**Ethical Algorithms for the Modern Clinician:**

An Introduction to Algorithmic Analysis for Patient Care

**Course Description**

With increasing adaptation of sophisticated algorithms such as artificial intelligence (AI) and generative machine learning models, it is crucial for future clinicians to have a working understanding of how these algorithms work and how they might impact patient care. While novel algorithms and technologies can improve the delivery of clinical medicine, research has shown that patients from minority backgrounds and those traditionally underrepresented in medicine may be inadvertently harmed in the process. It is important for future clinicians to be able to ***identify* these limitations** in clinical workflows and ***implement* meaningful solutions** to address them. To this end, the goal of this course is to evaluate modern algorithms with a critical lens, considering perspectives such as how algorithms may impact the fairness of clinical care, patient privacy, and physician decision making. The course website can be found [here](https://eamc-penn.github.io/).

**Learning Objectives**

1. **Define** an algorithm and what makes machine learning similar to and different from conventional algorithms;
2. **Identify** the key applications of machine learning and when computational algorithms might be helpful (or harmful) for patient care;
3. **Define** algorithmic bias as a property of an algorithm, and **characterize** bias mitigation strategies and how to better achieve fairness in clinical decision making;
4. **Define** privacy and anonymity, and **describe** existing anonymization techniques and pitfalls as they pertain to patient data;
5. **Analyze** how current data acquisition practices can affect minority patient populations;
6. **Describe** the accuracy-interpretability tradeoff and **reflect** on the role of interpretability in clinical algorithms and decision-making;
7. **Define** generative AI and **characterize** the new challenges facing patients and clinicians as a result of the adaption of generative AI software.

**Course Overview**

The course is broken down into **5 core modules** that are each **60 minutes** long. The topics for the modules are as follows:

* + - 1. [Introduction to Machine Learning](https://eamc-penn.github.io/ml.html): What is machine learning? How is it similar to and different from conventional software?
      2. [Bias and Fairness](https://eamc-penn.github.io/bias.html): How can algorithms be biased against different patient groups? How can we quantify, detect, and reduce bias in clinical decision making algorithms?
      3. [Privacy and Anonymization](https://eamc-penn.github.io/privacy.html): How can we anonymize patient data? Why does anonymization often fall short in protecting patient identities? How can we ensure that clinicians maintain patient privacy?
      4. [Algorithmic Interpretability](https://eamc-penn.github.io/interpretability.html): What does mean for an algorithm to be interpretable? Is it important for physicians to be able to explain how an algorithm works in order to use it in clinical practice?
      5. [Generative AI](https://eamc-penn.github.io/genai.html): What is generative AI, and how might it be used for patient care? What are the new challenges and opportunities associated with generative AI models?

Each of these 5 core modules have three components: (1) assigned student pre-reading; (2) synchronous learning and discussion in their doctoring groups; and (3) optional post-reading for the interested student.

**Student Pre-Reading**

Prior to each of the five modules above, each doctoring group will be broken up into two groups: Group A and Group B. Both groups will be tasked with reading two research papers prior to each module. One of the papers is an **Overview Paper** (same paper for both Groups A and B) that introduces topics that will be discussed in detail synchronously, and the other is a **Position Paper** (different papers for Groups A and B) that takes a position on the topic.

Here is an example for the [Intro to Machine Learning](https://eamc-penn.github.io/ml.html#evidence-based-medicine-discussion) module:

* **Discussion Question that will be asked during the doctoring session:** Should/Can AI be used to improve access to mental health resources?
* **Overview Article Pre-Reading for Both Groups A and B**: Stade EC, Stirman SW, Ungar LH, et al. Large language models could change the future of behavioral healthcare: A proposal for responsible development and evaluation. npj Mental Health Res 3(12). (2024). doi: 10.1038/s44184-024-00056-z. PMID: 38609507.
* **Position Paper Pre-Reading for Group A**: Yes, AI is more empathetic than physicians. Paper: Ayers JW, Poliak A, Dredze M, et al. Comparing physician and artificial intelligence chatbot responses to patient questions posted to a public social media forum. JAMA Intern Med 183(6): 589-96. (2023). doi: 10.1001/jamainternmed.2023.1838. PMID: 37115527
* **Position Paper Pre-Reading for Group B**: No, AI is too slow to escalate mental health scenarios. Heston TF. Safety of large language models in addressing depression. Cureus 15(12): e50729. (2023). doi: 10.7759/cureus.50729. PMID: 38111813

After completing the two assigned pre-readings, students should be prepared to participate in an **evidence-based medicine** **discussion** during the module discussing the discussion question posed above. No pre-reading assignment is expected to take students more than 30 minutes to complete.

A detailed list of all associated readings is included in **Appendix A** below.

**Course Format**

Each of the 5 course modules above follow roughly the same format:

**Overview of the Learning Objectives (1 min)**

**Introduction to the Topic (20 min)**: Students and preceptors will go over the basics of the topic at a level and depth specifically tailored for medical students of all backgrounds. This section is meant to be informative for students.

**Discussion Questions**: Throughout this section, each module contains a number of discussion questions to encourage students to reflect and think critically about the material in real-time. These discussion questions can be posed by the preceptors and students can answer and have a discussion. There are no right answers to these discussion questions.

**Hands-On Tutorial (10 min)**: Students and preceptors will gain hands-on experience in using state-of-the-art AI or ML models from the perspective of either clinicians or patients. For example, they might try using chatbots as a patient to interpret complex medical reports or to seek mental health care to better understand how and what AI resources are already being used by patients and physicians in practice.

**Evidence-Based Medicine Discussion (25 min)**: Students will be asked to leverage their pre-reading and personal experiences, and evidence-based insights to answer open-ended questions in their discussion/doctoring groups. Example discussion topics include (1) should AI be used to improve access to mental health resources?; and (2) do algorithms need to be interpretable in order for clinicians to leverage them for patient care? These open-ended questions have no right answer and should foster discussion between students.

**Summary (5 min)**: Students summarize and reflect on what they learned.

**Student Post-Reading**

Additional readings for the interested student are included at the end of each module. These post-reading resources are entirely optional.

**Course Material**

All of the course material is available to students and preceptors on the [course website](https://eamc-penn.github.io/) and also in PDF format (also available on the course website).

**Appendix A: Course Readings**

1. **Introduction to Machine Learning**
   1. **Overview Paper**: Stade EC, Stirman SW, Ungar LH, Boland CL, Schwartz HA, Yaden DB, Sedoc J, DeRubeis RJ, Willer R, Eichstaedt JC. Large language models could change the future of behavioral healthcare: A proposal for responsible development and evaluation. npj Mental Health Res 3(12). (2024). doi: [10.1038/s44184-024-00056-z](https://www.nature.com/articles/s44184-024-00056-z). PMID: 38609507
   2. **Position Paper A**: Ayers JW, Poliak A, Dredze M, Leas EC, Zhu Z, Kelley J, Faux DJ, Goodman AM, Longhurst CA, Hogarth M, Smith DM. Comparing physician and artificial intelligence chatbot responses to patient questions posted to a public social media forum. JAMA Intern Med 183(6): 589-96. (2023). doi: [10.1001/jamainternmed.2023.1838](https://jamanetwork.com/journals/jamainternalmedicine/fullarticle/2804309). PMID: 37115527
   3. **Position Paper B**: Heston TF. Safety of large language models in addressing depression. Cureus 15(12): e50729. (2023). doi: [10.7759/cureus.50729](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10727113/). PMID: 38111813
   4. **Optional Post-Readings**:
      1. Topol EJ. High-performance medicine: The convergence of human and artificial intelligence. Nat Med 25: 44-56. (2019). doi: [10.1038/s41591-018-0300-7](https://www.nature.com/articles/s41591-018-0300-7). PMID: 30617339
      2. Sidey-Gibbons JAM, Sidey-Gibbons CJ. Machine learning in medicine: A practical introduction. BMC Medical Research Methodology 19(64). (2019). doi: [10.1186/s12874-019-0681-4](https://bmcmedresmethodol.biomedcentral.com/articles/10.1186/s12874-019-0681-4). PMID: 30890124
      3. JAMA Podcast with Dr. Kevin Johnson from Penn Medicine. October 4, 2023. [Link](https://edhub.ama-assn.org/jn-learning/video-player/18820077).
2. **Bias and Fairness**
   1. **Optional Post-Readings:**
      1. Nicoletti L and Bass D. Humans are biased. Generative AI is even worse. Bloomberg. (2023). [Link to article](https://www.bloomberg.com/graphics/2023-generative-ai-bias/)
      2. Kearns M, Roth A. Responsible AI in the wild: Lessons learned at AWS. Amazon Science Blog. (2023). [Link to article](https://www.amazon.science/blog/responsible-ai-in-the-wild-lessons-learned-at-aws)
      3. Evaluating Model Fairness. Arize Blog. (2023). Accessed 19 May 2024. [Link to article](https://arize.com/blog/evaluating-model-fairness/)
      4. List JM, Palevsky P, Tamang S, et al. Eliminating algorithmic racial bias in clinical decision support algorithms: Use cases from the Veterans Health Administration. Health Equity 7(1): 809-16. (2023). doi: [10.1089/heq.2023.0037](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10698768/). PMID: 38076213
      5. Mittermaier M, Raza MM, Kvedar JC. Bias in AI-based models for medical applications: Challenges and mitigation strategies. npj Digit Med 6(113). (2023). doi: [10.1038/s41746-023-00858-z](https://pubmed.ncbi.nlm.nih.gov/37311802/). PMID: 37311802
3. **Privacy and Anonymization**
   1. **Overview Paper**: All of Us Research Program Overview. National Institutes of Health. Accessed 19 May 2024. [Link to article](https://allofus.nih.gov/about/program-overview)
   2. **Position Paper A**: Xia W, Basford M, Carroll R, Clayton EW, Harris P, Kantacioglu M, Liu Y, Nyemba S, Vorobeychik Y, Wan Z, Malin BA. Managing re-identification risks while providing access to the All of Us research program. J Am Med Inf Assoc 30(5): 907-14. (2023). doi: [10.1093/jamia/ocad021](https://doi.org/10.1093/jamia/ocad021). PMID: 36809550
   3. **Position Paper B**: Kaiser J. Million-person U.S. study of genes and health stumbles over including Native American groups. Science. (2019). [Link to article](https://www.science.org/content/article/million-person-us-study-genes-and-health-stumbles-over-including-native-american-groups)
   4. **Optional Post-Readings:**
      1. Gille F, Brall C. Limits of data anonymity: Lack of public awareness risks trust in health system activities. Life Sciences, Society and Policy 17(7). (2021). doi: [10.1186/s40504-021-00115-9](https://lsspjournal.biomedcentral.com/articles/10.1186/s40504-021-00115-9)
      2. Savage N. Privacy: The myth of anonymity. Nature 537: S70-2. (2016). [doi: 10.1038/537S70a](https://www.nature.com/articles/537S70a). PMID: 27602747
      3. Kapoor S. Revisiting HIPAA - Privacy concerns in healthcare tech. Berkeley Technology Law Journal. (2023). [Link to article](https://btlj.org/2023/01/revisiting-hipaa-privacy-concerns-in-healthcare-tech/)
      4. Ohm P. Broken promises of privacy: Responding to the surprising failure of anonymization. UCLA Law Review 57: 1701. (2010). [Link to article](https://ssrn.com/abstract=1450006)
      5. Pool J, Akhlaghpour S, Fatehi F, Burton-Jones A. A systematic analysis of failures in protecting personal health data: A scoping review. Int J Inf Manag 74: 102719. (2024). doi: [10.1016/j.ijinfomgt.2023.102719](https://doi.org/10.1016/j.ijinfomgt.2023.102719)
4. **Algorithmic Interpretability**
   1. **Overview Paper**: Imrie F, Davis R, van dr Schaar M. Multiple stakeholders drive diverse interpretability requirements for machine learning in healthcare. Nat Mach Intell 5: 824-9. (2023). doi: 10.1038/s42256-023-00698-2
   2. **Position Paper A**: Antony M, Kakileti ST, Shah R, Sahoo S, Bhattacharyya C, Manjunath G. Challenges of AI driven diagnosis of chest X-rays transmitted through smart phones: A case study in COVID-19. Sci Rep 13: 18102. (2023). doi: 10.1038/s41598-023-44653-y. PMID: 37872204
   3. **Position Paper B**: Ling J, Liao T, Wu Y, Wang Z, Jin H, Lu F, Fang M. Predictive value of red blood cell distribution width in septic shock patients with thrombocytopenia: A retrospective study using machine learning. J Clin Lab Anal 35(12): e24053. (2021). doi: 10.1002/jcla.24053. PMID: 34674393
   4. **Optional Post-Readings**:
      1. Model interpretability. Amazon Web Services Whitepapers. Accessed 20 May 2024. [Link to article](https://docs.aws.amazon.com/whitepapers/latest/ml-best-practices-healthcare-life-sciences/model-interpretability.html)
      2. Amann J, Blasimme A, Vayena E, Frey D, Madai VI. Explainability for artificial intelligence in healthcare: A multidisciplinary perspective. BMC Med Inform Decis Mak 20(1): 310. (2020). doi: [10.1186/s12911-020-01332-6](https://doi.org/10.1186/s12911-020-01332-6). PMID: 33256715
      3. Teng Q, Liu Z, Song Y, et al. A survey on the interpretability of deep learning in medical diagnosis. Multimed Syst 28(6): 2335-55. (2022). doi: [10.1007/s00530-022-00960-4](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9243744/). PMID: 35789785
5. **Generative AI**
   1. **Overview Paper**: Coming out to a chatbot? Researchers explore the limitations of mental health chatbots in LGBTQ+ communities. Science Daily. (2024). [Link to article](https://seas.harvard.edu/news/2024/05/coming-out-chatbot)
   2. **Position Paper A**: Alanzi T, Alsalem AA, Alzahrani H, Almudaymigh N, Alessa A, Mulla R, AlQahtani L, Bajonaid R, Alharthi A, Alnarhdi O, Alanzi N. AI-powered mental health virtual assistants’ acceptance: An empirical study on influencing factors among generations X, Y, and Z. Cureus 15(11): e49486. (2023). doi: [10.7759/cureus.49486](https://doi.org/10.7759/cureus.49486)
   3. **Position Paper B**: Rai S, Stade EC, Giorgi S, Guntuku SC. Key language markers of depression on social media depend on race. Proc Nat Acad Sci 121(14): e2319837121. (2024). doi: [10.1073/pnas.2319837121](https://doi.org/10.1073/pnas.2319837121)
   4. **Optional Post-Readings**:
      1. Kearns M. Responsible AI in the generative era. Amazon Blog. (2023). [Link to article](https://www.amazon.science/blog/responsible-ai-in-the-generative-era)
      2. Fournier-Tombs E, McHardy J. A medical ethics framework for conversational artificial intelligence. J Med Internet Res 25: e43068. (2023). doi: [10.2196/43068](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10373921/). PMID: 37224277
      3. Capraro V, Lentsch A, Acemoglu D, et al. The impact of generative artificial intelligence on socioeconomic inequalities and policy making. Proc Nat Acad Sci Nexus. (2024). doi: [10.2139/ssrn.4666103](https://dx.doi.org/10.2139/ssrn.4666103)