Viewpoint

AI Agents for Multimodal Oncology Diagnosis: Toward Transparent and Traceable Clinical Decision Support

Liuyang Yang[1*], PhD; Liyu Shan[2,3*], PhD; Xiangmei Yao[1*], PhD; Renbin Zhao[1], MD; Zengzheng Li[1], MD; Shuai Feng[1], MD; Yajie Wang[1], PhD

[1]Department of Hematology, The First People's Hospital of Yunnan Province, Affiliated Hospital of Kunming University of Science and Technology, Kunming, Yunnan, China

[2]The Third Affiliated Hospital of Kunming Medical University, Yunnan Cancer Hospital, Peking University Cancer Hospital Yunnan, Kunming, Yunnan, China

[3]The Second Affiliated Hospital of Kunming Medical University, Kunming, Yunnan, China

[*]these authors contributed equally

Corresponding Author:
Yajie Wang, PhD
Department of Hematology
The First People's Hospital of Yunnan Province, Affiliated Hospital of Kunming University of Science and Technology
No.157 Jinbi road
Kunming, Yunnan, 650032
China
Phone: 86 871 63639921
Email: kbb165wyj@sina.com

Abstract

Cancer diagnosis depends on data from radiology, digital pathology, molecular profiling, laboratory testing, and longitudinal clinical records. AI performs well in selected tasks, but most systems remain narrow and disconnected from the iterative reasoning required in oncology. This Viewpoint defines an AI agent as a feedback-driven system that maintains task state, selects among governed tools, observes results, and revises its plan under explicit safety constraints. This definition separates agents from multimodal foundation models, retrieval-augmented generation, and fixed workflow automation. We organize the discussion across multimodal data collection, preprocessing, fusion and representation learning, and diagnostic decision support. We distinguished agent-level evidence, component- or infrastructure-level evidence, and prospective propositions throughout. Clinical translation will require resilient failure handling, guideline version control, prospective evaluation, computational and workflow feasibility, and clinician authority over final decisions. The near-term opportunity is therefore transparent and traceable clinical decision support rather than autonomous cancer diagnosis.

(JMIR Cancer 2026;12:e103545) doi: 10.2196/103545

KEYWORDS

artificial intelligence agent; multimodal oncology diagnosis; clinical decision support; large language model; human–artificial intelligence collaboration; diagnostic workflow

Introduction

Cancer remains a leading cause of morbidity and mortality. Global Cancer Observatory (GLOBOCAN) 2022 estimated that there were nearly 20 million new cancer cases and 9.7 million cancer-related deaths worldwide in that year [1]. Treatment delays are associated with higher mortality across several cancer indications [2]. Timely diagnostic pathways are therefore important, but they must also account for genomic, histopathological, microenvironmental, and temporal heterogeneity [3].

Modern oncology diagnosis is intrinsically multimodal. Clinicians combine imaging, pathology, laboratory results, molecular assays, and longitudinal electronic records, often across separate specialties and information systems. This fragmentation creates cognitive and coordination burdens that motivate integrated workflows [4,5]. Task-specific AI has advanced lesion detection, pathology classification, molecular modeling, and risk prediction [4,6-8]. However, most deployed or evaluated models address a single bounded input-output task rather than the full diagnostic process.

Large language models (LLMs), multimodal foundation models, and tool-use frameworks create a technical basis for coordinating these tasks [9,10]. However, a model that accepts several modalities is not automatically an agent. Similarly, retrieval-augmented generation (RAG) adds external evidence to generation, and fixed workflow orchestration links predetermined modules, but neither property alone establishes runtime planning or feedback-driven control.

This Viewpoint examines AI agents as auditable orchestration layers for multimodal oncology diagnostics. We organized the field around 4 connected stages: data collection, preprocessing, multimodal fusion and representation learning, and diagnostic decision support. Our central position is deliberately bounded: agentic orchestration may improve coordination and auditability, but clinical benefit and safety remain deployment-specific questions.

From Task-Specific AI to Agentic Oncology Diagnostics

We used an operational definition of agentic behavior. In a conventional fixed pipeline, an output is generated through a predetermined composition, $y=f_n(...f_1(x))$, and the sequence of functions does not change at runtime. An AI agent instead maintains a state s_t and selects an action $a_t = \pi(g, s_t, A_t)$ from an allowed tool set A_t. For a clinical goal g, it observes each tool's result and updates its state before choosing the next action. The practical distinction therefore lies in conditional planning, tool selection, observation, and revision within a governed loop.

Four adjacent concepts should remain separate. A multimodal foundation model supplies reusable representations or generation across data types. RAG retrieves external sources to ground an answer. Fixed workflow orchestration executes a prescribed sequence of services. A feedback-driven AI agent can change the sequence, request missing evidence, retry or switch tools, abstain, and escalate to a clinician according to the evolving state. Long-term memory is optional and should be limited to an authorized, versioned, patient-specific or institutional state rather than treated as an unrestricted capability.

First, Figure 1 contrasts a conventional specialist synthesis with an agent-coordinated 4-stage workflow. Second, Figure 2 operationalizes the agentic layer as a plan-act-observe-update-check loop. Act coordinates the 4 stages in Figure 1B; update records state and provenance; and check evaluates evidence sufficiency, cross-modal consistency, and whether further action remains within the authorized scope. If these criteria are met, the synthesis proceeds to clinician review. Remediable gaps return the workflow to planning; gaps that cannot be supplemented trigger clinician escalation. The agent-coordinated elements in both figures are prospective design propositions and not validated clinical systems or performance claims.

We applied 3 evidence labels throughout this Viewpoint. Agent-level evidence evaluates a complete loop that plans, uses tools, and produces an integrated output, as in the external oncology evaluation by Ferber et al [11]. Component or infrastructure evidence evaluates a model, retrieval method, data pipeline, or governance mechanism that an agent could invoke [12]. Prospective propositions describe plausible clinical workflows that still require system-level and prospective validation. This hierarchy prevents strong component performance from being misread as proof of a clinically validated agent.

Figure 1. Conventional and AI agent–coordinated multimodal oncology workflows. (A) In the conventional workflow, the attending physician synthesizes modality-specific reports. (B) In the proposed agent-coordinated workflow, the agent coordinates multimodal data collection, preprocessing, fusion and feature extraction, and diagnostic decision support. CT: computed tomography; HIS: hospital information system; LIS: laboratory information system; LLM: large language model; PACS: picture archiving and communication system; PTC: papillary thyroid carcinoma; ROI: region of interest; TI-RADS 4C: Thyroid Imaging Reporting and Data System, category 4C.

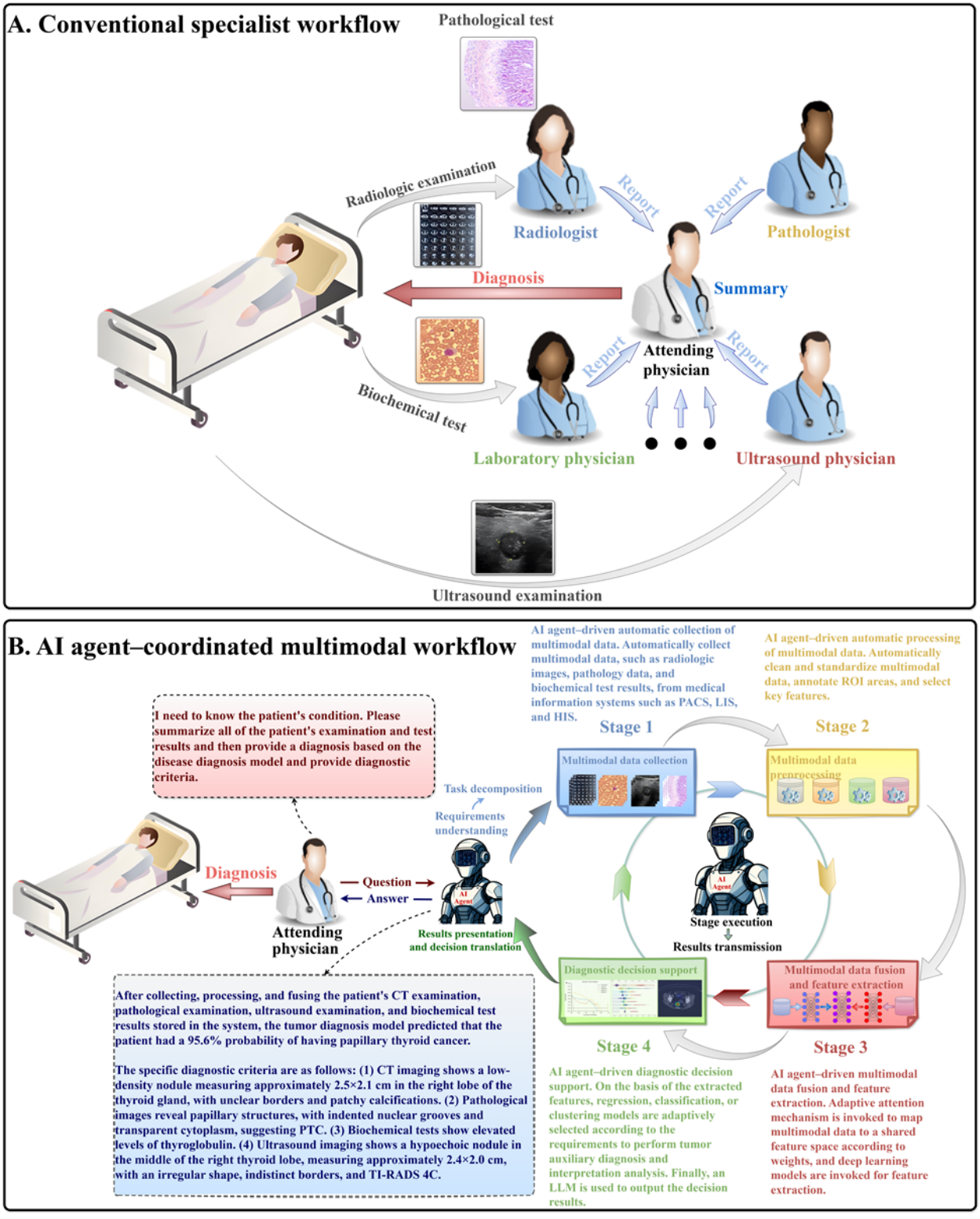
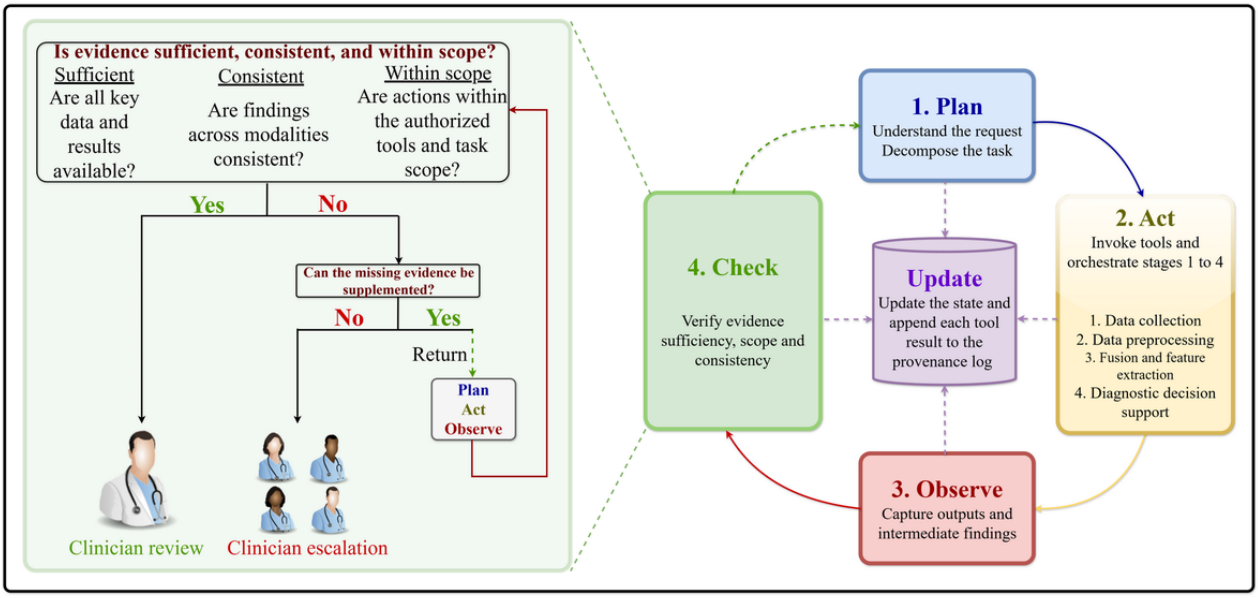


Figure 2. Feedback-driven orchestration and evidence-adequacy control for an AI agent in multimodal oncology diagnostics. The control architecture links plan, act, observe, update, and check. Act invokes governed tools across the 4 task stages shown in Figure 1B. Update appends each tool result to the current state and provenance log. Check evaluates whether evidence is sufficient and consistent across modalities and whether further action remains within the authorized scope. If these criteria are met, the synthesis proceeds to clinician review. Remediable gaps return the workflow to planning, whereas gaps that cannot be supplemented trigger clinician escalation.



Multimodal Data Collection

The first stage is the governed assembly of patient-specific evidence. Relevant data may be distributed across a hospital information system (HIS), laboratory information system (LIS), picture archiving and communication system (PACS), electronic medical record (EMR), pathology archives, and genomic platforms. These systems differ in format, ontology, temporal granularity, and access control.

Component-level infrastructure studies show that integration must be engineered. A single-center oncology data supply chain linked clinical, genomic, and imaging data for more than 170,000 patients across 11 cancer types and more than 800 features per case [13]. Its extract-transform-load (ETL), natural language processing (NLP), and quality control procedures support the feasibility of a data layer but not the validation of an agent. Standards such as Health Level Seven Fast Healthcare Interoperability Resources (HL7 FHIR) and Digital Imaging and Communications in Medicine (DICOM) can reduce site-specific translation [14]. Local mapping and governance nevertheless remain necessary.

An agent could use secure interfaces to identify, retrieve, and temporally align data for a specific diagnostic question. It could connect a new scan to prior pathology, biomarker trajectories, sequencing results, and notes [9,15]. Unstructured text and patient-reported or wearable data may add context in selected settings [16]. Each input should carry its source, acquisition and update times, missingness status, and freshness threshold. Contradictory or stale records should trigger clarification rather than silent fusion. Once evidence is assembled, the next risk is whether every modality is technically fit for analysis.

Multimodal Data Preprocessing

Multimodal preprocessing converts heterogeneous inputs into representations that downstream tools can use. Imaging varies by scanner and acquisition protocol. Digital pathology depends on tissue preparation, staining, and slide digitization [6]. Laboratory and genomic data may be incomplete or generated on different platforms, while clinical text often contains duplication and institution-specific shorthand.

A fixed workflow applies the same preprocessing sequence. An agentic workflow could select modality-specific tools according to file type, quality, and clinical purpose. It might invoke imaging quality control and segmentation or pathology stain normalization, tiling, tissue filtering, and feature extraction. Pathology foundation models provide reusable component representations [17,18], while self-supervised medical imaging models illustrate robustness and data-efficiency strategies [19]. Attribution tools such as gradient-weighted class activation mapping can support region-level inspection, although an attribution map is not itself a clinical explanation [20].

Every action should be recorded with the input source, time stamp, parser or model version, quality control result, and review status. If optical character recognition or parsing fails, the system should not fabricate a normalized record. It should retry within a bounded policy, switch to a validated alternate parser, return a partial result with an explicit missingness flag, or request human review. This provenance-aware preprocessing creates the conditions for fusion without concealing upstream uncertainty.

Multimodal Fusion and Representation Learning

Fusion is not the mere presence of multiple data types. Imaging describes macroscopic phenotype, pathology reveals tissue architecture, molecular assays characterize biological drivers, and longitudinal records provide clinical context. Multimodal models can combine these views for characterization and risk estimation [7], but many assume that all modalities are harmonized and simultaneously available.

At the molecular level, machine learning can integrate complementary omics layers and prioritize regulatory interactions. Omidi combined RNA sequencing and microRNA sequencing profiles from 2 public cohorts [21]. The analysis included 78 adrenocortical carcinoma samples from The Cancer Genome Atlas adrenocortical carcinoma (TCGA-ACC) and 250 normal adrenal samples from the Genotype-Tissue Expression (GTEx) project, reconstructing condition-specific competing endogenous RNA networks. Random forest models performed best for predicting microRNA-messenger RNA associations, and predicted interactions were cross-referenced with TargetScan and miRTarBase. This study is an in silico systems biology module that an agent could invoke. Batch effects, cohort imbalance, cross-cohort heterogeneity, and the absence of independent or experimental validation prevent these findings from being interpreted as validation of a multimodal agent.

An agent could select a fusion strategy according to the clinical question, available modalities, and uncertainty. If pathology and imaging are available while genomic testing is pending, it could use a modality-robust model, mark molecular evidence as unavailable, and revise the synthesis later. Explainable multimodal real-world models have identified patient-level marker contributions in large cohorts and undergone independent validation [22]. These results support component-level representation and explanation but not autonomous clinical orchestration. Fusion outputs must therefore pass an evidence-adequacy check before decision support. The evidence-adequacy gate and its return-or-escalate logic are shown in Figure 2.

Diagnostic Decision Support

The final stage translates analyzed evidence into a structured, reviewable diagnostic summary. The purpose is not autonomous diagnosis or treatment selection. It is to state the supported findings, unresolved conflicts, missing evidence, uncertainty, and provenance in a form that clinicians can audit.

A recent feasibility study evaluated Gemini 2.5 Pro (Alphabet Inc) and ChatGPT 4o (OpenAI) on 80 Turkish-language prostate-specific membrane antigen positron emission tomography–computed tomography (PSMA PET-CT) reports [23]. The prompts incorporated the American Joint Committee on Cancer (AJCC) staging, the Chemohormonal Therapy Versus Androgen Ablation Randomized Trial for Extensive Disease (CHAARTED) criteria, and few-shot examples. The models assigned T, N, and M categories and disease volume classes, with overall task accuracies of 93.8% and 91.3%, respectively. Errors clustered around equivocal findings and different interpretive thresholds, including overstaging of ambiguous lesions. The single-center design, 1 expert reference standard, and exploratory sample of 80 reports limit generalizability. The study supports an LLM-based report-to-staging component that requires expert oversight; it does not validate a complete AI agent.

Other component studies illustrate evaluable end points. Multimodal oncology chatbots have been tested on image-containing clinical cases [24]. Explanation types can alter physician trust and diagnostic performance [25]. On the contrary, fabricated details can induce hallucinated clinical elaboration despite simple mitigation attempts [26]. Agent evaluation should therefore include tool selection, evidence citation, contradiction handling, safe refusal, calibration, workflow burden, and final conclusion quality rather than fluency alone.

In practice, an agent might prepare a multidisciplinary team (MDT) case by linking a suspicious lung lesion to pathology, molecular markers, smoking history, and prior imaging. It could rank competing diagnoses, identify a missing confirmatory test, and provide an evidence-linked summary. Proposed MDT assistance and randomized source-attribution studies support case preparation and verifiable retrieval as useful components [27,28]. Clinicians must retain authority over the final judgment, with explicit responsibility assigned across users, institutions, and manufacturers [29].

Challenges and Future Directions

Limits of the Current Evidence

The evidence base remains uneven. Most cited studies validate models, infrastructure, or simulated workflows rather than prospectively deployed agentic systems. The following implementation requirements should therefore be read as a translational agenda and not as established clinical effectiveness.

Data Quality and Interoperability

Interoperability does not guarantee semantic correctness. A deployable system must map local fields to versioned ontologies, record lineage and freshness, and quantify missingness. It must also test performance across scanners, laboratories, institutions, demographic groups, and unavailable-modality patterns [13,14,19]. External validation and subgroup analysis remain necessary because an agent can amplify errors introduced by any upstream component.

Operational Resilience and Graceful Degradation

Real-time MDT use requires an explicit latency budget for retrieval, preprocessing, inference, and rendering. Institutions may precompute stable features, use asynchronous tasks, or choose smaller local models when the expected benefit does not justify the delay. Time-outs should activate circuit breakers rather than indefinite retries. A parser or API failure should trigger a typed error and one of the following governed responses: a bounded retry, use of a validated alternate tool, a partial-result flag, or abstention with human escalation. The audit record should preserve the failed attempt and the fallback

path. This fallback logic corresponds to the supplementation and clinician-escalation branches in Figure 2.

Knowledge Life Cycle and Guideline Synchronization

Oncology guidelines and drug knowledge change frequently. Retrieval stores should preserve source dates, jurisdiction, guideline version, and effective period. Updates require staged ingestion, regression tests against reference cases, comparison with the previous version, canary deployment, and rollback criteria. A response should identify the version used and warn when local policy conflicts with an external guideline. This reduces, but cannot eliminate, legacy hallucinations.

Evaluation and Clinical Integration

Evaluation should follow the clinical workflow and the evidence label. Component metrics include extraction accuracy, calibration, source verifiability, latency, throughput, and failure rate. Agent-level evaluation additionally requires plan appropriateness, tool-use accuracy, recovery from failed steps, contradiction resolution, abstention, clinician workload, and patient safety end points [11,24-28]. Prospective studies should compare the full human-agent workflow with current practice and not only the agent with a static model.

Governance, Cost, and Institutional Adoption

Deployment also depends on computing cost, local infrastructure, cybersecurity, procurement, workflow fit, regulatory status, reimbursement, and institutional adoption. Trustworthy AI principles require fairness, traceability, usability, robustness, and explainability [12]. Precision oncology governance should add access control, deidentification, audit logging, incident response, continuous monitoring, model and knowledge base change control, and explicit accountability [29,30]. These controls are prerequisites for a clinical service and not evidence that the service improves outcomes.

Table 1 summarizes the current evidence supporting AI agents in oncology diagnostics and the limitations that constrain their interpretation and clinical use.

Table 1. Evidence supporting and delimiting AI agents in oncology diagnostics.

Evidence domains	Study design or setting	Main empirical finding	Agentic relevance and boundary
Agent-level oncology decision support	External evaluation on 20 realistic multimodal oncology cases [11]	Reported 87.5% correct tool use, 91% correct clinical conclusions, and 75.5% accurate guideline citation	Agent-level evidence from simulated cases; a useful benchmark but not prospective clinical validation
Real-world multimodal infrastructure	Single-center data supply chain covering >170,000 patients, 11 cancer types, and >800 features per case [13]	Integrated clinical, genomic, and imaging data with extract-transform-load, natural language processing, quality control, and continuously updated records	Infrastructure evidence; supports a callable data layer but does not evaluate agent planning or clinical benefit
Multimodal oncology chatbot benchmarking	Case-based evaluation of multimodal chatbots on 79 image-containing oncology cases [24]	Demonstrated that image-text oncology reasoning can be measured with case-level accuracy	Component evidence; a chatbot benchmark does not establish workflow orchestration or safety
Large language model–assisted prostate cancer staging from unstructured imaging reports	Single-center feasibility study of 80 Turkish-language prostate-specific membrane antigen positron emission tomography–computed tomography reports with embedded American Joint Committee on Cancer and Chemohormonal Therapy Versus Androgen Ablation Randomized Trial for Extensive Disease criteria and few-shot prompting [23]	Overall task accuracy was 93.8% for Gemini 2.5 Pro and 91.3% for ChatGPT 4o; errors clustered around equivocal findings	Component evidence; supports a report-to-staging module requiring expert oversight and not a complete agent
Machine learning integration of molecular regulatory networks	In silico integration of 78 The Cancer Genome Atlas adrenocortical carcinoma tumors and 250 normal adrenal samples from the Genotype-Tissue Expression project using RNA and microRNA sequencing [21]	Random forest best predicted microRNA-messenger RNA associations; predictions were cross-referenced with TargetScan and miRTarBase.	Specialized analytic module; cohort imbalance, batch effects, and the absence of independent or experimental validation limit inference
Tumor MDT[a] workflow	Workflow-focused proposal for AI assistance in tumor MDTs [27]	Identified opportunities for case preparation, synthesis of unstructured data, and rationale capture	Prospective workflow proposition; final decisions remain clinician led and require comparative evaluation
Source-attributed uro-oncology decision support	Randomized reader evaluation of a retrieval-augmented uro-oncology chatbot [28]	Improved recommendation correctness, source attribution, and source verifiability compared with a general chatbot	Component evidence; supports governed retrieval and verifiable sources and not full agent autonomy
Explainable multimodal real-world oncology AI	Multimodal model in 15,726 patients across 38 solid cancer types with external validation in 3288 patients with lung cancer [22]	Estimated patient-level marker contributions and prognostic interactions	Model-level evidence; supports reusable prediction and explanation tools with external validity boundaries
Hallucination and safety	Multimodel simulation of adversarial hallucination during clinical decision support [26]	Fabricated details induced false elaboration despite simple mitigation prompts	Safety evidence supporting contradiction checks, retrieval constraints, abstention, and human escalation
Interoperability	Clinical interoperability analysis across the AI data life cycle [14]	Emphasized standardized information exchange for data entry, processing, sharing, replication, and analytics	Infrastructure evidence; local semantic mapping and provenance remain necessary
Governance, liability, and implementation	Consensus trustworthy AI guidance, liability analysis, and precision oncology implementation review [12,29,30]	Identified requirements for fairness, traceability, usability, robustness, explainability, responsibility, safety, governance, cost, and harmonization	Cross-cutting requirements; governance controls enable evaluation but are not evidence of clinical effectiveness

[a]MDT: multidisciplinary team.

Conclusions

AI agents should be understood as auditable, feedback-driven orchestration systems rather than as another name for multimodal models or RAG. Oncology-specific studies now support selected agent-level, component-level, and infrastructure capabilities, including tool use, report-to-staging, multimodal modeling, and source attribution. However, these findings do not establish the safety or benefit of autonomous diagnosis. The defensible near-term goal is a resilient human-agent workflow that exposes provenance, uncertainty, failures, guideline versions, and escalation decisions. Prospective comparison with current clinical practice is required before broader implementation claims are justified.

Acknowledgments

The authors thank colleagues and collaborators for valuable discussions and conceptual input and acknowledge institutional research platforms that facilitated literature retrieval and manuscript preparation. No other individuals or organizations contributed in a manner that met authorship criteria.

ChatGPT (version 5.6; OpenAI) was used during revision to support language polishing. The authors reviewed and approved all AI-assisted revisions and take full responsibility for the manuscript.

Funding

This work was supported by the National Natural Science Foundation of China (grant 72604378), Kunming Medical University Clinical Research Projects (grant MR-KYLY-6), Yunnan Fundamental Research Projects (grant 202501AU070024), Yunnan Provincial Department of Education Science Research Fund Project (grant 2025J0166), Yunnan Provincial Clinical Medical Center for Blood Disease and Thrombosis Prevention and Treatment (grant 2024YNLCYXZX0253), Project of Yunnan Province Clinical Research Center for Hematologic Disease (grant 202505AJ310003), Yunnan Province Major Difficult Diseases Chinese and Western Clinical Cooperation Pilot Project—Leukemia, and the Major Collaborative Project on Clinical Treatment of Critical Illnesses by Traditional Chinese and Western Medicine.

Data Availability

Data sharing is not applicable to this article as no datasets were generated or analyzed during this study.

Authors' Contributions

Conceptualization: LY, LS, XY, ZL
Critical revision of the manuscript: ZL, YW
Funding acquisition: LY, YW
Literature review: XY, RZ
Project administration: YW
Study design: LY
Supervision: YW
Visualization: LS
Writing—original draft: LY
Writing—review and editing: LS, XY, RZ, SF
All authors contributed to the intellectual development and final approval of the manuscript.

Conflicts of Interest

None declared.

References

1. Bray F, Laversanne M, Sung H, Ferlay J, Siegel RL, Soerjomataram I, et al. Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin. 2024;74(3):229-263. [FREE Full text] [doi: 10.3322/caac.21834] [Medline: 38572751]
2. Hanna TP, King WD, Thibodeau S, Jalink M, Paulin GA, Harvey-Jones E, et al. Mortality due to cancer treatment delay: systematic review and meta-analysis. BMJ. Nov 04, 2020;371:m4087. [FREE Full text] [doi: 10.1136/bmj.m4087] [Medline: 33148535]
3. Dagogo-Jack I, Shaw AT. Tumour heterogeneity and resistance to cancer therapies. Nat Rev Clin Oncol. Feb 2018;15(2):81-94. [doi: 10.1038/nrclinonc.2017.166] [Medline: 29115304]
4. Lotter W, Hassett MJ, Schultz N, Kehl KL, Van Allen EM, Cerami E. Artificial intelligence in oncology: current landscape, challenges, and future directions. Cancer Discov. May 01, 2024;14(5):711-726. [doi: 10.1158/2159-8290.CD-23-1199] [Medline: 38597966]
5. Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. Nat Med. Jan 2019;25(1):44-56. [doi: 10.1038/s41591-018-0300-7] [Medline: 30617339]
6. Bera K, Schalper KA, Rimm DL, Velcheti V, Madabhushi A. Artificial intelligence in digital pathology - new tools for diagnosis and precision oncology. Nat Rev Clin Oncol. Nov 2019;16(11):703-715. [FREE Full text] [doi: 10.1038/s41571-019-0252-y] [Medline: 31399699]
7. Lipkova J, Chen RJ, Chen B, Lu MY, Barbieri M, Shao D, et al. Artificial intelligence for multimodal data integration in oncology. Cancer Cell. Oct 10, 2022;40(10):1095-1110. [FREE Full text] [doi: 10.1016/j.ccell.2022.09.012] [Medline: 36220072]

8. Bhinder B, Gilvary C, Madhukar NS, Elemento O. Artificial intelligence in cancer research and precision medicine. Cancer Discov. Apr 2021;11(4):900-915. [FREE Full text] [doi: 10.1158/2159-8290.CD-21-0090] [Medline: 33811123]
9. Shah NH, Entwistle D, Pfeffer MA. Creation and adoption of large language models in medicine. JAMA. Sep 05, 2023;330(9):866-869. [doi: 10.1001/jama.2023.14217] [Medline: 37548965]
10. Huang J, Xu Y, Wang Q, Wang QC, Liang X, Wang F, et al. Foundation models and intelligent decision-making: progress, challenges, and perspectives. Innovation (Camb). May 12, 2025;6(6):100948. [FREE Full text] [doi: 10.1016/j.xinn.2025.100948] [Medline: 40528892]
11. Ferber D, El Nahhas OS, Wölflein G, Wiest IC, Clusmann J, Leßmann ME, et al. Development and validation of an autonomous artificial intelligence agent for clinical decision-making in oncology. Nat Cancer. Aug 2025;6(8):1337-1349. [doi: 10.1038/s43018-025-00991-6] [Medline: 40481323]
12. Lekadir K, Frangi AF, Porras AR, Glocker B, Cintas C, Langlotz CP, et al. FUTURE-AI: international consensus guideline for trustworthy and deployable artificial intelligence in healthcare. BMJ. Feb 05, 2025;388:e081554. [FREE Full text] [doi: 10.1136/bmj-2024-081554] [Medline: 39909534]
13. Chang JS, Kim H, Baek ES, Choi JE, Lim JS, Kim JS, et al. Continuous multimodal data supply chain and expandable clinical decision support for oncology. NPJ Digit Med. Feb 27, 2025;8(1):128. [FREE Full text] [doi: 10.1038/s41746-025-01508-2] [Medline: 40016534]
14. Rehburg F, Graefe A, Hübner M, Thun S. How interoperability can enable artificial intelligence in clinical applications. Stud Health Technol Inform. Aug 22, 2024;316:596-600. [doi: 10.3233/SHTI240485] [Medline: 39176813]
15. Esteva A, Robicquet A, Ramsundar B, Kuleshov V, DePristo M, Chou K, et al. A guide to deep learning in healthcare. Nat Med. Jan 2019;25(1):24-29. [doi: 10.1038/s41591-018-0316-z] [Medline: 30617335]
16. Manz CR, Schriver E, Ferrell WJ, Williamson J, Wakim J, Khan N, et al. Association of remote patient-reported outcomes and step counts with hospitalization or death among patients with advanced cancer undergoing chemotherapy: secondary analysis of the PROStep randomized trial. J Med Internet Res. May 17, 2024;26:e51059. [FREE Full text] [doi: 10.2196/51059] [Medline: 38758583]
17. Chen RJ, Ding T, Lu MY, Williamson DF, Jaume G, Song AH, et al. Towards a general-purpose foundation model for computational pathology. Nat Med. Mar 2024;30(3):850-862. [doi: 10.1038/s41591-024-02857-3] [Medline: 38504018]
18. Xu H, Usuyama N, Bagga J, Zhang S, Rao R, Naumann T, et al. A whole-slide foundation model for digital pathology from real-world data. Nature. Jun 2024;630(8015):181-188. [FREE Full text] [doi: 10.1038/s41586-024-07441-w] [Medline: 38778098]
19. Azizi S, Culp L, Freyberg J, Mustafa B, Baur S, Kornblith S, et al. Robust and data-efficient generalization of self-supervised machine learning for diagnostic imaging. Nat Biomed Eng. Jun 2023;7(6):756-779. [doi: 10.1038/s41551-023-01049-7] [Medline: 37291435]
20. Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D, Batra D. Grad-CAM: visual explanations from deep networks via gradient-based localization. Int J Comput Vis. Oct 11, 2020;128(2):336-359. [doi: 10.1007/s11263-019-01228-7]
21. Omidi J. Integrative machine learning-driven prioritization of ceRNA networks in adrenocortical carcinoma. Intell Oncol. Apr 2026;2(2):100056. [doi: 10.1016/j.intonc.2026.100056]
22. Keyl J, Keyl P, Montavon G, Hosch R, Brehmer A, Mochmann L, et al. Decoding pan-cancer treatment outcomes using multimodal real-world data and explainable artificial intelligence. Nat Cancer. Feb 2025;6(2):307-322. [doi: 10.1038/s43018-024-00891-1] [Medline: 39885364]
23. Ismayilov R, Aktas A, Gencoglu EA, Oguz A, Altundag O, Akcali Z. Staging prostate cancer with AI: a comparative study of large language models and expert interpretation on PSMA PET-CT reports. Mol Imaging Biol. Feb 2026;28(1):93-105. [doi: 10.1007/s11307-025-02072-7] [Medline: 41350971]
24. Chen D, Huang RS, Jomy J, Wong P, Yan M, Croke J, et al. Performance of multimodal artificial intelligence chatbots evaluated on clinical oncology cases. JAMA Netw Open. Oct 01, 2024;7(10):e2437711. [FREE Full text] [doi: 10.1001/jamanetworkopen.2024.37711] [Medline: 39441598]
25. Prinster D, Mahmood A, Saria S, Jeudy J, Lin CT, Yi PH, et al. Care to explain? AI explanation types differentially impact chest radiograph diagnostic performance and physician trust in AI. Radiology. Nov 2024;313(2):e233261. [doi: 10.1148/radiol.233261] [Medline: 39560483]
26. Omar M, Sorin V, Collins JD, Reich D, Freeman R, Gavin N, et al. Multi-model assurance analysis showing large language models are highly vulnerable to adversarial hallucination attacks during clinical decision support. Commun Med (Lond). Aug 02, 2025;5(1):330. [FREE Full text] [doi: 10.1038/s43856-025-01021-3] [Medline: 40753316]
27. Geukes Foppen RJ, Morkūnas M, Traverso A, Dienstmann R. AI assistance in tumor multidisciplinary teams. ESMO Real World Data Digit Oncol. Feb 12, 2026;11:100684. [FREE Full text] [doi: 10.1016/j.esmorw.2026.100684] [Medline: 41930313]
28. Carl N, Hetz MJ, Wies C, Haggenmüller S, Winterstein JT, Mangold MH, et al. Enhancing clinicians' trust in large language models via transparent source attribution: a randomized controlled evaluation in uro-oncology. Eur J Cancer. Jan 17, 2026;233:116168. [FREE Full text] [doi: 10.1016/j.ejca.2025.116168] [Medline: 41401634]
29. Gerke S, Simon DA, Roman BR. Liability risks of ambient clinical workflows with artificial intelligence for clinicians, hospitals, and manufacturers. JCO Oncol Pract. Mar 2026;22(3):357-361. [doi: 10.1200/OP-24-01060] [Medline: 40749149]

30. Hamamoto R, Koyama T, Takahashi S, Yasuda T, Kobayashi K, Akagi Y, et al. Implementing generative artificial intelligence in precision oncology: safety, governance, and significance. J Hematol Oncol. Feb 09, 2026;19(1):14. [FREE Full text] [doi: 10.1186/s13045-026-01781-y] [Medline: 41664164]

Abbreviations

AJCC: American Joint Committee on Cancer
CHAARTED: Chemohormonal Therapy Versus Androgen Ablation Randomized Trial for Extensive Disease
DICOM: Digital Imaging and Communications in Medicine
EMR: electronic medical record
ETL: extract-transform-load
GLOBOCAN: Global Cancer Observatory
GTEx: Genotype-Tissue Expression
HIS: hospital information system
HL7 FHIR: Health Level Seven Fast Healthcare Interoperability Resources
LIS: laboratory information system
LLM: large language model
MDT: multidisciplinary team
NLP: natural language processing
PACS: picture archiving and communication system
PSMA PET-CT: prostate-specific membrane antigen positron emission tomography–computed tomography
RAG: retrieval-augmented generation
TCGA-ACC: The Cancer Genome Atlas adrenocortical carcinoma

Edited by M Balcarras; submitted 04.Jun.2026; peer-reviewed by R Ismayilov, J Omidi; comments to author 13.Aug.2026; revised version received 19.Aug.2026; accepted 20.Aug.2026; published 15.Sep.2026

Please cite as:
Yang L, Shan L, Yao X, Zhao R, Li Z, Feng S, Wang Y
AI Agents for Multimodal Oncology Diagnosis: Toward Transparent and Traceable Clinical Decision Support
JMIR Cancer 2026;12:e103545
URL: https://cancer.jmir.org/2026/1/e103545
doi: 10.2196/103545
PMID: