Current Proteomics 22 (2025) 100057

Contents lists available at ScienceDirect

Current Proteomics

journal homepage: http://www.keaipublishing.com/en/journals/current-proteomics/

Applications of artificial intelligence and large language models in cancer immunotherapy

1. Introduction

Cancer immunotherapy has emerged in recent years as a major breakthrough in oncology, significantly improving the prognosis of patients with various types of cancer and achieving remarkable progress, particularly in the treatment of malignant tumors. However, substantial challenges such as therapeutic heterogeneity and drug resistance persist in clinical practice. The complexity and dynamic evolution of the tumor microenvironment not only influence the response to immune checkpoint inhibitors but also contribute to the development of primary or acquired resistance[1].

To address these challenges, artificial intelligence (AI) technologies, especially large language models (LLMs) based on deep learning, have become powerful tools for managing the complexity of cancer immunotherapy because of their strong capabilities in data processing and pattern recognition. AI can integrate and analyze multi source heterogeneous information, including high throughput sequencing, medical imaging, and clinical data, thereby providing critical support for precision oncology. As a cutting-edge branch of AI, LLMs possess the ability to deeply understand and generate natural language, enabling applications in medical text mining, clinical decision support, and multimodal data integration and analysis[2].

In summary, although cancer immunotherapy has provided substantial clinical benefits, tumor heterogeneity and therapeutic resistance continue to limit its overall effectiveness. AI technologies grounded in deep learning, particularly LLMs, are becoming important tools for overcoming these challenges due to their strong data processing capacity and advantages in cross modal information integration. This letter aims to summarize the latest applications of AI and LLMs in cancer immunotherapy and to offer theoretical and technical support for the advancement of precision cancer immunotherapy.

2. Applications of artificial intelligence in identifying targets for cancer immunotherapy

Integrating genomics, transcriptomics, proteomics, and other multi omics datasets using AI algorithms has become an important strategy for identifying targets in cancer immunotherapy[3]. Multi-omics data provide multi level information ranging from gene mutations and expression regulation to protein function. Through machine learning and deep learning techniques, AI can efficiently process and analyze these high dimensional and complex datasets to identify potential immunotherapy targets. For example, AI models can use feature selection and classification approaches to screen key genes and signaling pathways associated with tumor immune evasion, helping to reveal critical regulatory nodes within the tumor microenvironment[4]. In addition, by incorporating single cell sequencing data, AI can further dissect the heterogeneity of immune cells in the tumor microenvironment and assist in the precise localization of potential targets. Single cell RNA sequencing combined with AI has been shown to identify immune related molecular mechanisms, guide the discovery of new targets, and optimize immunotherapy strategies[5]. By integrating multi-dimensional omics information, AI-assisted approaches not only improve the accuracy and efficiency of target identification but also provide theoretical and technical support for personalized cancer immunotherapy (Fig. 1).

Deep learning models also demonstrate strong potential in analyzing large scale literature and clinical datasets to uncover novel immune checkpoints. Through natural language processing techniques, especially LLMs, AI can automatically extract and integrate massive amounts of textual information to rapidly identify new immunoregulatory targets and improve discovery efficiency. These AI assisted targets can be advanced toward clinical translation through strategies that combine bioinformatics prediction with experimental validation. In addition to supporting target discovery, AI can also assist in the design of new immunotherapies, thereby promoting the development of personalized cancer immunotherapy. With continued optimization of AI models and data integration methods, AI is expected to play an increasingly central role in the discovery and validation of novel immune checkpoints[6].

https://doi.org/10.1016/j.curpro.2025.100057
Received 5 November 2025; Received in revised form 20 November 2025; Accepted 1 December 2025

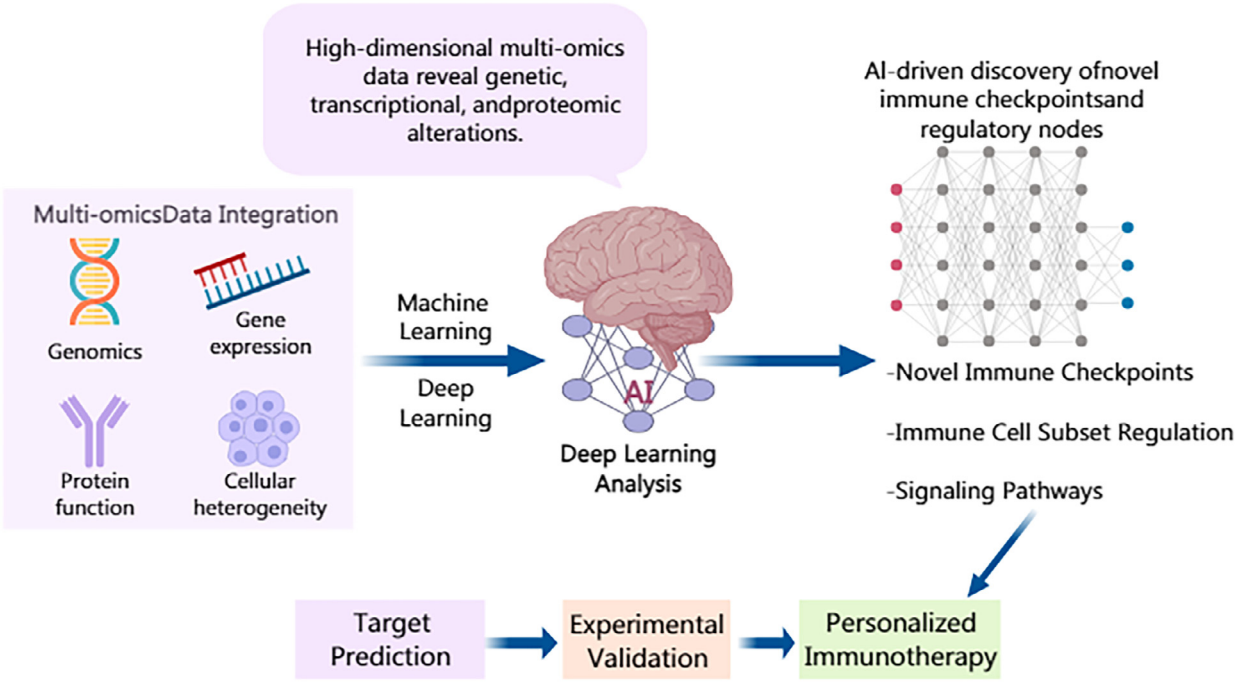

Fig. 1. Multi-omics data integration and AI-driven personalized immunotherapy development workflow.

3. The auxiliary role of large language models in cancer immunotherapy

The application of LLMs in cancer immunotherapy has greatly advanced the mining of clinical text data and the construction of knowledge graphs. LLMs are capable of processing massive volumes of medical literature and electronic health records (EHRs), extracting key information related to immunotherapy from both structured and unstructured data. For example, LLM-driven automated literature screening systems can efficiently identify biomarkers, drug targets, and therapeutic strategies in precision oncology, significantly improving both the accuracy and coverage of data retrieval[7]. In addition, LLMs can analyze patients' clinical notes, medical summaries, and laboratory reports to extract information on immunotherapy related adverse events, efficacy indicators, and patient feedback, thereby supporting the creation of comprehensive clinical knowledge bases[8].

Based on such information, constructing knowledge graphs for cancer immunotherapy becomes feasible. These knowledge graphs integrate multi-source data including literature, clinical records, genomics, and imaging, forming semantically rich networks of entities and relationships. In clinical trial design, drug target discovery, and adverse event mechanism studies, knowledge graphs can reveal complex biological relationships and potential causal mechanisms, supporting the development of intelligent question answering and clinical decision support systems[9].

In assisting physicians with personalized treatment planning, LLMs can deeply analyze clinical texts and, in combination with knowledge graphs, provide treatment recommendations and early warnings. For instance, in early warning systems for immune related adverse events, LLMs can automatically analyze patient language to detect potential neurotoxicities and other adverse reactions, enhancing patient safety. LLMs can also mine data from social media and patient forums to help clinicians understand patients' real world experiences with immunotherapy, facilitating personalized risk communication and psychosocial support[10].

LLMs demonstrate unique advantages in the design and optimization of cancer immunotherapy regimens. By integrating patients' clinical features, genomic information, and imaging data, LLMs can generate personalized immunotherapy recommendations to achieve precision medicine. For example, in non-small cell lung cancer (NSCLC) patients, LLMs combined with knowledge graphs and machine learning models can accurately predict immunotherapy responses and provide interpretable treatment strategies for clinicians. LLMs can also simulate the effects of different immunotherapy regimens, assisting in optimizing treatment strategies, such as the combination of surgery, radiotherapy, and immunotherapy, to minimize adverse effects and maximize efficacy[11] (Fig. 2).

Moreover, LLMs can continuously learn and update by integrating the latest clinical trial data and treatment guidelines, enabling dynamic optimization of therapeutic strategies. For instance, LLMs can help identify the optimal applications of different immune checkpoint inhibitors in combination with chemotherapy, offering NSCLC patients the best perioperative treatment plans[12].

In predicting and interpreting immune related biomarkers, LLMs can effectively identify potential markers associated with immunotherapy response by deeply mining gene expression data, clinical phenotypes, and pathological images. By integrating multi-omics data with machine learning, LLMs can screen for gene mutations, expression profiles, and immune cell infiltration features closely linked to immunotherapy outcomes. LLMs also enable the interpretation of complex biological information, translating high

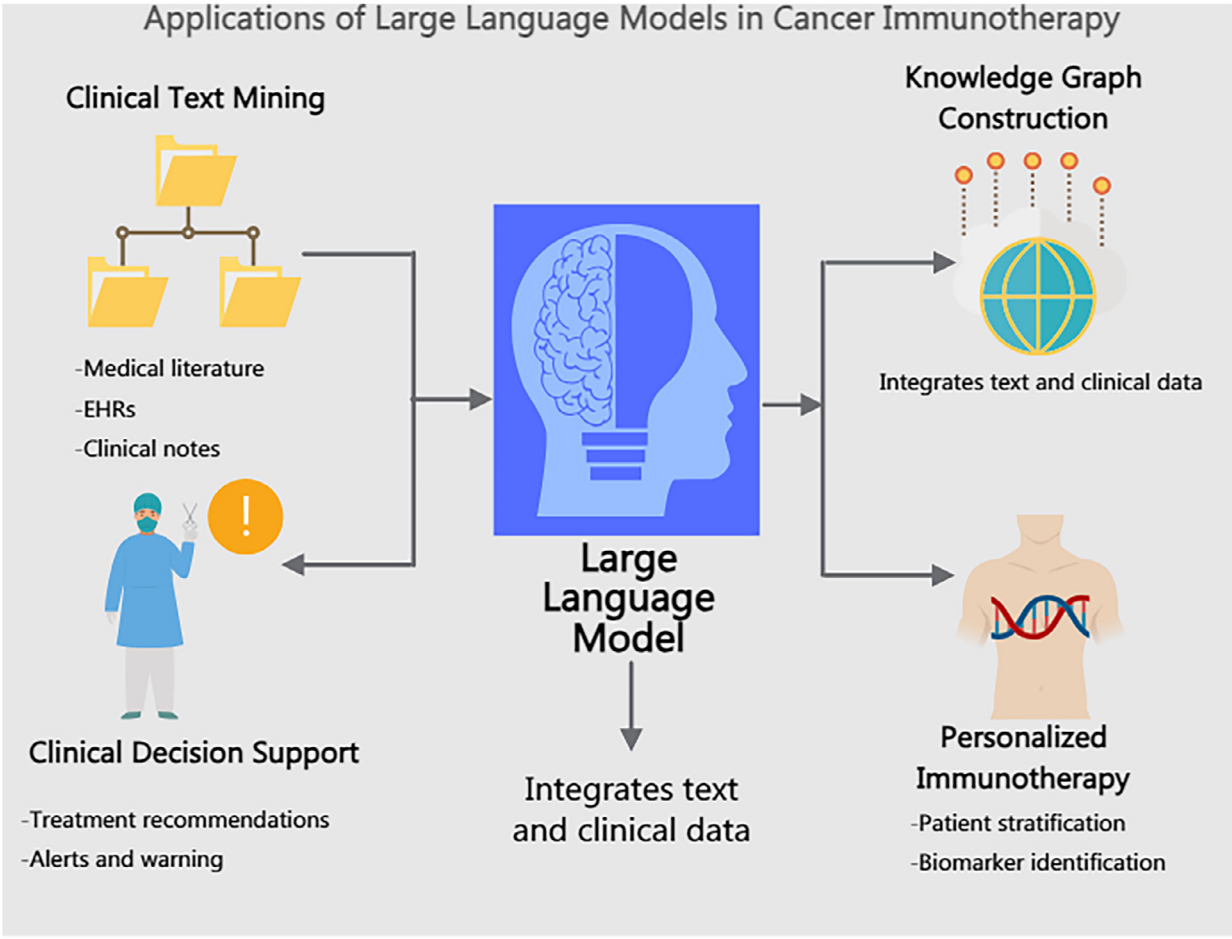

Fig. 2. Application scenarios of large language models in cancer immunotherapy.

dimensional omics data into clinically understandable insights and helping clinicians comprehend the biological significance of immune pathways and cellular interactions[13].

In early diagnosis and treatment monitoring, LLMs support real-time analysis of dynamic patient data, such as circulating tumor DNA, immune cell subset changes, and metabolomic indicators, predicting immunotherapy efficacy and providing early warnings of adverse events. By integrating imaging features and inflammation markers, LLMs can accurately predict early responses to immunotherapy in NSCLC patients, promoting the advancement of precision immunotherapy[14].

4. Challenges and future perspectives

High quality and diverse clinical and omics data are fundamental for training and optimizing AI models. In cancer immunotherapy, issues such as missing data, sample bias, and heterogeneity can significantly affect model performance. For instance, the complexity of multi-omics and clinical data in cancer immunotherapy means that incomplete or missing information can reduce the accuracy of predicting treatment responses. Furthermore, variations between institutions and isolated data silos limit data integration and sharing, further impacting the effectiveness of AI model training.

Patient privacy protection and data security represent another major challenge for AI applications in healthcare. Cancer patients' genomic information, imaging data, and electronic health records are highly sensitive and must comply with ethical and legal requirements. To ensure data security, techniques such as differential privacy, homomorphic encryption, and blockchain have been employed. In addition, federated learning, a distributed learning framework that does not require centralizing data, provides an effective solution to the conflict between cross institutional data sharing and privacy protection.

The "black box" nature of LLMs limits clinicians' trust and acceptance. AI models often lack transparency in their decision making, making it difficult to interpret predictions and hindering clinical adoption[15]. To improve model interpretability, researchers have developed various explainable AI (XAI) techniques, including feature importance based methods and counterfactual explanations. These approaches enable clinicians to better understand the driving factors behind treatment responses and the rationale for model inferences.

Looking ahead, research on AI model interpretability needs further strengthening. Combining multidisciplinary expert knowledge can help build intelligent support systems aligned with clinical workflows. In parallel, large scale, multicenter clinical validation trials are required to assess model safety, efficacy, and acceptability, thereby facilitating the clinical application of AI technologies in cancer immunotherapy.

5. Conclusion

Artificial intelligence, particularly large language models, has demonstrated remarkable potential in the field of cancer immunotherapy. The development of these technologies has not only facilitated the translation of basic research into clinical applications but also provided strong technical support for precision medicine. However, fully realizing their clinical value still requires overcoming challenges related to data quality, interpretability, and model reliability. In the future, efforts should focus on advancing research, promoting the integration of technological innovation with clinical needs, and driving the development of cancer immunotherapy to deliver more precise and effective treatment options for patients.

Funding statement

The author(s) declare financial support was received for the research and/or publication of this article.

Author contributions

YS: Conceptualization, Writing – original draft. JL: Conceptualization, Writing – original draft. JL: Writing – review & editing, Writing – original draft. XH: Project administration, Writing – review & editing. AH: Supervision. RF: Data curation. BX: Data curation. YW: Supervision. ZL: Supervision, Project administration, Writing – review & editing. QW: Supervision, Writing – review & editing.

Data availability statement

The datasets used and analysed during the current study are available from the corresponding author on reasonable request.

Ethics approval and consent to participate

Not applicable.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

1. Yang K, Halima A, Chan TA. Antigen presentation in cancer - mechanisms and clinical implications for immunotherapy. *Nature reviews. Clinical oncology*. 2023;20(9):604–623.
2. Bi WL, Hosny A, Schabath MB, Giger ML, Birkbak NJ, Mehrtash A, Allison T, Arnaout O, Abbosh C, Dunn IF, Mak RH, Tamimi RM, Tempany CM, Swanton C, Hoffmann U, Schwartz LH, Gillies RJ, Huang RY, Aerts H. Artificial intelligence in cancer imaging: Clinical challenges and applications. *CA: a cancer journal for clinicians,*. 2019;69(2):127–157.
3. Chen W, Zhang P, Tran TN, Xiao Y, Li S, Shah VV, Cheng H, Brannan KW, Youker K, Lai L, Fang L, Yang Y, Le NT, Abe JI, Chen SH, Ma Q, Chen K, Song Q, Cooke JP, Wang G. A visual-omics foundation model to bridge histopathology with spatial transcriptomics. *Nat Methods*. 2025;22(7):1568–1582.
4. Huang X, Rymbekova A, Dolgova O, Lao O, Kuhlwilm M. Harnessing deep learning for population genetic inference. *Nat Rev Genet*. 2024;25(1):61–78.
5. Tejada-Lapuerta A, Bertin P, Bauer S, Aliee H, Bengio Y, Theis FJ. Causal machine learning for single-cell genomics. *Nat Genet*. 2025;57(4):797–808.
6. Hu M, Alkhairy S, Lee I, Pillich RT, Fong D, Smith K, Bachelder R, Ideker T, Pratt D. Evaluation of large language models for discovery of gene set function. *Nat Methods*. 2025;22(1):82–91.
7. Wiest IC, Bhat M, Clusmann J, Schneider CV, Jiang X, Kather JN. Large language models for clinical decision support in gastroenterology and hepatology. *Nat Rev Gastroenterol Hepatol*. 2025;22(11):773–787.
8. Johri S, Jeong J, Tran BA, Schlessinger DI, Wongvibulsin S, Barnes LA, Zhou HY, Cai ZR, Van Allen EM, Kim D, Daneshjou R, Rajpurkar P. An evaluation framework for clinical use of large language models in patient interaction tasks. *Nat Med*. 2025;31(1):77–86.
9. Hu X, Gu L, Kobayashi K, Liu L, Zhang M, Harada T, Summers RM, Zhu Y. Interpretable medical image Visual Question Answering via multi-modal relationship graph learning. *Med Image Anal*. 2024;97:103279.
10. Elsaid MI, Meara AS, Owen DH. Role for Artificial Intelligence in the Detection of Immune-Related Adverse Events. *J Clin Oncol*. 2024;42(35):4119–4122.
11. Sun VH, Heemelaar JC, Hadzic I, Raghu VK, Wu CY, Zubiri L, Ghamari A, LeBoeuf NR, Abu-Shawer O, Kehl KL, Grover S, Singh P, Suero-Abreu GA, Wu J, Falade AS, Grealish K, Thomas MF, Hathaway N, Medoff BD, Gilman HK, Villani AC, Ho JS, Mooradian MJ, Sise ME, Zlotoff DA, Blum SM, Dougan M, Sullivan RJ, Neilan TG, Reynolds KL. Enhancing Precision in Detecting Severe Immune-Related Adverse Events: Comparative Analysis of Large Language Models and International Classification of Disease Codes in Patient Records. *J Clin Oncol*. 2024;42(35):4134–4144.
12. Carli F, Di Chiaro P, Morelli M, Arora C, Bisceglia L, De Oliveira Rosa N, Cortesi A, Franceschi S, Lessi F, Di Stefano AL, Santonocito OS, Pasqualetti F, Aretini P, Miglionico P, Diaferia GR, Giannotti F, Lio P, Duran-Frigola M, Mazzanti CM, Natoli G, Raimondi F. Learning and actioning general principles of cancer cell drug sensitivity. *Nat Commun*. 2025;16(1):1654.
13. Ligero M, El Nahhas OSM, Aldea M, Kather JN. Artificial intelligence-based biomarkers for treatment decisions in oncology. *Trends Cancer*. 2025;11(3):232–244.
14. Zhou R, Tang X, Wang Y. Emerging strategies to investigate the biology of early cancer. *Nat Rev Cancer*. 2024;24(12):850–866.
15. Temsah MH, Jamal A, Alhasan K, Temsah AA, Malki KH. OpenAI o1-Preview vs. ChatGPT in Healthcare: A New Frontier in Medical AI Reasoning. *Cureus*. 2024;16(10):e70640.

Yang Su[1], Ji'An Liu[1]
Shanghai Jiao Tong University School of Medicine, Shanghai 200011, China

Jing Li[1]
Shanghai Stomatological Hospital & School of Stomatology, Fudan University, Shanghai, China

Xufeng Huang
Faculty of Dentistry, University of Debrecen, Debrecen, Hungary
Department of Data Visualization, Faculty of Informatics, University of Debrecen, Debrecen, Hungary

András Hajdu
Department of Data Visualization, Faculty of Informatics, University of Debrecen, Debrecen, Hungary

Rao Fu
Shanghai Jiao Tong University School of Medicine, Shanghai 200011, China
Shanghai Xuhui District Dental Center, Shanghai, China

Bo Xu
Shanghai Jiao Tong University School of Medicine, Shanghai 200011, China

Yufeng Wang
College of Basic Medical Science, Yunnan University of Chinese Medicine, Kunming 650500, China

Xue Liu
School of Nursing, Shanghai JiaoTong University, Shanghai 200011, China

Zhengrui Li**
Shanghai Jiao Tong University School of Medicine, Shanghai 200011, China

Qi Wang*
Department of Oncology, Ruijin Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai, China

*Corresponding author at: Department of Oncology, Ruijin Hospital, Shanghai Jiao Tong University School of Medicine, No. 197, Ruijin 2nd Road, Huangpu District, Shanghai 200025. **Corresponding author at: Shanghai Jiao Tong University School of Medicine, No. 227, Chongqing South Road, Huangpu District, Shanghai 200025, China.
E-mail addresses: suyangxmu@163.com (Y. Su), lja2786@sjtu.edu.cn (J. Liu), jingj1031@163.com (J. Li), huangxufeng@mailbox.unideb.hu (X. Huang), hajdu.andras@inf.unideb.hu (A. Hajdu), frachel14@163.com (R. Fu), xb0115@sjtu.edu.cn (B. Xu), sgzh20000202@icloud.com (Y. Wang), xuexue11111031@163.com (X. Liu), lzr_0108@sjtu.edu.cn (Z. Li), drwang71111@163.com (Q. Wang)
[1] These authors contributed equally to this work.
Revised 20 November 2025