

PERSPECTIVE

## Can generative artificial intelligence enhance evidence-based and personalized medicine?

Arun J. Thirunavukarasu \*

International Centre for Eye Health, London School of Hygiene & Tropical Medicine, London, United Kingdom

\* [ajt205@cantab.ac.uk](mailto:ajt205@cantab.ac.uk)



*Should clinical applications of generative artificial intelligence (GAI) be designed to follow treatment guidelines rigidly, provide individualized recommendations, or be somewhere in-between? A recent study in PLOS Medicine comparing GAI versus physician treatment recommendations highlights the importance of context- and patient-specific circumstances in decision-making.*

### OPEN ACCESS

**Citation:** Thirunavukarasu AJ (2026) Can generative artificial intelligence enhance evidence-based and personalized medicine? *PLoS Med* 23(2): e1004931. <https://doi.org/10.1371/journal.pmed.1004931>

**Published:** February 17, 2026

**Copyright:** © 2026 Arun J. Thirunavukarasu. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Funding:** AJT is supported by the National Institute for Health and Care Research (ACF-2025-20-001). The funder had no role in the conception, preparation, or decision to publish this article.

**Competing interests:** I have read the journal's policy and the authors of this manuscript have the following competing interests: AJT is a member of the Editorial Board of PLOS Medicine.

The abilities of generative artificial intelligence (GAI) are continuing to expand, and clinicians and researchers are exploring and validating a growing number of clinical GAI applications [1]. The function of an application as well as the benchmark to which it is compared are active choices, and careful critical appraisal is essential to understand if a proposed solution is likely to be useful. In a recent *PLOS Medicine* study, Yang and colleagues [2] explored the potential impact of utilizing GAI to make treatment decisions for hepatocellular carcinoma (HCC) patients. Here, the authors assessed patients' long-term survival in a retrospective registry-based study, comparing GAI decisions to clinician decisions as well as guideline recommendations [2].

In this retrospective study of 13,614 untreated HCC patients, three GAI models were tasked with providing treatment recommendations. Survival outcomes were used to estimate the impact of different treatment regimes on the HCC patients with disease stages ranging from Barcelona-Clinic Liver Cancer (BCLC) stage A to BCLC-C [3]. The concordance of GAI and clinician decisions ranged from 26.8% to 32.7%. Concordant decisions were associated with improved survival in early disease (BCLC-A) but worse survival in later-stage disease (BCLC-B and BCLC-C) [2]. Further analyses indicated that discordance in BCLC-A HCC was associated with lower blood albumin and platelet count, as well as higher international normalized ratio (INR); while in BCLC-C HCC, discordance was associated with higher albumin, and lower bilirubin and INR [2].

**Abbreviations:** BCLC, Barcelona-Clinic Liver Cancer; GAI, generative artificial intelligence; HCC, hepatocellular carcinoma; INR, international normalized ratio.

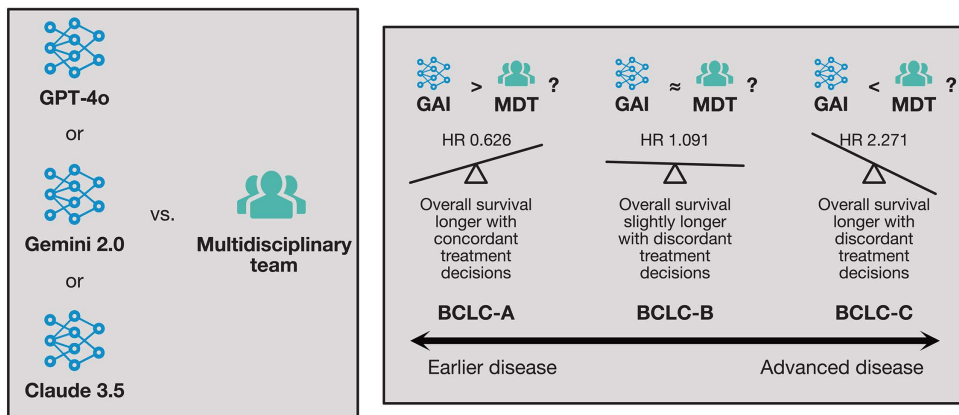
These results could mean that GAI decisions would improve outcomes in BCLC-A HCC patients, and worsen outcomes in BCLC-B and BCLC-C HCC (Fig 1). The discordance and survival results seem to be driven by the relative propensity of GAI and physicians to recommend curative therapy. Physicians appeared to place more emphasis on liver function, avoiding curative therapy in early HCC, where liver function was deemed limited and advocating curative therapy in later HCC when liver function was preserved (rather than adhering rigidly to treatment guidelines). In contrast, GAI was more likely to adhere to treatment guidelines—more intensive or local treatment of early disease and systemic or palliative therapy for late disease, with less influence of liver function. Of note, GAI did not have access to all of the data used by physicians, including raw imaging data which is critical for assessing the feasibility of surgical treatment. It is not clear whether multimodal GAI leveraging more data will behave in a similar fashion to the models tested in this study.

These inferences are tentative due to the inherent limitations of the observational study design. It is difficult to draw confident conclusions without prospective study, and randomization of HCC patients to GAI *versus* clinician-led treatment would be the best way to determine the true effect of GAI on overall survival—though currently ethically infeasible. Nevertheless, this study provides important insight into how GAI might fit into the clinical workflows, particularly as triadic care incorporating AI into the clinician-patient relationship is beginning to be discussed [1]. Guidelines are very helpful for promoting evidence-based care to patients, and can be used as a convenient ‘ground truth’ to assess the accuracy of GAI advice [4]. However, personalized care is also essential, accounting for the holistic clinical situation as well as individual circumstances and values. Do we want GAI to make an effort to mimic this idiosyncratic and nuanced process, or regurgitate guidelines with perfect recall?

There are situations where it is desirable for GAI to provide recommendations that are expected to be actioned, such as in efforts to broaden access to healthcare where access to clinicians is limited [5]. However, where clinicians are responsible for the care they provide, it may be more desirable for GAI to help improve efficiency by providing summaries of clinical information or mapping patients to relevant clinical guidelines. Physicians face a similar conundrum: Guidelines and clinical trials are designed to inform optimal decision-making, but treatments must be tailored to patients’ individual circumstances and preferences. For example, HCC experts cite various stratification tools and informative clinical studies to guide treatment, but also highlight the importance of multidisciplinary discussion and consideration of patient-specific factors [6]. The process of physician education and training should produce independent clinicians with an ability to weigh best available evidence with individual cases to provide the best healthcare possible [7].

A limiting factor for GAI developers is the information available to train and fine-tune models with. In medicine, the literature base skews towards ‘clinical research’ featuring groups sampled with the intention of drawing conclusions that generalize to broader populations. This includes randomized controlled trials, cohort studies, and case-control studies. In large part, this is because causal inference—such as

## Treatment recommendations in 13,614 treatment-naïve hepatocellular carcinoma patients



**Fig 1. Comparing GAI and physician treatment recommendations for HCC patients [2].** Higher overall survival was observed where clinician and generative artificial intelligence (GAI) decisions were concordant in earlier hepatocellular carcinoma (HCC), and discordant in later disease [2]. This might suggest that GAI decisions could improve outcomes in earlier disease, but worsen outcomes in later disease. Ideally, prospective study is necessary to interrogate these conclusions. BCLC, Barcelona Clinic Liver Cancer classification schema; MDT, multidisciplinary team; HR, hazard ratio.

<https://doi.org/10.1371/journal.pmed.1004931.g001>

determining the effect of a treatment—benefits from sufficient sample size and equilibration of factors that can bias results. GAI can draw on this evidence, as well as human synthesis projects such as clinical guidelines and systematic reviews, and may feasibly do better than any human at appraising the exponentially growing literature base [8]. However, the literature base alone does not govern best practice because individual circumstances matter: local resources, expertise, skills, values, and more [7]. For GAI to support individualized decision-making, it may benefit from accessing large corpora of idiosyncratic clinical narratives. This could require a change in priorities in clinicians' publications. While clinical research and literature synthesis currently garner the most attention and citations [9], it may become a priority to disseminate case reports, individual decision-making, and multidisciplinary team reasoning. This might manifest via conventional peer-reviewed routes, partnerships between industrial GAI developers and healthcare systems, or a more open system making information generally available. While the latter option is preferable for maximizing opportunities for all to benefit, dissemination must be balanced with the principles of patient consent and confidentiality.

It may not be likely that GAI, in its current form, becomes capable or trusted to take on the complex role of a physician [10]. Even applications as assistants or sense-checks of clinician decisions entail significant risks such as failure to challenge flawed premises leading to confident but nonsensical recommendations [11]. As GAI improves in handling imaging and other multimodal data, performance will depend heavily upon its designed function [12]. GAI is likely to work best when tasked with roles that play to its strengths: synthesizing large volumes of data, answering factual questions where the required information is available, and completing repetitive tasks at a rate beyond any human [1]. Therefore, choosing to prioritize application of algorithms, collection and summarization of disparate data, or interrogation of individual circumstances and values, seems most appropriate to ensure that applications perform a clear and understood role in care, and that these applications enhance rather than confuse thought processes.

Ultimately, the role GAI plays in healthcare is up to us. Its desired functionality is context-specific, and clinicians are well-placed to describe and decide which applications may improve the care provided to patients. Being explicit about what we want GAI to do is crucial: From this, development, validation, and implementation methods follow.

## References

1. Teo ZL, Thirunavukarasu AJ, Elangovan K, Cheng H, Moova P, Soetikno B, et al. Generative artificial intelligence in medicine. *Nat Med*. 2025;31(10):3270–82. <https://doi.org/10.1038/s41591-025-03983-2> PMID: [41053447](https://pubmed.ncbi.nlm.nih.gov/41053447/)
2. Yang K, Lee J, Jang JW, Sung PS, Han JW. Evaluating the clinical utility of large language models for hepatocellular carcinoma treatment recommendations: a nationwide retrospective registry study. *PLoS Med*. 2026;23(1):e1004855. <https://doi.org/10.1371/journal.pmed.1004855> PMID: [41528959](https://pubmed.ncbi.nlm.nih.gov/41528959/)
3. Calvet X, Bruix J, Ginés P, Bru C, Solé M, Vilana R, et al. Prognostic factors of hepatocellular carcinoma in the west: a multivariate analysis in 206 patients. *Hepatology*. 1990;12(4 Pt 1):753–60. <https://doi.org/10.1002/hep.1840120422> PMID: [2170267](https://pubmed.ncbi.nlm.nih.gov/2170267/)
4. CHART Collaborative. Reporting guidelines for chatbot health advice studies: explanation and elaboration for the Chatbot Assessment Reporting Tool (CHART). *BMJ*. 2025;390:e083305. <https://doi.org/10.1136/bmj-2024-083305> PMID: [40750271](https://pubmed.ncbi.nlm.nih.gov/40750271/)
5. Mateen BA, Menon V, Agweyu A, Korom R, Omoluabi E, McAfee D, et al. Trials for LLM-supported clinical decisions in African primary healthcare. *Nat Med*. 2025;31(9):2833–5. <https://doi.org/10.1038/s41591-025-03815-3> PMID: [40610804](https://pubmed.ncbi.nlm.nih.gov/40610804/)
6. Ducreux M, Abou-Alfa GK, Bekaii-Saab T, Berlin J, Cervantes A, de Baere T, et al. The management of hepatocellular carcinoma. Current expert opinion and recommendations derived from the 24th ESMO/World Congress on Gastrointestinal Cancer, Barcelona, 2022. *ESMO Open*. 2023;8(3):101567. <https://doi.org/10.1016/j.esmoop.2023.101567> PMID: [37263081](https://pubmed.ncbi.nlm.nih.gov/37263081/)
7. Sackett DL, Rosenberg WM, Gray JA, Haynes RB, Richardson WS. Evidence based medicine: what it is and what it isn't. *BMJ*. 1996;312(7023):71–2. <https://doi.org/10.1136/bmj.312.7023.71> PMID: [8555924](https://pubmed.ncbi.nlm.nih.gov/8555924/)
8. Druss BG, Marcus SC. Growth and decentralization of the medical literature: implications for evidence-based medicine. *J Med Libr Assoc*. 2005;93(4):499–501. PMID: [16239948](https://pubmed.ncbi.nlm.nih.gov/16239948/)
9. Patsopoulos NA, Analatos AA, Ioannidis JPA. Relative citation impact of various study designs in the health sciences. *JAMA*. 2005;293(19):2362–6. <https://doi.org/10.1001/jama.293.19.2362> PMID: [15900006](https://pubmed.ncbi.nlm.nih.gov/15900006/)
10. Thirunavukarasu AJ. Large language models will not replace healthcare professionals: curbing popular fears and hype. *J R Soc Med*. 2023;116(5):181–2. <https://doi.org/10.1177/01410768231173123> PMID: [37199678](https://pubmed.ncbi.nlm.nih.gov/37199678/)
11. Chen S, Gao M, Sasse K, Hartvigsen T, Anthony B, Fan L, et al. When helpfulness backfires: LLMs and the risk of false medical information due to sycophantic behavior. *NPJ Digit Med*. 2025;8(1):605. <https://doi.org/10.1038/s41746-025-02008-z> PMID: [41107408](https://pubmed.ncbi.nlm.nih.gov/41107408/)
12. Granata V, Fusco R, Setola SV, Santorsola M, Ottaiano A, Cerrone M, et al. An update of AI and radiomics in precision oncology: insights from liver tumors as case models. *Technol Cancer Res Treat*. 2025;24:15330338251387928. <https://doi.org/10.1177/15330338251387928> PMID: [41334723](https://pubmed.ncbi.nlm.nih.gov/41334723/)