Group 1 Dialogues Phase A

Jerry Xu, Seth Close, Fuad Chowdhury, Shreya Sethi, Taehun Lee

CMPUT 200: Ethics of Data Science and Artificial Intelligence

Dr. Nidhi Hegde

March 7, 2025

Argue for the negative side: Consider a hospital that uses AI to process patient medical data to make decisions for the doctor. This AI system uses complex algorithms and data from various sources - including your medical history, lab results, and other relevant data from the hospital and healthcare system, but also any data available for purchase on the data market - to assist doctors in making critical decisions. This raises concerns of privacy, security, bias, potential for error and other issues. Is the hospital's argument that such an AI system has benefits that outweigh risks correct?

Medical decision-making is a high-stakes process that demands accurate results even in challenging situations involving limited resources, uncertainty, and high pressure environments. Traditionally, hospitals and doctors rely on medical knowledge, test results, patient information, and experience to make these critical decisions about patient care. Doctors undergo years of education and hands-on practice to develop the expertise necessary to diagnose conditions, assess risk, and determine treatments based on the patient's unique medical history and symptoms (Masic, 2022). This approach relies on human doctors to correctly assess each patient's unique situation in the decision making process. However, within this process, there is opportunity for artificial intelligence (AI) systems to process patient medical data and assist doctors in making these critical decisions.

These systems would use complex algorithms and data gathered from various sources, including patient medical history, test results, hospital information to possibly predict outcomes and recommend treatment options. Although AI has only gained widespread adoption in recent years, the idea of using AI in the medical field is not new and has been used as early as the 1960s to help detect blood infections (Kulikowski, 2019). More recently, the use of AI in the medical field has steadily grown, fuelled by the digitalization of healthcare especially during the COVID-19 pandemic. The pandemic caused significant strain on medical services worldwide, and to combat the shortage of healthcare resources, AI tools were integrated in medical decision-making processes such as diagnosis, prognosis, and treatment. AI models were trained on data from clinical cases, including medical images and medical expert guide diagnosis, to improve accuracy of its predictions. These trained AI models not only supported physicians during the COVID-19 pandemic, but remain a valuable tool in the medical decision-making process. (Khosravi, 2024). Yet, the rapid development of AI tools in the medical field has raised concerns. AI tools rely heavily on large datasets, many of which contain biases that can disproportionately affect marginalized groups, leading to disparities in care. Furthermore, there are privacy concerns with using sensitive patient health data in AI models, especially when data or models are shared and sold. Lastly, AI tools often lack transparency, functioning as "black boxes" that doctors and hospitals use without fully understanding. This makes their decision-making process hard to interpret and validate. These risks outweigh the benefits of using AI systems in healthcare to assist in critical decision making, and as such they should not be used.

One of the most significant risks of AI in healthcare is bias. The amount and quality of medical data on marginalized groups creates biases and unfairness in any model that is trained on it. Historically, medical research has overlooked marginalized people - both racial minorities and women - leading to significant gaps in representation. As recent as 1993, it was uncommon for women to be included in clinical studies in the United States (Mastroianni, 1994). Even today, marginalized groups are often underrepresented in studies that directly affect their health (Sosinsky, 2022). This lack of relevant data increases the challenge of developing machine learning models that can accurately diagnose and help these people. Not having large and comprehensive datasets can lead to hard-to-track biases and inaccuracies. Many models may appear fair and unbiased from one group's perspective, but they can be deeply flawed and unreasonable when applied to other groups. An example is the Correctional Offender Management Profiling for Alternative Sanctions (COMPAS) model, which predicts an inmate's risk of recidivating - committing another crime in the future. It has deep biases in its predictions in regards to inmates' race. The biases resulted in incorrect predictions, leading to harsher punishments and unfair treatment (Larson, 2016). The bias was caused by the large disparity in the quantity of training data available for different groups (Hegde, 2025). A similar problem arises with models trained for health care because of the massive disparity in quality and quantity of data between men, women, and racial groups. It is simply unfair to use a system to make critical health decisions for people, when that system was not built for them. Even when accounting for the lack of data and known biases, AI system’s biases will still be less accurate due to having less training data and fewer opportunities to learn the specific health needs. Given the importance of health, it is an unacceptable risk for models trained without large amounts of accurate and fair data to be used in any hospital for general decision making.

Beyond bias, using AI systems in healthcare raises serious privacy concerns. AI systems process and store vast amounts of sensitive information, making them attractive targets for cyberattacks. A breach of such a system could expose personal health data, leading to identity theft, financial fraud, or reputational damage for patients (Stempel, 2024). Once such data is compromised, it may be difficult to fully mitigate the consequences, as stolen health records can be misused for deceitful activities. Another issue is the potential for patient data to be shared without explicit consent. This creates the possibility that personal health data could be used for purposes beyond medical care, such as marketing or research, further eroding privacy. For example, in 2019, Google partnered with Ascension Health, sharing millions of patients' health records without their knowledge or explicit consent (Alder, 2024). This collaboration aimed to leverage Google's artificial intelligence and cloud computing capabilities to analyze patient data and improve healthcare decision-making, such as predicting health risks. However, the secrecy surrounding the partnership raised significant ethical and legal concerns, as it also opened up the potential for the unauthorized use of data for commercial purposes. Such instances could lead to data being sold or shared with parties that may not have a direct role in the patient’s healthcare, raising concerns about the unauthorized use or exploitation of sensitive information. This erosion of privacy compromises patient trust in the healthcare system and its ability to safeguard personal health data.

AI systems can have unexplainable decision-making, which is a significant drawback when used in healthcare. When AI models have to handle enormous amounts of complex data, including redundant and irrelevant data, it leads to incorrect decision-making due to overfitting: when a machine learning algorithm learns the training data so well that it is only able to pass the required tests in the training set and fails to do so with unseen data. This is another significant disadvantage of employing AI in healthcare to support doctors; complex algorithms cause unexplainable decision-making. Another issue about this disadvantage is the “black-box” problem, which the deep learning algorithms used in the sector are claimed to be. The "black-box" problem is an AI system that fails to explain the result it outputted. The lack of interpretability in an industry like the healthcare sector could be fatal. For instance, doctors can not only make medical decisions but also be able to justify the medical decisions they made. Their reasoning is backed up by their education, experience in the medical field and critical thinking. Thus, when an AI offers a medical decision, without knowing how it drew conclusions, it is difficult for doctors to trust them (Kosinski, 2025). Typically, if a doctor makes a mistake in diagnosing a patient or provides incorrect treatment plans, the doctors themselves are held accountable and penalized, but if it is the AI system that made the mistake, who is to take responsibility for incorrectly treating the patient? The doctor who used the AI system to aid their diagnosis and treatment plan, the hospital who used the AI system, or the AI developers who built the system in the first place? Without strong safeguards, integrating AI into such an important and consequential industry as the healthcare sector is hazardous due to this legal ambiguity. An AI system cannot be correct 100% of the time as it has to handle a huge amount of data, and this may result in inaccurate diagnosis and unsuitable treatment solutions. These ethical and legal concerns make it very risky for doctors to use AI for support.

References

Alder, S. (2024, June 13). *Ascension Ransomware Attack: Initial Access Vector and Data Theft Confirmed*. The HIPAA Journal. https://www.hipaajournal.com/ascension-cyberattack-2024/

Khosravi, M., Zare, Z., Mojtabaeian, S. M., & Izadi, R. (2024, March 5). *Artificial Intelligence and decision-making in Healthcare: A thematic analysis of a systematic review of reviews*. Health services research and managerial epidemiology. https://pmc.ncbi.nlm.nih.gov/articles/PMC10916499/

Kosinski, M. (2025, January 15). *What is black box AI and how does it work?*. IBM. https://www.ibm.com/think/topics/black-box-ai (Accessed: 02 March 2025).

Kulikowski, C. A. (2019, August). *Beginnings of artificial intelligence in medicine (AIM): Computational Artifice Assisting Scientific Inquiry and clinical art - with reflections on present AIM Challenges*. Yearbook of medical informatics. https://pmc.ncbi.nlm.nih.gov/articles/PMC6697545/

Larson, J., Angwin, J., Kirchner, L., & Mattu, S. (2016, May 23). How we analyzed the compas recidivism algorithm. ProPublica. https://www.propublica.org/article/how-we-analyzed-the-compas-recidivism-algorithm

Masic, I. (2022). Medical decision making. Acta Informatica Medica, 30(3), 230. https://doi.org/10.5455/aim.2022.30.230-235

Mastroianni, AC. (1994, January 1). *NIH Revitalization Act of 1993 public law 103-43*. Women and Health Research: Ethical and Legal Issues of Including Women in Clinical Studies: Volume I. https://www.ncbi.nlm.nih.gov/books/NBK236531/

Hegde, N. (2025). Fairness: Introduction. Edmonton; University of Alberta. https://canvas.ualberta.ca/courses/18302/files/3165427

Sosinsky, A. Z., Rich-Edwards, J. W., Wiley, A., Wright, K., Spagnolo, P. A., & Joffe, H. (2022). Enrollment of female participants in United States drug and device phase 1–3 clinical trials between 2016 and 2019. *Contemporary Clinical Trials*, *115*, 106718. https://doi.org/10.1016/j.cct.2022.106718

Stempel, J. (2024, August 13). Enzo Biochem to pay $4.5 mln over cyberattack, NY attorney general says. *Reuters*. https://www.reuters.com/technology/cybersecurity/enzo-biochem-pay-45-mln-failing-safeguard-patient-data-2024-08-13/