UNIVERSITY OF WATERLOO

For Professor Maura Grossman

CS 492

How Biased Clinical Research Data Affects Healthcare

February 14, 2025

Jaskirat Pabla

Samin Sharar Nafi

Sahl Bakshi

Saboor Bakshi

Project Topic

Our project will explore the lack of diversity, including race and gender, in clinical research data and how these biases impact applications in healthcare and artificial intelligence (AI). AI models used for diagnostics, treatment recommendations, and predictive analytics rely heavily on large datasets. However, if these datasets lack representation from diverse ethnic groups, they risk producing biased and inaccurate medical predictions. This bias could lead to disparities in healthcare outcomes, disproportionately affecting underrepresented populations.

Relevance to the Course

This issue concerning diversity in data coverage is very important to the seminar as it borders computing, ethics, and social responsibility. AI is changing the landscape of healthcare, but if the data used is not accurate, then it will worsen the issue of systemic bias. These situations present some important challenges for justice, responsibility, and disclosure in medical systems. Solving this problem is what our project aims at, which is to provide insight into the ethical obligations of the developers and healthcare professionals to ensure that the medical services offered are dependable and available to all without discrimination.

Project Format

The project will be in the form of a report from research and interviews with healthcare workers, professors, and different researchers from SickKids and University of Waterloo. Steps to be taken include:
- Gather expert perspectives on what is currently being done to achieve diversification of clinical datasets.
- Discussing cases of healthcare bias, this will help identify boundaries of the problem.
- Propose ways to address bias that is evident in healthcare applications.
- The completed piece will take the form of an interview presentation. The interview will be recorded and transcribed, along with an analytical report highlighting the most relevant findings.

Methodology

- Literature Review: We will complete this literature review, this paper, and others to attempt to understand the biases associated with datasets for clinical research.
- Interview Preparation: Prepare a preliminary set of core questions for interviewees to evaluate the interviewees understanding regarding the diversity of datasets in AI and AI bias.
- Interviews with Experts: Execute recorded interviews with researchers, doctors, and professors from SickKids and the University of Waterloo.
- Analysis & Reporting: Examine what was revealed in the interview, compare it to what was available in the literature, and create a presentation that captures the thoughts of the conclusion.