] Chen J, Zhu L, Mou W, Lin A, Zeng D, Qi C, et al. STAGER checklist: Standardized testing and assessment guidelines for evaluating generative artificial intelligence reliability. IMetaOmics. 2024;1:e7. https://doi.org/10.1002/imo2.7

[147] Tam TYC, Sivarajkumar S, Kapoor S, Stolyar AV, Polanska K, McCarthy KR, et al. A framework for human evaluation of large language models in healthcare derived from literature review. NPJ Digit Med. 2024;7:258. https://doi.org/10.1038/s41746-024-01258-7

[148] Deshpande A, Sharp H. Responsible AI Systems: Who are the Stakeholders? Proceedings of the 2022 AAAI/ACM Conference on AI, Ethics, and Society [Internet]. Association for Computing Machinery; 2022 [cited 2025 Jul 17]. p. 227–36. https://doi.org/10.1145/3514094.3534187

[149] Contractor D, McDuff D, Haines JK, Lee J, Hines C, Hecht B, et al. Behavioral Use Licensing for Responsible AI. Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency. Association for Computing Machinery; 2022. p. 778–88. https://doi.org/10.1145/3531146.3533143

[150] The Lancet Regional Health – Europe. Embracing generative AI in health care. Lancet Reg Health Eur. 2023;30:100677. https://doi.org/10.1016/j.lanepe.2023.100677

[151] Engelke M, Baldini G, Kleesiek J, Nensa F, Dada A. FHIR-Former: enhancing clinical predictions through Fast Healthcare Interoperability Resources and large language models. J Am Med Inform Assoc. 2025;ocaf165. https://doi.org/10.1093/jamia/ocaf165

[152] Ong JCL, Chang SY-H, William W, Butte AJ, Shah NH, Chew LST, et al. Ethical and regulatory challenges of large language models in medicine. Lancet Digit Health. 2024;6:e428–32. https://doi.org/10.1016/S2589-7500(24)00061-X

[153] Strubell E, Ganesh A, McCallum A. Energy and Policy Considerations for Modern Deep Learning Research. Proceedings of the AAAI Conference on Artificial Intelligence. 2020;34:13693–6. https://doi.org/10.1609/aaai.v34i09.7123

[154] Gallifant J, Afshar M, Ameen S, Aphinyanaphongs Y, Chen S, Cacciamani G, et al. The TRIPOD-LLM reporting guideline for studies using large language models. Nat Med. Nature Publishing Group; 2025;31:60–9. https://doi.org/10.1038/s41591-024-03425-5

[155] Park SH, Suh CH, Lee JH, Kahn CE, Moy L. Minimum Reporting Items for Clear Evaluation of Accuracy Reports of Large Language Models in Healthcare (MI-CLEAR-LLM). Korean J Radiol. 2024;25:865–8. https://doi.org/10.3348/kjr.2024.0843

[156] Vasey B, Nagendran M, Campbell B, Clifton DA, Collins GS, Denaxas S, et al. Reporting guideline for the early-stage clinical evaluation of decision support systems driven by artificial intelligence: DECIDE-AI. Nat Med. Nature Publishing Group; 2022;28:924–33. https://doi.org/10.1038/s41591-022-01772-9

[157] Wang S, Tang Z, Yang H, Gong Q, Gu T, Ma H, et al. A novel evaluation benchmark for medical LLMs illuminating safety and effectiveness in clinical domains. NPJ Digit Med. 2026;9:91. https://doi.org/10.1038/s41746-025-02277-8

License: https://creativecommons.org/licenses/by-nd/4.0/