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

University of Cape Town School of IT Seminar
6 March 2025

William Hersh, MD
Professor
Department of Medical Informatics & Clinical Epidemiology
School of Medicine
Oregon Health & Science University
Portland, OR, USA
https://www.ohsu.edu/informatics
Email: hersh@ohsu.edu
Web: http://www.billhersh.info/
Blog: https://informaticsprofessor.blogspot.com/
X/Twitter: @williamhersh
BlueSky: @billhersh.bsky.social
LinkedIn: https://www.linkedin.com/in/billhersh/

References

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

Arvidsson, R., Gunnarsson, R., Entezarjou, A., Sundemo, D., Wikberg, C., 2024. ChatGPT (GPT-4) versus doctors on complex cases of the Swedish family medicine specialist examination: an observational comparative study. BMJ Open 14, e086148. https://doi.org/10.1136/bmjopen-2024-086148

Attia, Z.I., Friedman, P.A., Noseworthy, P.A., Lopez-Jimenez, F., Ladewig, D.J., Satam, G., Pellikka, P.A., Munger, T.M., Asirvatham, S.J., Scott, C.G., Carter, R.E., Kapa, S., 2019. Age and Sex Estimation Using Artificial Intelligence From Standard 12-Lead ECGs. Circ Arrhythm Electrophysiol 12, e007284. https://doi.org/10.1161/CIRCEP.119.007284

Avramovic, S., Avramovic, I., Wojtusiak, J., 2024. Exploring the Impact of GitHub Copilot on Health Informatics Education. Appl Clin Inform 15, 1121–1129. https://doi.org/10.1055/a-2414-7790

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

Ball, P., 2023. Is AI leading to a reproducibility crisis in science? Nature 624, 22–25. https://doi.org/10.1038/d41586-023-03817-6
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
Bensoussan, Y., Elemento, O., Rameau, A., 2024. Voice as an AI Biomarker of Health-Introducing Audiomics. JAMA Otolaryngol Head Neck Surg 150, 283–284. https://doi.org/10.1001/jamaoto.2023.4807
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
Cheng, L., Li, X., Bing, L., 2023. Is GPT-4 a Good Data Analyst?, in: Bouamor, H., Pino, J., Bali, K. (Eds.), Findings of the Association for Computational Linguistics: EMNLP 2023. Presented at the Findings 2023, Association for Computational Linguistics, Singapore, pp. 9496–9514. https://doi.org/10.18653/v1/2023.findings-emnlp.637
Chin, M.H., Afsar-Manesh, N., Bierman, A.S., Chang, C., Colón-Rodríguez, C.J., Dullabh, P., Duran, D.G., Fair, M., Hernandez-Boussard, T., Hightower, M., Jain, A., Jordan, W.B., Konya, S., Moore, R.H., Moore, T.T., Rodriguez, R., Shaheen, G., Snyder, L.P., Srinivasan, M., Umscheid, C.A., Ohno-Machado, L., 2023. Guiding Principles to Address the Impact of Algorithm Bias on Racial and Ethnic Disparities in Health and Health Care. JAMA Netw Open 6, e2345050. https://doi.org/10.1001/jamanetworkopen.2023.45050
Choi, J.H., Monahan, A., Schwarcz, D., 2023. Lawyering in the Age of Artificial Intelligence. https://doi.org/10.2139/ssrn.4626276
Cooper, A., Rodman, A., 2023. AI and Medical Education — A 21st-Century Pandora's Box. New England Journal of Medicine. https://doi.org/10.1056/NEJMp2304993
Coyner, A.S., Singh, P., Brown, J.M., Ostmo, S., Chan, R.V.P., Chiang, M.F., Kalpathy-Cramer, J., Campbell, J.P., Imaging and Informatics in Retinopathy of Prematurity Consortium, 2023. Association of Biomarker-Based Artificial Intelligence With Risk of Racial Bias in Retinal Images. JAMA Ophthalmol 141, 543–552. https://doi.org/10.1001/jamaophthalmol.2023.1310
Dell'Acqua, F., McFowland, E., Mollick, E.R., Lifshitz-Assaf, H., Kellogg, K., Rajendran, S., Krayer, L., Candelon, F., Lakhani, K.R., 2023. Navigating the Jagged Technological Frontier: Field Experimental Evidence of the Effects of AI on Knowledge Worker Productivity and Quality. https://doi.org/10.2139/ssrn.4573321
Denny, P., Prather, J., Becker, B.A., Finnie-Ansley, J., Hellas, A., Leinonen, J., Luxton-Reilly, A., Reeves, B.N., Santos, E.A., Sarsa, S., 2024. Computing Education in the Era of Generative AI. Commun. ACM 67, 56–67. https://doi.org/10.1145/3624720
Desaire, H., Chua, A.E., Isom, M., Jarosova, R., Hua, D., 2023. Distinguishing academic science writing from humans or ChatGPT with over 99% accuracy using off-the-shelf machine learning tools. Cell Rep Phys Sci 4, 101426. https://doi.org/10.1016/j.xcrp.2023.101426
Donzé, J., Aujesky, D., Williams, D., Schnipper, J.L., 2013. Potentially avoidable 30-day hospital readmissions in medical patients: derivation and validation of a prediction model. JAMA Intern Med 173, 632–638. https://doi.org/10.1001/jamainternmed.2013.3023

Dratsch, T., Chen, X., Rezazade Mehrizi, M., Kloeckner, R., Mähringer-Kunz, A., Püsken, M., Baeßler, B., Sauer, S., Maintz, D., Pinto Dos Santos, D., 2023. Automation Bias in Mammography: The Impact of Artificial Intelligence BI-RADS Suggestions on Reader Performance. Radiology 307, e222176. https://doi.org/10.1148/radiol.222176

Dubey, A., Jauhri, A., Yang, Z., Zhao, Z., 2024. The Llama 3 Herd of Models. https://doi.org/10.48550/arXiv.2407.21783

Edwards, C., 2024. Teaching Transformed. Commun. ACM 67, 12–13. https://doi.org/10.1145/3637208

Finlayson, S.G., Subbaswamy, A., Singh, K., Bowers, J., Kupke, A., Zittrain, J., Kohane, I.S., Saria, S., 2021. The Clinician and Dataset Shift in Artificial Intelligence. N Engl J Med 385, 283–286. https://doi.org/10.1056/NEJMc2104626

Gichoya, J.W., Banerjee, I., Bhimireddy, A.R., Burns, J.L., Celi, L.A., Chen, L.-C., Correa, R., Dullerud, N., Ghassemi, M., Huang, S.-C., Kuo, P.-C., Lungren, M.P., Palmer, L.J., Price, B.J., Purkayastha, S., Pyrros, A.T., Oakden-Rayner, L., Okechukwu, C., Seyyed-Kalantari, L., Trivedi, H., Wang, R., Zaiman, Z., Zhang, H., 2022. AI recognition of patient race in medical imaging: a modelling study. Lancet Digit Health 4, e406–e414. https://doi.org/10.1016/S2589-7500(22)00063-2

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

Gorecki, G.-P., Tomescu, D.-R., Pleș, L., Panaitescu, A.-M., Dragosloveanu, Șerban, Scheau, C., Sima, R.-M., Coman, I.-S., Grigorean, V.-T., Cochior, D., 2024. Implications of using artificial intelligence in the diagnosis of sepsis/sepsis shock. Germs 14, 77–84. https://doi.org/10.18683/germs.2024.1419

Greenhalgh, T., Fisman, D., Cane, D.J., Oliver, M., Macintyre, C.R., 2022. Adapt or die: how the pandemic made the shift from EBM to EBM+ more urgent. BMJ Evid Based Med 27, 253–260. https://doi.org/10.1136/bmjebm-2022-111952

Gupta, D., Demner-Fushman, D., Hersh, W., Bedrick, S., Roberts, K., 2024. Overview of TREC 2024 Biomedical Generative Retrieval (BioGen) Track. https://doi.org/10.48550/arXiv.2411.18069

Han, C., Kim, D.W., Kim, S., You, S.C., Park, J.Y., Bae, S., Yoon, D., 2023. Evaluation Of GPT-4 for 10-Year Cardiovascular Risk Prediction: Insights from the UK Biobank and KoGES Data. https://doi.org/10.2139/ssrn.4583995

Han, R., Acosta, J.N., Shakeri, Z., Ioannidis, J.P.A., Topol, E.J., Rajpurkar, P., 2024. Randomised controlled trials evaluating artificial intelligence in clinical practice: a scoping review. Lancet Digit Health 6, e367–e373. https://doi.org/10.1016/S2589-7500(24)00047-5

Hassan, C., Spadaccini, M., Mori, Y., Foroutan, F., Facciorusso, A., Gkolfakis, P., Tziatzios, G., Triantafyllou, K., Antonelli, G., Khalaf, K., Rizkala, T., Vandvik, P.O., Fugazza, A., Rondonotti, E., Glissen-Brown, J.R., Kamba, S., Maida, M., Correale, L., Bhandari, P., Jover, R., Sharma, P., Rex, D.K., Repici, A., 2023. Real-Time Computer-Aided Detection of Colorectal Neoplasia During Colonoscopy : A Systematic Review and Meta-analysis. Ann Intern Med. https://doi.org/10.7326/M22-3678

Hersh, W., 2024a. Translational AI: A Necessity and Opportunity for Biomedical Informatics and Data Science. NLM Musings from the Mezzanine. URL

https://nlmdirector.nlm.nih.gov/2024/02/07/translational-ai-a-necessity-and-opportunity-for-biomedical-informatics-and-data-science/ (accessed 2.10.24).
Hersh, W., 2024b. 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., 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. 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
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
Holmstrom, L., Christensen, M., Yuan, N., Weston Hughes, J., Theurer, J., Jujjavarapu, M., Fatehi, P., Kwan, A., Sandhu, R.K., Ebinger, J., Cheng, S., Zou, J., Chugh, S.S., Ouyang, D., 2023. Deep learning-based electrocardiographic screening for chronic kidney disease. Commun Med (Lond) 3, 73. https://doi.org/10.1038/s43856-023-00278-w
Hong, S., Lin, Y., Liu, Bang, Liu, Bangbang, Wu, B., Li, D., Chen, J., Zhang, J., Wang, J., Zhang, Li, Zhang, Lingyao, Yang, M., Zhuge, M., Guo, T., Zhou, T., Tao, W., Wang, W., Tang, X., Lu, X., Zheng, X., Liang, X., Fei, Y., Cheng, Y., Xu, Z., Wu, C., 2024. Data Interpreter: An LLM Agent For Data Science. https://doi.org/10.48550/arXiv.2402.18679
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).
Idrisoglu, A., Dallora, A.L., Anderberg, P., Berglund, J.S., 2023. Applied Machine Learning Techniques to Diagnose Voice-Affecting Conditions and Disorders: Systematic Literature Review. J Med Internet Res 25, e46105. https://doi.org/10.2196/46105
James, C.A., Wachter, R.M., Woolliscroft, J.O., 2022. Preparing Clinicians for a Clinical World Influenced by Artificial Intelligence. JAMA 327, 1333–1334. https://doi.org/10.1001/jama.2022.3580
Johnson, M., 2024. Generative AI and CS Education. Commun. ACM 67, 23–24. https://doi.org/10.1145/3632523
Johri, S., Jeong, J., Tran, B.A., Schlessinger, D.I., Wongvibulsin, S., Barnes, L.A., Zhou, H.-Y., Cai, Z.R., Van Allen, E.M., Kim, D., Daneshjou, R., Rajpurkar, P., 2025. An evaluation framework for clinical use of large language models in patient interaction tasks. Nat Med. https://doi.org/10.1038/s41591-024-03328-5
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

Kapoor, S., Narayanan, A., 2023. Leakage and the reproducibility crisis in machine-learning-based science. Patterns (N Y) 4, 100804. https://doi.org/10.1016/j.patter.2023.100804

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

King, M., 2023. How Search Generative Experience works and why retrieval-augmented generation is our future [WWW Document]. Search Engine Land. URL https://searchengineland.com/how-search-generative-experience-works-and-why-retrieval-augmented-generation-is-our-future-433393 (accessed 12.10.23).

Koopman, B., Zuccon, G., 2023. Dr ChatGPT tell me what I want to hear: How different prompts impact health answer correctness, in: Bouamor, H., Pino, J., Bali, K. (Eds.), Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing. Presented at the EMNLP 2023, Association for Computational Linguistics, Singapore, pp. 15012–15022. https://doi.org/10.18653/v1/2023.emnlp-main.928

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

Lewis, A.E., Weiskopf, N., Abrams, Z.B., Foraker, R., Lai, A.M., Payne, P.R.O., Gupta, A., 2023. Electronic health record data quality assessment and tools: a systematic review. J Am Med Inform Assoc ocad120. https://doi.org/10.1093/jamia/ocad120

Li, D., Gupta, K., Bhaduri, M., Sathiadoss, P., Bhatnagar, S., Chong, J., 2024. Comparing GPT-3.5 and GPT-4 Accuracy and Drift in Radiology Diagnosis Please Cases. Radiology 310, e232411. https://doi.org/10.1148/radiol.232411

Liang, W., Izzo, Z., Zhang, Y., Lepp, H., Cao, H., Zhao, X., Chen, L., Ye, H., Liu, S., Huang, Z., McFarland, D.A., Zou, J.Y., 2024. Monitoring AI-Modified Content at Scale: A Case Study on the Impact of ChatGPT on AI Conference Peer Reviews. https://doi.org/10.48550/arXiv.2403.07183

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, P., Yuan, W., Fu, J., Jiang, Z., Hayashi, H., Neubig, G., 2023. Pre-train, Prompt, and Predict: A Systematic Survey of Prompting Methods in Natural Language Processing. ACM Comput. Surv. 55, 195:1-195:35. https://doi.org/10.1145/3560815

Liu, X., Rivera, S.C., Moher, D., Calvert, M.J., Denniston, A.K., SPIRIT-AI and CONSORT-AI Working Group, 2020. Reporting guidelines for clinical trial reports for interventions involving artificial intelligence: the CONSORT-AI Extension. BMJ 370, m3164. https://doi.org/10.1136/bmj.m3164

Mangas-Sanjuan, C., de-Castro, L., Cubiella, J., Díez-Redondo, P., Suárez, A., Pellisé, M., Fernández, N., Zarraquiños, S., Núñez-Rodríguez, H., Álvarez-García, V., Ortiz, O., Sala-Miquel, N., Zapater, P., Jover, R., CADILLAC study investigators*, 2023. Role of Artificial Intelligence in Colonoscopy Detection of Advanced Neoplasias : A Randomized Trial. Ann Intern Med. https://doi.org/10.7326/M22-2619

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

Mitsuyama, Y., Matsumoto, T., Tatekawa, H., Walston, S.L., Kimura, T., Yamamoto, A., Watanabe, T., Miki, Y., Ueda, D., 2023. Chest radiography as a biomarker of ageing: artificial intelligence-based, multi-institutional model development and validation in Japan. The Lancet Healthy Longevity 0. https://doi.org/10.1016/S2666-7568(23)00133-2

Mollick, E., 2024. Post-apocalyptic education [WWW Document]. One Useful Thing. URL https://www.oneusefulthing.org/p/post-apocalyptic-education (accessed 9.4.24).

Mollick, E., 2023. The Homework Apocalypse [WWW Document]. One Useful Thing. URL https://www.oneusefulthing.org/p/the-homework-apocalypse (accessed 3.20.24).

Mukherjee, P., Humbert-Droz, M., Chen, J.H., Gevaert, O., 2023. SCOPE: predicting future diagnoses in office visits using electronic health records. Sci Rep 13, 11005. https://doi.org/10.1038/s41598-023-38257-9

Nam, J., 2023. 56% of College Students Have Used AI on Assignments or Exams | BestColleges [WWW Document]. BestColleges.com. URL https://www.bestcolleges.com/research/most-college-students-have-used-ai-survey/ (accessed 12.13.23).

Nelson, S.D., 2025. When AI Goes Astray - Understanding Model Drift. American Society of Health-System Pharmacists.

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

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).

Odri, G.-A., Yun Yoon, D.J., 2023. Detecting generative artificial intelligence in scientific articles: evasion techniques and implications for scientific integrity. Orthop Traumatol Surg Res 103706. https://doi.org/10.1016/j.otsr.2023.103706

Omiye, J.A., Gui, H., Rezaei, S.J., Zou, J., Daneshjou, R., 2024. Large Language Models in Medicine: The Potentials and Pitfalls : A Narrative Review. Ann Intern Med 177, 210–220. https://doi.org/10.7326/M23-2772

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

Owens, L.M., Wilda, J.J., Grifka, R., Westendorp, J., Fletcher, J.J., 2024. Effect of Ambient Voice Technology, Natural Language Processing, and Artificial Intelligence on the Patient-Physician Relationship. Appl Clin Inform 15, 660–667. https://doi.org/10.1055/a-2337-4739

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).

Plana, D., Shung, D.L., Grimshaw, A.A., Saraf, A., Sung, J.J.Y., Kann, B.H., 2022. Randomized Clinical Trials of Machine Learning Interventions in Health Care: A Systematic Review. JAMA Netw Open 5, e2233946. https://doi.org/10.1001/jamanetworkopen.2022.33946

Poldrack, R.A., Lu, T., Beguš, G., 2023. AI-assisted coding: Experiments with GPT-4. https://doi.org/10.48550/arXiv.2304.13187

Poplin, R., Varadarajan, A.V., Blumer, K., Liu, Y., McConnell, M.V., Corrado, G.S., Peng, L., Webster, D.R., 2018. Prediction of cardiovascular risk factors from retinal fundus photographs via deep learning. Nat Biomed Eng 2, 158–164. https://doi.org/10.1038/s41551-018-0195-0

Pyrros, A., Borstelmann, S.M., Mantravadi, R., Zaiman, Z., Thomas, K., Price, B., Greenstein, E., Siddiqui, N., Willis, M., Shulhan, I., Hines-Shah, J., Horowitz, J.M., Nikolaidis, P., Lungren, M.P., Rodríguez-Fernández, J.M., Gichoya, J.W., Koyejo, S., Flanders, A.E., Khandwala, N., Gupta, A., Garrett, J.W., Cohen, J.P., Layden, B.T., Pickhardt, P.J., Galanter, W., 2023. Opportunistic detection of type 2 diabetes using deep learning from frontal chest radiographs. Nat Commun 14, 4039. https://doi.org/10.1038/s41467-023-39631-x

Rajkomar, A., Oren, E., Chen, K., Dai, A.M., Hajaj, N., Hardt, M., Liu, P.J., Liu, X., Marcus, J., Sun, M., Sundberg, P., Yee, H., Zhang, K., Zhang, Y., Flores, G., Duggan, G.E., Irvine, J., Le, Q., Litsch, K., Mossin, A., Tansuwan, J., Wang, D., Wexler, J., Wilson, J., Ludwig, D., Volchenboum, S.L., Chou, K., Pearson, M., Madabushi, S., Shah, N.H., Butte, A.J., Howell, M.D., Cui, C., Corrado, G.S., Dean, J., 2018. Scalable and accurate deep learning with electronic health records. npj Digital Medicine 1, 1–10. https://doi.org/10.1038/s41746-018-0029-1

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

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

Sadasivan, V.S., Kumar, A., Balasubramanian, S., Wang, W., Feizi, S., 2023. Can AI-Generated Text be Reliably Detected? https://doi.org/10.48550/arXiv.2303.11156

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

Scarfe, P., Watcham, K., Clarke, A., Roesch, E., 2024. A real-world test of artificial intelligence infiltration of a university examinations system: A "Turing Test" case study. PLoS One 19, e0305354. https://doi.org/10.1371/journal.pone.0305354

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

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

Swanson, K., Liu, G., Catacutan, D.B., Arnold, A., Zou, J., Stokes, J.M., 2024. Generative AI for designing and validating easily synthesizable and structurally novel antibiotics. Nat Mach Intell 6, 338–353. https://doi.org/10.1038/s42256-024-00809-7

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

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.

Tu, T., Palepu, A., Schaekermann, M., Saab, K., Freyberg, J., Tanno, R., Wang, A., Li, B., Amin, M., Tomasev, N., Azizi, S., Singhal, K., Cheng, Y., Hou, L., Webson, A., Kulkarni, K., Mahdavi, S.S., Semturs, C., Gottweis, J., Barral, J., Chou, K., Corrado, G.S., Matias, Y., Karthikesalingam, A., Natarajan, V., 2024. Towards Conversational Diagnostic AI. https://doi.org/10.48550/arXiv.2401.05654

Tu, X., Zou, J., Su, W., Zhang, L., 2024. What Should Data Science Education Do With Large Language Models? Harvard Data Science Review 6. https://doi.org/10.1162/99608f92.bff007ab

Ueda, D., Matsumoto, T., Ehara, S., Yamamoto, A., Walston, S.L., Ito, A., Shimono, T., Shiba, M., Takeshita, T., Fukuda, D., Miki, Y., 2023. Artificial intelligence-based model to classify cardiac functions from chest radiographs: a multi-institutional, retrospective model development and validation study. Lancet Digit Health S2589-7500(23)00107–3. https://doi.org/10.1016/S2589-7500(23)00107-3

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

Vazquez, M.A., Oliver, G., Amarasingham, R., Amarasingham Ruben, Sundaram Venkatraghavan, Chan Kevin, Ahn Chul, Zhang Song, Bickel Perry, Parikh Samir M., Wells Barbara, Miller R. Tyler, Hedayati Susan, Hastings Jeffrey, Jaiyeola Adeola, Nguyen Tuan-Minh, Moran Brett, Santini Noel, Barker Blake, Velasco Ferdinand, Myers Lynn, Meehan Thomas P., Fox Chester, Toto Robert D., 2024. Pragmatic Trial of Hospitalization Rate in Chronic Kidney Disease. New England Journal of Medicine 390, 1196–1206. https://doi.org/10.1056/NEJMoa2311708

Walker, S.C., French, B., Moore, R.P., Domenico, H.J., Wanderer, J.P., Mixon, A.S., Creech, C.B., Byrne, D.W., Wheeler, A.P., 2023. Model-Guided Decision-Making for Thromboprophylaxis and Hospital-Acquired Thromboembolic Events Among Hospitalized Children and Adolescents: The CLOT Randomized Clinical Trial. JAMA Netw Open 6, e2337789. https://doi.org/10.1001/jamanetworkopen.2023.37789

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

Weiss, J., Raghu, V.K., Paruchuri, K., Zinzuwadia, A., Natarajan, P., Aerts, H.J.W.L., Lu, M.T., 2024. Deep Learning to Estimate Cardiovascular Risk From Chest Radiographs : A Risk Prediction Study. Ann Intern Med. https://doi.org/10.7326/M23-1898

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

Williams, C.Y.K., Zack, T., Miao, B.Y., Sushil, M., Wang, M., Kornblith, A.E., Butte, A.J., 2024. Use of a Large Language Model to Assess Clinical Acuity of Adults in the Emergency Department. JAMA Netw Open 7, e248895. https://doi.org/10.1001/jamanetworkopen.2024.8895

Wilson, F.P., 2023. Artificial Intelligence in Medicine Has a Major Cassandra Problem. Medium. URL https://fperrywilson.medium.com/artificial-intelligence-in-medicine-has-a-major-cassandra-problem-291140a124c6 (accessed 10.20.23).

Wu, K., Wu, E., Cassasola, A., Zhang, A., Wei, K., Nguyen, T., Riantawan, S., Riantawan, P.S., Ho, D.E., Zou, J., 2024. How well do LLMs cite relevant medical references? An evaluation framework and analyses. https://doi.org/10.48550/arXiv.2402.02008

Xu, S., Yang, L., Kelly, C., Sieniek, M., Kohlberger, T., Ma, M., Weng, W.-H., Kiraly, A., Kazemzadeh, S., Melamed, Z., Park, J., Strachan, P., Liu, Y., Lau, C., Singh, P., Chen, C., Etemadi, M., Kalidindi, S.R., Matias, Y., Chou, K., Corrado, G.S., Shetty, S., Tse, D., Prabhakara, S., Golden, D., Pilgrim, R., Eswaran, K., Sellergren, A., 2023. ELIXR: Towards a general purpose X-ray artificial intelligence system through alignment of large language models and radiology vision encoders [WWW Document]. arXiv.org. URL https://arxiv.org/abs/2308.01317v2 (accessed 9.26.23).

Yan, S., Knapp, W., Leong, A., Kadkhodazadeh, S., Das, S., Jones, V.G., Clark, R., Grattendick, D., Chen, K., Hladik, L., Fagan, L., Chan, A., 2024. Prompt engineering on leveraging large language models in generating response to InBasket messages. J Am Med Inform Assoc ocae172. https://doi.org/10.1093/jamia/ocae172

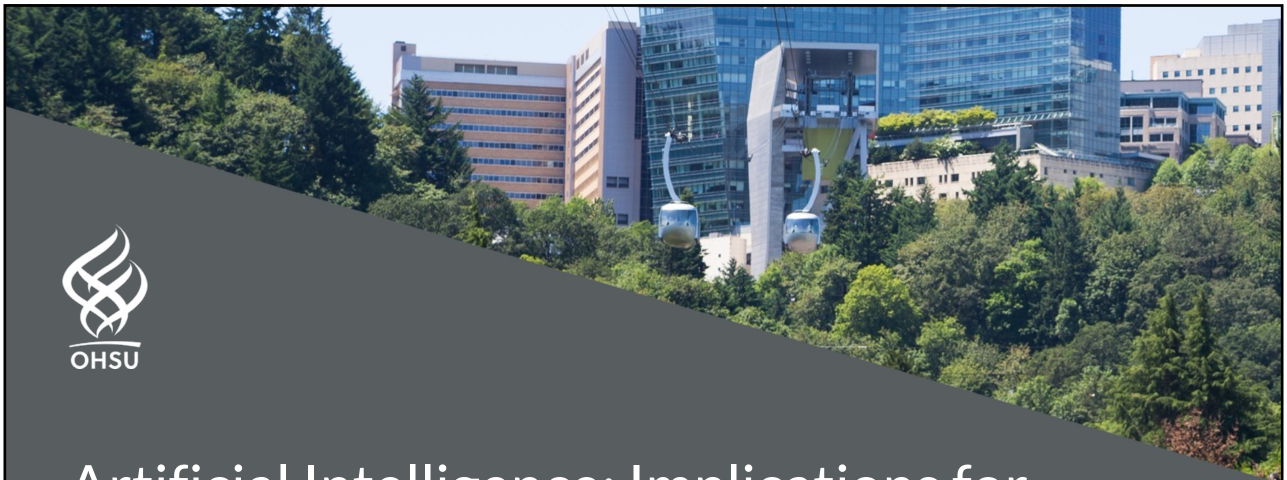
Yin, C., Chen, P.-Y., Yao, B., Wang, D., Caterino, J., Zhang, P., 2024. SepsisLab: Early Sepsis Prediction with Uncertainty Quantification and Active Sensing, in: Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, KDD '24. Association for Computing Machinery, New York, NY, USA, pp. 6158–6168. https://doi.org/10.1145/3637528.3671586

Yu, F., Moehring, A., Banerjee, O., Salz, T., Agarwal, N., Rajpurkar, P., 2024. Heterogeneity and predictors of the effects of AI assistance on radiologists. Nat Med 30, 837–849. https://doi.org/10.1038/s41591-024-02850-w

Zakka, C., Shad, R., Chaurasia, A., Dalal, A.R., Kim, J.L., Moor, M., Fong, R., Phillips, C., Alexander, K., Ashley, E., Boyd, J., Boyd, K., Hirsch, K., Langlotz, C., Lee, R., Melia, J., Nelson, J., Sallam, K., Tullis, S., Vogelsong, M.A., Cunningham, J.P., Hiesinger, W., 2024. Almanac — Retrieval-Augmented Language Models for Clinical Medicine. NEJM AI 1, AIoa2300068. https://doi.org/10.1056/AIoa2300068

Zanon, C., Toniolo, A., Bini, C., Quaia, E., 2023. ChatGPT Goes to The Radiology Department: A Pictorial Review. https://doi.org/10.20944/preprints202312.0714.v1

Zhou, Q., Chen, Z.-H., Cao, Y.-H., Peng, S., 2021. Clinical impact and quality of randomized controlled trials involving interventions evaluating artificial intelligence prediction tools: a systematic review. NPJ Digit Med 4, 154. https://doi.org/10.1038/s41746-021-00524-2



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

William Hersh, MD
Professor
Department of Medical Informatics & Clinical Epidemiology
Oregon Health & Science University
Portland, OR, USA

1

Overview

- About myself and my connection to UCT
- Uses, successes, and limitations of AI in medicine
- Evidence base for AI
- AI in medical education
- AI impacts on how we find and apply information

2

About myself and my connection to UCT

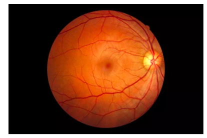
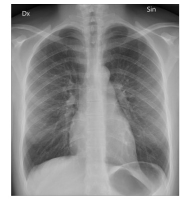
- Personal history
  - Trained in medicine (MD, residency in internal medicine)
  - Completed postdoc in medical informatics
  - At Oregon Health & Science University (OHSU) since 1990
    - Inaugural Chair, Department of Medical Informatics & Clinical Epidemiology
    - Inaugural Director, Biomedical Informatics Graduate Program
    - Research program focused on information retrieval (IR, aka search)
    - Development of educational programs and courses, including online
- Connections to UCT
  - Served on Scientific Advisory Board of H3ABionet (Nicky Mulder, UCT, PI)
  - Currently funded by US NIH in Computational Omics and Biomedical Informatics Program (COBIP – Hocine Bendou, contact PI)

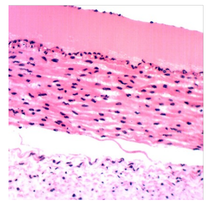
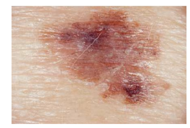
Artificial intelligence (AI) – biomedical perspective

- Biomedical applications played large role in first era of AI (20th century AI)
  - Early success of expert systems
  - Not clinically useful or scalable
- Included in resurgence in second era of AI (21st century AI)
  - Predictive AI – use of data and algorithms to predict some output (e.g., diagnosis, treatment recommendation, prognosis, etc.)
  - Generative AI – generates new output based on prompts (e.g., text, images, etc.)

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)
- 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

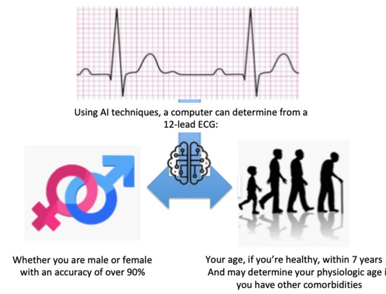
Predictive AI not limited to imaging

- Adverse events in hospitalizations from electronic health record (EHR) data (Rajkomar, 2018)
- Protein folding from amino acid sequences (Jumper, 2021)
- Model based on past ICD-10 codes and lab results to predict future diagnoses in office visits (Mukherjee, 2023)
- 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)
- Voice as a biomarker in Parkinson's Disease, Alzheimer's Disease, cognitive impairment, COVID-19, and others (Idrisoglu, 2023; Bensoussan, 2024)

- The list goes on and on, especially with addition of generative AI…

Also success in "seeing" where humans cannot (Topol, 2022)

- Retinal images
  - Age, biological sex, and cardiovascular risk determination from retinal images (Poplin, 2018)
  - Race (Coyner, 2023)
- Electrocardiograms (ECGs)
  - Age and biological sex determination (Attia, 2019)
  - Chronic kidney disease (Holmstrom, 2023)
- Chest x-rays
  - Race (Gichoya, 2022)
  - Cardiac function and valvular heart diseases (Ueda, 2023)
  - Diabetes (Pyrros, 2023)
  - Correlation with chronological age in healthy cohorts and, for various chronic diseases, difference between estimated age and chronological age (Mitsuyama, 2023)
  - Cardiac risk as accurately as common models, e.g., atherosclerotic cardiovascular disease (ASCVD) (Weiss, 2024)



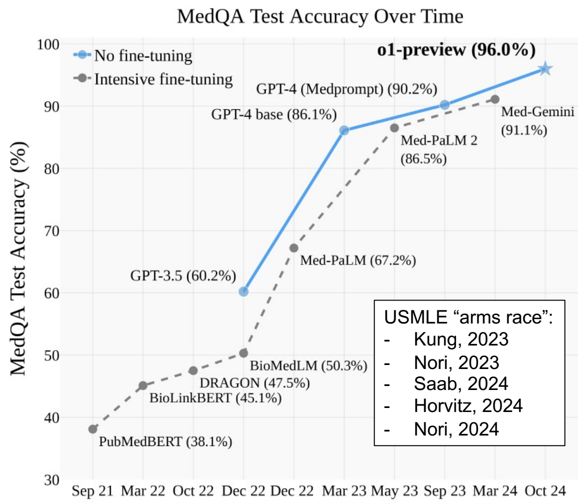
7

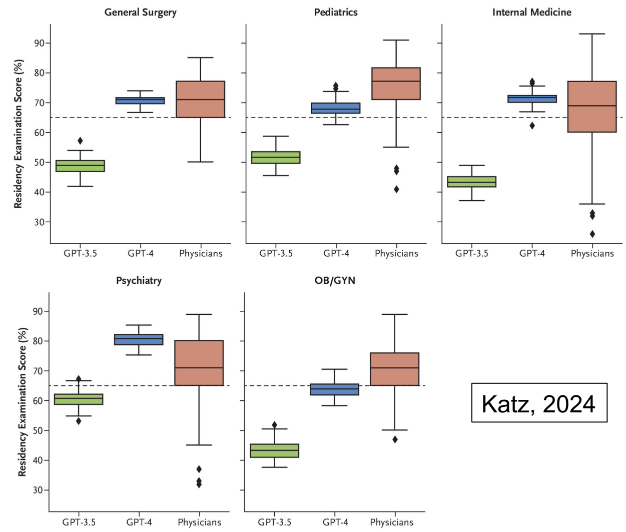
And now, generative AI

- 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 (Omiye, 2024)
  - 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 (Liu, 2023; Meskó, 2023)

8

Some successes of generative AI in medicine – medical board exams





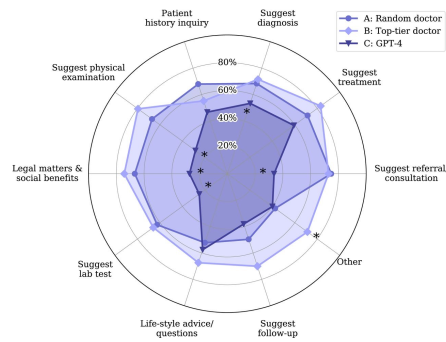
Katz, 2024

Successes of LLMs – solving clinical cases

- Comparable to but not better than expert humans (Kanjee, 2023; Rao, 2023; Chen, 2023)
- In simulated (text-based) objective structured clinical exam (OSCE) format, Google’s Articulate Medical Intelligence Explorer (AMIE) outperformed primary care physicians in text-based dialogue in history-taking, diagnostic accuracy, management reasoning, communication skills, and empathy (Tu, 2024)
- For 20 clinical cases, GPT-4 performed comparable to attending physicians and residents in diagnostic accuracy, correct clinical reasoning, and cannot-miss diagnosis inclusion (Cabral, 2024)
- In 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)

Successes of LLMs – solving cases, mixed results

- Comparison of responses from ChatGPT-4 and physicians for cases from Swedish family medicine specialist examination, scored by blinded reviewers (Arvidsson, 2024)
  - Higher scores for average physicians than GPT-4 or GPT-4o
- Conversational Reasoning Assessment Framework for Testing in Medicine (CRAFT-MD) focuses on natural dialogues, using simulated agents to interact with LLMs in controlled environment (Johri, 2025)
  - Performed worse in "conversational" than "examination-based" settings



Successes of LLMs (cont.)

- Communicating with patients
  - Answering questions in public forums (Sarraju, 2023; Ayers, 2023)
  - Writing letters with comparable or better empathy (Ali, 2023, Ayers, 2023)
- Closing the loop with predictive AI
  - Classifying CXR findings based on previous images and reports (Xu, 2023)
  - Predicting cardiovascular risk comparable to Framingham models (Han, 2023)
  - Designing and validating easily synthesizable antibiotics (Swanson, 2024)
  - Predicting acuity of patients in emergency department (Williams, 2024)

But some downsides to generative AI

- Dictionary.com 2023 word of year: *hallucinate* (Norlen, 2023)
- Fabrication and errors in the bibliographic citations – asked to produce short literature reviews on 42 multidisciplinary topics (Walters, 2023)
  - 55% of GPT-3.5 citations and 18% of GPT-4 citations fabricated
  - 43% of real (non-fabricated) GPT-3.5 citations and 24% of real GPT-4 citations include substantive errors

Downsides to generative AI (cont.)

- 8 clinical questions asked of 4 LLMs recapitulated "harmful, race-based medicine" (Omiye, 2023)
- Automated GPT detectors have mixed results (Sadasivan, 2023; Odri, 2023; Desaire, 2023; Tang, 2024)
  - More likely to classify non-native English writing as AI-generated (Liang, 2023)
  - Humans not able to discern AI writing either (Dell'Acqua, 2023)

And some 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
- ML algorithms, especially generative AI, have large carbon footprints, although details sometimes not known due to lack of company transparency (Kirkpatrick, 2023)
  - One estimate is that electricity consumption of AI request is 10-fold more than Google search (de Vries, 2023)

OHSU

Downsides to AI in general (cont.)

- Variable impacts on different levels of radiologists, leading to automation bias and detrimental effects of incorrect AI (Dratsch, 2023; Yu, 2024)
- Concerns about reproducibility (Ball, 2023)
  - Data bias (especially from EHR – Lewis, 2023; Chin, 2023)
  - Data leakage (Kapoor, 2023)
  - Data drift/shift (Finlayson, 2021; Li, 2024; Nelson, 2025)

OHSU

Will AI help or hinder healthcare?

- Real-world use still modest; most prominent applications include
  - Predictive models, e.g., sepsis (Gorecki, 2024; Yin, 2024)
  - Drafting replies to patient messages (Yan, 2024; Baxter, 2024; Small, 2024; Tai-Seale, 2024)
  - Ambient dictation (Owens, 2024)
- "AI won't replace radiologists, but radiologists who use AI will replace radiologists who don't," (Langlotz, 2019)
  - (Plug in your health profession)

17

AI impacts on evidence, education, and search

- Translational AI (Hersh, 2024)
  - Generating the evidence base is a necessity and opportunity
- Impact on education (Hersh, 2024)
  - How do we learn and assess what we learn?
- Search still matters (Hersh, 2024)
  - In many circumstances, who said what is more important than providing a generated answer


NIH National Library of Medicine
MUSINGS from the MEZZANINE
Innovations in Health Information from the National Library of Medicine
Translational AI: A Necessity and Opportunity for Biomedical Informatics and Data Science
Posted on February 7, 2024 by Guest Author

npj | digital medicine Article
https://doi.org/10.1038/s41746-024-01251-0
Results and implications for generative AI in a large introductory biomedical and health informatics course
William Hersh & Kate Fultz Hollis

Journal of the American Medical Informatics Association, 2024, 1–3
https://doi.org/10.1093/jamia/ocae014
Perspective
AMIA OXFORD
Perspective
Search still matters: information retrieval in the era of generative AI
William Hersh, MD*

18

Translational AI: how do we "show the evidence?"

- From evidence-based medicine (EBM), best evidence for any clinical intervention is from randomized controlled trials (RCTs) or systematic reviews of RCTs
- Although not as easy to carry out as RCTs of drugs or devices (and placebos), AI must demonstrate benefit for patient outcomes and/or healthcare delivery improvement
  - Additional issues for RCTs of AI (Liu, 2020)
- As with drugs and devices, we need to move from "basic science" to "clinical science"
- Not everything can be studied in an RCT and RCTs cannot be done for every last clinical question (Greenhalgh, 2022)

What is the evidence for the benefit of AI?

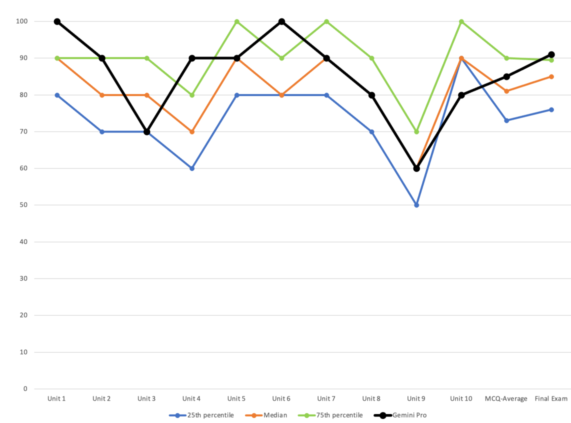
- Many, many papers published about models and simulated use (basic science), including systematic reviews of those papers
- Very few RCTs demonstrating value from real-world use (clinical science) – systematic reviews of RCTs show (Zhou, 2021; Plana, 2022; Han, 2024)
  - Much smaller numbers of RCTs – about 100, depending on how we count
  - 65-82% of RCTs showed positive outcomes
  - Many RCTs showed aspects of "risk of bias"

Perspective from some specific examples

- Computer-aided detection (CADe) of polyps in colonoscopy
  - One of earliest and most widely-studied applications of AI
  - Systematic review shows polyps missed by colonoscopists are discovered, but mostly small and clinically inconsequential (Hassan, 2023)
  - RCT of CADe found no increased detection of advanced neoplasias (Mangas-Sanjuan, 2023)
- RCT to assess whether use of previously validated hospital-acquired venous thromboembolism (HA-VTE) prognostic model, together with pediatric hematologist review, could reduce pediatric inpatient rates of HA-VTE (Walker, 2023)
  - No difference for intervention group randomized to use model
  - Reluctance to use model by primary physicians – used only 26% of time
  - For children in intervention arm, model mostly not used, AI's "Cassandra problem" (Wilson, 2023)?
- More failure of successful models to improve clinical outcomes
  - Hospital readmissions (Donzé, 2023)
  - Chronic kidney disease (Vazquez, 2024)

Impact on education: generative AI is a challenge

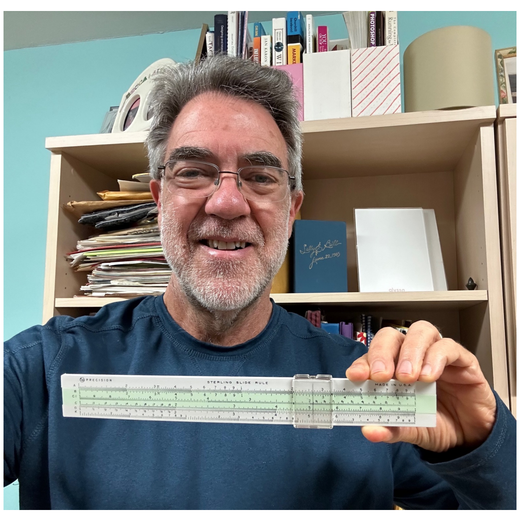
- Well-known, highly subscribed introductory course taught at graduate, continuing education, and medical student levels (Hersh, 2024)
  - Commercial LLMs prompted using interactive Web interface for multiple-choice and final exam questions from 2023 course materials
  - Highest score by Gemini Pro at about 75th percentile for 139 students, other LLMs close behind
- GitHub CoPilot in health informatics programming course (Avramovic, 2024)
  - For problems in SQL and Python, generated solutions worked well for simple tasks but less so for complex ones
  - Some solutions correct but not most efficient approach



Challenges for educators in many disciplines

- May be causing "homework apocalypse" (Mollick, 2023) but also provides opportunity to improve teaching and learning (Mollick, 2024)
- Impact in many disciplines beyond medicine, including
  - Passing college entrance and AP exams (Dubey, 2024)
  - Writing computer programs (Poldrack 2024; Denny, 2024; Johnson, 2024)
  - Creating data science pipelines (Cheng, 2024; Hong, 2024)
  - Writing legal briefs (Choi, 2023)
  - In 5 undergraduate psychology courses, scored above average among students on take-home exams with only 6% detection (Scarfe, 2024)

Now what? Educational cusps in my lifetime



Also critical is education of clinicians and informaticians

- AI should build on competencies in clinical informatics (Hersh, 2014; Hersh 2020; Hersh, 2023)
- Others note
  - Clinicians must be prepared to practice in a world of AI (James, 2022)
  - Medical schools face dual challenges of needing to teach about AI in practice but also adapt to its use by learners and faculty (Cooper, 2023)

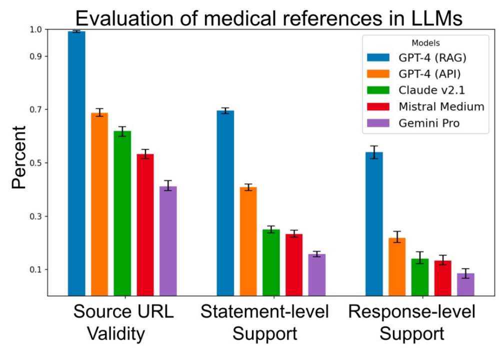
1. Find, search, and apply knowledge-based information to patient care and other clinical tasks
2. Effectively read from, and write to, the electronic health record (EHR) for patient care and other clinical activities
3. Use and guide implementation of clinical decision support (CDS)
4. Provide care 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 though personal health records and patient portals
10. Maintain professionalism in use of information technology tools, including social media
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. Use and critique artificial intelligence (AI) applications in clinical care

Search still matters (Hersh, 2024)

- Generative AI systems such as ChatGPT can be very useful but
  - For some tasks that many of us do, need more than answers, e.g.,
    - Clinical – patient-care questions
    - Research – methods and insights
    - Teaching – synthesizing knowledge for our students
  - Where the information comes from is as important what it says

How well do LLMs cite their sources?

- Best LLM with retrieval-augmented generation (RAG) (GPT-4 in CoPilot) achieved about 70% statement-level support and <50% for others (Wu, 2024)
- Further research in Text Retrieval Conference (TREC) Biomedical Generative Retrieval (BioGen) Track (Gupta, 2024)



Conclusions

- AI will profoundly impact the practice and education of all health professions
- Translational AI is a necessity and opportunity for clinicians, researchers and others
- Educators must develop new approaches to teaching and student assessment in era of generative AI
- Healthcare, informatics, and educational professionals must be competent with AI as much as any other tool in clinical practice
- Generative AI systems must provide attribution for their assertions

Questions?

William Hersh, MD
Professor
Department of Medical Informatics & Clinical Epidemiology
Oregon Health & Science University
Portland, OR, USA
Email
    hersh@ohsu.edu
Web
    http://www.billhersh.info
Blog
    https://informaticsprofessor.blogspot.com/
Textbook
    http://www.informaticsbook.info
What is Informatics?
    http://informatics.health