Health Problems of Civilization
eISSN: 2354-0265, ISSN: 2353-6942

REVIEW PAPER

Generative artificial intelligence in optimizing the quality of cancer care: potential, limitations, and future directions of development for Large Language Models. A narrative literature review

Joanna Gotlib-Małkowska[1(A,B,C,D,E,F)], **Kinga Włodarczyk**[1(A,B,F)], **Paweł Koczkodaj**[2(A,B,D)], **Marta Hreńczuk**[3(B,D,E)], **Adrian Nowakowski**[1(E,F)], **Mariusz Panczyk**[1(B,D,F)], **Ilona Cieślak**[1(B,E,F)]

[1]Department of Education and Research in Health Sciences, Medical University of Warsaw, Warsaw, Poland

[2]Department of Cancer Epidemiology and Primary Prevention, Maria Skłodowska Curie National Research Institute of Oncology, Warsaw, Poland

[3]Department of Surgical and Transplantation Nursing and Extracorporeal Therapies, Medical University of Warsaw, Warsaw, Poland

Gotlib-Małkowska J, Włodarczyk K, Koczkodaj P, Hreńczuk M, Nowakowski A, Panczyk M. Generative artificial intelligence in optimizing the quality of cancer care: potential, limitations, and future directions of development for Large Language Models. A narrative literature review. Health Prob Civil. https://doi.org/10.29316/hpc/218352

Tables: 3
Figures: 0
References: 46
Submitted: 2025 Dec 4
Accepted: 2026 Feb 20

Address for correspondence: Ilona Cieślak, Department of Education and Research in Health Sciences, Medical University of Warsaw, Litewska 14/16 Street, 00-581 Warsaw, Poland, e-mail: ilona.cieslak@wum.edu.pl

ORCID: Joanna Gotlib-Małkowska https://orcid.org/0000-0002-2717-7741, Kinga Włodarczyk https://orcid.org/0009-0004-1742-9492, Paweł Koczkodaj https://orcid.org/0000-0002-7525-0181, Marta Hreńczuk https://orcid.org/0000-0001-9103-8730, Adrian Nowakowski https://orcid.org/0009-0004-7247-4464, Mariusz Panczyk https://orcid.org/0000-0003-1830-2114, Ilona Cieślak https://orcid.org/0000-0001-7752-6527

Abstract

This literature review included scientific articles, published between 2022 and 2025, indexed in PubMed, Scopus, and Proquest. The article investigates the potential of generative artificial intelligence (GenAI), particularly Large Language Models (LLMs), to improve the quality of cancer care. LLMs have demonstrated effectiveness in patient education by simplifying complex medical terminology and tailoring content to the user's level of understanding. LLMs also assist physicians in clinical decision-making by analyzing medical data and supporting adherence to the latest guidelines. However, expert oversight is still necessary due to the risk of error. For cancer prevention, LLMs promote healthy lifestyle adoption, participation in screening programs, and vaccination. They also play an important role in reducing inequities in access to information. Another key feature of LLMs is their ability to translate complex diagnostic reports into patient-friendly language. LLMs have also shown promise in counteracting cancer-related misinformation. However, the article identifies certain LLMs limitations, such as model hallucination, incomplete personalization, and unresolved legal liability concerns. The article emphasizes ethical dilemmas, particularly those related to patient autonomy and the risk of dehumanizing care. For future progress, the article emphasizes the need to integrate LLMs into e-health systems and to develop specialized models supported by interdisciplinary teams.

Keywords: generative artificial intelligence, Large Language Models, clinical decision support, patient education as topic, oncology

Introduction

Generative Artificial Intelligence (GenAI), including Large Language Models (LLMs), is gaining prominence as a tool in health care, offering innovative solutions that improve access to services, increase the effectiveness of early disease detection, and facilitate population-level health education initiatives [1].

Artificial intelligence (AI)- and GenAI-based systems analyze epidemiological data, model population health risks, and predict the incidence of chronic diseases, including malignant neoplasms [2]. These capabilities enable more effective planning of preventive interventions and better allocation of healthcare resources [3].

AI facilitates the identification of high-risk populations, enabling targeted screening programs. For example, the early detection of breast, cervical, or colorectal cancer can be significantly improved through AI-driven algorithms, which may lead to better survival outcomes [4]. Moreover, AI-based systems are employed to monitor patient adherence to preventive and therapeutic recommendations [3,4].

A key application of AI in health care is to promote equity through reducing health disparities. Chatbots and voice assistants powered by LLMs can facilitate outreach to populations with low health literacy, limited access to information, and lower levels of formal education. By expanding access to health information, these technologies may help reduce disparities in the use of healthcare services [5].

Aim of the study

The aim of this study was to investigate the potential of GenAI, particularly LLMs, to improve the quality of cancer care.

Methods

Literature review methodology

This narrative literature review examines the potential, limitations, and future trajectories of LLMs applications in oncology.

Search strategy and data sources

To ensure comprehensive coverage of peer-reviewed clinical, health-science, social, and technical literature, the review incorporated articles indexed in the following databases: PubMed, EMBASE, Scopus, Web of Science, and ProQuest:

- PubMed and EMBASE (clinical and biomedical focus),

- Scopus and Web of Science (interdisciplinary and technical coverage),
- ProQuest (health-sciences and social impact).

Furthermore, to account for the rapid pace of innovation in AI, the arXiv preprint server was included in the search. This allowed for the identification of cutting-edge technical developments and emerging methodologies that may still be undergoing the formal peer-review process but are highly relevant to the current state of generative AI.

The search was initially conducted between March 10 and 14, 2025, and was subsequently updated in January 2026 to capture the most recent advancements. The January 2026 update was specifically used to bridge the gap between March 2025 and the end of the inclusion period.

Inclusion criteria and timeline

The inclusion period spans from November 30, 2022, to December 31, 2025. The start date of late 2022 was selected to coincide with the public release of GPT-3.5, which catalyzed a transition from traditional, non-generative machine learning toward generative AI research in clinical contexts.

Selection criteria

To ensure the analytical rigor of this review, the inclusion and exclusion criteria were structured as follows:

- inclusion criteria: the review incorporated original research articles, narrative reviews, and systematic reviews published in English between November 30, 2022, and December 31, 2025, that specifically addressed the application of generative AI in oncological settings;
- exclusion criteria: we excluded editorial comments, letters to the editor, opinion pieces, and conference abstracts to prioritize validated data and comprehensive syntheses. Furthermore, studies published in languages other than English or those focusing exclusively on traditional, non-generative machine learning algorithms were omitted from the final analysis.

Scope and keywords

The strategy targeted studies investigating LLMs utility across the oncological spectrum, including prevention, diagnosis, treatment planning, patient-clinician communication, and health disparities. The following Boolean search query was utilized: (*"Large Language Models" OR "LLM" OR "Generative AI" OR "ChatGPT"*) *AND* (*"oncology" OR "cancer care" OR "cancer diagnosis" OR "patient education"*).

Additional keywords used to refine the search included: "clinical decision support", "health communication", "misinformation", "digital health", and "AI ethics". A summary of the review methodology is presented in Table 1.

Table 1. Summary of review methodology

Parameter	Details
Review type	Narrative literature review
Timeframe	November 30, 2022 – December 31, 2025 (updated: Jan 2026)
Primary databases	PubMed, EMBASE, Scopus, Web of Science, ProQuest
Preprint server	arXiv (for cutting-edge technical developments)
Study types included	Original research articles, narrative reviews, systematic reviews
Study types excluded	Editorials, letters to the editor, opinion pieces, conference abstracts
Inclusion criteria	GenAI in oncology; published in English
Exclusion criteria	Non-generative ML; non-English publications; non-clinical focuses
Key search terms	Large Language Models (LLM), Generative AI, ChatGPT, oncology, cancer care
Clinical areas	Prevention, diagnosis, treatment planning, communication, health disparities

Methodological rigor and transparency

Although this study is designed as a narrative review, the search strategy and reporting process adhered to the core principles of the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines to ensure maximum transparency, reproducibility, and a systematic approach to literature identification (Table 2). By adopting a structured PRISMA-style framework, we minimized selection bias and ensured that the transition from broad database queries to the final inclusion of specific studies was documented with analytical clarity. This rigorous selection process ensures that the synthesized evidence represents a balanced and comprehensive overview of the current state of LLMs in oncology.

Table 2. PRISMA flow diagram

Stage	Activity	Number of records
Identification	Records identified from databases, included: PubMed EMBASE Scopus Web of Science ProQuest	1,204 683 54 377 71 19
Screening	Records after duplicates removed	759
	Records screened (title and abstract)	759
	Records excluded after title/abstract screening	401
	Full-text articles sought for retrieval	358
	Records identified from preprint servers (arXiv)	57
Selection	Total full-text articles assessed for eligibility	358 + 57 = 415
	Full-text articles excluded with reasons:	366
	Reason 1: Non-English language	92
	Reason 2: Focus on non-generative ML	178
	Reason 3: Editorials, letters, or opinion pieces	96
Included	**Total studies included in narrative review**	49

A total of 1,204 records were initially identified across five databases: PubMed, EMBASE, Scopus, Web of Science, and ProQuest. Following the removal of duplicates, 759 unique records remained for screening. During the title and abstract screening phase, 401 articles were excluded, leaving 358 studies for full-text retrieval. To capture emerging technical developments, an additional 57 records from the arXiv preprint server were included, resulting in 415 articles assessed for eligibility. Upon full-text review, 366 studies were excluded based on predefined criteria: 92 were non-English publications, 178 focused on non-generative machine learning, and 96 were editorials or opinion pieces. Consequently, 49 studies met all inclusion criteria and were synthesized in this narrative review.

Study selection process

The study selection process followed a structured screening protocol designed to minimize selection bias and ensure a rigorous evaluation of the literature. Two reviewers independently screened the titles and abstracts of all identified records against the predefined inclusion criteria. The selection process involved a consensus-driven assessment by the co-authors, who prioritized studies focusing on the practical implementation of LLMs, their impact on patient outcomes, health equity, and the broader technological and ethical implications of their use in oncology.

Any disagreements regarding the eligibility of a specific study were resolved through internal consensus or, where necessary, through consultation with a third senior reviewer to reach a final decision.

To minimize selection bias, a structured screening protocol was implemented. The process was conducted as follows:

- initial screening: two reviewers independently screened the titles and abstracts of all identified records against the predefined inclusion criteria;
- full-text review: studies that met the initial criteria underwent a comprehensive full-text evaluation to determine their final eligibility;
- conflict resolution: any disagreements regarding study eligibility were resolved through consensus-driven discussion among the co-authors or, when necessary, by consultation with a third senior reviewer.

Selection criteria: PICO framework

The selection process was guided by the PICO framework, prioritizing studies that focused on the practical implementation of LLMs, their impact on patient outcomes, health equity, and the associated technological and ethical implications (Table 3).

Table 3. Selection criteria: PICO framework

Criterion	Inclusion criteria	Exclusion criteria
Population (P)	Cancer patients, oncology professionals, medical students, and caregivers.	General populations not specifically related to oncology.
Intervention (I)	Applications of LLMs (e.g. GPT-4, Med-PaLM, Claude, Llama) in cancer care.	Traditional non-generative ML, rule-based algorithms, or AI without LLMs components.
Comparison (C)	Standard clinical practice, human expert performance, or non-AI digital tools.	Not mandatory for this narrative synthesis.
Outcomes (O)	Accuracy of diagnosis/treatment, patient education quality, clinician burnout, ethics, and privacy.	Purely technical metrics (e.g. perplexity, token speed) without clinical relevance.

Quality assessment and data synthesis

The SANRA scale was used as a framework to ensure the methodological quality of this narrative review. Each included study was evaluated by at least two authors.

Due to the heterogeneity of the literature – spanning technical benchmarks to ethical essays – a thematic synthesis was employed. The findings were categorized into three primary thematic pillars:

1. clinical decision support: LLMs utility in diagnosis and treatment planning;
2. patient-physician communication: impact on patient education and the clinical relationship;
3. ethical and regulatory challenges: implications for data privacy, health equity, and misinformation.

Approval of the Institutional Bioethics Committee

As the study did not involve human participants or the use of confidential clinical data, Bioethics Committee approval was not required.

Literature review results

Use of LLMs in clinical decision support in oncology

Clinical decision support in diagnostic and therapeutic processes is a major application area of AI in oncology [6]. Recent studies indicate that LLMs can generate summaries of diagnostic test results, analyze data from medical records, and produce reminders informed by the latest clinical guidelines for cancer treatment [7]. LLMs also assist physicians in selecting treatment regimens consistent with the recommendations of scientific societies and established clinical practice guidelines. However, as analyses show, LLMs may still produce incomplete or inconsistent recommendations, which makes ongoing expert supervision essential [6,7].

A fundamental advantage of AI integration into the clinical decision-making process is its ability to process and analyze large clinical datasets within a short period of time. This enables the personalization of oncology treatments, including the identification of patients most likely to respond to immunotherapy or targeted therapies [8]. As a result, more precise therapeutic strategies can be implemented, in line with the principles of personalized medicine.

AI and LLMs can also streamline administrative processes within healthcare institutions. These functionalities include the automation of medical documentation, patient data archiving, and relieving healthcare professionals of routine administrative tasks, among others [6-8].

As healthcare systems continue to evolve, the use of LLMs as tools to support interdisciplinary oncology teams seems inevitable. However, the safe implementation of these technologies requires strict adherence to data security, patient privacy, and medical ethics standards [5-8].

Use of LLMs in oncology patient education

LLMs offer a transformative approach to communicating medical information tailored to the recipient's knowledge level and specific needs [9,10]. Their ability to generate content of varying complexity enables accessible explanation of complex oncology concepts to individuals with no medical background. Thus, by adapting AI-generated messages to end-users' cognitive abilities and needs, these solutions improve health literacy at the population level [9-11].

In health education, LLMs are implemented via chatbots and voice assistants that provide patients with information on cancer prevention, recognition of red-flag symptoms, diagnostic procedures, and available treatment options [9-11]. These solutions can help reduce anxiety related to diagnostic and therapeutic uncertainty, supporting the process of making informed health decisions [9-11].

Research indicates that patient interaction with LLM-based systems can strengthen patients' understanding of medical advice and promote active involvement in therapeutic decision-making [12]. Additionally, LLMs play a crucial educational role for patients' families and caregivers. Access to reliable, up-to-date information on cancer progression, treatment side effects, and palliative care principles can substantially improve the quality of support provided by the patient's support network [9-12].

Role of LLMs in promoting healthy lifestyles as a component of primary cancer prevention

GenAI can contribute significantly to health education aimed at promoting healthy lifestyles, which is a fundamental aspect of primary prevention for many types of cancer [13]. Current recommendations from international institutions state that making lifestyle changes toward health-promoting behaviors is associated with a substantial reduction in cancer risk.

LLMs can provide users with personalized dietary recommendations, such as increasing fiber, fruit, and vegetables intake, limiting alcohol consumption, reducing red and processed meat intake, and increasing daily physical activity [14].

In primary cancer prevention, LLM-based chatbots provide coaching programs to support tobacco cessation. Smoking is the leading modifiable risk factor for lung cancer and a range of other malignancies, including cancers of the larynx, oral cavity, esophagus, and bladder [15].

For example, users seeking assistance with smoking cessation may receive tailored therapeutic guidance, including pharmacological interventions (e.g. nicotine replacement therapy, bupropion, varenicline), recommendations for behavioral therapies, and information about locally available psychological support services and cessation support groups. The effectiveness of these interventions is well-established in empirical evidence and endorsed by public health authorities [14,15].

Role of LLMs in reducing barriers to accessing reliable health information on cancer prevention, diagnosis, and treatment

LLMs play an important role in overcoming the obstacles that hinder access to reliable health information [16]. Their ability to process queries and generate natural language enables clear communication of complex medical content to a broad audience. This is particularly important for individuals with lower levels of formal education and limited literacy skills (i.e. low literacy).

Empirical evidence suggests that a significant proportion of the educational materials available online is written at a level higher than that recommended by the American Medical Association (AMA), i.e. a sixth-grade reading level [17]. In this context, LLMs provide substantial added value by enabling automatic adjustment of language complexity to users' specific needs.

For example, through appropriate query formulation (i.e. prompting), users can receive simplified explanations of diagnostic and screening procedures, such as mammography or colorectal cancer screening tests. A study evaluating the quality and readability of ChatGPT-generated responses to patient questions about breast cancer found that the model successfully simplified language to a sixth-grade level. This adaptation significantly improved the clarity, comprehensibility, and accessibility of the information provided [17-19].

Overcoming language barriers in health education through the use of LLMs

LLMs play a vital role in overcoming language barriers, ensuring equitable access to health information for patients who do not speak the dominant language of a given healthcare system. The multilinguality of LLMs enables medical content to reach a broad and diverse audience that would otherwise be excluded from traditional health education channels [20].

The use of LLMs in the translation and adaptation of educational content helps to eliminate systemic communication barriers resulting from the lack of translated health materials. These technologies promote equity in access to information, an objective reflected in the recommendations of health organizations such as the World Health Organization (WHO) and the European Centre for Disease Prevention and Control (ECDC) [12,21].

Role of LLMs in educating individuals with low health literacy

Another important way in which LLMs provide support is addressing health inequalities associated with low health literacy [22]. Individuals with limited health literacy often have difficulty understanding basic medical concepts. This hinders their ability to make informed decisions about preventive care, diagnostic, or therapeutic treatments, including cancer treatment.

LLMs offer the ability to adjust the level of detail and format of information dynamically. These functionalities include, among others, translating medical terminology into plain language, presenting content in a bullet-point format to facilitate information processing, and simulating a dialogue that resembles a conversation with a physician, all of which facilitate better patient understanding and knowledge retention [23].

The application of LLMs to translate and simplify diagnostic test results, such as mammography or breast biopsy reports, is a practical example of how they reduce communication barriers. A study on the effectiveness of ChatGPT in providing simplified explanations of medical results showed that 89.5% of participants considered the information to be clear, while the correct interpretation of the results after receiving a simplified explanation ranged from 63% to 87% [24].

Personalized health communication using LLMs to support patients with cognitive impairment and emotional distress

With their advanced capacity for content personalization, LLMs can support patients with cognitive impairments and those in emotional distress [25]. These groups may require simple, clear, and unambiguous messages to understand the information provided and make informed decisions regarding cancer diagnosis and treatment.

LLMs can simulate an empathetic tone in their responses. A communication style that incorporates emotional support may help reduce patient anxiety related to diagnostic and therapeutic procedures and support psychological adjustment to a cancer diagnosis [26,27].

In this context, practical examples of LLMs use include ChatGPT responses to patients seeking information about potential cancer treatment side effects, such as skin changes, hair loss, or altered taste. Not only do these responses include accurate medical information, but they also incorporate empathetic commentary, specific self-care advice, and suggestions for next steps, including specialist consultations. Such communication strategies may be particularly valuable in settings where access to professional psychological or psycho-oncology support is limited [25-27].

Simplified interpretation of diagnostic test results by LLMs

One of the most promising applications for LLMs in oncology is their capacity to translate complex and highly specialized diagnostic reports into patient-friendly language [28].

Diagnostic reports, which include radiological examinations (e.g. magnetic resonance imaging (MRI), computed tomography (CT), ultrasound) and pathological assessments (e.g. biopsy results, histopathological analyses), are typically written in technical language and contain specialist medical terminology, numerical data, classifications, and references to international staging systems, such as BI-RADS, PI-RADS, and TNM [29].

These reports are difficult for many patients to understand due to their high level of complexity. LLMs enable the automatic conversion of report content to match the recipient's level of health literacy. This involves explaining specialized terms in plain language, providing concise summaries of information, clarifying the clinical implications of the findings, and outlining recommended steps in diagnostic and therapeutic pathways, all while maintaining clarity, emotional neutrality, and an empathetic tone.

These solutions support the patient's right to understandable information and empower them to be active participants in their oncology care [30].

Role of LLMs in supporting oncology treatment planning

LLMs are gaining importance as tools to support oncology treatment planning. They are increasingly being used to provide patients with information about available treatment options aligned with latest clinical guidelines. The growing popularity of LLMs stems from their ability to rapidly generate clear, personalized, and understandable answers, helping patients navigate the complex decision-making processes associated with cancer therapies [31].

A study by Tsai et al. [32] evaluated the ability of ChatGPT to generate treatment recommendations for breast, prostate, and lung cancer that comply with the current National Comprehensive Cancer Network (NCCN) guidelines. The researchers used four distinct prompt templates to assess the impact of question wording on the quality and completeness of the model's responses [32]. The analysis revealed that in 98% of cases (102 of 104 responses), ChatGPT generated therapeutic recommendations that included at least one option compliant with the NCCN guidelines [32]. The recommendations were customized based on the stage of the disease and the histopathological type of cancer. For example, for early-stage (stage I) breast cancer confined to the breast, ChatGPT correctly recommended surgical treatment (e.g. lumpectomy or mastectomy), with the option of adjuvant radiotherapy and consideration of hormonal therapy if the cancer was estrogen receptor positive (ER+) [32].

Despite the high proportion of partially guideline-consistent responses, the study by Tsai et al. [32] identified significant limitations in ChatGPT's ability to accurately reflect current oncology treatment standards. In 34.3% of cases (35 of 102 responses), the generated treatment recommendations were either partially inconsistent with the NCCN guidelines or incomplete..

The most commonly identified inaccuracies were: recommendation of therapeutic interventions inappropriate for the disease stage and the omission of critical information regarding the need for sequential treatment. For example, the model recommended surgical treatment for stage IV non-small cell lung cancer (NSCLC), contrary to current NCCN guidelines that favor systemic palliative therapies, such as immunotherapy or targeted chemotherapy, for this advanced stage.

Additionally, "model hallucinations" were observed in 12.5% of cases, with the model generating responses not supported by evidence-based data or regulatory guidelines. The model recommended therapeutic regimens involving novel drugs not yet approved by regulatory agencies (e.g. the U.S. Food and Drug Administration – FDA) or not included in the NCCN guidelines at the time of the study [32].

The identified limitations clearly demonstrate the need for expert oversight mechanisms for LLM-generated content. Furthermore, solutions should be implemented to enable the provision of specialized knowledge sources to the model, which can be integrated with it, for example, through Retrieval-Augmented Generation, and to ensure their regular updates so that the generated information remains consistent with the latest clinical guidelines.

Another limitation is that LLMs can only personalize their responses based on the information provided by the user. While the models can adjust their output to patient-specific details included in the prompt, such as age, overall health and organ function, comorbidities, or prior oncological treatment, such personalization remains constrained by the quality, completeness, and accuracy of the input data. In clinical practice, these factors are critical for selecting appropriate therapeutic strategies and ensuring adherence to the principles of personalized medicine.

Evaluating the quality of LLM-generated content in health education and clinical decision support

LLMs utility in health education and clinical decision support is largely determined by the quality, clarity, and comprehensibility of the information they generate [33]. Research articles consistently use standardized tools to evaluate these parameters, including:

– DISCERN, a standardized tool for evaluating the quality of information on treatment options,
– PEMAT (Patient Education Materials Assessment Tool), a tool which evaluates the understandability and actionability of patient education materials,
– QUEST (Quality Evaluation Scoring Tool), a framework for assessing the quality of online health content.

These tools enable a comprehensive and systematic evaluation of health information in terms of its credibility, completeness, impartiality, comprehensibility, and ability to facilitate informed decision-making by both patients and healthcare professionals [24].

Musheyev et al. [17] analyzed the quality of information generated by ChatGPT with respect to the five most common malignant cancers: breast, prostate, lung, colorectal, and skin.

The assessment employed the DISCERN tool, which measures the quality of information on available treatment options across 16 criteria. The results showed that both ChatGPT 3.5 and ChatGPT 4.0 scored high in terms of information credibility, with an average DISCERN score of 4 out of 5. However, ChatGPT 4.0 was more likely to provide comprehensive and up-to-date content. Additionally, the PEMAT tool was applied to evaluate the understandability and usability of patient educational materials. ChatGPT scored over 80% in domains such as readability, message clarity, and appropriate content structure. Despite the study's positive results, the authors acknowledged a limitation of the model, i.e. its frequent omission of numerical data on the risks and benefits of specific therapies. This limitation may affect the model's ability to support patients in informed health decision-making [17]. Comparative analyses indicate significant differences in the quality of information generated by ChatGPT 3.5 and ChatGPT 4.0. According to Musheyev et al. [17], ChatGPT 4.0 outperformed ChatGPT 3.5 in terms of readability and adherence to user needs. In contrast, responses from ChatGPT 3.5 were often characterized by a higher level of linguistic complexity, which required additional clarification from medical personnel and hindered patient comprehension [17].

The way queries are formulated, or prompt engineering, is an important factor influencing the quality and usefulness of LLMs output. A study by Musheyev et al. [17] showed that clearly specifying expectations regarding the language level and response format significantly improves output readability and comprehensibility. For example, the use of the prompt: "Explain this at a sixth-grade reading level" led to a marked improvement in the readability of the output of both models. In ChatGPT 3.5, the Flesch Reading Ease score increased to 71.5, which corresponds to a text that is "easy to understand for elementary school students". In comparison, the responses not preceded by a detailed prompt scored around 52.6, indicating a moderate difficulty level that could be challenging for individuals with lower language proficiency.

Phenomenon of hallucination in LLMs as a challenge for medical information safety

One of the most critical challenges associated with LLMs use in medicine, including in oncology, is the risk of "model hallucinations", i.e. when the algorithm generates content that is not grounded in empirical data or scientific evidence [34]. Output may be generated that, while syntactically correct, coherent, and persuasive, is incorrect, incomplete, or non-compliant with current clinical guidelines.

Hallucinations may take different forms, including: the invention of non-existent drugs, therapeutic regimens, or diagnostic procedures; the misinterpretation of clinical trial data; or reinforcement of common medical myths and misinformation.

Chen et al. [35] provide a notable example of the limitations associated with model hallucinations. The researchers assessed the ability of ChatGPT to provide therapeutic recommendations for the treatment of breast, lung, and prostate cancer, based on adherence to NCCN guidelines [35]. The study showed that in 34.3% of cases (35 of 102 responses analyzed), ChatGPT generated therapeutic recommendations that contained at least one component inconsistent with current NCCN guidelines. Notably, 12.5% of the responses included information not supported by scientific evidence, e.g. incorrect treatment regimens and recommendations for non-existent or unapproved treatment options. For example, ChatGPT recommended immunotherapy for a cancer type with no approved indications at that time. In another case, the model incorrectly suggested surgery for a patient with stage IV non-small cell lung cancer, even though current clinical standards indicate that systemic palliative therapies, such as targeted chemotherapy or immunotherapy, are the preferred treatment options at this stage of disease progression [35].

Ethical and legal aspects of using LLMs in oncology

LLMs integration into clinical practice and health education in oncology presents a number of complex ethical and legal challenges. While these tools have significant potential to enhance patient education, support clinical decision-making, and improve access to information, careful risk assessment is essential before their use. Key concerns include accountability for medical errors generated by AI, ensuring the security of patient data under privacy regulations such as the GDPR, and addressing issues of patient autonomy and potential dehumanization in healthcare delivery [36,37].

Legal responsibility for errors and patient data security

One of the most pressing legal concerns is liability for LLM-generated errors. Although ChatGPT can provide therapeutic recommendations, it is not a certified medical device and is not subject to comprehensive regulatory requirements for AI systems in health care [38]. When a patient makes a treatment decision based on incorrect or outdated information generated by

AI, the question arises as to who bears the legal responsibility – the developer of the model, the service provider, the physician who relied on the recommendation, or the patient themselves? The current lack of clear legislative guidance in this area poses a significant legal risk, hindering clinical adoption of such technologies [38].

Patient privacy and the security of medical data are additional concerns. LLMs process large amounts of textual information, and when used to analyze medical data or generate reports based on clinical input, they may disclose sensitive information.

Patient autonomy and the risk of dehumanizing medical care

The use of LLMs in oncology also raises serious ethical questions concerning patient autonomy. Models such as ChatGPT can influence health-related decisions by controlling how information is presented, what content is selected, and even the language used in responses. Although access to simple, clear medical explanations can enhance patient understanding and enable them to participate more meaningfully in decision-making, there is a risk of information being manipulated or unintentionally distorted [39,40].

Although ChatGPT lacks awareness or intent, the way it generates responses can influence patients' perception of risk and subtly steer them towards solutions preferred by the algorithm instead of those that align with their personal values or preferences or, conversely, reinforce patients' own preferences regardless of their clinical validity or effectiveness.

Moreover, overreliance on LLMs in patient communication can lead to the dehumanization of medical care. Interactions with oncology patients are not only about providing information, they are also about providing empathic support, acknowledging patients' emotional states, and cultivating trust-based relationships. Substituting traditional doctor-patient interactions with AI chatbots could undermine such a therapeutic alliance by diminishing patients' sense of safety and emotional support, both of which are essential in the cancer treatment process [39-41].

Future prospects for the development of LLMs in oncology

As LLMs technology advances, its integration into existing healthcare systems becomes more feasible. A key area of progress involves embedding these tools into electronic health records (EHRs) and broader eHealth solutions. This integration could greatly improve clinical information management, optimize diagnostic and therapeutic pathways, and enhance communication between patients and healthcare professionals [41].

LLMs integration with EHR systems could transform the way physicians and other healthcare professionals access clinical information and process patient data. Tools such as ChatGPT can automatically generate medical history summaries, interpret diagnostic results, and create personalized treatment plans based on EHR data [40-42].

AI's ability to summarize complex visit notes into concise reports exemplifies its potential to support faster clinical decision-making. LLMs can also help create discharge instructions, dietary recommendations, and patient education materials, tailored to each individual's clinical profile.

Chung et al. [43] explored the potential use of ChatGPT in reporting and interpreting prostate MRI results. This study serves as an example of how LLMs can act as interfaces for eHealth systems [43-45]. Through integration with EHRs, ChatGPT was able to automatically retrieve data from diagnostic platforms, generate personalized reports for physicians, and provide simplified versions of results for patients, thereby promoting a deeper understanding of one's health status. In the future, these systems could be expanded to include symptom and adverse event monitoring by integrating data from telehealth platforms and wearable health monitoring devices.

Sallam [44] highlights the potential of LLMs to optimize patient support through eHealth portals. Such systems could enable users to search for information, schedule appointments, and access answers to frequently asked questions concerning cancer, diagnostic procedures, and treatment options [44-46].

Development of specialized models

As LLMs advance, the creation of specialized AI models dedicated to specific areas of oncology is becoming increasingly feasible. Instead of functioning as general-purpose assistants that cover a broad range of topics, these models focus on providing precise, in-depth, and up-to-date information related to specific cancers, their diagnosis, treatment, and patient care [45]. Specialized LLMs have the potential to significantly improve both the quality of health education and the efficiency of clinical processes by providing access to scientific and practical knowledge tailored to specific oncology subfields [45,46].

Specialization enables these models to be aligned more accurately with local clinical guidelines and the structure of national healthcare systems, thereby increasing their usefulness to both patients and physicians. Such models could be integrated with national cancer registries and local clinical databases, which would enable continuous content updates and greater compliance with clinical practice [45].

Interdisciplinary collaboration in content development and LLM model validation

The development of LLMs specialized for oncology requires close interdisciplinary collaboration, uniting clinicians, AI specialists, legal experts, bioinformaticians, and researchers from the social sciences and humanities. Only such collaboration would ensure that models are not only scientifically accurate but also ethically responsible, compliant with privacy regulations (e.g. GDPR), and aligned with end-user needs [45,46].

Technical teams and AI engineers are essential for designing model architectures, selecting appropriate training data, and implementing real-time knowledge update mechanisms. It is equally important to involve ethics and legal experts to ensure that the output produced by these models adheres to the principles of transparency, privacy, and information security [45,46].

Conclusions

The integration of GenAI, particularly LLMs, into oncology is a transformative advancement in cancer care. These technologies have the potential to significantly impact fields such as patient education, clinical decision support, personalized communication, and the promotion of health equity. LLMs can simplify complex medical terminology, facilitate informed decision-making, and help overcome language, literacy, and cognitive impairment barriers.

Despite their many benefits, LLMs have limitations. The risk of hallucinations, incomplete or outdated recommendations, and the current insufficient legal and ethical frameworks governing their use raise critical concerns. To ensure patient safety and output credibility, these issues require expert oversight, regular updates of training data, and comprehensive quality control mechanisms.

The future of LLMs in oncology lies in their integration with eHealth systems, the development of specialized models tailored to specific types of neoplasms and healthcare settings, and interdisciplinary collaboration to ensure ethical, legal, and clinical validity. Only a comprehensive approach will unlock the full potential of LLMs to deliver equitable, personalized, and high-quality cancer care.

Limitations of the study

This review includes select preprints to capture the most current technological advancements; however, readers should note that these works have not yet undergone formal peer review and should therefore be interpreted with appropriate caution. The preprints cited in this review were obtained from arXiv, an open-access repository widely used for the rapid dissemination of research manuscripts in fields such as physics, computer science, and biomedical informatics prior to journal peer review.

Disclosures and acknowledgements

The authors declare no conflicts of interest with respect to the research, authorship, and/or publication of this article.

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

During the preparation of this work, the authors used ChatGPT and Gemini (Google) in order to improve the linguistic quality, structural clarity, and grammatical accuracy of the manuscript. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication. The use of AI was limited to linguistic refinement and organizational support; all scientific synthesis, data interpretation, and conclusions were formulated by the human authors.

References:

1. Zheng Y, Wang L, Feng B, Zhao A, Wu Y. Innovating healthcare: the role of ChatGPT in streamlining hospital workflow in the future. Ann Biomed Eng. 2024; 52(4): 750-753. https://doi.org/10.1007/s10439-023-03323-w
2. Alessandri-Bonetti M, Liu HY, Giorgino R, Nguyen VT, Egro FM. The first months of life of ChatGPT and its impact in healthcare: a bibliometric analysis of the current literature. Ann Biomed Eng. 2024; 52(5): 1107-1110. https://doi.org/10.1007/s10439-023-03325-8
3. Ferdush J, Begum M, Hossain ST. ChatGPT and clinical decision support: scope, application, and limitations. Ann Biomed Eng. 2024; 52(5): 1119-1124. https://doi.org/10.1007/s10439-023-03329-4
4. Shaheen R, Salim H. Role of artificial intelligence in healthcare. In: Qidwai M, editor. Intersection of human rights and AI in healthcare. IGI Global Scientific Publishing; 2024. p. 173-200. https://doi.org/10.4018/979-8-3693-7051-3.ch007
5. Farhat R, Malik AK, Sheikh AM, Fatima A. The role of AI in enhancing healthcare access and service quality in resource-limited settings. International Journal of Artificial Intelligence. 2024; 11(2): 70-79. https://doi.org/10.36079/lamintang.ijai-01102.709
6. Liao C, Chu C, Lien M, Wu Y, Wang T. AI-enhanced healthcare: integrating ChatGPT-4 in ePROs for improved oncology care and decision-making: a pilot evaluation. Current Oncology. 2025; 32(1): 7. https://doi.org/10.3390/curroncol32010007
7. Kola MB. Integration of large language models in clinical decision support: a framework for human-AI collaboration in healthcare. International Journal of Scientific Research in Computer Science, Engineering and Information Technology. 2024; 10(6): 2352-2363. https://doi.org/10.32628/cseit2410612432
8. Gupta B, Ta P, Ram K, Sivaprakasam M. Comprehensive modeling and question answering of cancer clinical practice guidelines using LLMs. In: IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology (CIBCB), Natal, Brazil. New York: IEEE; 2024. p. 1-8. https://doi.org/10.1109/cibcb58642.2024.10702112
9. Carl N, Schramm F, Haggenmüller S, Kather JN, Hetz MJ, Wies C, et al. Large language model use in clinical oncology. NPJ Precision Oncology. 2024; 8(1): 240. https://doi.org/10.1038/s41698-024-00733-4

10. Aydın S, Karabacak M, Vlachos V, Margetis K. Large language models in patient education: a scoping review of applications in medicine. Frontiers in Medicine. 2024; 11. https://doi.org/10.3389/fmed.2024.1477898
11. Gao M, Varshney AS, Chen S, Goddla V, Gallifant J, Doyle P, et al. The use of large language models to enhance cancer clinical trial educational materials. arXiv. Forthcoming 2024. https://doi.org/10.48550/arxiv.2412.01955
12. Zitu M, Le TT, Duong TV, Haddadan S, Garcia MAC, Amorrortu RP, et al. Large language models in cancer: potentials, risks, and safeguards. BJR Artificial Intelligence. 2024; 2(1): ubae019. https://doi.org/10.1093/bjrai/ubae019
13. Subramanian CR, Yang DA, Khanna R. Enhancing health care communication with large language models—the role, challenges, and future directions. JAMA Netw Open. 2024; 7(3): e240347. https://doi.org/10.1001/jamanetworkopen.2024.0347
14. Ferrara E. Large language models for wearable sensor-based human activity recognition, health monitoring, and behavioral modeling: a survey of early trends, datasets, and challenges. Sensors. 2024; 24(15): 5045. https://doi.org/10.3390/s24155045
15. Caglayan A, Slusarczyk W, Rabbani RD, Ghose A, Papadopoulos V, Boussios S. Large language models in oncology: revolution or cause for concern?. Current Oncology. 2024; 31(4): 1817-1830. https://doi.org/10.3390/curroncol31040137
16. Deroy A, Maity S. Cancer-answer: empowering cancer care with advanced large language models. arXiv. Forthcoming 2025. https://doi.org/10.48550/arxiv.2411.06946
17. Musheyev D, Pan A, Loeb S, Kabarriti AE. How well do artificial intelligence chatbots respond to the top search queries about urological malignancies?. Eur Urol. 2024; 85(1): 13-16. https://doi.org/10.1016/j.eururo.2023.07.004
18. Lim S, Schmälzle R. Artificial intelligence for health message generation: an empirical study using a large language model (LLM) and prompt engineering. Front Commun. 2023; 8: 1129082. https://doi.org/10.3389/fcomm.2023.1129082
19. 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(11): e2343689. https://doi.org/10.1001/jamanetworkopen.2023.43689
20. Saadi N, Raha T, Clément C, Pimentel MA, Rajan R, Kanithi PK. Bridging language barriers in healthcare: a study on Arabic LLMs. arXiv. Forthcoming 2025. https://doi.org/10.48550/arxiv.2501.09825

21. Li M, Huang J, Yeung J, Blaes A, Johnson S, Liu H. CancerLLM: a large language model in cancer domain. arXiv. Forthcoming 2024. https://doi.org/10.48550/arxiv.2406.10459
22. Siu AHY, Gibson DP, Chiu C, Kwok A, Irwin M, Christie A, et al. ChatGPT as a patient education tool in colorectal cancer-An in-depth assessment of efficacy, quality and readability. Colorectal Disease. 2025; 27(1): e17267. https://doi.org/10.1111/codi.17267
23. Hao Y, Holmes J, Waddle M, Davis B, Yu N, Vickers K, et al. Personalizing prostate cancer education for patients using an HER-Integrated LLM agent. arXiv. Forthcoming 2025. https://doi.org/10.48550/arXiv.2409.19100
24. 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. 2025; 28(2): 394-399. https://doi.org/10.1038/s41391-024-00826-y
25. Thapa S. Adhikari S. Leveraging ChatGPT-Like large language models for Alzheimer's disease: enhancing care, advancing research, and overcoming challenges. In: Bhambri P, Soni R, Tran TA, editors. Smart healthcare systems. Boca Raton: CRC Press; 2024. p. 265-275. https://doi.org/10.1201/9781032698519-19
26. Ferrario A, Sedlakova J, Trachsel M. The role of humanization and robustness of large language models in conversational artificial intelligence for individuals with depression: a critical analysis. JMIR Ment Health. 2024; 11: e56568. https://doi.org/10.2196/56569
27. Geantă M, Bădescu D, Chirca N, Nechita OC, Radu CG, Rascu S, et al. The potential impact of large language models on doctor–patient communication: a case study in prostate cancer. Healthcare (Basel). 2024; 12(15): 1548. https://doi.org/10.3390/healthcare12151548
28. Wang A, Zhou J, Zhang P, Cao H, Xin H, Xu X, et al. Large language model answers medical questions about standard pathology reports. Front Med (Lausanne). 2024; 11: 1402457. https://doi.org/10.3389/fmed.2024.1402457
29. Güneş YC, Cesur T, Çamur E, Karabekmez LG. Evaluating text and visual diagnostic capabilities of large language models on questions related to the Breast Imaging Reporting and Data System Atlas 5th edition. Diagn Interv Radiol. 2025; 31(2): 111-129. https://doi.org/10.4274/dir.2024.242876

30. Trapp C, Schmidt-Hegemann N, Keilholz M, Brose S, Marschner S, Schönecker S, et al. Patient- and clinician-based evaluation of large language models for patient education in prostate cancer radiotherapy. Strahlentherapie Und Onkologie. 2025; 201: 333-342. https://doi.org/10.1007/s00066-024-02342-3
31. Garg A, Gupta S, Vats S, Handa P, Goel N. Prospect of large language models and natural language processing for lung cancer diagnosis: a systematic review. Expert Systems. 2024; 41(11): e13697. https://doi.org/10.1111/exsy.13697
32. Tsai CY, Cheng PY, Deng JH, Jaw FS, Yii SC. ChatGPT v4 outperforming v3.5 on cancer treatment recommendations in quality, clinical guideline, and expert opinion concordance. Digit Health. 2024; 10: 20552076241269538. https://doi.org/10.1177/20552076241269538
33. Ho CN, Tian T, Ayers AT, Aaron RE, Phillips VJ, Wolf RM, et al. Qualitative metrics from the biomedical literature for evaluating large language models in clinical decision-making: a narrative review. BMC Medical Informatics and Decision Making. 2024; 24(1): 357. https://doi.org/10.1186/s12911-024-02757-z
34. Floyd W, Kleber T, Carpenter DJ, Pasli M, Qazi J, Huang C, et al. Current strengths and weaknesses of ChatGPT as a resource for radiation oncology patients and providers. Int J Radiat Oncol Biol Phys. 204; 118(4): 905-915. https://doi.org/10.1016/j.ijrobp.2023.10.020
35. Chen LC, Zack T, Demirci A, Sushil M, Miao B, Kasap C, et al. Assessing large language models for oncology data inference from radiology reports. JCO Clin Cancer Inform. 2024; 8: e2400126. https://doi.org/10.1200/cci.24.00126
36. Chow JC, Li K. Ethical considerations in human-centered AI: Advancing oncology chatbots through large language models. JMIR Bioinform Biotechnol. 2024; 5: e64406. https://doi.org/10.2196/64406
37. Ong JCL, Chang SYH, 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(6): e428-e432. https://doi.org/10.1016/s2589-7500(24)00061-x
38. Naik N, Hameed BM, Shetty DK, Swain D, Shah M, Paul R, et al. Legal and ethical consideration in artificial intelligence in healthcare: who takes responsibility?. Front Surg. 2022; 9: 862322. https://doi.org/10.3389/fsurg.2022.862322

39. Zhou Z, Qin P, Cheng X, Shao M, Ren Z, Zhao Y, et al. ChatGPT in oncology diagnosis and treatment: applications, legal and ethical challenges. Curr Oncol Rep. 2025; 27(4): 336-354. https://doi.org/10.1007/s11912-025-01649-3
40. Verlingue L, Boyer C, Olgiati L, Mairesse CB, Morel D, Blay JY. Artificial intelligence in oncology: ensuring safe and effective integration of language models in clinical practice. Lancet Reg Health Eur. 2024; 46: 101064. https://doi.org/10.1016/j.lanepe.2024.101064
41. Kolla L, Parikh RB. Uses and limitations of artificial intelligence for oncology. Cancer. 2024; 130(12): 2101-2107. https://doi.org/10.1002/cncr.35307
42. Rydzewski NR, Dinakaran D, Zhao SG, Ruppin E, Turkbey B, Citrin DE, et al. Comparative evaluation of LLMs in clinical oncology. NEJM AI. 2024; 1(5). https://doi.org/10.1056/aioa2300151
43. 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
44. Sallam M. ChatGPT utility in healthcare education, research, and practice: Systematic review on the promising perspectives and valid concerns. Healthcare. 2023; 11(6): 887. https://doi.org/10.3390/healthcare11060887
45. Iannantuono GM, Bracken-Clarke D, Floudas CS, Roselli M, Gulley JL, Karzai F. Applications of large language models in cancer care: current evidence and future perspectives. Front Oncol. 2023; 13: 1268915. https://doi.org/10.3389/fonc.2023.1268915
46. Li L, Dinh L, Hu S, Hemphill L. Academic collaboration on large language model studies increases overall but varies across disciplines. arXiv. Forthcoming 2024. https://doi.org/10.48550/arXiv.2408.04163