

AI and Biomedical Informatics Education: Competencies and Challenges

Illinois Eye-TEACH
June 4-5, 2026

William Hersh, MD
Professor

Division of Informatics, Clinical Epidemiology, and Translational Data Science
Department of Medicine
School of Medicine
Oregon Health & Science University
Portland, OR, USA

Email: hersh@ohsu.edu

Web: <https://billhersh.info/>

LinkedIn: <https://www.linkedin.com/in/bilhersh/>

BlueSky: [@billhersh.bsky.social](https://bsky.app/profile/bilhersh.bsky.social)

References

- 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>
- 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>
- 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](https://doi.org/10.1016/S0140-6736(25)02075-6)
- Brodeur, P.G., Buckley, T.A., Kanjee, Z., Goh, E., Ling, E.B., Jain, P., Cabral, S., Abdunour, 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>
- 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>
- Hersh, W., 2023. Physician and Medical Student Competence in AI Must Include Broader Competence in Clinical Informatics. *Informatics Professor*. URL <https://informaticsprofessor.blogspot.com/2023/09/physician-and-medical-student.html> (accessed 9.15.23).

- Hersh, W., Ehrenfeld, J., 2020. Clinical Informatics, in: Health Systems Science, 2nd Edition. Elsevier, pp. 156–170.
- 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>
- Hersh, W.R., Gorman, P.N., Biagioli, F.E., Mohan, V., Gold, J.A., Mejicano, G.C., 2014. Beyond information retrieval and electronic health record use: competencies in clinical informatics for medical education. Adv Med Educ Pract 5, 205–212. <https://doi.org/10.2147/AMEP.S63903>
- 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>
- Hunt, V.M., Sprehe, L.K., de Lomba, W.C., Gameiro, R.R., Woite, N.L., Wade, C.G., Chen, J.H., Rougas, S.C., Fraser, H.S.F., Sullivan, A.M., Fregni, F., 2026. What the AI era doctor should know: a scoping review of proposed artificial intelligence competencies for medical education. NPJ Digit Med. <https://doi.org/10.1038/s41746-026-02761-9>
- 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>
- 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>
- 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>
- 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>
- Ng, F.Y.C., Thirunavukarasu, A.J., Cheng, H., Tan, T.F., Gutierrez, L., Lan, Y., Ong, J.C.L., Chong, Y.S., Ngiam, K.Y., Ho, D., Wong, T.Y., Kwek, K., Doshi-Velez, F., Lucey, C.,

- Coffman, T., Ting, D.S.W., 2023. Artificial intelligence education: An evidence-based medicine approach for consumers, translators, and developers. *Cell Rep Med* 4, 101230. <https://doi.org/10.1016/j.xcrm.2023.101230>
- Nori, H., Lee, Y.T., Zhang, S., Carignan, D., Edgar, R., Fusi, N., King, N., Larson, J., Li, Y., Liu, W., Luo, R., McKinney, S.M., Ness, R.O., Poon, H., Qin, T., Usuyama, N., White, C., Horvitz, E., 2023. Can Generalist Foundation Models Outcompete Special-Purpose Tuning? Case Study in Medicine. <https://doi.org/10.48550/arXiv.2311.16452>
- 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>
- 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).
- 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>
- Russell, R.G., Lovett Novak, L., Patel, M., Garvey, K.V., Craig, K.J.T., Jackson, G.P., Moore, D., Miller, B.M., 2023. Competencies for the Use of Artificial Intelligence-Based Tools by Health Care Professionals. *Acad Med* 98, 348–356. <https://doi.org/10.1097/ACM.0000000000004963>
- 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>
- 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>



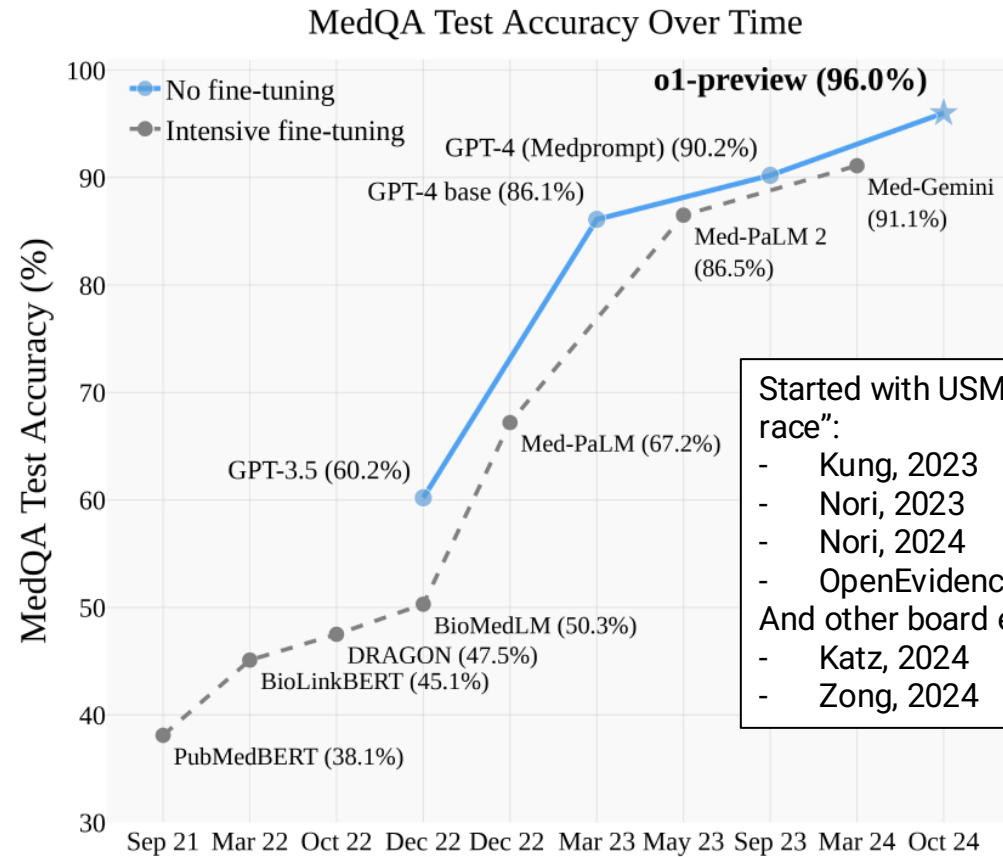
AI and Biomedical Informatics Education – Competencies and Challenges

William Hersh, MD
Professor
Division of Informatics, Clinical Epidemiology, and Translational Data Science
Oregon Health & Science University
Portland, OR, USA

Overview

- Accomplishments of AI
- Competencies for clinicians and students
- Challenges for AI in education

We face daily tsunami of impressive accomplishments of AI in medicine



Started with USMLE “arms race”:

- Kung, 2023
- Nori, 2023
- Nori, 2024
- OpenEvidence, 2025

And other board exams:

- Katz, 2024
- Zong, 2024

naturemedicine

Explore content ▾ About the journal ▾ Publish with us ▾ Subscribe

[nature](#) > [nature medicine](#) > [articles](#) > [article](#)

Article | Published: 05 February 2025

GPT-4 assistance for improvement of physician performance on patient care tasks: a randomized controlled trial

[Ethan Goh](#), [Robert J. Gallo](#), [Eric Strong](#), [Yingjie Weng](#), [Hannah Kerman](#), [Jason A. Freed](#), [Joséphine A. Cool](#), [Zahir Kanjee](#), [Kathleen P. Lane](#), [Andrew S. Parsons](#), [Neera Ahuja](#), [Eric Horvitz](#), [Daniel Yang](#), [Arnold Milstein](#), [Andrew P. J. Olson](#), [Jason Hom](#), [Jonathan H. Chen](#) & [Adam Rodman](#)

Nature Medicine 31, 1233–1238 (2025) | [Cite this article](#)

Science

Current Issue First release papers Archive About ▾ [Submit manuscript](#)

HOME > SCIENCE > VOL. 392, NO. 6797 > PERFORMANCE OF A LARGE LANGUAGE MODEL ON THE REASONING TASKS OF A PHYSICIAN

RESEARCH ARTICLE | MACHINE LEARNING

[f](#) [X](#) [b](#) [in](#) [G+](#) [v](#) [e](#) [m](#)

Performance of a large language model on the reasoning tasks of a physician

PETER G. BRODEUR [ORCID](#), THOMAS A. BUCKLEY [ORCID](#), ZAHIR KANJEE [ORCID](#), ETHAN GOH [ORCID](#), [...], AND ADAM RODMAN [ORCID](#) +20 authors [Authors Info & Affiliations](#)

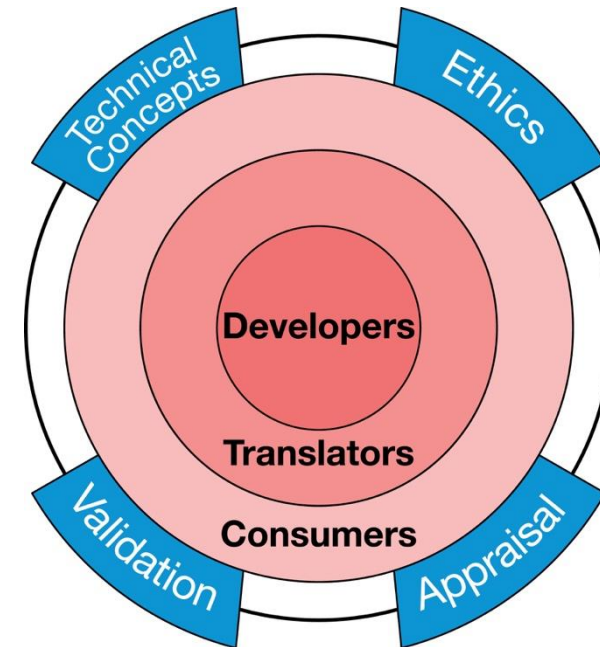
SCIENCE • 30 Apr 2026 • Vol 392, Issue 6797 • pp. 524-527 • DOI: 10.1126/science.adz4433



Education and training of clinicians and informaticians essential

- Should build on competencies in informatics (Hersh, 2014; Hersh 2020; Hersh, 2023)
- Different needs for diverse learners (Ng, 2023)
- Emerging AI competencies (Russell, 2023; Hunt, 2026)

1. Find, search, and apply knowledge-based information to patient care and other clinical tasks
2. Effectively read from, and write to, the electronic health record for patient care and other clinical activities
3. Use and guide implementation of clinical decision support (CDS)
4. Provide care using population health management approaches
5. Protect patient privacy and security
6. Use information technology to improve patient safety
7. Engage in quality measurement selection and improvement
8. Use health information exchange (HIE) to identify and access patient information across clinical settings
9. Engage patients to improve their health and care delivery through personal health records and patient portals
10. Maintain professionalism through use of information technology tools
11. Provide clinical care via telemedicine and refer patients as indicated
12. Apply personalized/precision medicine
13. Participate in practice-based clinical and translational research
14. Apply and evaluate artificial intelligence applications in patient care and other clinical activities

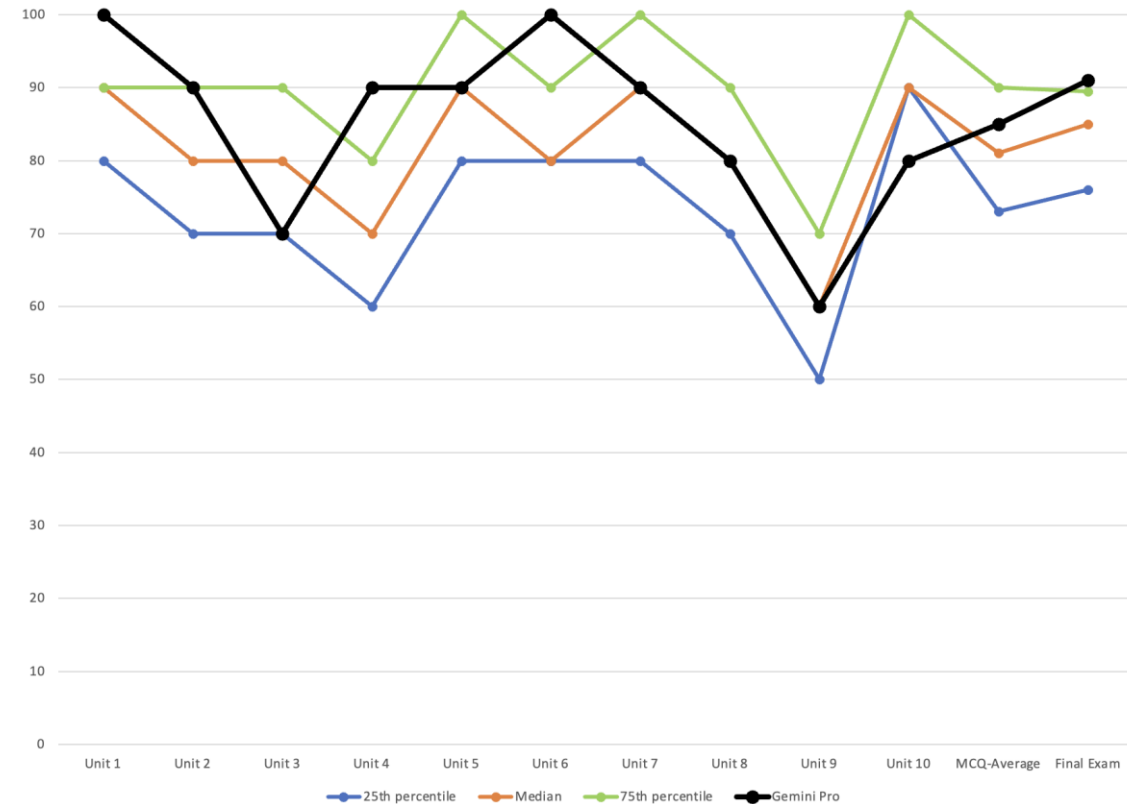


Concerns for generative AI in clinical practice and education

- How do we assess students in era of AI?
- Does AI really “outperform” physicians?
- AI performance worsens when used by humans
- AI may undermine cognition and learning
- When and how to introduce AI into medical education and practice

Student assessment more challenging – my story (Hersh, 2024)

- Large introductory biomedical and health informatics course
 - Curriculum and assessments in courses taught to 133 graduate students, medical students, and continuing education students in 2023
- Assessments include
 - Multiple-choice questions (MCQs) – 10 questions in each of 10 units
 - Final exam – 33 short-answer questions
 - Term paper – not assessed in this study
- Generative AI
 - Use of off-the-shelf LLMs for course assessments



Does AI really “outperform” physicians?

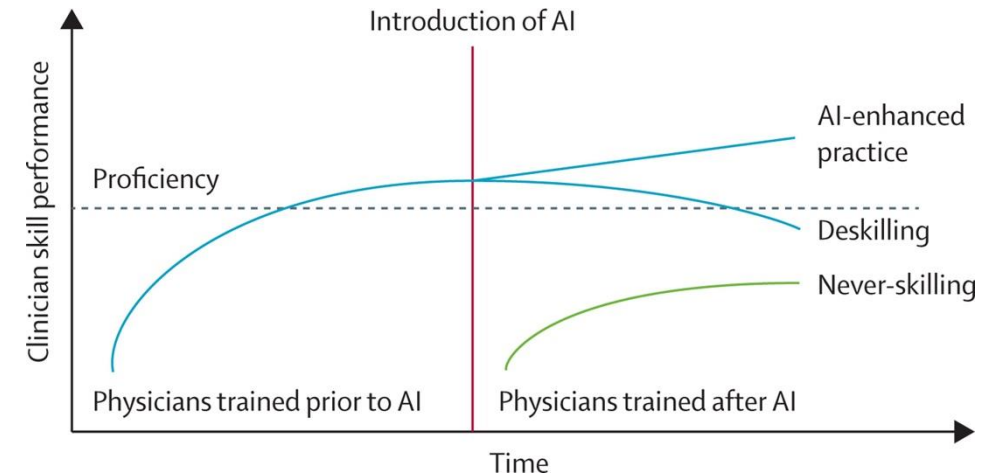
- Performance of LLMs on reasoning tasks of physicians on curated case-based benchmarks, including (Brodeur, 2026)
 - NEJM clinicopathological conferences
 - Cases from (Goh, 2024)
 - 76 ED cases at three diagnostic touchpoints – triage in ED, initial evaluation by physician, and admission to hospital or ICU
- On MSD Manual vignettes, frontier LLMs achieved high accuracy on final diagnoses but performed less well at generating differential diagnoses and navigating uncertainty at earlier reasoning stages (Rao, 2026)
- These show human intervention still required for LLMs on patient cases

AI tools work less well in hands of humans

- LLMs in physician diagnostic reasoning (Goh, 2024)
 - 50 physicians randomized to LLM or conventional information resources and assessed with diagnostic performance rubric for 6 diagnostic cases
 - No statistical difference between physicians using LLM (76%) vs. conventional (74%) resources; LLM alone scored better than either (92%)
- LLMs as medical assistants for general public (Bean, 2026)
 - 1298 people assisted on 10 scenarios randomized to LLM or information source of choice
 - Used alone, LLMs completed scenarios accurately, correctly identifying conditions in 94.9% of cases and disposition in 56.3%
 - Participants using same LLMs identified relevant conditions in 34.5% of cases and disposition in 44.2%, similar to control group
- Growing body of research raising concerns about reduced cognition and learning with use of LLMs compared to searching and synthesis (Lee, 2025; Kosmyna, 2025; Melamud, 2025; Barcaui, 2025; Liu, 2026)

When and how to introduce AI into medical education and practice

- Concerns for de-skilling and never-skilling (Berzin, 2025)
 - Scoping review found risk of deskilling and loss of expertise in several specialties (Heudel, 2026)
 - Students should be calibrated and supervised to avoid never-skilling in medical education (Ke, 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)
- 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)



Final thoughts

- Generative AI has profoundly impacted biomedicine and will continue to do so
- Some unanswered questions
 - What is its optimal role in education?
 - How do we assess students in era of AI?
 - What is the best way for future clinicians, informaticians, researchers, and others to learn about use of AI?
- Many opportunities for students and faculty to answer these questions!

Thank you!

William Hersh, MD, FACP, FACMI,
FAMIA, FIAHSI
Professor
Division of Informatics, Clinical
Epidemiology, and Translational Data
Science
Department of Medicine
School of Medicine
Oregon Health & Science University
Portland, OR, USA
Email: hersh@ohsu.edu
Web: <https://www.billhersh.info>

