ACCSCIENCE PUBLISHING

Tumor Discovery

PERSPECTIVE ARTICLE

Large language models in oncodermatology

Mansak Shishak[1†], Sorun Shishak[2†*], and Sameer Rastogi[3]

[1]Department of Dermatology, Fortis Hospital and Research Center, New Delhi, India

[2]Department of Radiation Oncology, Medanta Hospital, Gurugram, Haryana, India

[3]Department of Medical Oncology, All India Institute of Medical Sciences, New Delhi, India

Abstract

With the release of ChatGPT, a generative artificial intelligence-based chatbot, the potential for multiple applications in medicine seems not just plausible but inevitably intertwined with oncology science. Their rapid fine-tuning and evolution into updated versions that incorporate different modes—gradually enhancing "reasoning" and contextual abilities—offer promise in bridging critical gaps in healthcare delivery. In the evolving landscape of precision oncology and dermatology, the skin has long served as both a mirror and a messenger, revealing malignancies through primary lesions, paraneoplastic signs, metastases, or therapy-related adverse effects. Large language models, deep learning systems trained on vast corpora that include open-source biomedical data, are poised to redefine how we process, contextualize, and act upon complex, language-rich, and algorithm-rich clinical datasets.

Keywords: Large language models; Artificial intelligence; ChatGPT; Oncology; Oncodermatology

†These authors contributed equally to this work.

***Corresponding author:** Sorun Shishak (sorun.shishak@medanta.org)

Citation: Shishak M, Shishak S, Rastogi S. Large language models in oncodermatology. *Tumor Discov*. 2025;4(4):118-122. doi: 10.36922/TD025310074

Received: July 31, 2025

Revised: September 23, 2025

Accepted: October 11, 2025

Published online: November 4, 2025

1. Introduction

Conventional artificial intelligence (AI)-based tools in oncology are trained for and focused on specific tasks, such as the classification of tumors, imaging, segmentation, and histopathology analysis. By design, adaptive abilities are not built into these systems. Large language models (LLMs), on the other hand, are more versatile, with increasing capacities to handle diverse and non-specific tasks.[1] They are chatbots that generate textual answers, without being specifically trained for medical purposes. LLMs are neural networks that generate words using predictive power on semantics and syntax, following self-supervised or unsupervised learning. Using AI-based algorithms, these models are trained on a large corpus of text data, often obtained from the web. Updated and newer LLMs can encode domain-agnostic tasks and self-train by learning in-context scenarios without pre-training. This enables a dynamic interaction with systems for constant dialogues between man and digital interface. Due to their easily usable format and likeness to search engines, LLMs have gained unique traction and rapid adoption by all. In this perspective review, we analyze whether hype meets real-world gaps and utility in the nascent stage of science-meets-chatbots in cutaneous oncology. Initially considered foundation models, they may hold value and potential for knowledge retrieval, clinical decision support, precision care, and research, provided their use is guided by careful application and assessment.[2]

2. Role of language in cutaneous oncology

Impactful healthcare outcomes require ongoing communication between various components of the healthcare ecosystem. Attempts to develop conversational tools using a computer program were first tried in patient counseling.[3] Patient education is a cornerstone of effective management in medicine. The idea of chatbots and instant replies, and the ability to chat endlessly, has shifted the perception of the manual one-on-one counseling process.

Oncodermatology is a specialized subset of oncology that addresses cutaneous malignancies (e.g., malignant melanoma and basal cell carcinoma), paraneoplastic syndromes, dermatology concerns in cancer care (e.g., toxicity-related skin reactions and supportive care for patients), and preventive oncology for skin and mucocutaneous cancers. LLMs relay information that enables communication; this not only bridges the semantic gap between understanding descriptive dermatology and oncologic decision-making but also provides valuable information to patients and care providers. It has been shown that conversational LLMs, such as ChatGPT, can expedite patient care via public education materials and simplify the process of breaking down medical jargon to edible pieces of important, useful information in melanoma prevention.[4]

Where image-based AI diagnosis has dominated oncodermatology innovation, LLMs unlock a different layer of meaning beyond diagnostics—the textual skin of oncology, capable of improving the struggles of healthcare records, patient counseling, emerging pattern recognition, drug discovery, and monitoring of therapeutic management. Future models may serve not only as diagnostic adjuncts but also as tools of epistemic discovery—surfacing correlations, syndromic archetypes, and treatment pathways that have long eluded traditional rule-based systems.

Language processing tools function via mathematical models of statistical probability by providing the best linguistic output, arranging one word after the other. There is no doubt that word-generating software can certainly influence and impact the end user or receiver through the messages received. This is seen in day-to-day life while interacting with automated chat support systems, outside of healthcare scenarios. Following the initial fervor on the abilities of LLMs, questions have been raised about their "thinking" and "feeling" abilities. In the context of cancer care, outputting words associated with care and empathy can presumably make one feel heard and understood, blurring the lines between human support and a mechanical chat machine. Do machines really process emotions? The answer is no, as it lacks the fundamental biological and physiological pathways required for a human's central nervous system. LLMs are inherently devoid of conceptual understanding, but perfectly capable of delivering the best language support. This highlights how LLMs are not just another technology upgrade or static repositories of knowledge, but a dynamic and evolving system that can hold more influence than intended. It could be inevitable that LLMs outgrow their parroting nature and become vital interpreters of the skin-cancer dialogue, where AI augments not only diagnostics but also the language of clinical reasoning.[5]

3. Translating promises into clinical applicability

3.1. Digital record keeping

Although the shift from manual record-keeping to digital spaces has eased many administrative tasks, integrating electronic health records (EHRs) with LLMs carries both potential and limitations. For the assessment of clinical trial eligibility in melanoma, the performance of a medically specialized LLM (Synapsis LLM) was compared to that of oncology research nurses for data extraction using manual chart review (MCR) from electronic medical record (EMR).[6] While the LLM proved to be swifter at data extraction (task completion in 2.5 min vs. 427 min), seemingly underscoring the need to automate tasks using appropriate AI tools, the use and interpretation of such EMR-LLM synergy is not linear. In research and clinical trials, the task of a study team involves repeated reviews of medical records to carefully interpret data. Beyond simple data entry, it requires contextual interpretation—clarifying ambiguities—and incorporating additional clinical notes that automated systems may overlook. While technology could accelerate collation, it is the intellectual effort invested in the gold standard process of MCR that safeguards scientific accuracy to ensure that the dataset reflects clinical complexity and builds the authenticity of the research.

In addition, EMRs pose additional clinical challenges that LLMs are not trained to solve. Unstructured EMRs carry precious and invaluable data, including images of handwritten prescriptions, scanned documents with poor resolution, and vital information embedded in physician notes. The use of optical character recognition software on such images enables text extraction and natural language processing. For example, in 189 patients with diffuse large B-cell lymphoma, such an approach allowed for assessment of performance status, prognostication, and survival evaluation.[7]

The data in EHRs are static and carry fixed/historical information, being devoid of real-time updates that are crucial for relevant monitoring. Kabak *et al.*[8] proposed a Health Level 7 Fast Healthcare Interoperability Resources–retrieval-augmented generation system to deliver real-time access to updated servers (clinical guidelines) that can navigate current challenges of medical LLMs. The incorporation of updated supportive and semantic segmentation tools into the record-keeping database provides a targeted and focused approach to the retrieval of research data. Furthermore, details on laboratory tests, biomarkers, and response to therapy can serve as templates for research frontiers.

3.2. Toward better care

In a systematic review and meta-analysis, Carl *et al.*[9] highlighted the role of transformer architecture in newer LLM models, showing a more efficient and accurate method to generate texts and visuals. Most studies lean toward presenting the correctness of oncology questions alone without fully capturing LLMs' holistic role. What needs addressing are standardized and validated tools in LLM-based research. Overfitting will augment the purported utility of applications while underperforming outside of a specific training dataset. Language model sources are heterogeneous by nature and can output less relevant data. In cutaneous oncology, with the integration of text-to-image and image-based output decisions, patient autonomy is gaining new ground. With self-upload of texts, documents, and images on popular LLM platforms, suspicious lesions or lesions such as actinic keratosis, seborrheic keratosis, keratoacanthoma, naevi, and pre-malignant conditions can be possibly screened or given directives to guide them toward specialist care at the right time. Consider a situation where malignancies with poor prognosis, such as acral lentiginous melanoma, are detected early via LLM tools. Prudent and prompt expert care is what can truly be labeled as man-and-machine teamwork, leading to meaningful applications in oncology care delivery.

The current limiting step in healthcare delivery lies in the cognitive overload by primary care physicians, who are required to triage all cases, often compounded by prolonged waiting times for specialist care. This has been a major impediment in many regions globally. Smart integration of LLMs, guided by stringent foresight and supported by robust regulatory mechanisms, represents a promising pathway to reform existing practices and advance healthcare delivery.[10]

3.3. Multimodal LLMs: The future is nearer than it seems

Multitasking models, called multimodal LLMs, can synthesize and analyze data, integrating dermoscopic images, histopathology reports, biomarkers, pharmacovigilance systems, and clinical registries into cohesive and explainable predictions or management recommendations. Akin to a tumor board, the incorporation of LLMs as decision-making adjuncts within a neo-multidisciplinary team seems not just plausible but increasingly essential.[8] As multimodal tools evolve toward specialized LLMs, their efforts, recommendations, discoveries, and predictions of oncocare could merge with human expertise through humans-in-the-loop frameworks. Such interdisciplinary collaboration, encompassing dermatology, dermatopathology, medical, surgical, and radiation oncology, alongside AI experts and LLM tools, could redefine precision oncology. With multiomics approaches, including genomics, transcriptomics, proteomics, and metabolomics, at the forefront of discovery, the future of cutaneous malignancies may radically shift.[11]

4. Not all is rosy

Many factors can influence the quality of AI-generated outputs. Practical and common errors include improper prompts—illustrated by the garbage in, garbage out principle—along with poor-quality smartphone images, inconsistencies in image pre- and post-processing, and variations in angles or lighting. A particularly critical concern lies in the representation of skin disease severity and diversity, which is profoundly influenced by pigment content and can exacerbate inequities in care. Errors owing to variations in clinical presentation across "melanin-rich" and "melanin-deficient" skin phototypes can lead to misdiagnosis and, consequently, inappropriate management or treatment decisions.

Most AI-based tools, including LLM-driven models, have been developed using datasets that underrepresent darker skin tones and rare diseases. For example, dermatofibrosarcoma protuberans may closely resemble several benign lesions, both in online datasets and in clinical practice. Technical assessment by trained individuals is the key driver of successful and efficient application. Emphasizing the uniqueness of health and diseases in humans requires the highest standards of clinical oversight, irrespective of digital tools. It is pertinent to recognize LLMs as devices, and their regulatory approval processes must acknowledge this.[12]

4.1. AI stewardship in practice

Risk mitigation in LLM applications remains crucial and complex. It is often argued that no test or diagnosis is 100% safe and reliable; likewise, LLMs are not exempt from errors analogous to those made by humans. Hallucinated data in the form of misleading and inaccurate medical information, such as inappropriate drug or procedure-related recommendations, requires stringent safety guardrails, expert validation, prompt rectification, and human review/approval. The dynamic nature of medicine further compounds these challenges, with constant updates in drug approvals, evolving guidelines, region-specific medical policies, and variable drug availability. Digitization of healthcare bears the risk of perpetuating and augmenting existing inequities in healthcare delivery.[13] Resource allocation decisions driven by commercial AI tools may inadvertently deepen such disparities. Moreover, privacy and security concerns, particularly the long-term implications of digital footprints that remain even after anonymization and pseudo-anonymization, pose significant ethical challenges. Ultimately, fairness and trust in AI cannot be seen as trade-offs in the pursuit of faster and more efficient health care.

5. Conclusion

Oncodermatology-specific LLMs that are interpretable, ethically deployed, and clinically validated—if integrated thoughtfully—may transform care at the skin–cancer interface, where nuanced symptoms, hidden patterns, and early warning signs often converge. Defining the scope of care and establishing timely, ethically driven regulatory foresight are urgently needed. In their current form, LLMs require rigorous assessment to delineate their strengths and limitations, explore avenues for purposeful applications, and optimize their use across administrative, clinical, educational, and research settings.

Acknowledgments

None.

Funding

None.

Conflict of interest

The authors declare that they have no competing interests.

Author contributions

Conceptualization: Mansak Shishak
Writing–original draft: Mansak Shishak, Sorun Shishak
Writing–review & editing: Sorun Shishak, Sameer Rastogi

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data

Not applicable.

References

1. Liu F, Zhou H, Gu B, *et al*. Application of large language models in medicine. *Nat Rev Bioeng.* 2025;3:445-464.

   doi: 10.1038/s44222-025-00279-5

2. Singhal K, Azizi S, Tu T, *et al*. Large language models encode clinical knowledge. *Nature.* 2023;620:172-180.

   doi: 10.1038/s41586-023-06291-3

3. Lane R, Hay A, Schwarz A, Berry DM, Shrager J. *Eliza Reanimated: The World's First Chatbot Restored On The World's First-Time Sharing System*. *arXiv.* 2025. Available from: https://arxiv.org/abs/2501.06707 [Last accessed on 2025 Oct 21].

   doi: 10.48550/arXiv.2501

4. Roster K, Kann RB, Farabi B, Gronbeck C, Brownstone N, Lipner SR. Readability and health literacy scores for ChatGPT-generated dermatology public education materials: Cross-sectional analysis of sunscreen and melanoma questions. *JMIR Dermatol.* 2024;7:e50163.

   doi: 10.2196/50163

5. Bender EM, Gebru T, McMillan-Major A, Shmitchell S. On the dangers of stochastic parrots: Can language models be too big? In: *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency (FAccT '21).* New York, NY: Association for Computing Machinery; 2021. p. 610-623.

   doi: 10.1145/3442188.3445922

6. Veccio C, Barley S, Kennedy LB, *et al*. Analysis of a large language model-based system versus manual review in clinical data abstraction and deduction from real-world medical records of patients with melanoma for clinical trial eligibility assessment. *J Clin Oncol.* 2025;43:1571.

   doi: 10.1200/JCO.2025.43.16_suppl.1571

7. Hom J, Nikowitz J, Ottesen R, Niland JC. Facilitating clinical research through automation: Combining optical character recognition with natural language processing. *Clin Trials.* 2022;19(5):504-511.

   doi: 10.1177/17407745221093621

8. Kabak Y, Erturkmen F, Gokce B, *et al*. F*HIR-RAG-MEDS: Integrating HL7 FHIR with Retrieval-Augmented Large Language Models for Enhanced Medical Decision Support*. arXiv Preprint. 2025.

   doi: 10.48550/arXiv.2509.07706

9. Carl N, Schramm F, Haggenmüller S, *et al*. Large language model use in clinical oncology. *NPJ Precis Oncol.* 2024;8:240.

   doi: 10.1038/s41698-024-00733-4

10. Chen D, Parsa R, Swanson K, *et al*. Large language models in oncology: A review. *BMJ Oncol.* 2025;4(1):e000759.

doi: 10.1136/bmjonc-2025-000759

11. Wei L, Niraula D, Gates EDH, *et al*. Artificial intelligence (AI) and machine learning (ML) in precision oncology: A review on enhancing discoverability through multiomics integration. *Br J Radiol.* 2023;96:20230211.

doi: 10.1259/bjr.20230211

12. Gilbert S, Kather JN, Hogan A. Augmented non-hallucinating large language models as medical information curators. *NPJ Digit Med.* 2024;7:100.

doi: 10.1038/s41746-024-01081-0

13. Crowe B, Rodriguez JA. Identifying and addressing bias in artificial intelligence. *JAMA Netw Open.* 2024; 7(8):e2425955.

doi: 10.1001/jamanetworkopen.2024.25955