Artificial Intelligence (AI): Promise and Peril

Department of Pathology Grand Rounds – October 4, 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., 2023. 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

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

Bulten, W., Kartasalo, K., Chen, P.-H.C., Ström, P., Pinckaers, H., Nagpal, K., Cai, Y., Steiner, D.F., van Boven, H., Vink, R., Hulsbergen-van de Kaa, C., van der Laak, J., Amin, M.B., Evans, A.J., van der Kwast, T., Allan, R., Humphrey, P.A., Grönberg, H., Samaratunga, H., Delahunt, B., Tsuzuki, T., Häkkinen, T., Egevad, L., Demkin, M., Dane, S., Tan, F., Valkonen, M., Corrado, G.S., Peng, L., Mermel, C.H., Ruusuvuori, P., Litjens, G., Eklund, M., PANDA challenge consortium, 2022. Artificial intelligence for diagnosis and Gleason grading of prostate cancer: the PANDA challenge. Nat Med. https://doi.org/10.1038/s41591-021-01620-2

Campanella, G., Hanna, M.G., Geneslaw, L., Miraflor, A., Werneck Krauss Silva, V., Busam, K.J., Brogi, E., Reuter, V.E., Klimstra, D.S., Fuchs, T.J., 2019. Clinical-grade computational pathology using weakly supervised deep learning on whole slide images. Nature Medicine 25, 1301–1309. https://doi.org/10.1038/s41591-019-0508-1

Chang, W.L., Grady, N., 2019. NIST Big Data Interoperability Framework: Volume 1, Definitions.

Chen, C., Lu, M.Y., Williamson, D.F.K., Chen, T.Y., Schaumberg, A.J., Mahmood, F., 2022. Fast and scalable search of whole-slide images via self-supervised deep learning. Nat Biomed Eng 6, 1420–1434. https://doi.org/10.1038/s41551-022-00929-8

Chen, S., Kann, B.H., Foote, M.B., Aerts, H.J.W.L., Savova, G.K., Mak, R.H., Bitterman, D.S., 2023. Use of Artificial Intelligence Chatbots for Cancer Treatment Information. JAMA Oncol e232954. https://doi.org/10.1001/jamaoncol.2023.2954

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.

Cooper, A., Rodman, A., 2023. AI and Medical Education — A 21st-Century Pandora's Box. New England Journal of Medicine. https://doi.org/10.1056/NEJMp2304993

Coyner, A.S., Singh, P., Brown, J.M., Ostmo, S., Chan, R.V.P., Chiang, M.F., Kalpathy-Cramer, J., Campbell, J.P., Imaging and Informatics in Retinopathy of Prematurity Consortium, 2023. Association of Biomarker-Based Artificial Intelligence With Risk of Racial Bias in Retinal Images. JAMA Ophthalmol 141, 543–552. https://doi.org/10.1001/jamaophthalmol.2023.1310

Davenport, T.H., Harris, J., Abney, D., 2017. Competing on Analytics: The New Science of Winning; With a New Introduction, Revised Edition. ed. Harvard Business Review Press, Boston, Massachusetts.

DeCamp, M., Lindvall, C., 2023. Mitigating bias in AI at the point of care. Science 381, 150–152. https://doi.org/10.1126/science.adh2713

Donoho, D., 2017. 50 Years of Data Science. Journal of Computational and Graphical Statistics 26, 745–766. https://doi.org/10.1080/10618600.2017.1384734

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

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, K.E., Rodman, A.M., Morgan, D.J., 2023. Preparing Physicians for the Clinical Algorithm Era. N Engl J Med. https://doi.org/10.1056/NEJMp2304839

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.

Halaweh, M., 2023. ChatGPT in education: Strategies for responsible implementation. Contemporary Educational Technology 15, 1–11.

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

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., 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., 2022. Health Informatics: Practical Guide, 8th Edition. Lulu.com.

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

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

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

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

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
Meyer, J., Khademi, A., Têtu, B., Han, W., Nippak, P., Remisch, D., 2022. Impact of artificial intelligence on pathologists' decisions: an experiment. J Am Med Inform Assoc 29, 1688–1695. https://doi.org/10.1093/jamia/ocac103
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
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
Moskvichev, A., Odouard, V.V., Mitchell, M., 2023. The ConceptARC Benchmark: Evaluating Understanding and Generalization in the ARC Domain. https://doi.org/10.48550/arXiv.2305.07141
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
Omiye, J.A., Lester, J., Spichak, S., Rotemberg, V., Daneshjou, R., 2023. Beyond the hype: large language models propagate race-based medicine. https://doi.org/10.1101/2023.07.03.23292192
Ötleş, E., James, C.A., Lomis, K.D., Woolliscroft, J.O., 2022. Teaching artificial intelligence as a fundamental toolset of medicine. Cell Rep Med 3, 100824. https://doi.org/10.1016/j.xcrm.2022.100824
Ozyoruk, K.B., Can, S., Darbaz, B., Başak, K., Demir, D., Gokceler, G.I., Serin, G., Hacisalihoglu, U.P., Kurtuluş, E., Lu, M.Y., Chen, T.Y., Williamson, D.F.K., Yılmaz, F., Mahmood, F., Turan, M., 2022. A deep-learning model for transforming the style of tissue images from cryosectioned to formalin-fixed and paraffin-embedded. Nat Biomed Eng 6, 1407–1419. https://doi.org/10.1038/s41551-022-00952-9
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).

Ratliff, M., Sya'ban, S.N., Wazir, A., Haidar, S., Keeth, S., 2023. AI and ChatGPT in Health Professions Education [WWW Document]. Lecturio. URL https://www.lecturio.com/pulse/using-chatgpt-in-medical-education-for-virtual-patient-and-cases/ (accessed 7.26.23).

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

Sahni, N.R., Carrus, B., 2023. Artificial Intelligence in U.S. Health Care Delivery. N Engl J Med 389, 348–358. https://doi.org/10.1056/NEJMra2204673

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
Seth, P., Hueppchen, N., Miller, S.D., Rudzicz, F., Ding, J., Parakh, K., Record, J.D., 2023. Data Science as a Core Competency in Undergraduate Medical Education in the Age of Artificial Intelligence in Health Care. JMIR Med Educ 9, e46344. https://doi.org/10.2196/46344
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
Song, A.H., Jaume, G., Williamson, D.F.K., Lu, M.Y., Vaidya, A., Miller, T.R., Mahmood, F., 2023. Artificial intelligence for digital and computational pathology. Nat Rev Bioeng 1–20. https://doi.org/10.1038/s44222-023-00096-8
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
Steiner, D.F., MacDonald, R., Liu, Y., Truszkowski, P., Hipp, J.D., Gammage, C., Thng, F., Peng, L., Stumpe, M.C., 2018. Impact of Deep Learning Assistance on the Histopathologic Review of Lymph Nodes for Metastatic Breast Cancer. Am J Surg Pathol 42, 1636–1646. https://doi.org/10.1097/PAS.0000000000001151
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
Wagner, S.J., Reisenbüchler, D., West, N.P., Niehues, J.M., Veldhuizen, G.P., Quirke, P., Grabsch, H.I., Brandt, P.A. van den, Hutchins, G.G.A., Richman, S.D., Yuan, T., Langer, R., Jenniskens, J.C.A., Offermans, K., Mueller, W., Gray, R., Gruber, S.B., Greenson, J.K., Rennert, G., Bonner, J.D., Schmolze, D., James, J.A., Loughrey, M.B., Salto-Tellez, M., Brenner, H., Hoffmeister, M., Truhn, D., Schnabel, J.A., Boxberg, M., Peng, T., Kather, J.N., 2023. Fully transformer-based biomarker prediction from colorectal cancer histology: a large-scale multicentric study. https://doi.org/10.48550/arXiv.2301.09617
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
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 (AI): Promise and Peril

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

1

Artificial Intelligence: Promise and Peril

- Definitions
- Historical perspectives
- Current accomplishments, including in Pathology
- Evidence base for AI interventions
- Role in medical education
- Future directions

2

Warning at 11:20 am

gina shirah
@GinaShirah81815

Turn off your cell phones on October 4th. The EBS is going to "test" the system using 5G. This will activate the Marburg virus in people who have been vaccinated. And sadly turn some of them into zombies.

8:55 PM · 2023-09-30 from Earth · **25.2K** Views

Elohiym are fallen angels_Lucifer Father of Cain
@ElohiymOf33770

Make no mistake the Mark of the Beast Vaxx is being activated by the EBS Test on October 4th!

The results of the zombie 5g activation may be immediate or within a days.....

> **TheBibleIsAnIdol** @bible_an5915 · Sep 27
> ⚡The Zombie 5g Vaxx Activation⚡
> 💥Emergency Alert System Test💥
> 🎯Oct 4th: 2:20 PM-2:50 PM (EDT)🎯
> 🚫USE NO TECHNOLOGY🚫
> ⚫3 Days of Darkness⚫
> Zombies possessed by AI Demons from Bottomless Pit
> twitter.com/then_great1223...

5:51 PM · Oct 3, 2023 · **164** Views

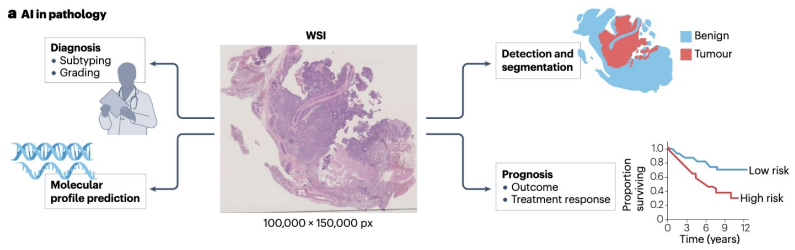
3

The challenge of giving a talk on AI…
https://www.nature.com/articles/s44222-023-00096-8

nature reviews bioengineering
https://doi.org/10.1038/s44222-023-00096-8
Review article
Check for updates

Artificial intelligence for digital and computational pathology

Andrew H. Song[1,2,3,4,8], Guillaume Jaume[1,2,3,4,8], Drew F. K. Williamson[1,2,3,4], Ming Y. Lu[1,2,3,4,5], Anurag Vaidya[1,2,3,4,6], Tiffany R. Miller[1] & Faisal Mahmood[1,2,3,4,7]



4

Definitions and terminology related to artificial intelligence (AI)

- AI – "information systems and algorithms capable of performing tasks associated with human intelligence" (Rajpurkar, 2022; Sahni, 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 – "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
- Other terms
  - Data science – science of learning from data (Donoho, 2017)
  - Data analytics – use of data and statistical analysis to build explanatory and predictive models and drive decisions and actions (Davenport, 2017)
  - Big Data – data characterized by large volume, velocity, variety and variability (Chang, 2019)

5

Machine learning (ML) (Shah, 2023)

- Overall goal is to build models that learn from data
- Initially two categories of models
  - Supervised – learn to predict from labeled data
  - Unsupervised – learn from naturally occurring patterns or groupings within data
- Now variations
  - Reinforcement learning – learning from experience with existing models
  - Transfer learning – applying learning trained for one task to another
  - Self-supervised learning – identify labels from patterns in data

6

Functionality of ML

Classification

- Models typically trained via supervised learning
  - Clinical functions may include diagnosis, treatment, outcomes, etc.
- Methods include
  - Logistic regression
  - k-nearest neighbors (kNN)
  - Naïve Bayes
  - Support vector machine (SVM)
  - Neural networks

Regression

- Models typically trained via unsupervised learning
  - Clinical functions may include risk stratification, clustering data, etc.
- Methods include
  - Linear regression
  - Random forest
  - Neural networks

7

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

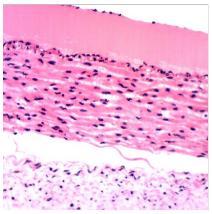
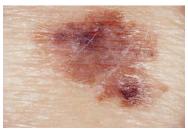
- Earliest paper related to AI and biomedical informatics attributed to Ledley and Lusted (1959) aiming to model physician reasoning through symbolic logic and probability
- Warner (1961) developed mathematical model for diagnosing congenital heart disease
- In 1960s-1970s, emergence of "expert systems" – 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)

8

Re-emergence of AI in 21st century

- "Predictive AI" driven by advances in ML, increasing availability of data, and more powerful computers and networks (Topol, 2019; Rajpurkar, 2022)
  - Deep learning in imaging advanced 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



Additional results in pathology

- Gleason grading for prostate cancer comparable to pathologists (Bulten, 2022)
- Cryosectioned images transformed to formalin-fixed and paraffin-embedded views (Ozyoruk, 2022)
- Augment efficiency of pathology whole-slide searching (Chen, 2022)
- Pathology slides predict colorectal cancer biopsy biomarkers and outcomes (Wagner, 2023)

Assisting pathologists

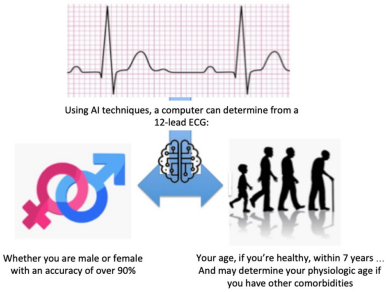
- Algorithm-assisted pathologists demonstrated higher accuracy than either DL algorithm or pathologist alone (Steiner, 2018)
  - Assistance significantly increased sensitivity of detection for micrometastases (91% vs. 83% alone)
  - Reduced time compared to pathologist alone for positive (61 vs. 116 sec) and negative images (111 vs. 137 sec)
- "Weakly supervised" (using clinical diagnoses) interpretation of pathology slides would allow pathologists to exclude 65–75% of slides while retaining 100% sensitivity (Campanella, 2019)
- In simulated study, decisions with AI decision-aid improved performance, even when it was not used (Meyer, 2022)

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)



13

And now, "generative AI"

- Introduction of ChatGPT in November, 2022 brought new type of AI into focus: generative AI
  - Initially based on GPT-3.5 model; added larger GPT-4 soon after
- 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 ChatGPT (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 for specific tasks (Chung, 2022)
  - Activated by (and importance of) prompting (Liu, 2023)

14

Results of ChatGPT and other LLMs

- Medical board exams
  - USMLE "arms race," starting with (Kung, 2023)
    - Claimed best – https://www.openevidence.com/blog/openevidence-ai-first-ai-score-above-90-percent-on-the-usmle
    - Even on "soft skills" (e.g., communication skills, ethics, empathy, and professionalism) questions (Brin, 2023)
  - Passing level on some board exams (clinical informatics – Kumah-Crystal, 2023; radiology – Bhayana, 2023) but not others (neonatology – Beam, 2023)
- 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)

15

Results of ChatGPT and other LLMs (cont.)

- Communicating with patients
  - Answer questions in public forums and write letters with comparable or better empathy (Sarraju, 2023; Ali, 2023, Ayers, 2023)
- Use of predictive AI (closing the AI loop)
  - Classifying CXR findings (Xu, 2023)
  - Predicting cardiovascular risk comparable to Framingham models (Han, 2023)

16

But there are downsides to generative AI

- Equally compelling disinformation – humans cannot distinguish between true and false tweets generated by GPT-3 and written by real Twitter users (Spitale, 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 dermatology questions asked of 4 LLMs recapitulated "harmful, race-based medicine" (Omiye, 2023)
- Performs worse than humans in abstraction and analogy problems (Moskvichev, 2023)
- GPT detectors more likely to classify non-native English writing as AI-generated (Liang, 2023)

17

What is the evidence base for AI?

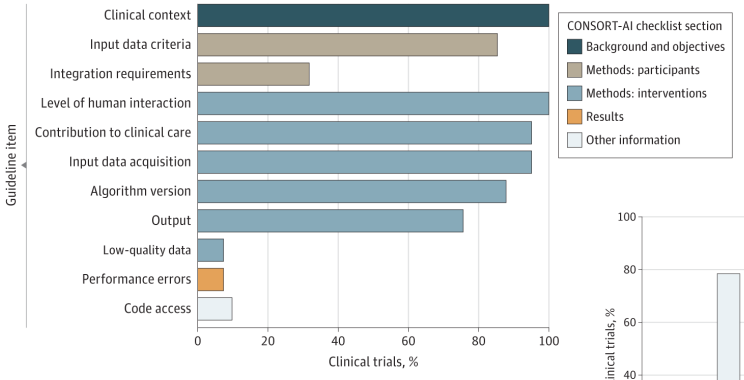
- Models are an important foundation (basic science) but need evidence of improved patient outcomes, care delivery, or other benefits (clinical science)
- Best evidence for any clinical intervention is randomized controlled trial (RCT) or systematic review of RCTs
- Three recent systematic reviews, all with limitations
  - *Clinical impact and quality of randomized controlled trials involving interventions evaluating artificial intelligence prediction tools* (Zhou, 2021)
  - *Randomized Clinical Trials of Machine Learning Interventions in Health Care* (Plana, 2022)
  - *Randomized Controlled Trials Evaluating AI in Clinical Practice* (Han, 2023 – scoping review, preprint)

18

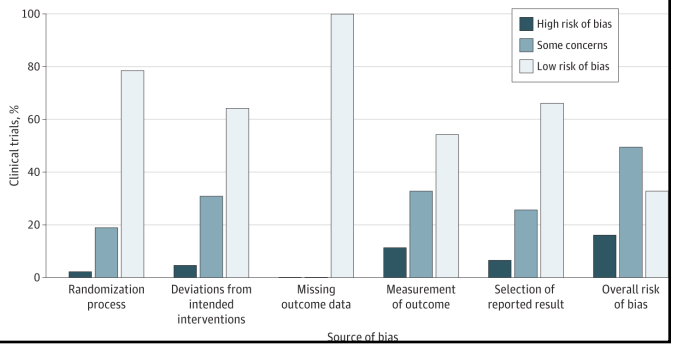
Zhou, 2021

- Systematic review of all RCTs using
  - Traditional statistical methods (TS) – mostly regression
  - Machine learning (ML) – all but deep learning
  - Deep learning (DL) – neural networks
- TS and ML tools focused on assistive treatment decisions, assistive diagnosis, and risk stratification, whereas DL tools only focused on assistive diagnosis
- Key results
  - 81 RCTs identified, 50 with "positive" results (61.5%)
  - Majority of RCTs had indeterminate or high risk of bias
  - Trials of DL methods mostly focused on endoscopic procedures

Plana, 2022



- 41 RCTs identified
- No assessment of study outcomes
- Assessment for adherence to CONSORT-AI and risk of bias



Han, 2023

- 84 RCTs identified
  - 64 positive; 5 non-inferior
  - No assessment of risk of bias or other aspects of study quality
- RCTs predominantly DL systems for medical imaging (57/84,67.9%)
  - Mostly video-based (40/57, 70.2%) driven by large number of endoscopy trials (32/40,80.0%)
- Outside imaging, systems operated on structured data, e.g.,
  - EHR (14/27,51.9%)
  - Waveform (10/27,37.0%)
  - Free text (3/27,11.1%)

	Primary Result				
Comparison	Significant improvement	No significant improvement	Demonstrated noninferiority	Significant deterioration	Grand Total
AI vs Clinician	3	1	3	1	8
AI vs routine care	14	4	..	..	18
Assisted vs Unassisted	47	9	2	..	58
Grand Total	64	14	5	1	84

Additional examples of why we need to build evidence base

- Analysis of commercial radiology AI tools versus radiologists showed (Plesner, 2023)
  - Moderate to high sensitivity for detecting airspace disease, pneumothorax, and pleural effusion on chest radiographs
  - More false-positive findings than radiology reports
  - Decreased performance for smaller-sized target findings and when multiple findings were present
- RCT of mobile phone app using two AI skin lesion algorithms previously shown to perform comparably to specialists and superior to novices (Tschandl, 2019) assessed decisions to biopsy or dismiss patient and was significantly inferior to specialists (Menzies, 2023)

Need to build the evidence base (cont.)

- Computer-aided detection (CADe) for polyp detection during colonoscopy
  - Systematic review found increased detection of adenomas but not advanced adenomas and higher rates of unnecessary removal of nonneoplastic polyps (Hassan, 2023)
  - Most 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)

Need to build the evidence base (cont.)

- 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)
  - Although model provides more accuracy, benefit to those in this (early) 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 RCTs to validate use of model (Sanghavi, 2023)
- Hospital readmissions
  - 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)

AI and health professions education

- Mostly physician-based but applies to all health professions
- Before generative AI there was recognition of need for competencies in clinical informatics for medical education (Hersh, 2014; Hersh 2020; Hersh, 2023)
- Others noted
  - AI should be taught as a "fundamental toolset" (Ötleş, 2022)
  - Clinicians must be prepared to practice in a world of AI (James, 2022)
  - Medical schools face dual challenges of needing to teach about AI in practice but also adapt to its use by learners and faculty (Cooper, 2023)
- New AI-competency frameworks
  - Use of AI-based tools by healthcare professionals (Russell, 2023; Liaw, 2023; Seth, 2023)
  - We must prepare physicians for the "clinical algorithm era" (Goodman, 2023)

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

25

ChatGPT will change education, not destroy it (Heaven, 2023)

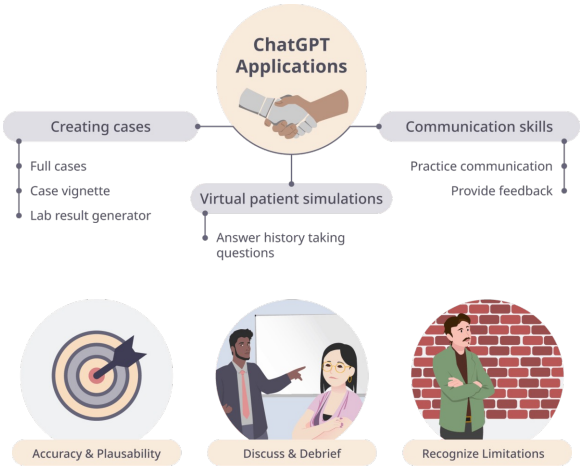
- Assessment already broken: "if ChatGPT makes it easy to cheat on an assignment, teachers should throw out the assignment rather than ban the chatbot"
- Change focus: use ChatGPT to generate an argument and then annotate it according to how effective argument was for a specific audience; then turn in rewrite based on their criticism
- Overcome misinformation and bias: ask students to use ChatGPT to generate text on a topic and then point out flaws
- Interact with ChatGPT to debate and generate counterarguments

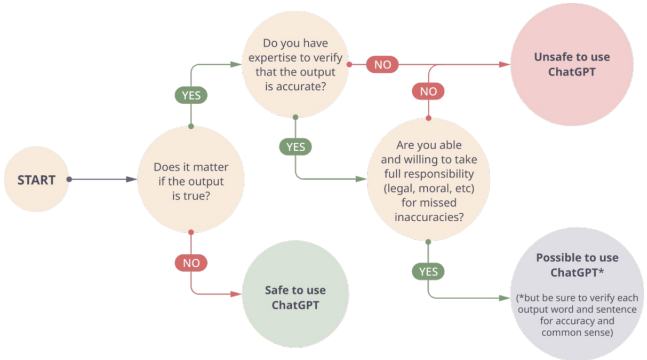
26

Strategies for responsible use of ChatGPT in education (Halaweh, 2023)

- Explicit policy in course syllabi or assessments
- Student reflection
- Audit trail of queries
- Use AI detector tools
- Swap student and instructor roles in use

ChatGPT in medical education (Ratliff, 2023)





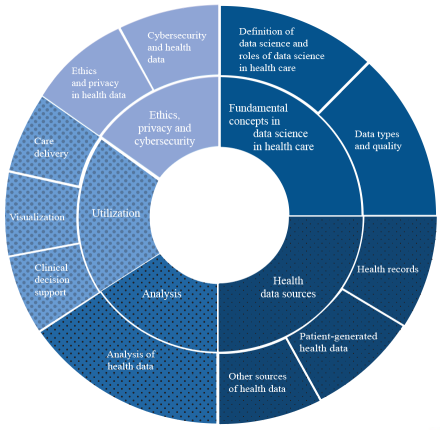
Competencies for use of AI-based tools by healthcare professionals (Russell, 2023)

Domains	Details
Basic knowledge of AI	Explain what AI is and describe its healthcare applications
Social and ethical implications of AI	Explain how social, economic, and political systems influence AI-based tools and how these relationships impact justice, equity, and ethics
AI-enhanced clinical encounters	Carry out AI-enhanced clinical encounters that integrate diverse sources of information in creating patient-centered care plans
Evidence-based evaluation of AI-based tools	Evaluate the quality, accuracy, safety, contextual appropriateness, and biases of AI-based tools and their underlying datasets in providing care to patients and populations
Workflow analysis for AI-based tools	Analyze and adapt to changes in teams, roles, responsibilities, and workflows resulting from implementation of AI-based tools
Practice-based learning and improvement regarding AI-based tools	Participate in continuing professional development and practice-based improvement activities related to use of AI tools in healthcare

29

Core competencies in data science for medical education in age of AI in healthcare (Seth, 2023)

- Fundamental concepts in data science in health care
  - Definition of data science and roles of data science in health care
  - Data types and quality
- Health data sources
  - Health records
  - Patient-generated health data
  - Other sources of health data
- Analysis
  - Analysis of health data
- Usage
  - Visualization
  - Care delivery
  - Clinical decision support
- Ethics, privacy, and cybersecurity
  - Ethics and privacy in health data
  - Cybersecurity and health data



30

Preparing physicians for the "clinical algorithm era" (Goodman, 2023)

Preclinical medical education

- Teach probability in medical school using intuitive, modern approaches
- Teach probabilistic clinical reasoning
- Assess probability and probabilistic reasoning skills
- Teach core, foundational working knowledge of CDS and EHR implementation, relevant to clinical use
- Practice interpreting CDS output in applied learning

Clinical training

- Reinforce probabilistic training and application
- Build CDS interpretation into curricula
- Reinforce working knowledge of CDS and EHR implementation, relevant to clinical use
- Include working knowledge of CDS in ACGME core competencies

Will AI help or hinder medicine?

- Real-world use and evidence base still modest – need "translational" perspective
- "AI won't replace radiologists, but radiologists who use AI will replace radiologists who don't," (Langlotz, 2019)
- Must also address bias in data and algorithms
  - AI may compromise care if not used properly (DeCamp, 2023)
  - Must be implemented in responsible (Dorr, 2023) and fair (Chen, 2023) ways

Questions?

William Hersh, M.D.
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/
Also on
Facebook
LinkedIn
Twitter – @williamhersh

33