Artificial Intelligence: Implications for Health Professions Education

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

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

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).

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

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

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
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.R., Mollick, L., 2023. Using AI to Implement Effective Teaching Strategies in Classrooms: Five Strategies, Including Prompts. https://doi.org/10.2139/ssrn.4391243
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., 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
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
Preiksaitis, C., Rose, C., 2023. Opportunities, Challenges, and Future Directions of Generative Artificial Intelligence in Medical Education: Scoping Review. JMIR Med Educ 9, e48785. https://doi.org/10.2196/48785

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

Sangha, V., Nargesi, A.A., Dhingra, L.S., Khunte, A., Mortazavi, B.J., Ribeiro, A.H., Banina, E., Adeola, O., Garg, N., Brandt, C.A., Miller, E.J., Ribeiro, A.L.J., Velazquez, E.J., Giatti, L., Barreto, S.M., Foppa, M., Yuan, N., Ouyang, D., Krumholz, H.M., Khera, R., 2023. Detection of Left Ventricular Systolic Dysfunction From Electrocardiographic Images. Circulation. https://doi.org/10.1161/CIRCULATIONAHA.122.062646

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

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

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.

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

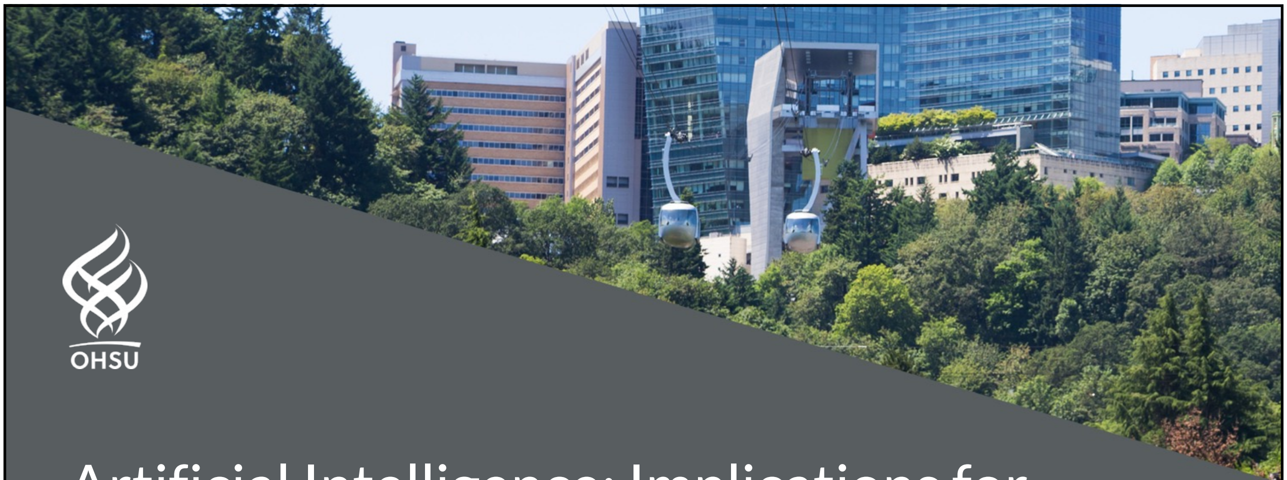
Walters, W.H., Wilder, E.I., 2023. Fabrication and errors in the bibliographic citations generated by ChatGPT. Sci Rep 13, 14045. https://doi.org/10.1038/s41598-023-41032-5

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

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

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).
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 Health Professions Education

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

1

Objectives

- After this talk, you will be able to
  - Define major types of AI and their successes and limitations
  - Describe the impact of AI in health professions education
  - Discuss best practices going forward for using AI in health professions education

- Disclosure
  - None

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

OHSU

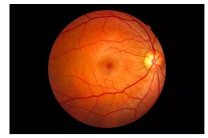
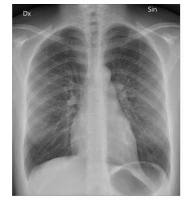
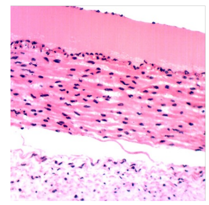
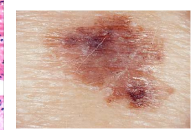
History of AI – first era in mid-20th century

- 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)

OHSU

Re-emergence of AI in 21st century

- "Predictive AI" driven by advances in machine learning, 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

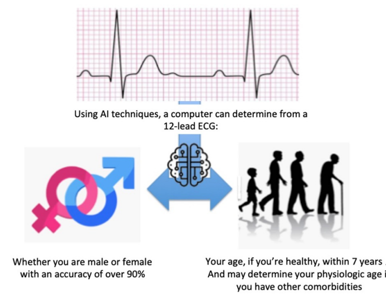





Predictive AI not limited to imaging

- 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)
- ML 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)

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)
  - Left ventricular systolic dysfunction from ECG images (Sangha, 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)



And now, "generative AI"

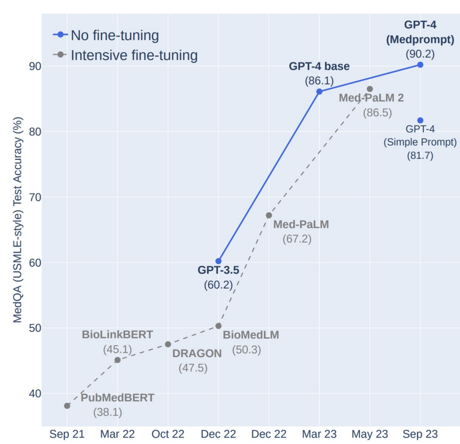
- Introduction of ChatGPT on November 30, 2022 brought new type of AI into focus: generative AI
- Based on large language models (LLMs) processed by deep neural networks using large amounts of training data and tuned for specific tasks
  - Trained on massive amounts of text and other content, e.g., large 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, e.g.,
  - CoPilot – version of GPT-4 integrated into Microsoft Bing
  - Google – with Bard/PaLM
- Plugins enhance functionality
  - BrowserPilot – access live Web sites
  - ScholarAI – search PubMed and other research databases
  - SmartSlides – generate (short) Powerpoint presentations
  - SciSummary – summarize scientific papers
- Retrieval augmented generation (RAG) – merging generative AI and information retrieval (IR)
- "Small" language models – Phi-2, Mistral, etc.

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 ChatGPT and other LLMs (cont.)

- Communicating with patients
  - Answer questions in public forums (Sarraju, 2023; Ayers, 2023)
  - Write letters with comparable or better empathy (Ali, 2023, Ayers, 2023)
  - Generating surgical consent forms better than surgeons (Decker, 2023)
- Use of predictive AI (closing the AI loop)
  - 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)

OHSU

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
- 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)
- LLMs reflect content (and bias) of text used for training (Schaul, 2023)
- Automated GPT detectors have mixed results (Sadasivan, 2023; Odri, 2023; Desaire, 2023)
  - More likely to classify non-native English writing as AI-generated (Liang, 2023)
  - Humans not able to discern AI writing either (Dell'Acqua, 2023)

OHSU

And some downsides to AI in general

- After clinical models deployed, performance may decline due to actual real-world use (Vaid, 2023; Palmer, 2023)
- Inexperienced, moderately experienced, and very experienced radiologists reading mammograms are prone to different types of automation bias when being supported by an AI-based system (Dratsch, 2023)
- Implementing diabetic retinopathy screening in rural Thailand and India found challenges related to equipment operation, workflows, and image quality, and need for training and attention to human factors (Widner, 2023)
- Concerns about reproducibility crisis (Ball, 2023) from problems such as data bias (especially from EHR – Lewis, 2023) and data leakage (Kapoor, 2023)

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)
- Evidence base still small – systematic reviews of randomized clinical trials (RCTs) of predictive AI systems (Zhou, 2021; Plana, 2022; Han, 2023) show
  - Small numbers of RCTs (dozens) – especially relative to predictive model papers (thousands)
  - Suboptimal methodologies leading to risk of bias
  - Mix of positive/negative results
- "AI won't replace radiologists, but radiologists who use AI will replace radiologists who don't," (Langlotz, 2019)
  - (Plug in your health profession)

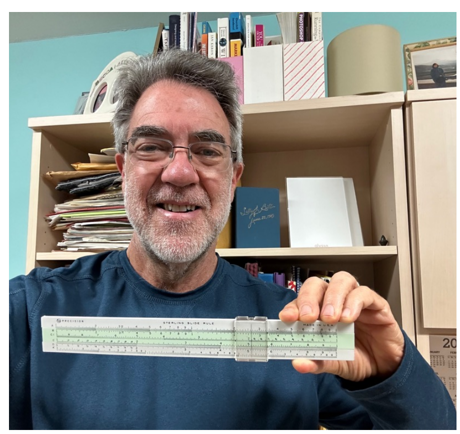
AI and health professions education

- Before generative AI there was recognition of need for competencies in clinical informatics for medical education (Hersh, 2014; Hersh 2020; Hersh, 2023)
- Clinicians must be prepared to practice in a world of AI (James, 2022)
- New AI-competency frameworks highlight what health professions students must master (Russell, 2023; Liaw, 2023)
  - Mostly physician-based but applies to all health professions

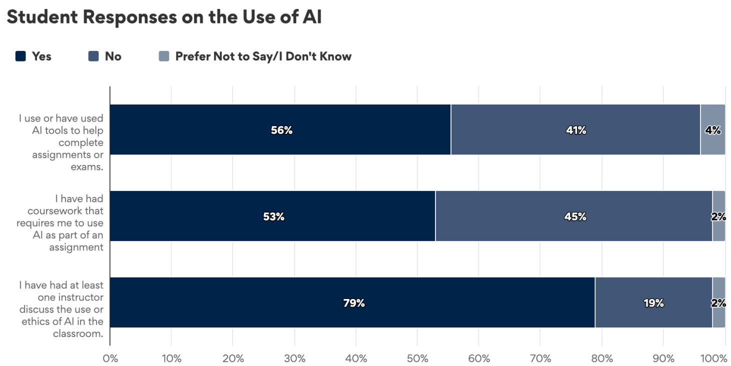
1. Find, search, and apply knowledge-based information to patient care and other clinical tasks
2. Effectively read from, and write to, the electronic health record (EHR) for patient care and other clinical activities
3. Use and guide implementation of clinical decision support (CDS)
4. Provide care 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

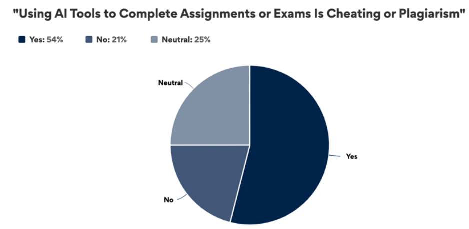
Generative AI and teaching

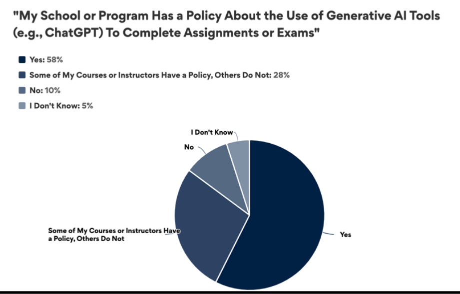
- A transformative cusp?
  - Anyone under 30 know what I am holding?
- Much in our courses can be done by generative AI
- My approach (so far)
  - Gen AI policy from OHSU Provost generally and my course specifically
  - Allowing explicit use in certain assignments



Use already widespread – (non-scientific) survey of 1000 college students (Nam, 2023)







Faculty should be "assigning AI" (Mollick, 2023)

AI USE	ROLE	PEDAGOGICAL BENEFIT	PEDAGOGICAL RISK
MENTOR	Providing feedback	Frequent feedback improves learning outcomes, even if all advice is not taken.	Not critically examining feedback, which may contain errors.
TUTOR	Direct instruction	Personalized direct instruction is very effective.	Uneven knowledge base of AI. Serious confabulation risks.
COACH	Prompt metacognition	Opportunities for reflection and regulation, which improve learning outcomes.	Tone or style of coaching may not match student. Risks of incorrect advice.
TEAMMATE	Increase team performance	Provide alternate viewpoints, help learning teams function better.	Confabulation and errors. "Personality" conflicts with other team members.
STUDENT	Receive explanations	Teaching others is a powerful learning technique.	Confabulation and argumentation may derail the benefits of teaching.
SIMULATOR	Deliberate practice	Practicing and applying knowledge aids transfer.	Inappropriate fidelity.
TOOL	Accomplish tasks	Helps students accomplish more within the same time frame.	Outsourcing thinking, rather than work.

Risks:
- Confabulation
- Bias – from training content
- Privacy – policies not always clear
- Instructional – student over-reliance

Personal experience so far with generative AI in teaching

- Must address head-on
- Policy
  - Explicit on proper use and when discouraged or prohibited
  - Concerns for when benefits student vs. undermines learning
- Use in an assignment
  - Student term paper can be a conversation with GenAI/LLM about several course topics with dialogue and critique of output

Need policy for generative AI: mine for introductory course
https://dmice.ohsu.edu/hersh/introcourse-generativeAI-policy.html

OHSU *Introduction to Biomedical & Health Informatics* Course Policy for Use of ChatGPT and Generative AI

William Hersh, MD
Professor
Department of Medical Informatics & Clinical Epidemiology
School of Medicine
Oregon Health & Science University

This page reflects course policy for the Oregon Health & Science University (OHSU) course that I teach called **Introduction to Biomedical & Health Informatics**. I teach versions of this course in several OHSU programs, including:

- Biomedical Informatics Graduate program - BMI 510/610 - Introduction to Biomedical & Health Informatics
- AMIA 10x10 ("ten by ten") course - OHSU-AMIA 10x10 course
- MD curriculum course, MINF 705B/709A

ChatGPT and generative AI systems based on large language models (LLMs) can be a useful tool for learning all kinds of topics, including in biomedical and health informatics. These tools should not, however, be used to substitute one's own knowledge. Students can "converse" with ChatGPT or generative AI systems to get ideas for answers to questions, but the final responses to discussion forums, quiz and test questions, and the term paper, should reflect their own thinking, judgment, and language.

This policy is derived from the overall OHSU policy for academic integrity, including the use of AI. The OHSU Biomedical Informatics Graduate Program is developing a general policy for use of generative AI in courses, but in the meantime, I am adopting the following guidelines for course activities:

- **Discussion forums** - the purpose of the discussion forums is for students to discuss issues that elaborate on unit course materials. Individual forum postings are not graded, although a component of the course grade is based on participation in the forums, comparable to what used to be participation in live classrooms. While students can "converse" with generative AI to get ideas for responses to forum questions, what is actually posted in the forum by students should represent their own ideas and thought processes.
- **Homework self-assessment** - students can ask generative AI about topics mentioned in the multiple-choice questions but are expected to answer the questions based on their own knowledge of materials covered in the lectures.
- **Term paper/project** - students can ask generative AI for help in brainstorming about their term paper. Generative AI systems do not write long papers, and their output tends to focus on generalities and may be prone to confabulation, especially in generating references. The 10-15 page term paper should have a focus on a specific topic, and delve into it with coherent discussion and ample references, including recent ones, as outlined in the course syllabus.
- **Final exam** - students must not access generative AI during the final exam.

If you are a student and have a question on whether use of generative AI is appropriate, please reach out directly to me (email is best for initial contact).

As a guiding principle, we expect and require that all work submitted be the student's own, original work. When considering using such a generative AI tool, students should ask themselves: Will the tool's output be something I will be turning in directly? In general, students may use such tools as a source of information, but not to produce output that they intend to turn in or as a replacement for a traditional cited reference.

Most ethical and conduct policies in our informatics educational programs, and in the work we subsequently do as professionals, are enforced through an **honor code**. We recognize we cannot police all inappropriate use of AI or other activities. We hope that students will find ways to use LLMs to enhance their learning but not substitute for or become dependent on it.

Last update: January 8, 2024

Opportunities, challenges, and directions for Gen AI in medical education (Preiksaitis, 2023)

- Theme 1: Test performance and preparation
  - Licensing examination performance
  - Specialty exam performance
  - Undergraduate exam performance
  - Improving understanding
  - Self-directed learning
  - Exam preparation/practice
- Theme 2: Novel learning strategies
  - Development of personalized learning plans
  - Creation of learning materials
  - Providing feedback
  - Communication skills training
  - Clinical image generation for learning
  - Medical humanities exercises
- Theme 3: Writing and research assistance
  - Assisting non-native speakers
  - Translations
  - Literature review/summarization
  - Fabricated references/hallucinations
- Theme 4: Academic integrity concerns
  - Cheating on examinations
  - Reduced effectiveness of learning exercises
  - Technological plagiarism
  - Need for policy development
  - Guidance for disclosure and transparency
- Theme 5: Accuracy and dependability
  - Reliance on training data
  - Lack of up-to-date information
  - Hallucination
  - Confidence expressed by models
  - Misinformation propagation
  - Limited accuracy in specific areas
  - Need for further training in limitations
- Theme 6: Potential detriments to learning
  - Overdependence
  - Challenges with assessment
  - Propagating inaccurate information
  - Inequities in access

Conclusions

- AI will profoundly impact the practice and education of all health professions
- Healthcare professionals must be competent with AI as much as any other tool in their clinical practice
- Educators and students must adapt to generative AI for writing, examination, and other pedagogic tasks

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