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

Michigan State University College of Osteopathic Medicine Research Day
April 11, 2024

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

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

AMA: Physicians enthusiastic but cautious about health care AI [WWW Document], 2023. . American Medical Association. URL https://www.ama-assn.org/press-center/press-releases/ama-physicians-enthusiastic-cautious-about-health-care-ai (accessed 1.16.24).

Antaki, F., Touma, S., Milad, D., El-Khoury, J., Duval, R., 2023. Evaluating the Performance of ChatGPT in Ophthalmology: An Analysis of Its Successes and Shortcomings. Ophthalmol Sci 3, 100324. https://doi.org/10.1016/j.xops.2023.100324

Artificial Intelligence in Health, Health Care, and Biomedical Science: An AI Code of Conduct Principles and Commitments Discussion Draft [WWW Document], 2024. . NAM Perspectives. URL https://nam.edu/artificial-intelligence-in-health-health-care-and-biomedical-science-an-ai-code-of-conduct-principles-and-commitments-discussion-draft/ (accessed 4.9.24).

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

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

Barnett, G.O., Cimino, J.J., Hupp, J.A., Hoffer, E.P., 1987. DXplain. An evolving diagnostic decision-support system. JAMA 258, 67–74. https://doi.org/10.1001/jama.258.1.67

Beam, K., Sharma, P., Kumar, B., Wang, C., Brodsky, D., Martin, C.R., Beam, A., 2023. Performance of a Large Language Model on Practice Questions for the Neonatal Board Examination. JAMA Pediatr e232373. https://doi.org/10.1001/jamapediatrics.2023.2373

Benoit, J.R.A., 2023. ChatGPT for Clinical Vignette Generation, Revision, and Evaluation. https://doi.org/10.1101/2023.02.04.23285478

Bhayana, R., Bleakney, R.R., Krishna, S., 2023. GPT-4 in Radiology: Improvements in Advanced Reasoning. Radiology 307, e230987. https://doi.org/10.1148/radiol.230987

Brin, D., Sorin, V., Vaid, A., Soroush, A., Glicksberg, B.S., Charney, A.W., Nadkarni, G., Klang, E., 2023. Comparing ChatGPT and GPT-4 performance in USMLE soft skill assessments. Sci Rep 13, 16492. https://doi.org/10.1038/s41598-023-43436-9

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

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

Clancey, W.J., Shortliffe, E.H., 1984. Readings in medical artificial intelligence: the first decade. Addison-Wesley Longman Publishing Co., Inc., USA.

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

Decker, H., Trang, K., Ramirez, J., Colley, A., Pierce, L., Coleman, M., Bongiovanni, T., Melton, G.B., Wick, E., 2023. Large Language Model-Based Chatbot vs Surgeon-Generated Informed Consent Documentation for Common Procedures. JAMA Netw Open 6, e2336997. https://doi.org/10.1001/jamanetworkopen.2023.36997

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., John, G., Genné, D., Mancinetti, M., Gouveia, A., Méan, M., Bütikofer, L., Aujesky, D., Schnipper, J., 2023. Effects of a Multimodal Transitional Care Intervention in Patients at High Risk of Readmission: The TARGET-READ Randomized Clinical Trial. JAMA Intern Med 183, 658–668. https://doi.org/10.1001/jamainternmed.2023.0791
Dorr, D.A., Adams, L., Embí, P., 2023. Harnessing the Promise of Artificial Intelligence Responsibly. JAMA 329, 1347–1348. https://doi.org/10.1001/jama.2023.2771
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
Edwards, C., 2024. Teaching Transformed. Commun. ACM 67, 12–13. https://doi.org/10.1145/3637208
Embi, P.J., 2021. Algorithmovigilance-Advancing Methods to Analyze and Monitor Artificial Intelligence-Driven Health Care for Effectiveness and Equity. JAMA Netw Open 4, e214622. https://doi.org/10.1001/jamanetworkopen.2021.4622
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
Goodman, R.S., Patrinely, J.R., Stone, C.A., Zimmerman, E., Donald, R.R., Chang, S.S., Berkowitz, S.T., Finn, A.P., Jahangir, E., Scoville, E.A., Reese, T.S., Friedman, D.L., Bastarache, J.A., van der Heijden, Y.F., Wright, J.J., Ye, F., Carter, N., Alexander, M.R., Choe, J.H., Chastain, C.A., Zic, J.A., Horst, S.N., Turker, I., Agarwal, R., Osmundson, E., Idrees, K., Kiernan, C.M., Padmanabhan, C., Bailey, C.E., Schlegel, C.E., Chambless, L.B., Gibson, M.K., Osterman, T.J., Wheless, L.E., Johnson, D.B., 2023. Accuracy and Reliability of Chatbot Responses to Physician Questions. JAMA Netw Open 6, e2336483. https://doi.org/10.1001/jamanetworkopen.2023.36483
Greenes, R., Del Fiol, G. (Eds.), 2023. Clinical Decision Support and Beyond: Progress and Opportunities in Knowledge-Enhanced Health and Healthcare, 3rd edition. ed. Academic Press.
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

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., Topol, E., Rajpurkar, P., 2023. Randomized Controlled Trials Evaluating AI in Clinical Practice: A Scoping Evaluation. https://doi.org/10.1101/2023.09.12.23295381

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

Heneghan, J.A., Walker, S.B., Fawcett, A., Bennett, T.D., Dziorny, A.C., Sanchez-Pinto, L.N., Farris, R.W.D., Winter, M.C., Badke, C., Martin, B., Brown, S.R., McCrory, M.C., Ness-Cochinwala, M., Rogerson, C., Baloglu, O., Harwayne-Gidansky, I., Hudkins, M.R., Kamaleswaran, R., Gangadharan, S., Tripathi, S., Mendonca, E.A., Markovitz, B.P., Mayampurath, A., Spaeder, M.C., Pediatric Data Science and Analytics (PEDAL) subgroup of the Pediatric Acute Lung Injury and Sepsis Investigators (PALISI) Network, 2023. The Pediatric Data Science and Analytics Subgroup of the Pediatric Acute Lung Injury and Sepsis Investigators Network: Use of Supervised Machine Learning Applications in Pediatric Critical Care Medicine Research. Pediatr Crit Care Med. https://doi.org/10.1097/PCC.0000000000003425

Hersh, W., 2024a. Search still matters: information retrieval in the era of generative AI. J Am Med Inform Assoc ocae014. https://doi.org/10.1093/jamia/ocae014

Hersh, W., 2024b. 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., 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.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

Huang, J., Neill, L., Wittbrodt, M., Melnick, D., Klug, M., Thompson, M., Bailitz, J., Loftus, T., Malik, S., Phull, A., Weston, V., Heller, J.A., Etemadi, M., 2023. Generative Artificial

Intelligence for Chest Radiograph Interpretation in the Emergency Department. JAMA Netw Open 6, e2336100. https://doi.org/10.1001/jamanetworkopen.2023.36100
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
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
Khare, Y., 2023. Generative AI vs Predictive AI: What is the Difference? Analytics Vidhya. URL https://www.analyticsvidhya.com/blog/2023/09/generative-ai-vs-predictive-ai/ (accessed 12.12.23).
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).
Kumah-Crystal, Y., Mankowitz, S., Embi, P., Lehmann, C.U., 2023. ChatGPT and the clinical informatics board examination: the end of unproctored maintenance of certification? J Am Med Inform Assoc ocad104. https://doi.org/10.1093/jamia/ocad104
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).
Lancaster, F.W., 1979. Information retrieval systems: Characteristics, testing, and evaluation, 2nd ed edition. ed. John Wiley & Sons, New York.
Langlotz, C.P., 2019. Will Artificial Intelligence Replace Radiologists? Radiol Artif Intell 1, e190058. https://doi.org/10.1148/ryai.2019190058
Lea, A.S., 2023. Digitizing Diagnosis. Johns Hopkins University Press. https://doi.org/10.56021/9781421446813
Ledley, R.S., Lusted, L.B., 1960. The use of electronic computers in medical data processing: aids in diagnosis, current information retrieval, and medical record keeping. IRE Trans Med Electron ME-7, 31–47. https://doi.org/10.1109/iret-me.1960.5008003
Ledley, R.S., Lusted, L.B., 1959. Reasoning foundations of medical diagnosis; symbolic logic, probability, and value theory aid our understanding of how physicians reason. Science 130, 9–21. https://doi.org/10.1126/science.130.3366.9

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

Levine, D.M., Tuwani, R., Kompa, B., Varma, A., Finlayson, S.G., Mehrotra, A., Beam, A., 2023. The Diagnostic and Triage Accuracy of the GPT-3 Artificial Intelligence Model. https://doi.org/10.1101/2023.01.30.23285067

Levkovich, I., Elyoseph, Z., 2023. Identifying depression and its determinants upon initiating treatment: ChatGPT versus primary care physicians. Fam Med Community Health 11, e002391. https://doi.org/10.1136/fmch-2023-002391

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., 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

Liaw, W., Kueper, J.K., Lin, S., Bazemore, A., Kakadiaris, I., 2022. Competencies for the Use of Artificial Intelligence in Primary Care. Ann Fam Med 20, 559–563. https://doi.org/10.1370/afm.2887

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

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

Medical groups taking their time to adopt the right set of AI tools [WWW Document], 2023. . Medical Group Management Association. URL https://www.mgma.com/mgma-stat/medical-groups-taking-their-time-to-adopt-the-right-set-of-ai-tools (accessed 11.22.23).

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

Meyer, A., Benn, R., 2023. Hype Cycle for Healthcare Providers, 2023 [WWW Document]. Gartner. URL https://www.gartner.com/en/documents/4534899 (accessed 1.6.23).

Miller, R.A., Pople, H.E., Myers, J.D., 1982. Internist-1, an experimental computer-based diagnostic consultant for general internal medicine. N Engl J Med 307, 468–476. https://doi.org/10.1056/NEJM198208193070803

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. Co-Intelligence: Living and Working with AI. Portfolio.

Mollick, E.R., Mollick, L., 2023. Using AI to Implement Effective Teaching Strategies in Classrooms: Five Strategies, Including Prompts. https://doi.org/10.2139/ssrn.4391243

Moon, I., LoPiccolo, J., Baca, S.C., Sholl, L.M., Kehl, K.L., Hassett, M.J., Liu, D., Schrag, D., Gusev, A., 2023. Machine learning for genetics-based classification and treatment response prediction in cancer of unknown primary. Nat Med 29, 2057–2067. https://doi.org/10.1038/s41591-023-02482-6

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).

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

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

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., Kannan, A., Chen, K., Vardoulakis, L., Chou, K., Cui, C., Dean, J., 2019. Automatically Charting Symptoms From Patient-Physician Conversations Using Machine Learning. JAMA Intern Med 179, 836–838. https://doi.org/10.1001/jamainternmed.2018.8558
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
Raschka, S., 2023. Understanding Encoder And Decoder LLMs. Ahead of AI. URL https://magazine.sebastianraschka.com/p/understanding-encoder-and-decoder (accessed 9.6.23).
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).
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
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
Schaul, K., Chen, S.Y., Tiku, N., 2023. Inside the secret list of websites that make AI like ChatGPT sound smart [WWW Document]. Washington Post. URL https://www.washingtonpost.com/technology/interactive/2023/ai-chatbot-learning/ (accessed 11.2.23).
Schubert, M.C., Wick, W., Venkataramani, V., 2023. Performance of Large Language Models on a Neurology Board-Style Examination. JAMA Netw Open 6, e2346721. https://doi.org/10.1001/jamanetworkopen.2023.46721
Shah, C., 2022. A Hands-On Introduction to Machine Learning. Cambridge University Press.

Shortliffe, E.H., Davis, R., Axline, S.G., Buchanan, B.G., Green, C.C., Cohen, S.N., 1975. Computer-based consultations in clinical therapeutics: explanation and rule acquisition capabilities of the MYCIN system. Comput Biomed Res 8, 303–320. https://doi.org/10.1016/0010-4809(75)90009-9

Spitale, G., Biller-Andorno, N., Germani, F., 2023. AI model GPT-3 (dis)informs us better than humans. Sci Adv 9, eadh1850. https://doi.org/10.1126/sciadv.adh1850

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.

Triola, M.M., Reinstein, I., Marin, M., Gillespie, C., Abramson, S., Grossman, R.I., Rivera, R., 2023. Artificial Intelligence Screening of Medical School Applications: Development and Validation of a Machine-Learning Algorithm. Acad Med 98, 1036–1043. https://doi.org/10.1097/ACM.0000000000005202

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

Wang, S., Scells, H., Koopman, B., Zuccon, G., 2023. Can ChatGPT Write a Good Boolean Query for Systematic Review Literature Search? https://doi.org/10.48550/arXiv.2302.03495

Warner, H.R., Toronto, A.F., Veasey, L.G., Stephenson, R., 1961. A mathematical approach to medical diagnosis. Application to congenital heart disease. JAMA 177, 177–183. https://doi.org/10.1001/jama.1961.03040290005002

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

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

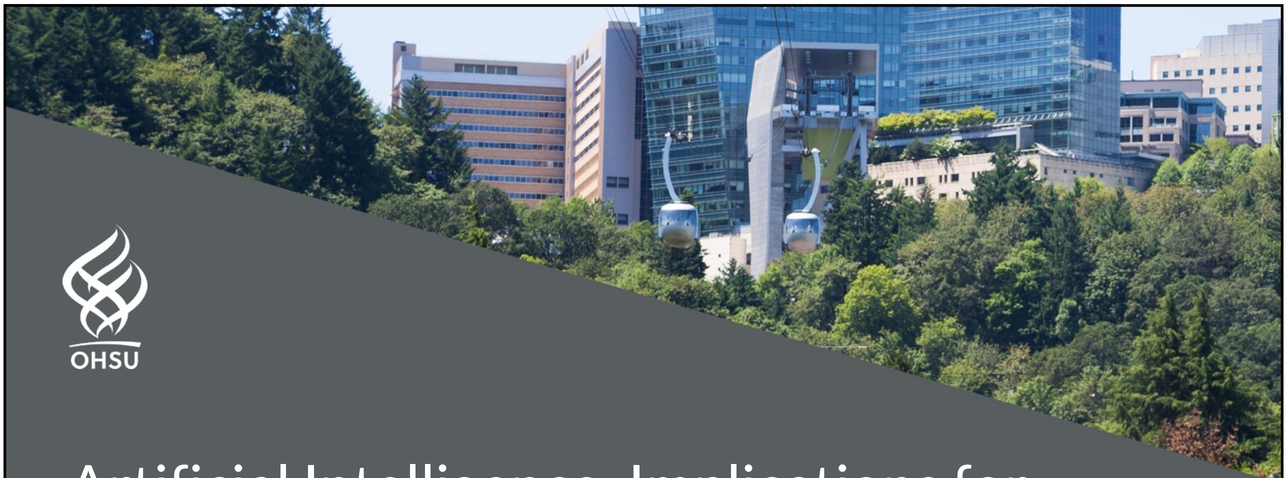
Xu, S., Yang, L., Kelly, C., Sieniek, M., Kohlberger, T., Ma, M., Weng, W.-H., Kiraly, A., Kazemzadeh, S., Melamed, Z., Park, J., Strachan, P., Liu, Y., Lau, C., Singh, P., Chen, C., Etemadi, M., Kalidindi, S.R., Matias, Y., Chou, K., Corrado, G.S., Shetty, S., Tse, D., Prabhakara, S., Golden, D., Pilgrim, 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).

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. 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



OHSU

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

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

1

Objectives and disclosures

- After this talk, you will to be able to
  - Define the major types of AI and their applications, successes, and limitations
  - Discuss the evidence base for AI, its limitations, and how to improve it
  - Follow a pathway for application of AI and its role in clinical practice, research, and education
- Disclosures
  - None

OHSU

2

Artificial intelligence (AI) defined

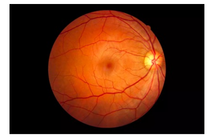
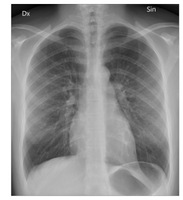
- AI – "information systems and algorithms capable of performing tasks associated with human intelligence" (Rajpurkar, 2022)
- Some classify AI into two broad categories (Khare, 2023)
  - Predictive AI – use of data and algorithms to predict some output (e.g., diagnosis, treatment recommendation, prognosis, etc.)
  - Generative AI – generates new output based on prompts (e.g., text, images, etc.)
- A large part of modern success of AI due to machine learning (ML) – "computer programs that learn without being explicitly programmed" (McCarthy, 1990, attributed to Samuel, 1959; Shah, 2023)
  - Most success with deep learning, based on many-layered neural networks

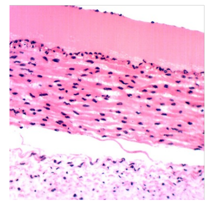
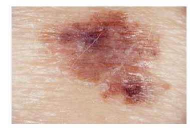
History of AI – first era in mid-20th century

- Earliest paper related to AI and biomedical informatics attributed to Ledley and Lusted (1959, 1960) aiming to model physician reasoning through symbolic logic and probability
- Warner (1961) developed mathematical model for diagnosing congenital heart disease
- In 1960s-1970s, emergence of "expert systems" – computer programs aiming to mimic human expertise (historical overview – Lea, 2023)
  - Rule-based systems – PhD dissertation of Shortliffe (1975) and subsequent work (Clancey, 1984)
  - Disease profiles and scoring algorithms – INTERNIST-1 (Miller, 1982) and DxPlain (Barnett, 1987)
- Limited by approach of manual construction and maintenance of knowledge
  - Not scalable or sustainable
  - Led to "AI winter" between 1990-2010
  - Main remnant is clinical decision support (CDS) for electronic health records (EHRs) that emerged in 1990s for electronic health records (Greenes, 2023)

Re-emergence of AI in 21st century

- "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

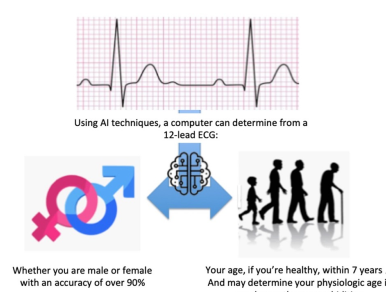
- Adverse events in hospitalizations from EHR data (Rajkomar, 2018)
- Generating clinical notes from patient and physician verbal interaction (Rajkomar, 2019)
- Protein folding from amino acid sequences (Jumper, 2021)
- Model based on past ICD-10 codes and lab results to predict future diagnoses in office visits (Mukherjee, 2023)
- Semantic reconstruction of continuous language from fMRI brain recordings (Tang, 2023)
- Map chemicals to odors perceived by humans (Lee, 2023)

Predictive AI (cont.)

- Assist in detection of inappropriate 16% duplicate images in papers in toxicology journal (David, 2023)
- Perform initial screening of medical school applicants comparable to faculty (Triola, 2023)
- Next-generation sequencing (NGS) data able to predict cancer of unknown primary with high confidence for 41% of tumors, leading to improved survival for those patients (Moon, 2023)
- Predict Alzheimer's Disease from EHR data up to 7 years before diagnosis (Tang, 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)



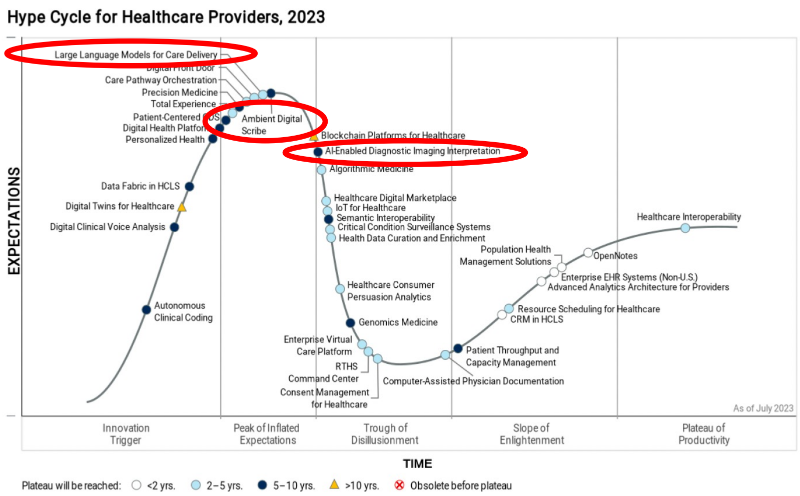
And now, "generative AI"

- 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 (Raschka, 2023)
  - Fine-tuned with reinforcement learning from human feedback (RLHF) (Lambert, 2022)
  - Activated by (and importance of) prompting (Liu, 2023; Meskó, 2023)

Generative AI is more than ChatGPT

- Adding generative AI to search, including retrieval-augmented generation (RAG) (King, 2023)
  - CoPilot – GPT-4 integrated into Microsoft Bing
  - Google – Bard and now Gemini
- Many products adding generative AI, e.g., Microsoft Office, Adobe Acrobat, etc.
- "Small" language models – Phi-2, Mistral, etc.
  - Clinically-oriented models, e.g., Almanac (Zakka, 2024)
- Many roles and challenges in academia, business, etc. (Mollick, 2024)

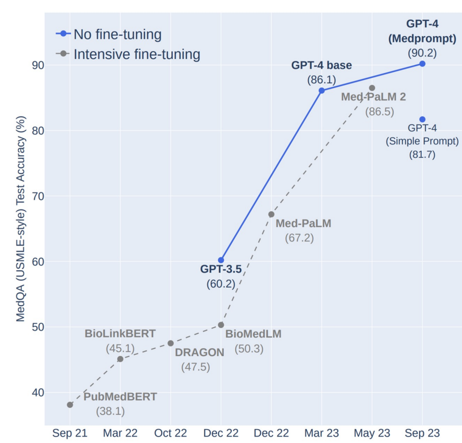
Generative AI at peak of inflated expectations (Meyer, 2023)



Gartner

Results of ChatGPT and other LLMs

- Medical board exam questions
  - USMLE "arms race," starting with (Kung, 2023)
    - Now best with GPT-4 and specific types of prompting (Nori, 2023)
    - Even on "soft skills" (e.g., communication skills, ethics, empathy, and professionalism) questions (Brin, 2023)
  - Passing level on most board exam questions (clinical informatics – Kumah-Crystal, 2023; radiology – Bhayana, 2023; neurology – Schubert, 2023) but not others (neonatology – Beam, 2023, used only GPT-3.5)
- Answering questions
  - Vary by subject domain and type, but sometimes wrong and/or incomplete (e.g., Antaki, 2023; Chen, 2023; Goodman, 2023)
- Solving clinical cases
  - Comparable to but not better than expert humans (e.g., Levine, 2023; Kanjee, 2023; Rao, 2023; Benoit, 2023; Levkovich, 2023)



Results of LLMs (cont.)

- In simulated (text-based) objective structured clinical exam (OSCE) format, 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)

Results 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)
  - Generating surgical consent forms better than surgeons (Decker, 2023)
- Closing the loop with predictive AI
  - Classifying CXR findings based on previous images and reports (Xu, 2023)
  - Generating CXR reports from new images in ED from prior images and reports (Huang, 2023)
  - Predicting cardiovascular risk comparable to Framingham models (Han, 2023)
  - Designing and validating easily synthesizable and structurally novel antibiotics (Swanson, 2024)

But there are some downsides to generative AI

- Dictionary.com 2023 word of year: hallucinate
  - https://content.dictionary.com/word-of-the-year-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
- LLMs reflect content (and bias) of text used for training (Schaul, 2023)

Downsides to generative AI (cont.)

- 8 clinical questions asked of 4 LLMs recapitulated "harmful, race-based medicine" (Omiye, 2023)
- Equally compelling disinformation – humans cannot distinguish between true and false tweets generated by GPT-3 and written by real Twitter users (Spitale, 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

17

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)

18

Will AI help or hinder healthcare?

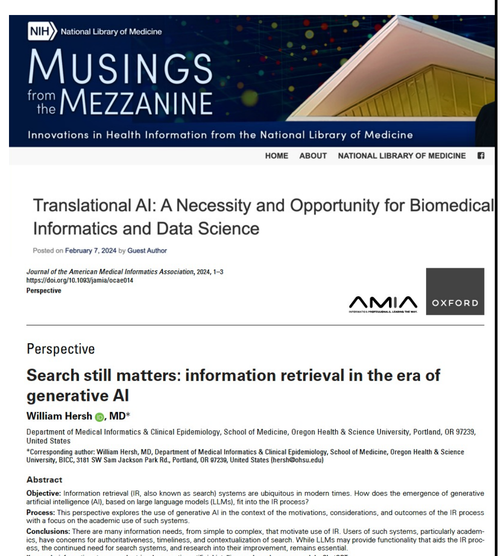
- Real-world use still modest
  - As of Sept 2023, only 21% of medical groups using AI applications in practice (MGMA, 2023)
  - EHR usability, patient communications, and billing outrank AI as top tech priorities among medical groups (MGMA, 2023)
  - AI tools used by only 38% of physicians (AMA, 2023)
- "AI won't replace radiologists, but radiologists who use AI will replace radiologists who don't," (Langlotz, 2019)
  - (Plug in your health profession)

AI Implications 19

19

What do we need for AI applications to make it to the plateau of productivity?

- Translational AI (Hersh, 2024)
  - Show us the evidence
  - How do we learn about it
- 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

HOME ABOUT NATIONAL LIBRARY OF MEDICINE

Translational AI: A Necessity and Opportunity for Biomedical Informatics and Data Science

Posted on February 7, 2024 by Guest Author

Journal of the American Medical Informatics Association, 2024, 1–3
https://doi.org/10.1093/jamia/ocae014
Perspective

AMIA INFORMATICS PROFESSIONALS. LEADING THE WAY. OXFORD

Perspective

Search still matters: information retrieval in the era of generative AI

William Hersh, MD*

Department of Medical Informatics & Clinical Epidemiology, School of Medicine, Oregon Health & Science University, Portland, OR 97239, United States

*Corresponding author: William Hersh, MD, Department of Medical Informatics & Clinical Epidemiology, School of Medicine, Oregon Health & Science University, BICC, 3181 SW Sam Jackson Park Rd., Portland, OR 97239, United States (hersh@ohsu.edu)

Abstract

Objective: Information retrieval (IR, also known as search) systems are ubiquitous in modern times. How does the emergence of generative artificial intelligence (AI), based on large language models (LLMs), fit into the IR process?

Process: This perspective explores the use of generative AI in the context of the motivations, considerations, and outcomes of the IR process with a focus on the academic use of such systems.

Conclusions: There are many information needs, from simple to complex, that motivate use of IR. Users of such systems, particularly academics, have concerns for authoritativeness, timeliness, and contextualization of search. While LLMs may provide functionality that aids the IR process, the continued need for search systems, and research into their improvement, remains essential.

Key words: information storage and retrieval; generative artificial intelligence; large language models; ChatGPT.

AI Implications 20

20

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 so far?

- 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, 2023)
  - Much smaller numbers of RCTs – about 100, depending on how we count
  - 65-82% of RCT showed positive outcomes
  - Many RCTs showed aspects of "risk of bias"

Learning from some specific examples

- Computer-aided detection (CADe) of polyps in colonoscopy
  - One of earliest and most widely-studied applications of AI
  - Recent 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)
- 30-day hospital readmissions
  - After implementation of CMS penalty, proliferation of highly accurate predictive models published in mid-2010s
  - Recent RCT showed use of high-quality model and implementation of program around it did not reduce readmissions (Donzé, 2023)

Examples (cont.)

- 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)?
- Pragmatic trial of predictive algorithm and practice facilitators in outpatient practices did not reduce hospitalization rate in chronic kidney disease over 1 year compared with usual care (Vazquez, 2024)

How do we get to "translational AI?"

- In pediatric critical care ML research, "the literature demonstrates incomplete reporting, absence of external validation, and infrequent clinical implementation" (Heneghan, 2023)
- Singh, X/Twitter, Feb 8 2024: "Researched models aren't implemented. Implemented models aren't researched."
- Postmarket surveillance, e.g., algorithmovigilance (Embi, 2021)
- Responsible use (Dorr, 2023) and code of conduct (Adams, NAM, 2024) for AI
- Building the evidence base (Hersh, 2024)

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)
- AI-specific competency frameworks (Russell, 2023; Liaw, 2023)

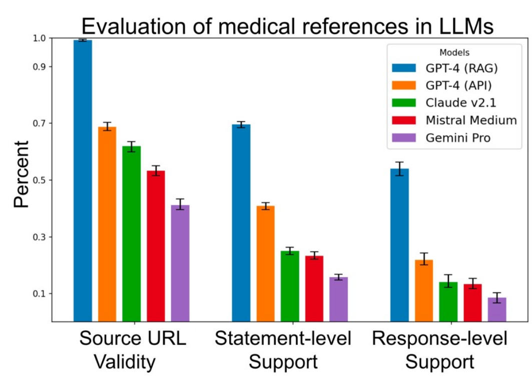
1. Find, search, and apply knowledge-based information to patient care and other clinical tasks
2. Effectively read from, and write to, the electronic health record (EHR) for patient care and other clinical activities
3. Use and guide implementation of clinical decision support (CDS)
4. Provide care 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 are cool and fun, 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
- Information retrieval (IR) systems "do not inform user about a subject; indicate the existence (or nonexistence) and whereabouts of documents related to an information request" (Lancaster, 1978)

Search in the era of generative AI

- Another adage of EBM?
  - Gen AI for background questions
  - Search and critical appraisal for foreground questions
- Retrieval-augmented generation (RAG) for improving Gen AI but do we need "generation-augmented retrieval" for LLMs to aid search?
  - Evidence modest so far, e.g., using ChatGPT for generating Boolean queries did not improve search results (Wang, 2023)
  - Best LLM with RAG (GPT-4 in CoPilot) achieved about 70% statement-level support and <50% for others (Wu, 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
- 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