Artificial Intelligence: Implications for Informatics Education

December 14, 2023

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/
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

Al-Zaiti, S.S., Martin-Gill, C., Zègre-Hemsey, J.K., Bouzid, Z., Faramand, Z., Alrawashdeh, M.O., Gregg, R.E., Helman, S., Riek, N.T., Kraevsky-Phillips, K., Clermont, G., Akcakaya, M., Sereika, S.M., Van Dam, P., Smith, S.W., Birnbaum, Y., Saba, S., Sejdic, E., Callaway, C.W., 2023. Machine learning for ECG diagnosis and risk stratification of occlusion myocardial infarction. Nat Med. https://doi.org/10.1038/s41591-023-02396-3

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

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

Chung, H.W., Hou, L., Longpre, S., Zoph, B., Tay, Y., Fedus, W., Li, Y., Wang, X., Dehghani, M., Brahma, S., Webson, A., Gu, S.S., Dai, Z., Suzgun, M., Chen, X., Chowdhery, A., Castro-Ros, A., Pellat, M., Robinson, K., Valter, D., Narang, S., Mishra, G., Yu, A., Zhao, V., Huang, Y., Dai, A., Yu, H., Petrov, S., Chi, E.H., Dean, J., Devlin, J., Roberts, A., Zhou, D., Le, Q.V., Wei, J., 2022. Scaling Instruction-Finetuned Language Models. https://doi.org/10.48550/arXiv.2210.11416

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

Clune, M.W., 2023. AI Means Professors Need to Raise Their Grading Standards [WWW Document]. The Chronicle of Higher Education. URL https://www.chronicle.com/article/ai-means-professors-need-to-raise-their-grading-standards (accessed 9.13.23).

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

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

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

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.

Gregorich, M., Kammer, M., Heinzel, A., Böger, C., Eckardt, K.-U., Heerspink, H.L., Jung, B., Mayer, G., Meiselbach, H., Schmid, M., Schultheiss, U.T., Heinze, G., Oberbauer, R., BEAt-DKD Consortium, 2023. Development and Validation of a Prediction Model for Future Estimated Glomerular Filtration Rate in People With Type 2 Diabetes and Chronic Kidney Disease. JAMA Netw Open 6, e231870. https://doi.org/10.1001/jamanetworkopen.2023.1870

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

Heaven, W.D., 2023. ChatGPT is going to change education, not destroy it [WWW Document]. MIT Technology Review. URL https://www.technologyreview.com/2023/04/06/1071059/chatgpt-change-not-destroy-education-openai/ (accessed 7.17.23).

Hersh, W., 2020. Information Retrieval: A Biomedical and Health Perspective, 4th ed, Health Informatics. Springer International Publishing.

Hersh, W., Pentecost, J., Hickam, D., 1996. A task-oriented approach to information retrieval evaluation. Journal of the American Society for Information Science 47, 50–56. https://doi.org/10.1002/(SICI)1097-4571(199601)47:1<50::AID-ASI5>3.0.CO;2-1

Hersh, W.R., Crabtree, M.K., Hickam, D.H., Sacherek, L., Friedman, C.P., Tidmarsh, P., Mosbaek, C., Kraemer, D., 2002. Factors associated with success in searching MEDLINE and applying evidence to answer clinical questions. J Am Med Inform Assoc 9, 283–293. https://doi.org/10.1197/jamia.m0996

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

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

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., 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

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

Lou, S., Du, F., Song, W., Xia, Y., Yue, X., Yang, D., Cui, B., Liu, Y., Han, P., 2023. Artificial intelligence for colorectal neoplasia detection during colonoscopy: a systematic review and meta-analysis of randomized clinical trials. eClinicalMedicine 66. https://doi.org/10.1016/j.eclinm.2023.102341

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

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. Assigning AI: Seven Approaches for Students, with Prompts. https://doi.org/10.2139/ssrn.4475995

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).

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

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).

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

Sanghavi, S.F., 2023. Modeling Future Estimated Glomerular Filtration Rate in Patients With Diabetes-Are There Risks to Early Risk Stratification? JAMA Netw Open 6, e238652. https://doi.org/10.1001/jamanetworkopen.2023.8652

Sarraju, A., Bruemmer, D., Van Iterson, E., Cho, L., Rodriguez, F., Laffin, L., 2023. Appropriateness of Cardiovascular Disease Prevention Recommendations Obtained From a Popular Online Chat-Based Artificial Intelligence Model. JAMA. https://doi.org/10.1001/jama.2023.1044

Shah, 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

Shung, D.L., 2023. From Tool to Team Member: A Second Set of Eyes for Polyp Detection. Ann Intern Med. https://doi.org/10.7326/M23-2022

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., 2019. Deep Medicine: How Artificial Intelligence Can Make Healthcare Human Again, Illustrated Edition. ed. Basic Books, New York.

Topol, E.J., 2023. As artificial intelligence goes multimodal, medical applications multiply. Science 381, adk6139. https://doi.org/10.1126/science.adk6139

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

Wachter, R.M., 2023. The Disappointing Impact of Interventions to Prevent Hospital Readmissions. JAMA Intern Med 183, 668–669. https://doi.org/10.1001/jamainternmed.2023.0804

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 Informatics 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
  - Discuss key issues for AI in informatics education
  - Describe opportunities in AI for informatics learners
- Note
  - This talk is a major revision of a talk given earlier this year about AI for health professions education
  - Aim to leave last 10-15 minutes for discussion

2

Financial disclosures

- I pay $20/month to use ChatGPT 4

Definitions related to artificial intelligence (AI)

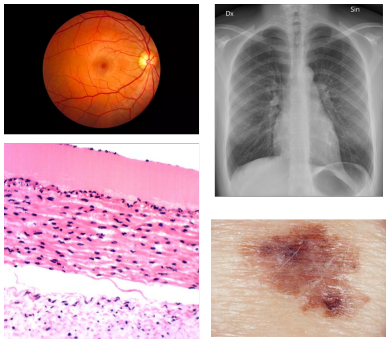
- AI – "information systems and algorithms capable of performing tasks associated with human intelligence" (Rajpurkar, 2022)
- Some classify AI into two broad categories (Khare, 2023)
  - Predictive AI – use of data and algorithms to predict some output (e.g., diagnosis, treatment recommendation, prognosis, 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) 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 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

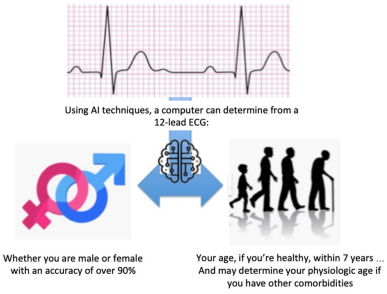


Predictive AI not limited to imaging

- 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)
- Next-generation sequencing (NGS) data on >36K tumors able to predict cancer of unknown primary with high confidence for 41% of tumors, leading to improved survival for those patients (Moon, 2023)
- 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, 2023)

- 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)
  - Diagnosis and risk stratification in occlusive myocardial infarction (Al-Zaiti, 2023)
  - 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)
- Voice data
  - Differences in presence of Type 2 diabetes mellitus (Kaufman, 2023)



And now, "generative AI"

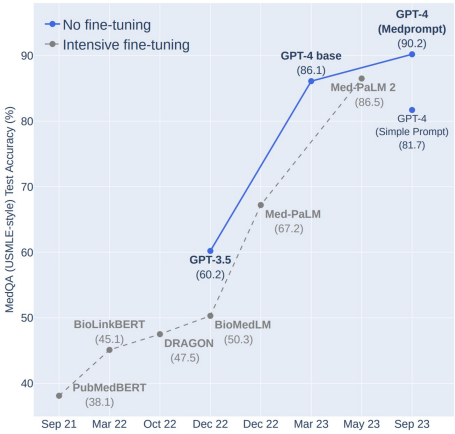
- Introduction of ChatGPT on November 30, 2022 brought new type of AI into focus: generative AI
- Based on large language models (LLMs) processed by deep neural networks using large amounts of training data and tuned for specific tasks
  - Trained on massive amounts of text and other content, e.g., large 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.,
  - Bing – with version of GPT-4
  - 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; Lekovich, 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)

But there are downsides to generative AI

- 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; some argue violate copyright
  - Several use Colossal Clean Crawled Corpus (C4) dataset (Dodge, 2021), which includes all content on dmice.ohsu.edu
- 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)

And 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 (Kapoor, 2023) from problems such as data bias (especially from EHR – Lewis, 2023) and data leakage (Ball, 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)
- 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)
- Must be implemented in responsible ways (Dorr, 2023)

Examples of need to build evidence base

- Hospital readmissions – major topic from mid-2010s
  - RCT on a hospital readmissions initiative showed no improvement (Donzé, 2023)
  - Accompanying editorial noted "just because we can predict something does not mean we can necessarily do something about it" (Wachter, 2023)
- Predicting future trajectories in estimated glomerular filtration rate (eGFR) in adults with type 2 diabetes and chronic kidney disease
  - New model excels over previous models in more accurate estimation of risk earlier in the disease course (Gregorich, 2023)
  - While model provides more accuracy and benefit in this phase of disease, might be outweighed by "inappropriate avoidance of intravenous contrast, patient anxiety, and unnecessary testing with its associated costs and what is really needed are clinical trials to validate use of the model" (Sanghavi, 2023)

Evidence base (cont.)

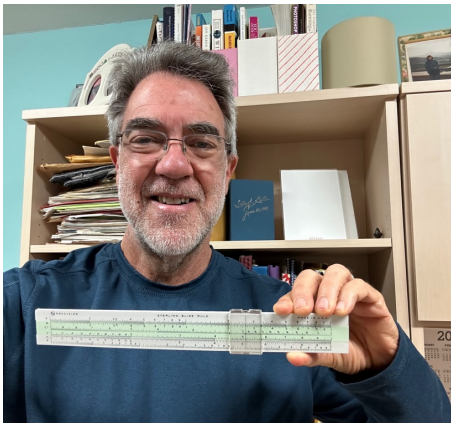
- Computer-aided detection (CADe) for polyp detection during colonoscopy
  - Recent systematic reviews show increased detection of adenomas but not advanced adenomas and higher rates of unnecessary removal of nonneoplastic polyps (Hassan, 2023; Lou, 2023)
  - Recent RCT showed no significant difference in advanced colorectal neoplasia detection rate (34.8% with intervention vs. 34.6% for controls) or mean number of advanced colorectal neoplasias detected per colonoscopy (Mangas-Sanjuan, 2023)
  - Many challenges in implementing AI in real world, which may impact RCT results, but must build evidence base to support use (Shung, 2023)

Issues in informatics education

- What is optimal use of generative AI in learning?
  - When might generative AI use by students be detrimental or unethical?
- What opportunities are there in AI for informatics students?

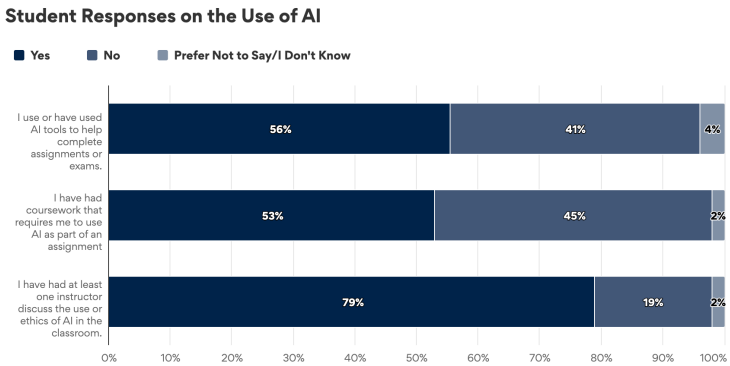
Generative AI and teaching

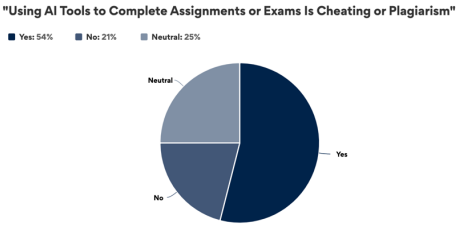
- A transformative cusp?
  - Anyone under 30 know what I am holding?
- Much in our courses can be done by generative AI
- Thought leader in this regard is Ethan Mollick, U of Penn
  - https://www.oneusefulthing.org/
- My approach (so far)
  - Gen AI policy from Provost generally and my course specifically
  - Allowing explicit use in certain assignments

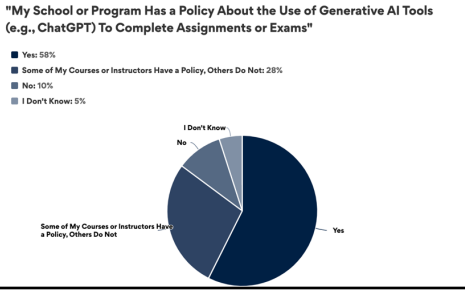


19

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







20

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

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

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. This page reflects specific course policy for the Oregon Health & Science University (OHSU) course, BMI 510/610 - Introduction to Biomedical & Health Informatics. This policy applies to courses derived from BMI 510/610, including the AMIA 10x10 course and the MD curriculum course, MINF 705B/709A.

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 BMI 510/610 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 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 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: December 11, 2023

OHSU Academic Integrity & AI

https://o2.ohsu.edu/teaching-and-learning-center/academic-integrity

Artificial intelligence (AI) chatbots and writing tools are here to stay. ChatGPT and other tools will only get better at automatically creating output that appear to be written by humans. These tools have been trained on vast amounts of written data found online and will only improve as more data are digested. Unfortunately, the efficacy of these tools can undermine the purposes of assignments. So, what are faculty to do?

First, don't be afraid of AI. Acknowledge its existence and use. Second, be explicit from the beginning of the course about the importance of academic integrity and integrity in general in learners' professional fields. Some faculty add an honor pledge that students must acknowledge when submitting assignments.

Here are ways to address and utilize AI tools in your courses:

1. Be explicit as to what kind of AI assistance is acceptable in your course, what isn't, and why. Put it in your syllabus so students understand what you will accept and how you will act if students are found to be violating your rules.
2. Require citations and/or credit for AI usage.
3. Use specific and targeted prompts for assignments. Assignments that require students to engage in specific tasks or to answer specific questions based on course material will make it more difficult for them to rely on an AI tool. Additionally, you can provide students a detailed rubric that shows how you will grade their assignments. AI tools will not know the specific items you are grading on and will not know to include those items in the responses.
4. Encourage students to provide their personal opinions and experiences on the topic being studied.
5. Use assessments that require students to apply their knowledge in potential professional scenarios.
6. Include presentations. Assignments that require students to present their work orally or in a group setting can also be effective in preventing the use of chatbots.
7. Create assignments that view and/or critique responses from AI tools such as ChatGPT.
8. Consider using proctored assessments to verify students are able to perform skills without AI assistance.

My first foray into generative AI assignments

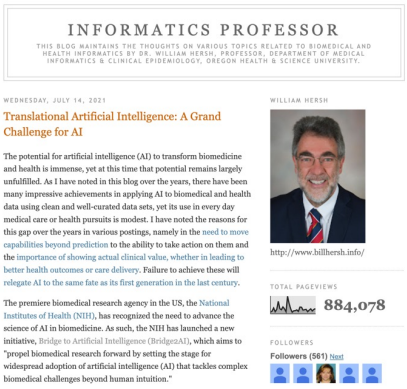
- Option for short (10x10) and long (510/610) papers
- For 510: Having a Conversation with GenAI/LLM About Informatics Topics
  - You should choose at least three topics in which you have an interest or questions beyond what is covered in the course
  - You should then use at least two different GenAI/LLM systems (ChatGPT, Perplexity.AI, or those embedded in search engines Google or Bing)
  - Your conversation with the GenAI/LLM systems should involve at least 3 iterations, i.e., initial prompt and follow-up with at least 2 more prompts
  - For each topic, you should follow up with a 1-2 page analysis of the output of the two or more systems, noting the accuracy, completeness, and errors in their analysis.
  - If the systems generate references, you should discuss whether the reference are relevant, whether they actually exist, and if any other sources would be better to cite. You will probably need to do conventional searching to help with this
- For 10x10: one topic, one LLM, three iterations, and analysis

Results

- Uptake in Fall 2023
  - 10x10 – 6/21 students
  - BMI 510 – 3/15 students
- Assessment
  - For the most part, carried out as assigned
  - For 510, most common second LLM was Perplexity.AI
  - Reasonable assessments of output
  - For 510, all had thoughtful analyses
    - One student had copious additional suggested references
    - Another student created a nice table of attributes to evaluate each LLM interaction

Opportunities in AI for students

- Despite the hype, these are early days
- Still much to know about implementation, evaluation, and risks
- Translational research analogy applies
  - Much "basic science" shows promise in the "lab"
  - Still need "translation" and "evidence" to demonstrate clinical value for patients and clinicians



INFORMATICS PROFESSOR

THIS BLOG MAINTAINS THE THOUGHTS ON VARIOUS TOPICS RELATED TO BIOMEDICAL AND HEALTH INFORMATICS BY DR. WILLIAM HERSH, PROFESSOR, DEPARTMENT OF MEDICAL INFORMATICS & CLINICAL EPIDEMIOLOGY, OREGON HEALTH & SCIENCE UNIVERSITY.

WEDNESDAY, JULY 14, 2021

Translational Artificial Intelligence: A Grand Challenge for AI

The potential for artificial intelligence (AI) to transform biomedicine and health is immense, yet at this time that potential remains largely unfulfilled. As I have noted in this blog over the years, there have been many impressive achievements in applying AI to biomedical and health data using clean and well-curated data sets, yet its use in every day medical care or health pursuits is modest. I have noted the reasons for this gap over the years in various postings, namely in the need to move capabilities beyond prediction to the ability to take action on them and the importance of showing actual clinical value, whether in leading to better health outcomes or care delivery. Failure to achieve these will relegate AI to the same fate as its first generation in the last century.

The premiere biomedical research agency in the US, the National Institutes of Health (NIH), has recognized the need to advance the science of AI in biomedicine. As such, the NIH has launched a new initiative, Bridge to Artificial Intelligence (Bridge2AI), which aims to "propel biomedical research forward by setting the stage for widespread adoption of artificial intelligence (AI) that tackles complex biomedical challenges beyond human intuition."

WILLIAM HERSH

http://www.billhersh.info/

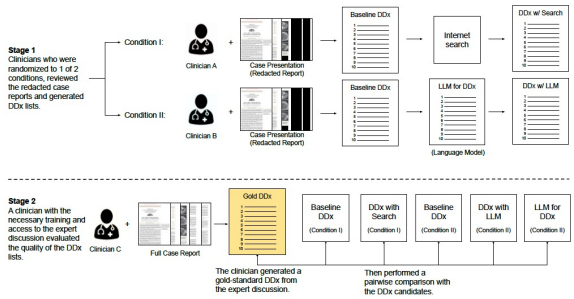
TOTAL PAGEVIEWS

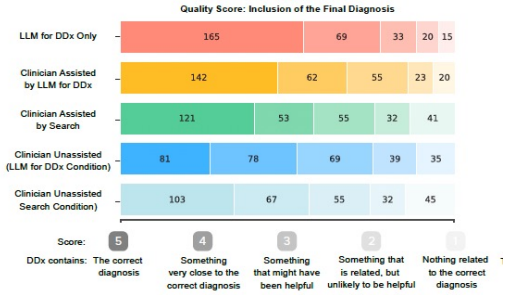
884,078

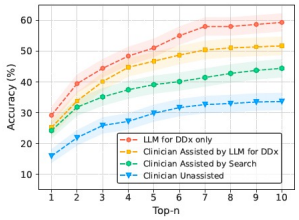
FOLLOWERS

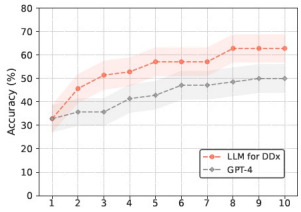
Followers (561) Next

Starting to see user studies, e.g., Google LLM for DDx (McDuff, 2023)









Some lessons learned from IR user studies (Hersh, 1996; Hersh, 2002)

- Differences in user metrics (e.g., questions answered) overwhelm system metrics (e.g., recall and precision)
- Just because you show users the correct information does not mean they complete the task correctly
- Sometimes user performance worsens with an information intervention, e.g., they go from a correct to incorrect answer

Conclusions

- Predictive and generative AI will profoundly impact the practice and education of all health professions, including informatics
  - Day-to-day impact, especially in clinical settings, small so far but likely to grow
  - Need real-world implementation and evaluation for safety and efficacy just like all other clinical interventions
- Educators and students must adapt to generative AI for writing, examination, and other pedagogic tasks
  - What constitutes appropriate use?
- There are many opportunities for informatics students to advance AI along the translational spectrum

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

Some other recent happenings

Appearance on episode of *Health and Explainable AI Podcast* from University of Pittsburgh HexAI Research Laboratory

- https://podcasters.spotify.com/pod/show/hexailab/episodes/William-Hersh-on-Pitt-HexAI-e2d7j14

My textbook, *Health Informatics: Practical Guide, 8th Edition* (Lulu.com, 2022), now available for viewing via OHSU Library

- https://librarysearch.ohsu.edu/permalink/01ALLIANCE_OHSU/1ff99el4/alma99900459559401858

Upcoming appearance on Informatics in the Round Podcast by Kevin Johnson

- Details forthcoming