Artificial Intelligence: Implications for Healthcare Practice, Research, and Education

William Hersh, MD, FACMI, FAMIA, FIAHSI, FACP
Professor
Oregon Health & Science University
Portland, OR, USA
https://billhersh.info/presentations

Overview

- Uses, successes, and limitations of AI in medicine
- AI in education
- AI impacts on how we find and apply information

OHSU

Artificial intelligence (AI)

- AI – information systems and algorithms capable of performing tasks associated with human intelligence (Stanford, 2024)
  - Origin of term attributed to summer workshop held at Dartmouth College in 1956 (McCarthy, 2007)
  - A large part of modern success due to machine learning (ML) – "computer programs that learn without being explicitly programmed" (McCarthy, 1990, attributed to Samuel, 1959)
    - Most success with deep learning, based on many-layered neural networks

Three "epochs" of AI (Howell, 2024), leading to current (21st century) classification of

- Predictive AI
- Generative AI

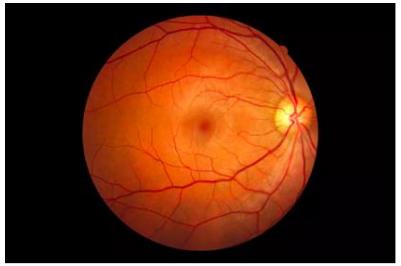
Next epoch may be agentic AI (Powell, 2026)?

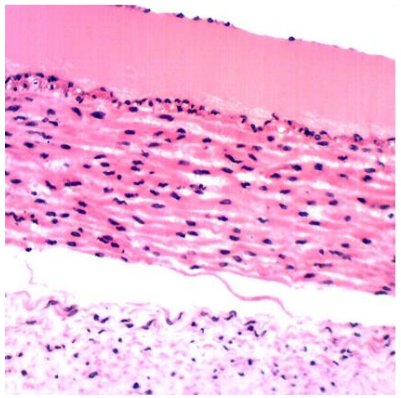
Approximate beginning year	1950s **AI 1.0** Symbolic AI and probabilistic models	2011 **AI 2.0** Deep learning	2018-2022 **AI 3.0** Foundation models
Core functionality and key features	Follows directly encoded rules (if-then rules or decision trees)	Predicts and/or classifies information Task-specific (1 task at a time); requires new data and retraining to perform new tasks	Generates new content (text, sound, images) Performs different types of tasks without new data or retraining; prompt creates new model behaviors
Training method	Rules based on expert knowledge are hand-encoded in traditional programming	Learning patterns based on examples labeled as ground truth	Self-supervised learning from large datasets to predict the next word or sentence in a sequence
Performance capabilities	Follows decision path encoded in its rules. *Eg, ask a series of questions to determine whether a picture is a cat or a dog.*	Classifies information based on training: *"Is this a cat or a dog?"* *"How many dogs will be in the park at noon?"*	Interprets and responds to complex questions: *"Explain the difference between a cat and a dog."*
Examples of performance	IBM's Deep Blue beat the world champion in chess **Health care:** Rule-based clinical decision support tools	Photo searching without manual tagging, voice recognition, language translation **Health care:** diabetic retinopathy detection, breast cancer and lung cancer screening, skin condition classification, predictions based on electronic health records	Writing assistants in word processors, software coding assistants, chatbots **Health care:** Med-PaLM and Med-PaLM-2, medically tuned large language models, PubMedGPT, BioGPT
Examples of challenges and risks	Human logic errors and bias in encoded rules lead to limited capability with real-world situations	Out-of-distribution problems (real-time data differs from training data) Catastrophic forgetting (not remembering early parts of a long sequence of text) Bias related to underlying training data	Hallucinations (plausible but incorrect responses based solely on predictions) Grounding and attribution Bias related to underlying training data and semantics of language in datasets

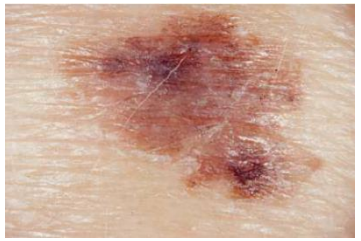
OHSU

Predictive AI in medicine

- "Predictive AI" driven by advances in ML, increasing availability of data, and more powerful computers and networks (Topol, 2019; Rajpurkar, 2022)
  - Deep learning in imaging breakthroughs by Hinton (2006) – led to Nobel Prize for Physics 2024
- Most success in image interpretation (Rajpurkar, 2023); examples include
  - Radiology – chest x-rays for diagnosis of pneumonia and tuberculosis
  - Ophthalmology – retinal images for diagnosis of diabetic retinopathy
  - Dermatology – skin lesions for diagnosis of cancer
  - Pathology – breast cancer slides to predict metastasis







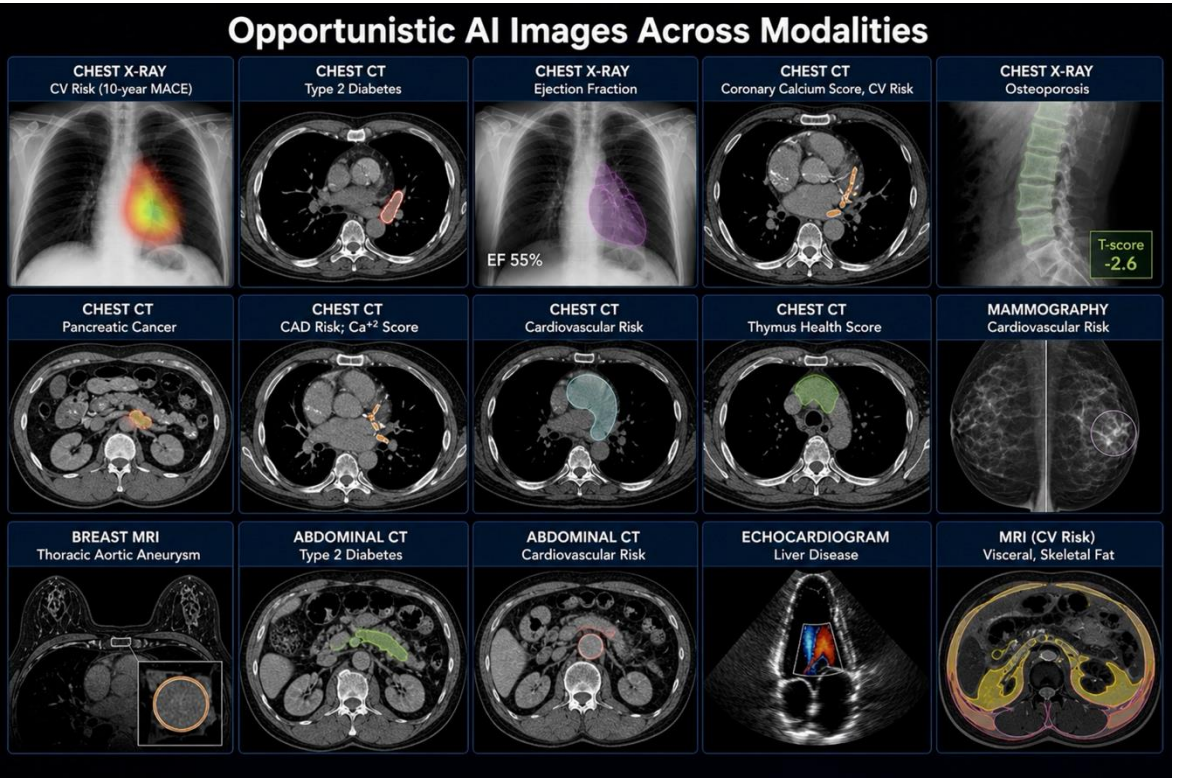
OHSU

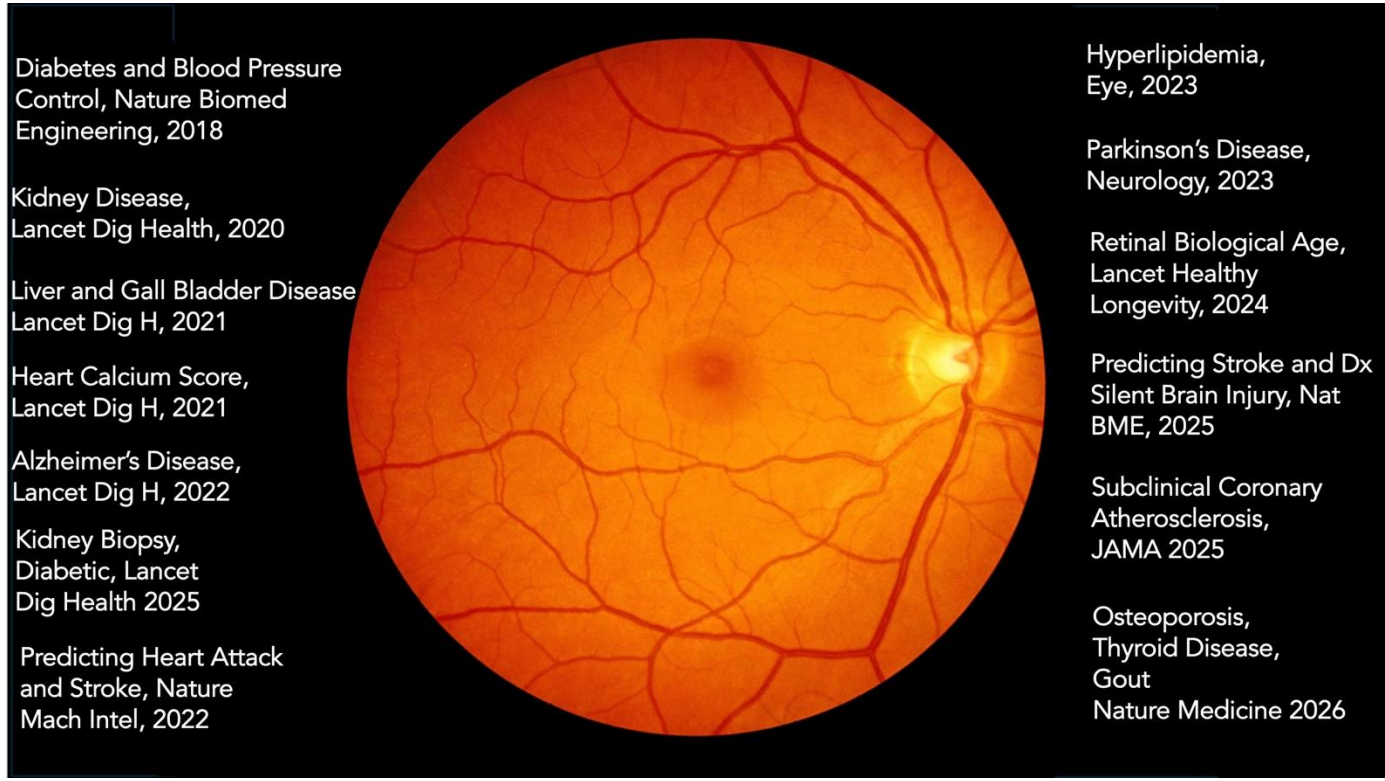
Predictive AI not limited to imaging

- Many promising results, e.g.,
  - Protein folding from amino acid sequences (Jumper, 2021) – led to Nobel Prize for Chemistry 2024
  - Semantic reconstruction of continuous language from fMRI brain recordings (Tang, 2023)
  - Map chemicals to odors perceived by humans (Lee, 2023)
  - Predict Alzheimer's Disease from EHR data up to 7 years before diagnosis (Tang, 2024)
  - AI-supported mammography screening showed consistently favorable outcomes compared with standard double reading (Gommers, 2026)
- Mixed results in practice, e.g.,
  - Computer-aided detection (CAD) in colonoscopy did not improve detection of significant lesions (Hsu, 2026) – may lead to "de-skilling" (Budzyń, 2025)
  - Externally-validated CDS tools from Epic show mixed accuracy (Patel, 2026)

OHSU

Can also “see” where humans cannot (Topol, 2022) – “opportunistic” AI (Topol, 2026)





OHSU

And now, generative AI

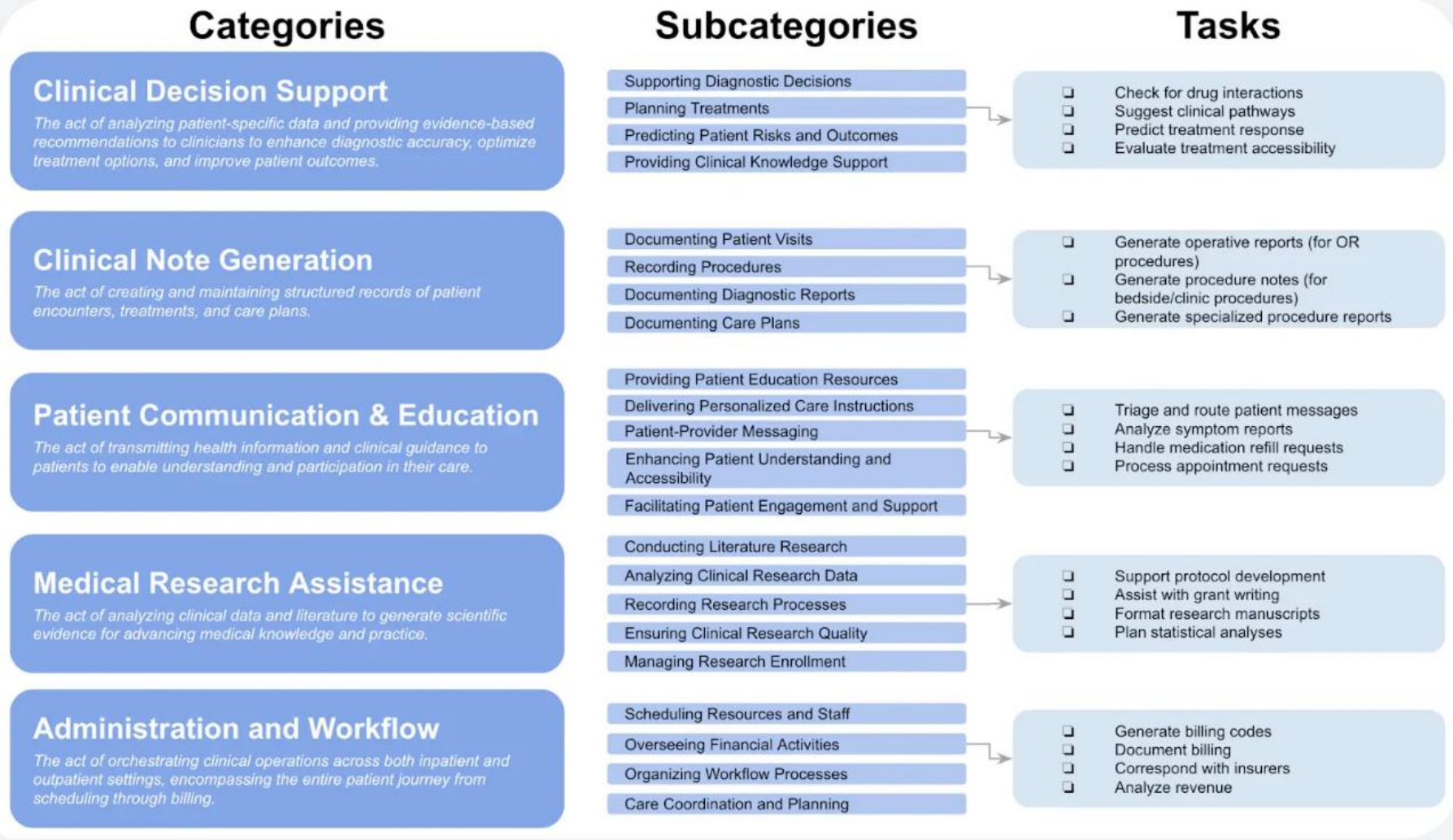
- Introduction of ChatGPT on November 30, 2022 brought new type of AI into focus: generative AI
- Based on large language models (LLMs) processed by deep neural networks using large amounts of training data and tuned for specific tasks (Teo, 2025; Wachter, 2026, Topol, 2026)
  - Trained on massive amounts of text and other content, e.g., large Web crawls, books, Wikipedia, and more for GPT (Roberts, 2022)
  - Use transformer models that predict words in sequence from billions/trillions of words and add measure of importance to “attention” words (Shao, 2024)
  - Fine-tuned with reinforcement learning from human feedback (RLHF) (Lambert, 2022)
  - Activated by (and importance of) prompting (Meskó, 2023; Liu, 2025)
  - Think “differently” from humans (Perc, 2025)

Recent developments in generative AI

- Growing use by
  - Patients
    - 32% of US adults for physical or mental health (Montero, 2026)
    - 20% of adolescents for mental health advice (McBain, 2026)
  - Consumers
    - 49% of US adults (Gottfried, 2026)
    - 55% of US adults over 50 (Brewer, 2025)
  - Clinicians – 81% of US physicians (AMA, 2026)
- From chatbots to agents – agentic AI (Zou, 2025; Powell, 2026)
  - From LLM responses to agents that sense, plan, act, and learn
  - Interaction with systems (e.g., EHRs), tools (e.g., knowledge access), and other agents to solve specific tasks, e.g., medical diagnoses, treatment recommendations, etc.

OHSU

Uses of generative AI in healthcare (Shah, 2025)



OHSU

Successes and caveats from AI research

Successes and caveats

- Passing board exams
- Solving clinical cases
- Summarizing care
- Patient communication
- Conducting research
- Coordinating care

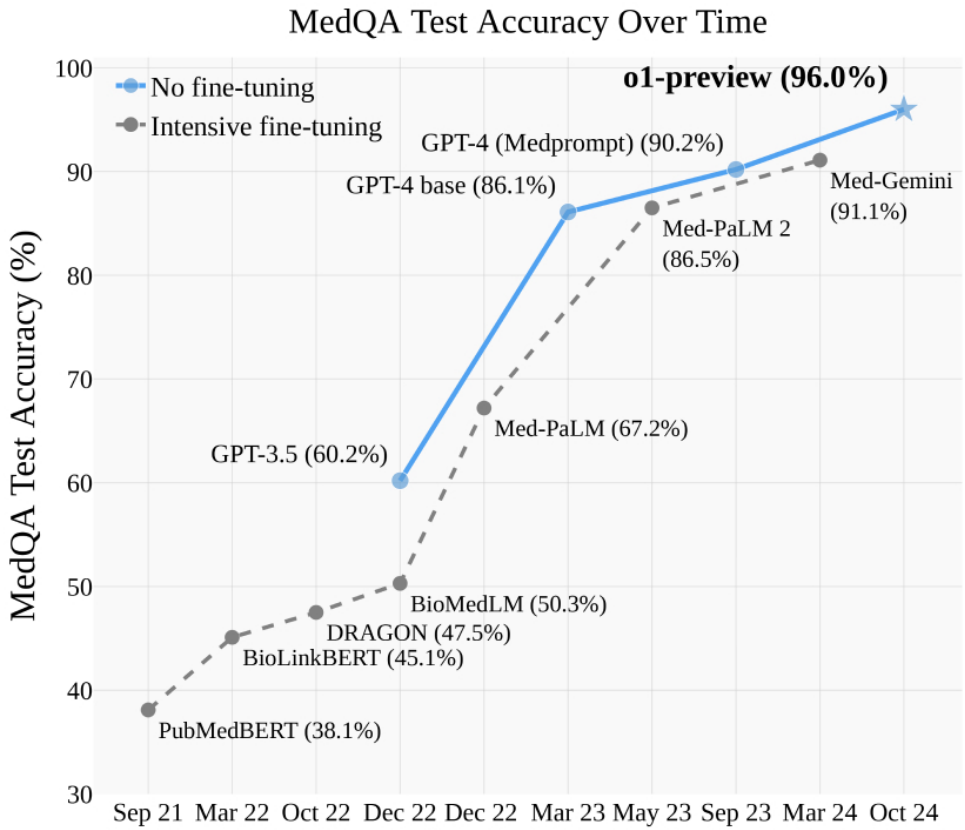
Concerns

- Tsunami of research papers
- Models are “black boxes,” proprietary and constantly changing
- Mostly models using datasets; little on real-world patient care and other tasks
- Predominantly text-based

OHSU

Early success – passing medical board exams

- MedQA dataset based on sample multiple-choice questions (12,273 in English, more in Chinese) from three-part medical exam required for licensing in US (Jin, 2021)
- USMLE "arms race" (Kung, 2023; Nori, 2023; Saab, 2024; Horvitz, 2024; Nori, 2024; OpenEvidence, 2025)
- Systematic review found 16 LLMs on 198 medical exams conducted in 28 countries across 15 languages (Zong, 2024)
  - Including post-residency board exams in Israel (Katz, 2024)
- Concerns for MedQA dataset
  - Time to bench USMLE benchmark? (Raji, 2025)
  - Changing correct answer to "none of the above" resulted in up to halving of scores (Bedi, 2025)



OHSU

Solving clinical cases – many studies, datasets, models, etc.

- Early LLMs comparable to expert humans (Kanjee, 2023; Rao, 2023; Chen, 2023; Cabral, 2024)
- Randomized vignette study of diagnostic reasoning (Goh, 2024)
  - Physicians scored comparably with or without GPT-4 (76.3% vs. 73.7%, NS)
  - GPT-4 alone did better (92.1%, SS)
- Outperformed physicians in emergency medicine at three diagnostic touchpoints (triage in ER, initial evaluation by physician, and admission to hospital or ICU) (Brodeur, 2026)
- Frontier models (GPT, Gemini, Claude) vs. specialized clinical models (Open Evidence, UpToDate AI)
  - Frontier models better in MedQA questions, HealthBench dataset items, and physician-generated queries (Vishwanath, 2026)
  - OE better in 620 real-world point-of-care queries (Real-POCQi) (Feng, 2026)

OHSU

Caveats for solving clinical cases – does AI "outperform" physicians?

- "Cases" entered in research studies typically highly curated and well-written – different from workflow of clinical practice (Pines, 2026)
- Frontier LLMs achieved high accuracy on final diagnoses but performed less well at generating differential diagnoses and navigating uncertainty at other reasoning stages (Rao, 2026)
- Performance declines with human use of LLMs
  - Physicians
    - For written cases, LLMs alone better (Goh, 2024)
    - Even with AI literacy training, still show automation bias (Qazi, 2026)
  - Patients – for 10 scenarios, LLMs alone identified conditions and dispositions better than use by general public (Bean, 2026)

OHSU

Summarizing care

- Widely adopted use is “ambient dictation”
  - Solution for increased clinician time spent in EHR
  - High physician satisfaction but small objective time savings (Shah, 2025; Ma, 2025; Duggan, 2025; Stults, 2025)
- Growing use for summarizing
  - Clinic notes (Van Veen, 2024)
  - Emergency department (You, 2026)
  - Discharge summaries (Grolleau, 2026)
  - In general, high satisfaction but small time savings

OHSU

Patient communication

- Early successes
  - Answering questions in public forums (Sarraju, 2023; Ayers, 2023)
  - Writing letters with comparable or better empathy (Ali, 2023, Ayers, 2023)
- Patient chatbots for health questions reviewed by physicians found 76% accuracy (Mingole, 2026)
  - Experience in health system patient portal showed high usability and satisfaction, especially among patients with greater comorbidities (Tai-Seale, 2026)
- Drafting replies to patient messages (Baxter, 2024; Small, 2024)
  - High clinician satisfaction; no time savings (Tai-Seale, 2024)

OHSU

Conducting research

- Growing ability to conduct aspects of research from hypothesis generation to experiments and data analysis – especially with use of agents (Li, 2026; Lu, 2026)
- Demonstrated success in identifying therapeutic agents for
  - Novel SARS-CoV-2 nanobodies (Swanson, 2025)
  - Dry age-related macular degeneration (dAMD) (Ghareeb, 2026)
- Concerns
  - Increasing volume of lower-quality publications, overwhelming peer-review system (Gartenberg, 2026)
  - Growing concern for “AI slop” in scientific papers (Thorp, 2026)
  - Detection of fabricated citations – increased from 4 per 10K papers through 2024 to 57 per 10K papers in 2026, 1 in 277 papers have 1+ (Topaz, 2026)

Coordinating care

- Growing application of agents for coordination of care (Jiang, 2025; Xu, 2026)
  - ChatEHR to navigate within Epic EHR (McElfresh, 2025)
  - Personal health agents (Heydari, 2025)
  - Deploying in clinical practice (Gallifant, 2025)
- Uptake by EHR vendors, e.g. Epic EHR agents for patients, clinicians, and healthcare operations

OHSU

Some downsides to generative AI

- Dictionary.com 2023 word of year: *hallucinate* (Norlen, 2023)
- Fabrication and errors in bibliographic citations (Walters, 2023; Wu, 2025)
  - Search still matters – authority of information as important as content (Hersh, 2024)
- Perpetuating bias
  - 8 clinical questions asked of 4 LLMs recapitulated “harmful, race-based medicine” (Omiye, 2023)
  - Give variable advice dependent upon race, socioeconomic status, and LGBT status not supported by clinical reasoning (Omar, 2025; Ji, 2025)

OHSU

Downsides to generative AI (cont.)

- Neither algorithms (Tang, 2024) nor humans (Cooke, 2025) can reliably detect
  - More likely to classify non-native English writing as AI-generated (Liang, 2023)
  - Clinicians unable to distinguish human-generated vs. ChatGPT-generated medical manuscripts (Helgeson, 2025)
- Deepfakes
  - Threaten clinician identity, data in record, and evidence (Bair, 2026)
  - Need to protect physician identity, patient safety, and trust (AMA, 2026)

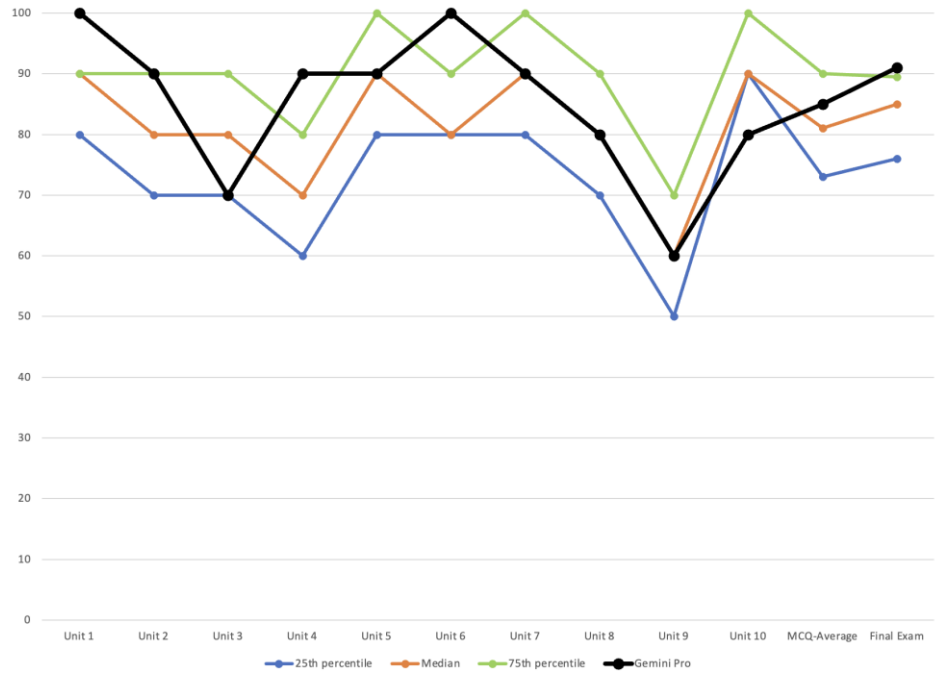
Additional downsides to AI in general

- After clinical models deployed, performance may decline due to actual real-world use (Vaid, 2023; Palmer, 2023)
- Implementing diabetic retinopathy screening in rural Thailand and India found (Widner, 2023)
  - Challenges related to equipment operation, workflows, and image quality
  - Need for training and attention to human factors
- Environmental impact
  - Data center usage of US electricity was 4.4% in 2023, expected to rise to 6.7-12% by 2028 (Shehabi, 2024)
  - Context of energy use important (O'Donnell, 2025)
    - Many other human uses of energy and water are comparable (Masley, 2025)
    - Is AI prompt 10-fold better than Google search (de Vries, 2023)?
  - Data center needs can be met by renewable energy (Clemmer, 2026)

OHSU

Challenges in education

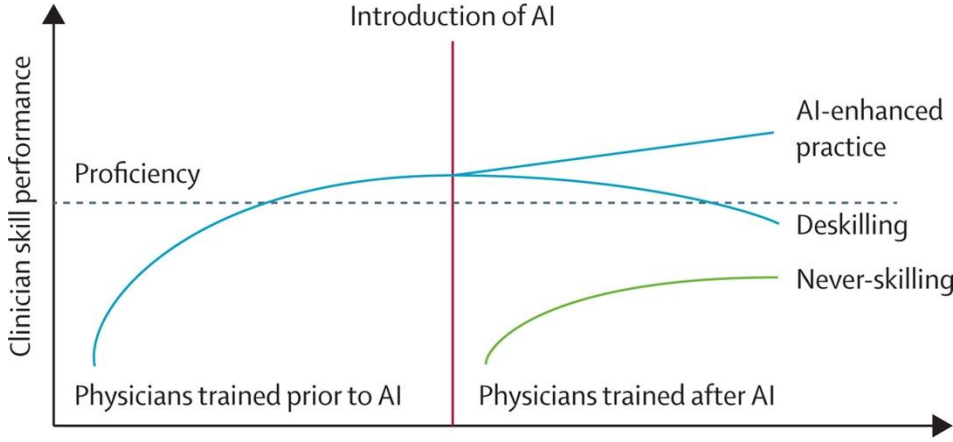
- Early recognition of “homework apocalypse” (Mollick, 2023)
- Studies showing
  - Diminished skill for independent problem-solving (Lee, 2025)
  - Reduced cognitive connection and later recall (Kosmyna, 2025)
  - Use of LLMs resulted in ”shallower knowledge” than learned via Web search, even when links provided in LLM output (Melumad, 2025)
  - Diminished Long-term knowledge retention in undergraduates learning AI (Barcaui, 2025; Liu, 2026)
- Challenges for student assessment in an online introductory informatics course (Hersh, 2024)

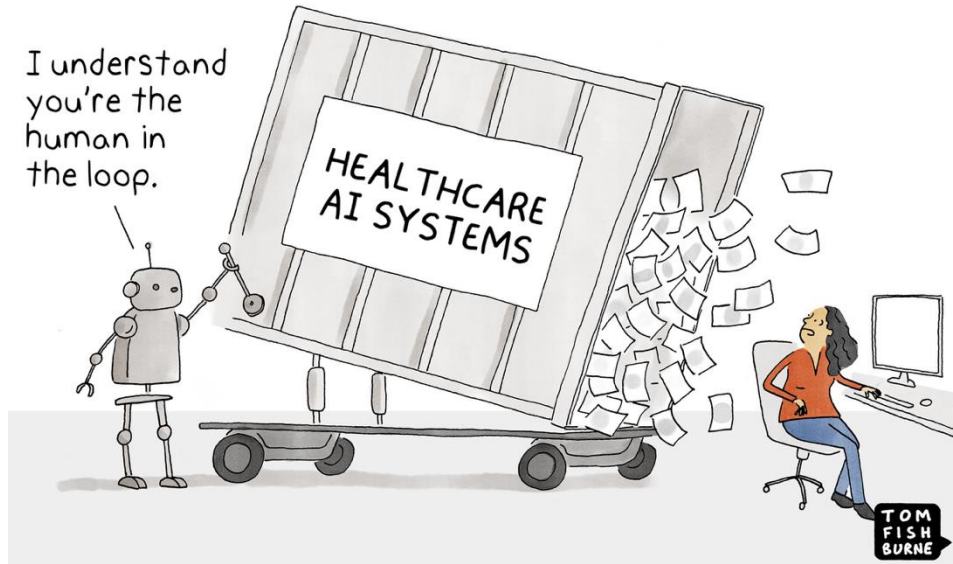


OHSU

Challenges in medical education

- Concerns for de-skilling and never-skilling (Berzin, 2025)
- Scoping review found risk of deskilling and loss of expertise among multiple specialties (Heudel, 2026)
- Focus on “clinician in the loop” may shift responsibility for AI safety from developers to clinicians and not be failsafe for patients (Toro-Tobon, 2026)
- Concern about failure to acquire the cognitive foundations of clinical expertise because AI tools provided the answers before the struggle that produces mastery could occur (Keren, 2026)
- Students should be calibrated and supervised for optimal use in medical education and practice (Ke, 2026)





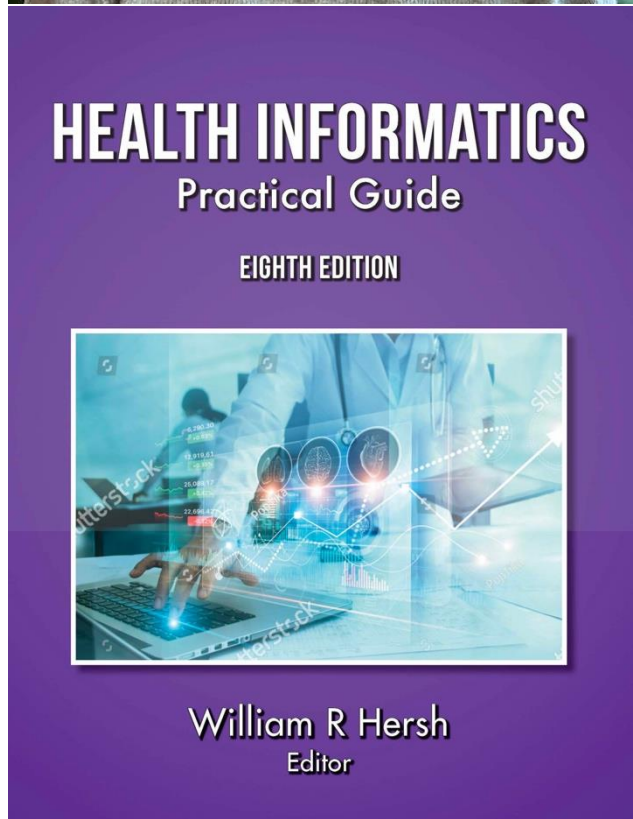
Conclusions

- AI profoundly impacting healthcare, but ultimate role unclear
  - Need real-world evidence base (Goldenberg, 2026)
  - 21st-century clinicians must be critical consumers (Hersh, 2025)
- "AI won't replace radiologists, but radiologists who use AI will replace radiologists who don't," (Langlotz, 2019)
  - (Plug in your health profession)

OHSU

Questions?

William Hersh, MD, FACMI, FAMIA, FIAHSI, FACP
Professor
Oregon Health & Science University
Portland, OR, USA
Email
hersh@ohsu.edu
Web
https://billhersh.info/
Textbook
https://informaticsbook.info/
What is Informatics?
https://informatics.health/



References Cited

AI Demystified: Introduction to AI [WWW Document], 2024. . Stanford University IT. URL https://uit.stanford.edu/service/techtraining/ai-demystified/introduction (accessed 3.16.26).

AI-generated deepfakes: Key policy principles and proposed protections [WWW Document], 2026. . American Medical Association. URL https://www.ama-assn.org/practice-management/digital-health/ai-generated-deepfakes-key-policy-principles-and-proposed (accessed 7.9.26).

Ali, S.R., Dobbs, T.D., Hutchings, H.A., Whitaker, I.S., 2023. Using ChatGPT to write patient clinic letters. Lancet Digit Health 5, e179–e181. https://doi.org/10.1016/S2589-7500(23)00048-1

AMA: AI usage among doctors doubles as confidence in technology grows [WWW Document], 2026. . American Medical Association. URL https://www.ama-assn.org/press-center/ama-press-releases/ama-ai-usage-among-doctors-doubles-confidence-technology-grows (accessed 6.26.26).

Ayers, J.W., Poliak, A., Dredze, M., Leas, E.C., Zhu, Z., Kelley, J.B., Faix, D.J., Goodman, A.M., Longhurst, C.A., Hogarth, M., Smith, D.M., 2023a. Comparing Physician and Artificial Intelligence Chatbot Responses to Patient Questions Posted to a Public Social Media Forum. JAMA Intern Med 183, 589–596. https://doi.org/10.1001/jamainternmed.2023.1838

Ayers, J.W., Zhu, Z., Poliak, A., Leas, E.C., Dredze, M., Hogarth, M., Smith, D.M., 2023b. Evaluating Artificial Intelligence Responses to Public Health Questions. JAMA Netw Open 6, e2317517. https://doi.org/10.1001/jamanetworkopen.2023.17517

Bair, H., 2026. It looks like your doctor and talks like your doctor. But it's not your doctor. STAT. URL https://www.statnews.com/2026/07/14/deepfakes-ai-generated-medicine-misinformation-regulation-video/ (accessed 7.15.26).

Barcaui, A., 2025. ChatGPT as a cognitive crutch: Evidence from a randomized controlled trial on knowledge retention. Social Sciences & Humanities Open 12, 102287. https://doi.org/10.1016/j.ssaho.2025.102287

Baxter, S.L., Longhurst, C.A., Millen, M., Sitapati, A.M., Tai-Seale, M., 2024. Generative artificial intelligence responses to patient messages in the electronic health record: early lessons learned. JAMIA Open 7, ooae028. https://doi.org/10.1093/jamiaopen/ooae028

Bean, A.M., Payne, R.E., Parsons, G., Kirk, H.R., Ciro, J., Mosquera-Gómez, R., Hincapié M, S., Ekanayaka, A.S., Tarassenko, L., Rocher, L., Mahdi, A., 2026. Reliability of LLMs as medical assistants for the general public: a randomized preregistered study. Nat Med. https://doi.org/10.1038/s41591-025-04074-y

Bedi, S., Jiang, Y., Chung, P., Koyejo, S., Shah, N., 2025. Fidelity of Medical Reasoning in Large Language Models. JAMA Netw Open 8, e2526021. https://doi.org/10.1001/jamanetworkopen.2025.26021

Berzin, T.M., Topol, E.J., 2025. Preserving clinical skills in the age of AI assistance. The Lancet 406, 1719. https://doi.org/10.1016/S0140-6736(25)02075-6

Brewer, R., 2025. How Older Adults Use and Think About AI [WWW Document]. U-M Institute for Healthcare Policy & Innovation. URL https://ihpi.umich.edu/national-poll-healthy-aging/national-findings/how-older-adults-use-and-think-about-ai (accessed 1.16.26).

Brodeur, P.G., Buckley, T.A., Kanjee, Z., Goh, E., Ling, E.B., Jain, P., Cabral, S., Abdulnour, R.-E., Haimovich, A.D., Freed, J.A., Olson, A., Morgan, D.J., Hom, J., Gallo, R., McCoy, L.G., Mombini, H., Lucas, C., Fotoohi, M., Gwiazdon, M., Restifo, D., Restrepo, D., Horvitz, E., Chen, J., Manrai, A.K., Rodman, A., 2026. Performance of a large language model on the reasoning tasks of a physician. Science 392, 524–527. https://doi.org/10.1126/science.adz4433

Budzyń, K., Romańczyk, M., Kitala, D., Kołodziej, P., Bugajski, M., Adami, H.O., Blom, J., Buszkiewicz, M., Halvorsen, N., Hassan, C., Romańczyk, T., Holme, Ø., Jarus, K., Fielding, S., Kunar, M., Pellise, M., Pilonis, N., Kamiński, M.F., Kalager, M., Bretthauer, M., Mori, Y., 2025. Endoscopist deskilling risk after exposure to artificial intelligence in colonoscopy: a multicentre, observational study. Lancet Gastroenterol Hepatol S2468-1253(25)00133–5. https://doi.org/10.1016/S2468-1253(25)00133-5

Cabral, S., Restrepo, D., Kanjee, Z., Wilson, P., Crowe, B., Abdulnour, R.-E., Rodman, A., 2024. Clinical Reasoning of a Generative Artificial Intelligence Model Compared With Physicians. JAMA Intern Med. https://doi.org/10.1001/jamainternmed.2024.0295

Chen, A., Chen, D.O., Tian, L., 2023. Benchmarking the symptom-checking capabilities of ChatGPT for a broad range of diseases. J Am Med Inform Assoc ocad245. https://doi.org/10.1093/jamia/ocad245

Clemmer, S., Dotson, S., Gignac, J., Sattler, S., Shaver, L., 2026. Data Center Power Play | Union of Concerned Scientists [WWW Document]. Union of Concerned Scientists. URL https://www.ucs.org/resources/data-center-power-play (accessed 1.21.26).

Cooke, D., Edwards, A., Barkoff, S., Kelly, K., 2025. As Good as a Coin Toss: Human Detection of AI-Generated Content. Commun. ACM 68, 100–109. https://doi.org/10.1145/3729417

de Vries, A., 2023. The growing energy footprint of artificial intelligence. Joule 7, 2191–2194. https://doi.org/10.1016/j.joule.2023.09.004

Duggan, M.J., Gervase, J., Schoenbaum, A., Hanson, W., Howell, J.T., Sheinberg, M., Johnson, K.B., 2025. Clinician Experiences With Ambient Scribe Technology to Assist With Documentation Burden and Efficiency. JAMA Netw Open 8, e2460637. https://doi.org/10.1001/jamanetworkopen.2024.60637

Feng, J., Patel, V., Heagerty, P., Mai, Y., Sivaraman, V., Vossler, P., Ouyang, J., Jena, A.B., 2026. Expert Evaluation of Clinical AI Tools on Real Point-of-Care Clinical Queries. https://doi.org/10.48550/arXiv.2606.28960

Gallifant, J., Kellogg, K., Butler, M., Centi, A., Chen, S., Doyle, P., Dutta, S., Guo, J., Hadfield, M., Kim, E., Kozono, D., Aerts, H.J., Landman, A., Mak, R.H., Mishuris, R., Nelson, T., Savova, G., Sharon, E., Silverman, B., Topaloglu, U., Warner, J.L., Bitterman, D., 2025. A Field Guide to Deploying AI Agents in Clinical Practice. https://doi.org/10.2139/ssrn.5732262

Gartenberg, C., Hasan, S., Murray, A., Pierce, L., 2026. More Versus Better: Artificial Intelligence, Incentives, and the Emerging Crisis in Peer Review. Organization Science.

Ghareeb, A.E., Chang, B., Mitchener, L., Yiu, A., Szostkiewicz, C.J., Shved, D., Gyimesi, G.J., Laurent, J.M., Wright, S.M., Razzak, M.T., White, A.D., Finnemann, S.C., Hinks, M.M., Rodriques, S.G., 2026. A multi-agent system for automating scientific discovery. Nature. https://doi.org/10.1038/s41586-026-10652-y

Goh, E., Gallo, R., Hom, J., Strong, E., Weng, Y., Kerman, H., Cool, J.A., Kanjee, Z., Parsons, A.S., Ahuja, N., Horvitz, E., Yang, D., Milstein, A., Olson, A.P.J., Rodman, A., Chen, J.H., 2024. Large Language Model Influence on Diagnostic Reasoning: A Randomized Clinical

Trial. JAMA Netw Open 7, e2440969. https://doi.org/10.1001/jamanetworkopen.2024.40969
Goldenberg, A., Wiens, J., 2026. Is AI actually improving healthcare? Nat Med 32, 1182–1183. https://doi.org/10.1038/s41591-026-04329-2
Gommers, J., Hernström, V., Josefsson, V., Sartor, H., Schmidt, D., Hjelmgren, A., Larsson, A.-M., Hofvind, S., Andersson, I., Rosso, A., Hagberg, O., Lång, K., 2026. Interval cancer, sensitivity, and specificity comparing AI-supported mammography screening with standard double reading without AI in the MASAI study: a randomised, controlled, non-inferiority, single-blinded, population-based, screening-accuracy trial. Lancet 407, 505–514. https://doi.org/10.1016/S0140-6736(25)02464-X
Gottfried, J., Bishop, W., Anderson, M., Faverio, M., Park, E., McClain, C., 2026. Americans and AI 2026: Chatbots, Smart Devices and Views on Impact. Pew Research Center. URL https://www.pewresearch.org/internet/2026/06/17/americans-and-ai-2026-chatbots-smart-devices-and-views-on-impact/ (accessed 6.26.26).
Grolleau, F., Liang, A.S., Keyes, T., Ma, S.P., Lew, T., Huynh, T.R., Steele, N., Chung, P., Qin, P., Chandra, G., Wang, S.F., Mullen, E., Carpenter, L., Hoppenfeld, M., Morrin, M., Kyerematen, B.A., Ambers, N., Kotecha, N., Alsentzer, E., Hom, J., Shah, N.H., Schulman, K., Chen, J.H., 2026. Physician-Reported Safety Outcomes of AI-Generated Hospital Course Summaries. JAMA Netw Open 9, e2616556. https://doi.org/10.1001/jamanetworkopen.2026.16556
Helgeson, S.A., Johnson, P.W., Gopikrishnan, N., Koirala, T., Moreno-Franco, P., Carter, R.E., Quicksall, Z.S., Burger, C.D., 2025. Human Reviewers' Ability to Differentiate Human-Authored or Artificial Intelligence-Generated Medical Manuscripts: A Randomized Survey Study. Mayo Clin Proc 100, 622–633. https://doi.org/10.1016/j.mayocp.2024.08.029
Hersh, W., 2025. Generative Artificial Intelligence: Implications for Biomedical and Health Professions Education. Annu Rev Biomed Data Sci. https://doi.org/10.1146/annurev-biodatasci-103123-094756
Hersh, W., 2024. Search still matters: information retrieval in the era of generative AI. J Am Med Inform Assoc 31, 2159–2161. https://doi.org/10.1093/jamia/ocae014
Hersh, W., Fultz Hollis, K., 2024. Results and implications for generative AI in a large introductory biomedical and health informatics course. NPJ Digit Med 7, 247. https://doi.org/10.1038/s41746-024-01251-0
Heudel, P.E., Crochet, H., Filori, Q., Bachelot, T., Blay, J.Y., 2026. Artificial intelligence in medicine: a scoping review of the risk of deskilling and loss of expertise among physicians. ESMO Real World Data Digit Oncol 12, 100693. https://doi.org/10.1016/j.esmorw.2026.100693
Heydari, A.A., Gu, K., Srinivas, V., Yu, H., Zhang, Z., Zhang, Y., Paruchuri, A., He, Q., Palangi, H., Hammerquist, N., Metwally, A.A., Winslow, B., Kim, Y., Ayush, K., Yang, Y., Narayanswamy, G., Xu, M.A., Garrison, J., Lee, A.A., Vafeiadou, J., Graef, B., Galatzer-Levy, I.R., Schenck, E., Barakat, A., Perez, J., Shreibati, J., Hernandez, J., Faranesh, A.Z., Prieto, J.L., Heneghan, C., Liu, Y., Zhan, J., Malhotra, M., Patel, S., Althoff, T., Liu, X., McDuff, D., Xu, X. "Orson," 2025. The Anatomy of a Personal Health Agent. https://doi.org/10.48550/arXiv.2508.20148
Hinton, G.E., Osindero, S., Teh, Y.-W., 2006. A fast learning algorithm for deep belief nets. Neural Comput 18, 1527–1554. https://doi.org/10.1162/neco.2006.18.7.1527

Horvitz, E., Nori, H., Usuyama, N., 2024. Run-time strategies in foundation models: from Medprompt to OpenAI o1-preview. Microsoft Research. URL https://www.microsoft.com/en-us/research/blog/advances-in-run-time-strategies-for-next-generation-foundation-models/ (accessed 12.2.24).

Howell, M.D., Corrado, G.S., DeSalvo, K.B., 2024. Three Epochs of Artificial Intelligence in Health Care. JAMA 331, 242–244. https://doi.org/10.1001/jama.2023.25057

Hsu, W.-F., Kuo, C.-Y., Yen, H.-H., Lin, Y.-M., Chen, Y.-N., Sun, C.-K., Chang, W.-Y., Lin, H.-H., Wu, H.-Y., Chang, L.-C., Chang, C.-Y., Wu, M.-S., Chiu, H.-M., 2026. Computer-Assisted Colonoscopy in High-Adenoma Detection Rate Settings in a High-Risk Population: A Randomized Clinical Trial. JAMA Netw Open 9, e264881. https://doi.org/10.1001/jamanetworkopen.2026.4881

Ji, Y., Ma, W., Sivarajkumar, S., Zhang, H., Sadhu, E.M., Li, Z., Wu, X., Visweswaran, S., Wang, Y., 2025. Mitigating the risk of health inequity exacerbated by large language models. NPJ Digit Med 8, 246. https://doi.org/10.1038/s41746-025-01576-4

Jiang, Y., Black, K.C., Geng, G., Park, D., Zou, J., Ng, A.Y., Chen, J.H., 2025. MedAgentBench: A Virtual EHR Environment to Benchmark Medical LLM Agents. NEJM AI 2, AIdbp2500144. https://doi.org/10.1056/AIdbp2500144

Jin, D., Pan, E., Oufattole, N., Weng, W.-H., Fang, H., Szolovits, P., 2021. What Disease Does This Patient Have? A Large-Scale Open Domain Question Answering Dataset from Medical Exams. Applied Sciences 11, 6421. https://doi.org/10.3390/app11146421

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohl, S.A.A., Ballard, A.J., Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R., Adler, J., Back, T., Petersen, S., Reiman, D., Clancy, E., Zielinski, M., Steinegger, M., Pacholska, M., Berghammer, T., Bodenstein, S., Silver, D., Vinyals, O., Senior, A.W., Kavukcuoglu, K., Kohli, P., Hassabis, D., 2021. Highly accurate protein structure prediction with AlphaFold. Nature 596, 583–589. https://doi.org/10.1038/s41586-021-03819-2

Kanjee, Z., Crowe, B., Rodman, A., 2023. Accuracy of a Generative Artificial Intelligence Model in a Complex Diagnostic Challenge. JAMA 330, 78–80. https://doi.org/10.1001/jama.2023.8288

Katz, U., Cohen, E., Shachar, E., Somer, J., Fink, A., Morse, E., Shreiber, B., Wolf, I., 2024. GPT versus Resident Physicians — A Benchmark Based on Official Board Scores. NEJM AI 0, AIdbp2300192. https://doi.org/10.1056/AIdbp2300192

Ke, Y., Jin, L., Ong, J.C.L., Thirunavukarasu, A.J., Car, J., Cheung, C.Y., Tham, Y.C., Ting, D.S.W., Ong, M.E.H., Compton, S., Narayan, A., Keane, P.A., Wong, T.Y., Bates, D.W., Tan, P., Liu, N., 2026. AI-induced never-skilling in medical education. Nat Med. https://doi.org/10.1038/s41591-026-04438-y

Keren, R., Desai, B.R., West, D.C., 2026. Promoting Clinical Expertise in the Age of AI: No Struggle, No Mastery. JAMA. https://doi.org/10.1001/jama.2026.6097

Kosmyna, N., Hauptmann, E., Yuan, Y.T., Situ, J., Liao, X.-H., Beresnitzky, A.V., Braunstein, I., Maes, P., 2025. Your Brain on ChatGPT: Accumulation of Cognitive Debt when Using an AI Assistant for Essay Writing Task. https://doi.org/10.48550/arXiv.2506.08872

Kung, T.H., Cheatham, M., Medenilla, A., Sillos, C., De Leon, L., Elepaño, C., Madriaga, M., Aggabao, R., Diaz-Candido, G., Maningo, J., Tseng, V., 2023. Performance of ChatGPT on USMLE: Potential for AI-assisted medical education using large language models. PLOS Digit Health 2, e0000198. https://doi.org/10.1371/journal.pdig.0000198

Lambert, N., Castricato, L., von Werra, L., Havrilla, A., 2022. Illustrating Reinforcement Learning from Human Feedback (RLHF) [WWW Document]. Hugging Face. URL https://huggingface.co/blog/rlhf (accessed 12.10.23).
Langlotz, C.P., 2019. Will Artificial Intelligence Replace Radiologists? Radiol Artif Intell 1, e190058. https://doi.org/10.1148/ryai.2019190058
Lee, B.K., Mayhew, E.J., Sanchez-Lengeling, B., Wei, J.N., Qian, W.W., Little, K.A., Andres, M., Nguyen, B.B., Moloy, T., Yasonik, J., Parker, J.K., Gerkin, R.C., Mainland, J.D., Wiltschko, A.B., 2023. A principal odor map unifies diverse tasks in olfactory perception. Science 381, 999–1006. https://doi.org/10.1126/science.ade4401
Lee, H.-P. (Hank), Sarkar, A., Tankelevitch, L., Drosos, I., Rintel, S., Banks, R., Wilson, N., 2025. The Impact of Generative AI on Critical Thinking: Self-Reported Reductions in Cognitive Effort and Confidence Effects From a Survey of Knowledge Workers, in: Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems, CHI '25. Association for Computing Machinery, New York, NY, USA, pp. 1–22. https://doi.org/10.1145/3706598.3713778
Li, B., Saini, A.K., Hernandez, J.G., Moore, J.H., 2026. Agentic AI and the rise of in silico team science in biomedical research. Nat Biotechnol 44, 711–725. https://doi.org/10.1038/s41587-026-03035-1
Liang, W., Yuksekgonul, M., Mao, Y., Wu, E., Zou, J., 2023. GPT detectors are biased against non-native English writers. Patterns (N Y) 4, 100779. https://doi.org/10.1016/j.patter.2023.100779
Liu, G., Christian, B., Dumbalska, T., Bakker, M.A., Dubey, R., 2026. AI Assistance Reduces Persistence and Hurts Independent Performance. https://doi.org/10.48550/arXiv.2604.04721
Liu, J., Liu, F., Wang, C., Liu, S., 2025. Prompt Engineering in Clinical Practice: Tutorial for Clinicians. J Med Internet Res 27, e72644. https://doi.org/10.2196/72644
Lu, Chris, Lu, Cong, Lange, R.T., Yamada, Y., Hu, S., Foerster, J., Ha, D., Clune, J., 2026. Towards end-to-end automation of AI research. Nature 651, 914–919. https://doi.org/10.1038/s41586-026-10265-5
Ma, S.P., Liang, A.S., Shah, S.J., Smith, M., Jeong, Y., Devon-Sand, A., Crowell, T., Delahaie, C., Hsia, C., Lin, S., Shanafelt, T., Pfeffer, M.A., Sharp, C., Garcia, P., 2025. Ambient artificial intelligence scribes: utilization and impact on documentation time. J Am Med Inform Assoc 32, 381–385. https://doi.org/10.1093/jamia/ocae304
Masley, A., 2025. Using ChatGPT is not bad for the environment - a cheat sheet [WWW Document]. Andy Masley. URL https://blog.andymasley.com/p/a-cheat-sheet-for-conversations-about (accessed 7.12.26).
McBain, R.K., Cantor, J.H., Breslau, J., Diliberti, M., Zhang, L.A., Zhang, F., Burnett, A., Kofner, A., Rader, B., Pataranutaporn, P., Stein, B.D., Mehrotra, A., Yu, H., 2026. AI Chatbot Use and Disclosure for Mental Health Among US Adolescents and Young Adults. JAMA Pediatr e262015. https://doi.org/10.1001/jamapediatrics.2026.2015
McCarthy, J., 2007. What is Artificial Intelligence? Stanford University.
McCarthy, J., Feigenbaum, E.A., 1990. In Memoriam: Arthur Samuel: Pioneer in Machine Learning. AIMag 11, 10–10. https://doi.org/10.1609/aimag.v11i3.840
McElfresh, D., Sharma, A., Masariya, P., Dennis, C., Jones, E., Kumar, V., Wang, X., Jasti, K., Mouniswamy, S., Revri, A., Kotecha, N., 2025. How To Build a Safe, Secure Medical AI Platform [WWW Document]. Stanford Institute for Human-Centered AI (HAI). URL

https://hai.stanford.edu/news/how-to-build-a-safe-secure-medical-ai-platform (accessed 1.19.26).
Melumad, S., Yun, J.H., 2025. Experimental evidence of the effects of large language models versus web search on depth of learning. PNAS Nexus 4, pgaf316. https://doi.org/10.1093/pnasnexus/pgaf316
Meskó, B., 2023. Prompt Engineering as an Important Emerging Skill for Medical Professionals: Tutorial. J Med Internet Res 25, e50638. https://doi.org/10.2196/50638
Mingole, B., Majumdar, A., Choudhury, F.A., Kraschnewski, J.L., Sundar, S.S., Yadav, A., 2025. Dr. GPT Will See You Now, but Should It? Exploring the Benefits and Harms of Large Language Models in Medical Diagnosis using Crowdsourced Clinical Cases. https://doi.org/10.48550/arXiv.2506.13805
Mollick, E., 2023. The Homework Apocalypse [WWW Document]. One Useful Thing. URL https://www.oneusefulthing.org/p/the-homework-apocalypse (accessed 3.20.24).
Montero, A., Montalvo, J., Kearney, Valdes, I., Hamel, L., 2026. KFF Tracking Poll on Health Information and Trust: Use of AI For Health Information and Advice. KFF. URL https://www.kff.org/public-opinion/kff-tracking-poll-on-health-information-and-trust-use-of-ai-for-health-information-and-advice/ (accessed 6.26.26).
Nori, H., King, N., McKinney, S.M., Carignan, D., Horvitz, E., 2023. Capabilities of GPT-4 on Medical Challenge Problems. https://doi.org/10.48550/arXiv.2303.13375
Nori, H., Usuyama, N., King, N., McKinney, S.M., Fernandes, X., Zhang, S., Horvitz, E., 2024. From Medprompt to o1: Exploration of Run-Time Strategies for Medical Challenge Problems and Beyond. https://doi.org/10.48550/arXiv.2411.03590
Norlen, N., Barrett, G., 2023. Word of the Year 2023 [WWW Document]. Dictionary.com. URL https://content.dictionary.com/word-of-the-year-2023/ (accessed 10.9.24).
O'Donnell, J. archive, Crownhart, C., 2025. We did the math on AI's energy footprint. Here's the story you haven't heard. [WWW Document]. MIT Technology Review. URL https://www.technologyreview.com/2025/05/20/1116327/ai-energy-usage-climate-footprint-big-tech/ (accessed 1.17.26).
Omar, M., Soffer, S., Agbareia, R., Bragazzi, N.L., Apakama, D.U., Horowitz, C.R., Charney, A.W., Freeman, R., Kummer, B., Glicksberg, B.S., Nadkarni, G.N., Klang, E., 2025. Sociodemographic biases in medical decision making by large language models. Nat Med 31, 1873–1881. https://doi.org/10.1038/s41591-025-03626-6
Omiye, J.A., Lester, J.C., Spichak, S., Rotemberg, V., Daneshjou, R., 2023. Large language models propagate race-based medicine. npj Digit. Med. 6, 1–4. https://doi.org/10.1038/s41746-023-00939-z
OpenEvidence Creates the First AI in History to Score a Perfect 100% on the United States Medical Licensing Examination (USMLE) [WWW Document], 2025. . OpenEvidence. URL https://www.openevidence.com/announcements/openevidence-creates-the-first-ai-in-history-to-score-a-perfect-100percent-on-the-united-states-medical-licensing-examination-usmle (accessed 11.29.25).
Palmer, K., 2023. The 'model-eat-model world' of clinical AI: How predictive power becomes a pitfall. STAT. URL https://www.statnews.com/2023/10/10/the-model-eat-model-world-of-clinical-ai-how-predictive-power-becomes-a-pitfall/ (accessed 11.28.23).
Patel, H., Crusco, S., Hansen, D., Morales, J., Patale, T., Chiuzan, C., Sands, M., Naidich, J., Myers, K., Davidson, K.W., Hirsch, J.S., Zanos, T.P., 2026. A Systematic Review and

Meta-analysis of Externally Validated Epic Clinical Decision Support Tools. J Gen Intern Med. https://doi.org/10.1007/s11606-026-10381-y
Perc, M., 2025. Counterfeit judgments in large language models. Proc Natl Acad Sci U S A 122, e2528527122. https://doi.org/10.1073/pnas.2528527122
Pines, J., 2026. Did AI Really Beat ER Doctors At Diagnosis? Here's What The Study Showed [WWW Document]. Forbes. URL https://www.forbes.com/sites/jessepines/2026/05/22/did-ai-really-beat-er-doctors-at-diagnosis-heres-what-the-study-showed/ (accessed 6.29.26).
Powell, K., Fishman, E.K., Chu, L.C., Rowe, S.P., Crawford, C.K., 2026. Agentic Artificial Intelligence: The Power to Change Medicine and Our World. J Am Coll Radiol 23, 306–308. https://doi.org/10.1016/j.jacr.2025.06.032
Qazi, I.A., Ali, A., Khawaja, A.U., Akhtar, M.J., Sheikh, A.Z., Alizai, M.H., 2026. Automation Bias in Large Language Model–Assisted Diagnostic Reasoning among Physicians Trained in AI Literacy — A Randomized Clinical Trial. NEJM AI.
Raji, I.D., Daneshjou, R., Alsentzer, E., 2025. It's Time to Bench the Medical Exam Benchmark. NEJM AI 2, AIe2401235. https://doi.org/10.1056/AIe2401235
Rajpurkar, P., Chen, E., Banerjee, O., Topol, E.J., 2022. AI in health and medicine. Nat Med 1–8. https://doi.org/10.1038/s41591-021-01614-0
Rajpurkar, P., Lungren, M.P., 2023. The Current and Future State of AI Interpretation of Medical Images. N Engl J Med 388, 1981–1990. https://doi.org/10.1056/NEJMra2301725
Rao, A., Pang, M., Kim, J., Kamineni, M., Lie, W., Prasad, A.K., Landman, A., Dreyer, K., Succi, M.D., 2023. Assessing the Utility of ChatGPT Throughout the Entire Clinical Workflow: Development and Usability Study. J Med Internet Res 25, e48659. https://doi.org/10.2196/48659
Rao, A.S., Esmail, K.P., Lee, R.S., Jiang, S., Arraiza Carlo, B., Gill, J., Khanna, P., Kalmowitz, E., Montagnese, B., Heydari, K., Jiao, Q., Bott, E., Nguyen, D., Wang, G., Hood, M., Landman, A.B., Succi, M.D., 2026. Large Language Model Performance and Clinical Reasoning Tasks. JAMA Netw Open 9, e264003. https://doi.org/10.1001/jamanetworkopen.2026.4003
Roberts, G., 2022. AI Training Datasets: the Books1+Books2 that Big AI eats for breakfast - Musings of Freedom. Musings of Freedom. URL https://gregoreite.com/drilling-down-details-on-the-ai-training-datasets/ (accessed 9.6.23).
Saab, K., Tu, T., Weng, W.-H., Tanno, R., Stutz, D., Wulczyn, E., Zhang, F., Strother, T., Park, C., Vedadi, E., Chaves, J.Z., Hu, S.-Y., Schaekermann, M., Kamath, A., Cheng, Y., Barrett, D.G.T., Cheung, C., Mustafa, B., Palepu, A., McDuff, D., Hou, L., Golany, T., Liu, L., Alayrac, J., Houlsby, N., Tomasev, N., Freyberg, J., Lau, C., Kemp, J., Lai, J., Azizi, S., Kanada, K., Man, S., Kulkarni, K., Sun, R., Shakeri, S., He, L., Caine, B., Webson, A., Latysheva, N., Johnson, M., Mansfield, P., Lu, J., Rivlin, E., Anderson, J., Green, B., Wong, R., Krause, J., Shlens, J., Dominowska, E., Eslami, S.M.A., Chou, K., Cui, C., Vinyals, O., Kavukcuoglu, K., Manyika, J., Dean, J., Hassabis, D., Matias, Y., Webster, D., Barral, J., Corrado, G., Semturs, C., Mahdavi, S.S., Gottweis, J., Karthikesalingam, A., Natarajan, V., 2024. Capabilities of Gemini Models in Medicine. https://doi.org/10.48550/arXiv.2404.18416
Sarraju, A., Bruemmer, D., Van Iterson, E., Cho, L., Rodriguez, F., Laffin, L., 2023. Appropriateness of Cardiovascular Disease Prevention Recommendations Obtained From a Popular Online Chat-Based Artificial Intelligence Model. JAMA. https://doi.org/10.1001/jama.2023.1044

Shah, N., Pfeffer, M., Liang, P., 2025. Holistic Evaluation of Large Language Models for Medical Applications | Stanford HAI [WWW Document]. Stanford Institute for Human-Centered AI. URL https://hai.stanford.edu/news/holistic-evaluation-of-large-language-models-for-medical-applications (accessed 3.14.25).

Shah, S.J., Devon-Sand, A., Ma, S.P., Jeong, Y., Crowell, T., Smith, M., Liang, A.S., Delahaie, C., Hsia, C., Shanafelt, T., Pfeffer, M.A., Sharp, C., Lin, S., Garcia, P., 2025. Ambient artificial intelligence scribes: physician burnout and perspectives on usability and documentation burden. J Am Med Inform Assoc 32, 375–380. https://doi.org/10.1093/jamia/ocae295

Shao, M., Basit, A., Karri, R., Shafique, M., 2024. Survey of Different Large Language Model Architectures: Trends, Benchmarks, and Challenges. IEEE Access 12, 188664–188706. https://doi.org/10.1109/ACCESS.2024.3482107

Shehabi, A., Smith, S.J., Hubbard, A., Newkirk, A., Lei, N., Siddik, M.A., Holecek, B., Koomey, J.G., Masanet, E.R., Sartor, D.A., 2024. 2024 United States Data Center Energy Usage Report | Energy Technologies Area [WWW Document]. Lawrence Berkeley National Laboratory. URL https://eta.lbl.gov/publications/2024-lbnl-data-center-energy-usage-report (accessed 3.20.25).

Small, W.R., Wiesenfeld, B., Brandfield-Harvey, B., Jonassen, Z., Mandal, S., Stevens, E.R., Major, V.J., Lostraglio, E., Szerencsy, A., Jones, S., Aphinyanaphongs, Y., Johnson, S.B., Nov, O., Mann, D., 2024. Large Language Model-Based Responses to Patients' In-Basket Messages. JAMA Netw Open 7, e2422399. https://doi.org/10.1001/jamanetworkopen.2024.22399

Stults, C.D., Deng, S., Martinez, M.C., Wilcox, J., Szwerinski, N., Chen, K.H., Driscoll, S., Washburn, J., Jones, V.G., 2025. Evaluation of an Ambient Artificial Intelligence Documentation Platform for Clinicians. JAMA Netw Open 8, e258614. https://doi.org/10.1001/jamanetworkopen.2025.8614

Swanson, K., Wu, W., Bulaong, N.L., Pak, J.E., Zou, J., 2025. The Virtual Lab of AI agents designs new SARS-CoV-2 nanobodies. Nature 646, 716–723. https://doi.org/10.1038/s41586-025-09442-9

Tai-Seale, M., Baxter, S.L., Vaida, F., Walker, A., Sitapati, A.M., Osborne, C., Diaz, J., Desai, N., Webb, S., Polston, G., Helsten, T., Gross, E., Thackaberry, J., Mandvi, A., Lillie, D., Li, S., Gin, G., Achar, S., Hofflich, H., Sharp, C., Millen, M., Longhurst, C.A., 2024. AI-Generated Draft Replies Integrated Into Health Records and Physicians' Electronic Communication. JAMA Netw Open 7, e246565. https://doi.org/10.1001/jamanetworkopen.2024.6565

Tai-Seale, M., Millen, M., Vaida, F., Barbato, C., Diaz, J., Walker, A., Singh, K., Sitapati, A., Pham, D., Tran, A., Ficklin-Badaloni, K., Soberg, C., Dang, L., Mandvi, A., Graham, R., Tang, A., Allen, N., Seale, T., Longhurst, C.A., 2026. Early experience with a patient-facing AI chatbot integrated in a patient portal. JAMIA Open 9, ooag083. https://doi.org/10.1093/jamiaopen/ooag083

Tang, A.S., Rankin, K.P., Cerono, G., Miramontes, S., Mills, H., Roger, J., Zeng, B., Nelson, C., Soman, K., Woldemariam, S., Li, Y., Lee, A., Bove, R., Glymour, M., Aghaeepour, N., Oskotsky, T.T., Miller, Z., Allen, I.E., Sanders, S.J., Baranzini, S., Sirota, M., 2024. Leveraging electronic health records and knowledge networks for Alzheimer's disease prediction and sex-specific biological insights. Nat Aging 4, 379–395. https://doi.org/10.1038/s43587-024-00573-8

Tang, J., LeBel, A., Jain, S., Huth, A.G., 2023. Semantic reconstruction of continuous language from non-invasive brain recordings. Nat Neurosci 26, 858–866. https://doi.org/10.1038/s41593-023-01304-9
Tang, R., Chuang, Y.-N., Hu, X., 2024. The Science of Detecting LLM-Generated Text. Commun. ACM 67, 50–59. https://doi.org/10.1145/3624725
Teo, Z.L., Thirunavukarasu, A.J., Elangovan, K., Cheng, H., Moova, P., Soetikno, B., Nielsen, C., Pollreisz, A., Ting, D.S.J., Morris, R.J.T., Shah, N.H., Langlotz, C.P., Ting, D.S.W., 2025. Generative artificial intelligence in medicine. Nat Med 31, 3270–3282. https://doi.org/10.1038/s41591-025-03983-2
Thorp, H.H., 2026. Resisting AI slop. Science 391, 5. https://doi.org/10.1126/science.aee8267
Topaz, M., Roguin, N., Gupta, P., Zhang, Z., Peltonen, L.-M., 2026. Fabricated citations: an audit across 2·5 million biomedical papers. Lancet 407, 1779–1781. https://doi.org/10.1016/S0140-6736(26)00603-3
Topol, E., 2026. The Paradox of Medical AI Implementation. Ground Truths. URL https://erictopol.substack.com/p/the-paradox-of-medical-ai-implementation (accessed 5.14.26).
Topol, E., 2022. The amazing power of "machine eyes." Ground Truths. URL https://erictopol.substack.com/p/the-amazing-power-of-machine-eyes (accessed 10.14.22).
Topol, E., 2019. Deep Medicine: How Artificial Intelligence Can Make Healthcare Human Again, Illustrated Edition. ed. Basic Books, New York.
Topol, E.J., 2026. The Future of AI-Facilitated Medicine. Dædalus.
Toro-Tobon, D., Ponte, O.P., Montori, V.M., Brito, J.P., 2026. Clinician in the loop: a flawed solution for AI oversight. BMJ 393, e089213. https://doi.org/10.1136/bmj-2025-089213
Vaid, A., Sawant, A., Suarez-Farinas, M., Lee, J., Kaul, S., Kovatch, P., Freeman, R., Jiang, J., Jayaraman, P., Fayad, Z., Argulian, E., Lerakis, S., Charney, A.W., Wang, F., Levin, M., Glicksberg, B., Narula, J., Hofer, I., Singh, K., Nadkarni, G.N., 2023. Implications of the Use of Artificial Intelligence Predictive Models in Health Care Settings : A Simulation Study. Ann Intern Med. https://doi.org/10.7326/M23-0949
Van Veen, D., Van Uden, C., Blankemeier, L., Delbrouck, J.-B., Aali, A., Bluethgen, C., Pareek, A., Polacin, M., Reis, E.P., Seehofnerová, A., Rohatgi, N., Hosamani, P., Collins, W., Ahuja, N., Langlotz, C.P., Hom, J., Gatidis, S., Pauly, J., Chaudhari, A.S., 2024. Adapted large language models can outperform medical experts in clinical text summarization. Nat Med 30, 1134–1142. https://doi.org/10.1038/s41591-024-02855-5
Vishwanath, K., Alyakin, A., Ghosh, M., Hage, A., Neifert, S.N., Orillac, C., Mandelberg, N.J., Khan, H.A., Lee, J.V., Yao, J.J., Small, W.R., Varma, A., Hewitt, D.B., Aphinyanaphongs, Y., Alber, D.A., Oermann, E.K., 2026. General-purpose large language models outperform specialized clinical AI tools on medical benchmarks. Nat Med. https://doi.org/10.1038/s41591-026-04431-5
Wachter, R., 2026. A Giant Leap: How AI Is Transforming Healthcare and What That Means for Our Future. Portfolio, New York.
Walters, W.H., Wilder, E.I., 2023. Fabrication and errors in the bibliographic citations generated by ChatGPT. Sci Rep 13, 14045. https://doi.org/10.1038/s41598-023-41032-5
Widner, K., Virmani, S., Krause, J., Nayar, J., Tiwari, R., Pedersen, E.R., Jeji, D., Hammel, N., Matias, Y., Corrado, G.S., Liu, Y., Peng, L., Webster, D.R., 2023. Lessons learned from translating AI from development to deployment in healthcare. Nat Med 29, 1304–1306. https://doi.org/10.1038/s41591-023-02293-9

Wu, K., Wu, E., Wei, K., Zhang, A., Casasola, A., Nguyen, T., Riantawan, S., Shi, P., Ho, D., Zou, J., 2025. An automated framework for assessing how well LLMs cite relevant medical references. Nat Commun 16, 3615. https://doi.org/10.1038/s41467-025-58551-6

Xu, G., Li, X., Chen, Y., Duan, Y., Wu, S., Yu, H., Chiu, C.-H., Ni, J., Tang, N., Li, T.J.-J., Yuille, A., Jin, W., Shi, Y., 2026. A comprehensive survey of AI agents in healthcare. J Biomed Inform 179, 105045. https://doi.org/10.1016/j.jbi.2026.105045

You, A.X., Kahl, N.M., Patel, A., Cao, J., Castillo, E.M., Bedmutha, P.S., Chan, T., Singh, K., Longhurst, C.A., 2026. Artificial Intelligence Note Summarization in the Emergency Department. JAMA Netw Open 9, e2618021. https://doi.org/10.1001/jamanetworkopen.2026.18021

Zong, H., Wu, R., Cha, J., Wang, J., Wu, E., Li, J., Zhou, Y., Zhang, C., Feng, W., Shen, B., 2024. Large Language Models in Worldwide Medical Exams: Platform Development and Comprehensive Analysis. J Med Internet Res 26, e66114. https://doi.org/10.2196/66114

Zou, J., Topol, E.J., 2025. The rise of agentic AI teammates in medicine. Lancet 405, 457. https://doi.org/10.1016/S0140-6736(25)00202-8