https://doi.org/10.1038/s41591-025-03836-y

Rapid deployment of large language model DeepSeek in Chinese hospitals demands a regulatory response

Tianyi Shen, Yuxi Li, Yanlin Cao, Xin Du, Xinru Wang, Yajuan Zhang & Yi Zhang

DeepSeek's potential for on-premises deployment in hospitals reveals a regulatory 'gray area' that requires enhanced guidance.

DeepSeek, a Chinese technology startup, has released a series of open-source large language models (LLMs), notably its flagship reasoning LLM: DeepSeek-R1. Introduced in January 2025, DeepSeek-R1 rivals dominant US models such as OpenAI's o1 while substantially reducing operational costs[1,2]. Recent research highlights the potential for LLMs to improve efficiency and effectiveness in various medical tasks, including clinical decision support, medical documentation, research and education, and chronic care support[3,4]. However, deployment of LLMs in physical hospitals, which involves actual physicians and patients, has faced persistent challenges, including data privacy, high computational costs, integration complexity and limited explainability[5].

Prior to DeepSeek-R1, these challenges restricted the adoption of LLMs to a limited number of large hospitals in China. However, this landscape is rapidly evolving. As of 8 May 2025, we had identified more than 755 healthcare institutions, ranging from leading hospitals to primary care institutions, that had deployed DeepSeek-R1. Among these institutions, over 500 had achieved on-premises deployment (OPD), in which one or more DeepSeek models are deployed and operated within the hospital's own infrastructure rather than using external cloud services or third-party applications. These institutions have applied DeepSeek-R1 in various scenarios, including clinical services, hospital operations, research and education, and personal health-management services (Table 1 provides actual OPD scenarios). Notably, deployments have also been observed in underdeveloped provinces with lower per-capita gross domestic product and limited healthcare expenditure. Nonetheless, these deployments currently exist in regulatory 'gray areas' and pose risks to both patients and hospitals, which highlights the urgent need for enhanced regulatory guidance.

DeepSeek's technical advantages

DeepSeek has emerged in China as the first widely adopted hospital LLM solution, owing to several technical advantages, notably its high price-to-performance ratio[6] and open-source customizability, both crucial for reducing hospital deployment costs. Currently, two commercial all-in-one systems dominate for DeepSeek-R1, and both configurations remain affordable for many hospitals. Relative to the reasoning performance of earlier LLMs, DeepSeek-R1's multi-stage training approach achieves higher reasoning performance, comparable to that of OpenAI's o1, which is crucial for complex medical tasks[2,7].

DeepSeek-R1 also adopts the Massachusetts Institute of Technology's open-source license, which grants substantial flexibility for use, modification and distribution. This ensures compatibility with existing hospital systems. By offering smaller models distilled from DeepSeek-R1 with lower hardware requirements, DeepSeek brings OPD costs below US $100,000 in practice. Moreover, hospitals can select customized solutions on the basis of their specific needs (Table 1), hardware availability and budget constraints. Consequently, DeepSeek substantially lowers the threshold for OPD in hospitals.

Prior to the release of DeepSeek, many top-tier hospitals in China had already begun experimenting with open-source LLMs developed by companies such as Qwen and Llama. For these institutions, achieving rapid OPD was feasible even within a short time frame. However, we acknowledge the potential bias inherent in using institutional announcements to infer actual deployment of LLMs. In the absence of a government registry or official deployment database, we relied on the Tencent Official Account platform, where Chinese healthcare institutions maintain official accounts, as our authoritative source.

To minimize the discrepancy between announcements and actual deployments, we applied a refined screening methodology, identifying 755 health institutions with actual deployments. To further ensure the dataset's authenticity, we conducted a survey of 70 randomly selected members of the China Hospital Information Management Association who are responsible for information technology (IT) departments at their respective institutions and therefore have first-hand knowledge of DeepSeek's actual deployment status. The survey achieved an 83% response rate, with responses received from 58 healthcare institutions across 15 provinces, municipalities and autonomous regions. This dual-source validation confirmed the accuracy and reliability of our dataset.

Why hospitals favor DeepSeek

First, most medical applications impose stringent security and confidentiality demands on both training data and user inputs (Table 1). The OPD enabled by DeepSeek ensures that all data remain on in-house servers, which minimizes the risk of data leaks and safeguards patient privacy. This approach complies with current legal standards for medical data protection and reduces compliance risks when hospitals integrate LLMs into internal systems and use proprietary data for domain-specific development and fine-tuning. Fig. 1 illustrates the real-world system architecture of an operational DeepSeek OPD, wherein the on-premises infrastructure module and control layer ensure the security requirements described above.

Second, DeepSeek facilitates the integration of LLMs into existing workflows. Hospitals can develop and fine-tune LLMs tailored specifically to their needs and integrate these models smoothly with their core internal systems (Fig. 1). Crucially, hospital-led model development, combined with training for staff and healthcare professionals, may help

Table 1 | Scenarios and demands for real-world deployment of DeepSeek healthcare solutions in hospitals

Scenario	Task	Applications and functions	Data privacy and security demands (5-point scale)	Reasoning capability and explainability demands (5-point scale)
Clinical service	Pre-clinical assistance	Chatbot for patient navigation and triage: (1) clarifies patient symptoms and needs via Q&A; and (2) provides non-emergency triage and care pathway and department recommendations on the basis of information from (1), with ancillary functions that include route planning and hospital policy queries	III The inference inputs include only vague patient symptom information	III Although it targets laypeople, it involves preliminary inference of symptoms, which affects patients' healthcare-seeking behaviors
		Pre-consultation inquiry: enables patients to record their medical history and symptoms before consultation, automatically generating a preliminary diagnosis and structured report to equip physicians with comprehensive background information	IV The inference inputs include certain types of patent privacy data	
	Clinical decision support	Diagnostic decision support: leverages a fine-tuning database of extensive medical documents to comprehensively analyze patients' chief complaints, demographic information, test reports and medical records history, providing diagnostic and further examination recommendations (some systems incorporate multi-modal modules to process medical images and symptoms for auxiliary diagnosis)	V In addition to publicly available medical documents, the fine-tuning datasets also include highly confidential and sensitive medical data, including EMRs, medical tests, and biological information; moreover, the inference inputs also include detailed patient privacy data and demographic information	V Designed to assist HPs and directly impact patient health and safety, it requires highly rigorous reasoning, self-reflection and CoT presentation for HPs' reference to enhance explainability
		Therapeutic planning decision support: integrates complete diagnostic information with a patient's medical history, treatment outcomes, demographic information and individualized needs, offering physicians multiple therapeutic planning recommendations and benefit–risk analyses for references		
		Clinical dialogue and report interpretation: provides patient-friendly interpretations of clinical dialogues and medical test reports (synchronized online) to enhance doctor–patient communication and information symmetry	V The inference inputs include detailed medical documents and patient privacy data	III Designed to help laypeople understand specialized clinical decisions
	Clinical documentation generation	Standardized EMR generation and error correction: (1) synchronizes internal systems data (for example, HIS, PACS and LIS) to generate structured EMRs with error correction; and (2) transcribes doctor–patient dialogues in real time, extracting key information into structured records, enabling physicians to focus on patient communication	V Fine-tuning datasets and inference inputs access the hospital's core internal databases, which include confidential medical data, documents, and patient privacy data	IV Used to generate professional medical documents or perform reviews and quality control, which requires robust reasoning capabilities for error correction and explanations to HPs
		Generation, error correction and optimization of other clinical documents: synchronizes internal system data to generate structured case reports, adverse event reports, discharge reports and specific post-treatment patient information, which are subsequently edited by physicians		
	Pharmaceutical service	Prescription review: verifies prescription compliance, drug–drug interactions and medication administration processes while providing medication risk warnings and alerts	III The inference inputs include only medication information	
Hospital operation	Hospital management	Administrative assistant: (1) consolidates hospital regulations, workflows and inspection feedback to build an internal office knowledge base that supports staff consultation (within OA); and (2) generates structured office documents, enabling clinical and administrative staff to focus on core hospital functions instead of tedious paperwork	IV The knowledge base includes non-public administrative information and document resources	II Used for general conversation, information retrieval and non-specialized text generation
		Medical documentation quality control: Performs bulk monitoring of the completion and quality of key clinical documents (for example, EMRs, case reports and adverse event reports) to support hospital quality management	V Fine-tuning training and inference inputs access the hospital's core internal databases, which include confidential medical data, documents and patient privacy data	IV Used mainly for reviewing professional medical documents and processing specialized medical data
	Data management	Medical data processor: utilizes authoritative disease classifications, guidelines and consensus documents to clean, standardize and annotate extensive volumes of unstructured historical data, including abbreviation conversion, to support the construction of standardized medical databases and EMR repositories for large model development and disease surveillance		
Research and education	Scientific research	Research assistant: supports semantic analysis of unstructured clinical documents and EMRs, automates data cleaning and feature extraction, and builds disease-prediction and subtyping models to identify potential clinical research directions and assist with clinical trial design		IV Designed for HPs, requiring robust reasoning capabilities and interpretability to meet advanced scientific research requirements
	Professional agent	Medical chatbots: embedded in internal systems and activated in real time during routine workflows, these chatbots support professional medical Q&A, natural-language retrieval and analysis of clinical documents and guidelines and medical education through simulated patient interviews and case-based training	II–V Sensitivity increases with customization; determined mainly by the presence of confidential internal data in the knowledge base	

Table 1 (continued) | Scenarios and demands for real-world deployment of DeepSeek healthcare solutions in hospitals

Scenario	Task	Applications and functions	Data privacy and security demands (5-point scale)	Reasoning capability and explainability demands (5-point scale)
Personal health management service	Chronic disease monitoring	Health log and disease management: supports patients with chronic disease to upload personal medical records and daily health data for intelligent risk assessment, follow-ups and health recommendations	IV	II–IV
	Health reports analysis	Personal health check-up report interpretation and recommendations: enables itemized interpretation and interactive Q&A of online-synchronized personal health check-up reports, and provides tailored recommendations for follow-up clinical consultations or health management.	The inference inputs include certain types of patient privacy data.	Determined by the system's specific functions available to patients

Our questionnaire survey with IT personnel from 32 healthcare institutions with completed deployments confirmed that all applications listed here are currently deployed and operational. We evaluated the data privacy and security demand level (on a five-point scale) on the basis of the privacy and sensitivity of the data in the fine-tuning dataset and user inputs; and evaluated reasoning capability and explainability demands level (on a five-point scale) on the basis of the intended functions and targeting users. CoT, chain of thought; EMR, electronic medical record; HIS, hospital information system; HPs, healthcare professionals; OA, office automation; Q&A, question and answer.

to mitigate organizational inertia, reduce resistance to technological changes and alleviate trust deficits.

Third, the reasoning and chain-of-thought capabilities of DeepSeek are highly valuable in hospital settings, especially for applications that target healthcare professionals, such as diagnostic and therapeutic decision support, for which the highest standards of reasoning and explainability are required (Table 1). For example, DeepSeek can perform comprehensive self-verification, reflection and extended chain-of-thought reasoning. Although the underlying inference remains a 'black box', this approach enables healthcare professionals to trace DeepSeek's reasoning and review referenced sources. In turn, these features unleash the LLM's ability to integrate extensive datasets with patient-specific information in a comprehensive and transparent manner, which allows users to understand, interrogate and build greater trust in the model's outputs.

Finally, DeepSeek-R1 has a multi-stage training pipeline that can bypass the traditional and costly supervised fine-tuning phase. This approach has been successfully replicated on artificial intelligence (AI) open-science platforms such as Hugging Face, where researchers are currently recreating DeepSeek-R1 (ref. 8). Enhanced model efficiency is crucial, given the substantial energy demands associated with conventional LLMs[9]. Notably, for organizations and countries that lack substantial energy resources, the DeepSeek approach offers cost-effective, high-performance LLM development, democratizing the technology.

Risks associated with LLMs in hospitals

Despite its substantial benefits, increased deployment of LLMs in hospitals introduces specific risks for healthcare systems and patients, especially given the speed of deployment.

First, accuracy and reliability require particular attention. LLMs can produce inaccurate or fabricated information ('hallucinations')[10], and models with advanced reasoning capabilities such as DeepSeek-R1 can embellish these erroneous outputs, making them seem highly plausible. This can jeopardize patient safety through inaccurate diagnoses, inappropriate treatment recommendations, misguided patient care or contamination of medical data. Risks are amplified as hospitals increasingly develop multimodal LLMs that interface with imaging archives and communication systems to assist in diagnoses. In such multimodal settings, physicians might become overly reliant on LLM outputs, while identifying errors in non-linguistic modalities becomes increasingly challenging, which elevates the risk of misdiagnosis.

Second, for pre-clinical assistance or chronic disease monitoring aimed at patients (rather than physicians), seemingly plausible but inaccurate advice could lead to treatment delays or inappropriate self-treatment. Moreover, in early-adopting hospitals, uncertainties remain about the quality of the datasets used for fine-tuning and the competencies of IT departments involved in the integration and alignment of multimodal medical data. Finally, limited expertise in mitigating potential security vulnerabilities can make LLM solutions susceptible to misuse, including 'jailbreaking', which poses ethical challenges, legal compliance issues or internal data-leakage risks.

Regulatory gaps

Due to the unexpectedly rapid deployment of DeepSeek, China currently lacks appropriate regulations specifically tailored to the use of LLMs in hospitals. This regulatory gap has allowed rapid deployment of unregulated DeepSeek applications, which has raised expert concerns about patient safety, clinical efficacy and data security[11].

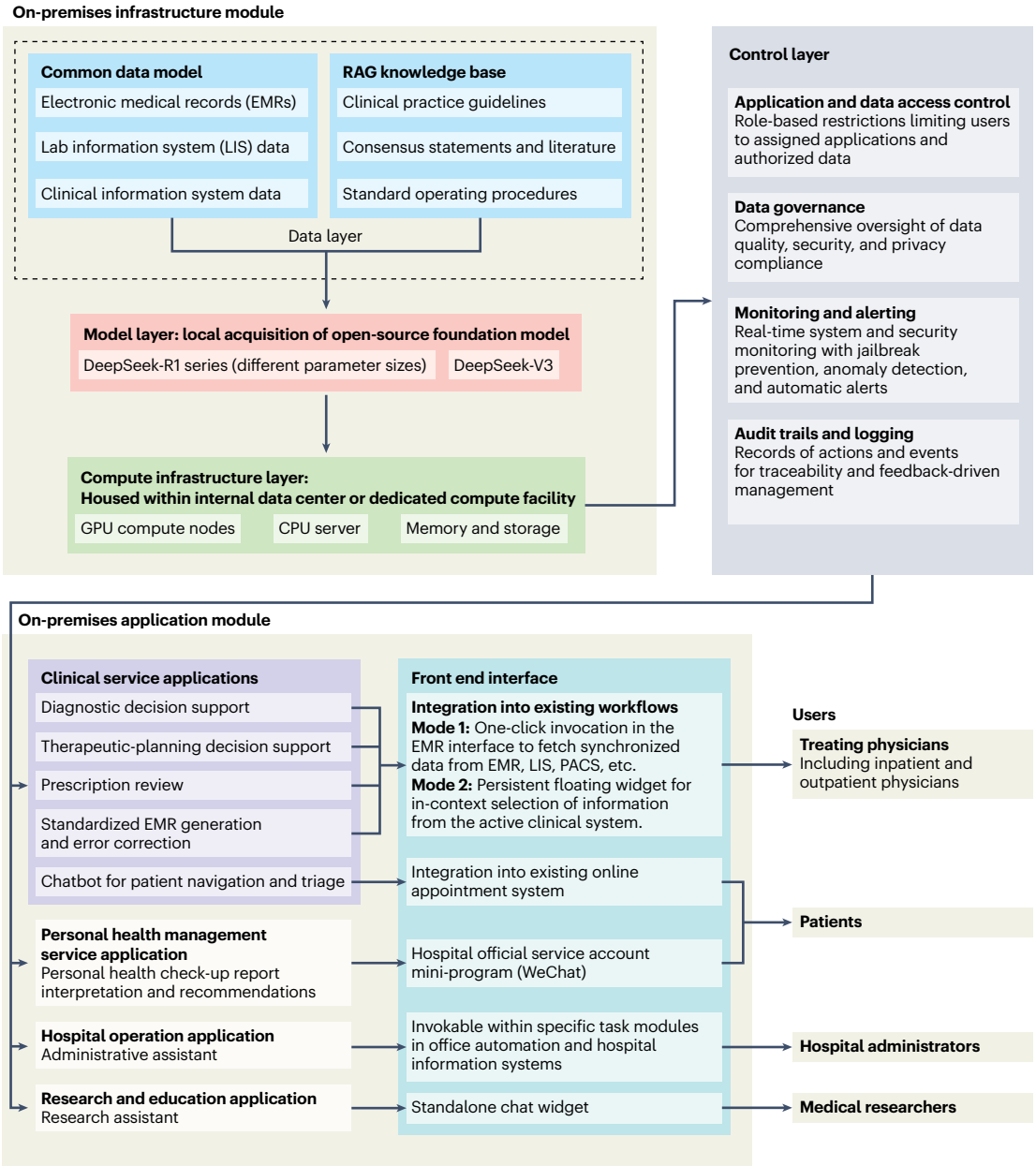

Fig. 1 | A real-world system architecture of DeepSeek OPD at Hospital P in Beijing. CPU, central processing unit; EMR, electronic medical record; GPU, graphics processing unit; LIS, lab information system; PACS, picture archiving and communication system; RAG, retrieval-augmented generation (retrieves relevant external passages and integrates them into a model's context to improve response accuracy).

Regulatory gaps arise mainly in the two following areas: (1) the absence of clear, risk-based classification standards for different LLM applications; and (2) the lack of dedicated regulatory bodies and clear pathways for validating, approving and monitoring these applications.

First, on the basis of the risks outlined above, LLM-based deployments for clinical diagnostic tasks or those that directly serve patients exhibit higher risk levels. Under existing international regulatory frameworks (such as those in the USA and European Union), these high-risk LLM-based deployments would probably require classification as medical devices (Table 2). In contrast, China's current classification standards remain ambiguous, which has led to confusion in hospitals, where higher-risk applications are frequently conflated with medical chatbots intended solely for research and education.

Second, considerable regulatory gaps remain in the oversight of both medical devices and non-device applications. In China, hospital activities are supervised by the National Health Commission (NHC), whereas medical-device development falls under the National Medical Products Administration (NMPA). The dual role of hospitals as both developers and users of LLMs, coupled with insufficient coordination between the NHC and NMPA, leaves LLM solutions inadequately regulated.

Table 2 | Regulatory status of LLM-based healthcare applications for physicians and patients

Task	Applications	Status in USA	Status in EU	Status in China
Clinical decision support	Diagnostic decision support	Non-device (medical device for (1) analysis of medical images, IVD signals or signal patterns; or (2) provision of specific instruction and time-critical cases)	Medical device	Non-device (medical device when used for analysis of medical images or signals)
	Therapeutic-planning decision support	Non-device (medical device when used for the provision of specific instruction and in time-critical cases)	Medical device	Medical device (lack of clear classification standard for medical devices versus non-devices)
Pre-clinical assistance (patient-oriented tasks)	Chatbot for patient navigation and triage	Medical device	Medical device	Medical device (ambiguous)
	Pre-consultation inquiry	Medical device (non-device if the preliminary diagnosis is visible only to HPs)	Non-device (medical device when used for the provision of information for diagnostic or therapeutic decision-making)	Medical device (lack of clear classification standard for medical devices versus non-devices)
Chronic disease monitoring (patient-oriented tasks)	Health log and disease management	Medical device	Medical device	Non-device (medical device when used for the analysis of wearable device data)

EU, European Union; IVD, in vitro diagnostic device.

Moreover, China's current oversight of AI in hospitals is confined to tools that use medical device data for specific purposes, such as assisted diagnosis, detection, triage or quantification, which leverage deep learning mainly for the analysis of medical images and signals. Such applications follow clear NMPA pathways for validation, approval and surveillance. However, LLM-based solutions differ substantially from traditional AI applications and present three distinct regulatory challenges. First, they have a nearly infinite range of training datasets, inputs and outputs, including potential hallucinations, that far exceeds the scope of conventional, purpose-specific evaluation tools. Second, they have high output variability, as reasoning-capable LLMs generate responses dynamically in response to user interactions and thereby complicate performance verification and surveillance[12,13]. Third, they have ongoing adaptability, as hospitals can fine-tune these models in response to evolving medical guidelines and updated hospital guidelines, which renders conventional regulatory pathways inadequate.

There are several large-scale benchmarking tools (such as MedDX-Bench[14], MedBench[15] and CMB) available for the evaluation of medical LLM performance, including hallucination detection and safety control metrics. Although these benchmarks could provide regulatory tools for validating and monitoring LLM-based solutions, few hospitals with OPD of LLMs have publicly reported benchmark evaluations, and Chinese regulatory authorities have yet to formally recognize or systematize these evaluation methods.

Toward a framework for safe and responsible deployment

Given the identified risks and regulatory shortcomings, a forward-looking framework is urgently needed to guide the safe and responsible deployment of DeepSeek and other LLMs in China's healthcare system. We recommend the following three priority actions.

First, establish clear, risk-based standards for categorizing LLM applications. LLMs are deployed across diverse scenarios, from hospital operations to clinical-decision support, which pose varying risks. Health authorities should define classification criteria for LLM applications on the basis of the intended function and associated risk level. Meanwhile, regulatory responsibilities within health authorities, mainly the NHC and NMPA, should be clearly delineated according to these categories, with coordination mechanisms established to ensure coherent oversight.

Second, define functional thresholds that trigger medical-device regulation for high-risk LLM applications. Certain use cases, such as those intended for diagnostic and therapeutic decision support, require rigorous regulatory oversight, owing to their direct impact on patient health. As the NMPA has yet to issue LLM-specific guidance, it should establish explicit criteria that specify when an LLM crosses the threshold into a regulated medical device and is thereby subject to formal approval and regulatory control.

Third, establish a lifecycle-management pathway for LLM applications regulated as medical devices. These applications should be architecturally and operationally partitioned from non-device applications within the system, using a dedicated, independent on-premises infrastructure module and subject to the following: (1) LLM-tailored evaluation tools, including cross-metric benchmark suites that span broad operational ranges, which may contain specific benchmarks (for example, MedBench or CMB), depending on their focus, along with transparency requirements for public disclosure of evaluation results; (2) real-world validation and mandatory post-deployment monitoring, given the high variability of LLMs in human–AI interactions; and (3) a change-control mechanism with re-evaluation and surveillance triggered by updates that materially alter model behavior.

Conclusion

DeepSeek's more efficient training and open-source solutions for cost-effective LLM deployment explain its rapid adoption in hospitals across China. From a technological standpoint, these advances will encourage healthcare-oriented LLMs, which will facilitate widespread innovation. However, from a patient-centered perspective, the associated health risks must be critically addressed. Currently, many

countries, including China, lack sufficient regulatory preparedness for the OPD of LLMs in hospitals, especially around the balance between innovation and patient safety. Consequently, establishing a robust and comprehensive regulatory framework to standardize and promote the safe deployment of LLMs in healthcare settings has become both urgent and indispensable.

Tianyi Shen[1,2], Yuxi Li[3,4], Yanlin Cao[5], Xin Du[1], Xinru Wang[1], Yajuan Zhang[6,7] ✉ & Yi Zhang[1] ✉

[1]Vanke School of Public Health, Tsinghua University, Beijing, China. [2]School of Public Policy and Management, Tsinghua University, Beijing, China. [3]Department of Cardiology, Peking University First Hospital, Beijing, China. [4]Information Center, Peking University First Hospital, Beijing, China. [5]Institute of Medical Information, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China. [6]School of Pharmaceutical Sciences, Tsinghua University, Beijing, China. [7]Key Laboratory of Innovative Drug Research and Evaluation, National Medical Products Administration, Beijing, China.
✉e-mail: zhangyajuan@mail.tsinghua.edu.cn; |yi_zhang@mail.tsinghua.edu.cn

Published online: 30 July 2025

References

1. Gibney, E. *Nature* **638**, 13–14 (2025).
2. DeepSeek-AI et al. Preprint at http://arxiv.org/abs/2501.12948 (2025).
3. Thirunavukarasu, A. J. et al. *Nat. Med.* **29**, 1930–1940 (2023).
4. Dong, B. et al. *Sci. Bull.* **70**, 283–286 (2025).
5. Xiao, H. et al. *Inf. Fusion* **117**, 102888 (2025).
6. DeepSeek-AI et al. Preprint at http://arxiv.org/abs/2412.19437 (2025).
7. Tordjman, M. et al. *Nat. Med.* https://doi.org/10.1038/s41591-025-03726-3 (2025).
8. Conroy, G. & Mallapaty, S. *Nature* **638**, 300–301 (2025).
9. *Nat. Electron.* **8**, 95 (2025).
10. Lee, P., Bubeck, S. & Petro, J. *N. Engl. J. Med.* **388**, 1233–1239 (2023).
11. Zeng, D., Qin, Y., Sheng, B. & Wong, T. Y. *JAMA* **333**, 1866–1869 (2025).
12. Gilbert, S. et al. *Nat. Med.* **29**, 2396–2398 (2023).
13. Freyer, O., Wiest, I. C., Kather, J. N. & Gilbert, S. *Lancet Digit. Health* **6**, e662–e672 (2024).
14. Liu, X. et al. *Nat. Med.* **31**, 932–942 (2025).
15. Liu, M. et al. *Big Data Mining Analytics* **7**, 1116–1128 (2024).

Competing interests

The authors declare no competing interests.