EVALUATIONS OF AI APPLICATIONS IN HEALTHCARE

WRITTEN BY MILDRED CHO

BEST ETHICAL PRACTICES – PROBLEM FORMULATION

In this section, we will learn about specific actions that have been recommended as best ethical practices in the development and deployment of machine learning-based AI in health care applications.

Our thinking about best ethical practices will start at the stage of **problem formulation** and determining the purpose of the AI system that you want to develop.

First, is the intent ethical, is its purpose to enhance the health and well-being of patients? Even if the intended purpose seems ethical, could there be negative consequences if misused? Increasingly we are seeing data scientists challenging the purpose and the uses of the products they are developing, whether the purpose is to target advertising to individuals using social media, or to identify individuals through facial recognition algorithms. For example, the Association of Computing Machinery, one of the leading professional organizations of computer scientists, has urged a suspension of the use of facial recognition technologies because of their potential for prejudicial impact on human and legal rights. The Association has also asserted that developers and operators, as well as users of facial recognition technology are accountable for these systems' use and misuse.

While these might seem like questions that don't need to be asked in the health care AI context because the answers are obvious, they are especially important to ask for AI developed by teams that have conflicting or competing interests. Health care settings are rife with such competing interests because they are operating under resource constraints, including constraints on finances, personnel, equipment and supplies, not to mention financial incentives to improve care. Together, these interests and incentives all create pressure to avoid certain kinds of patients, or avoid providing certain kinds of treatments. AI developers have to be vigilant about mitigating conflicts of interest. But how? We'll talk about this more specifically, but first, let's consider the issue of problem formulation.

Formulating the AI problem involves translating a high-level goal such as "improving patient care" into actionable questions that can be answered by available data. The challenge is that actionable questions and available data are usually limited, so a lot can get lost in translation, and practical constraints on problem formulation can lead to undetected errors and bias. A fairly common and seemingly straightforward task such as risk stratification is often fraught because risk can be defined in many different ways, and usually in ways that are not measured directly but through proxy variables. We have seen that the use of health care costs as a proxy for risk or health care needs has led to racial bias because risk scores generated from predictive models based on costs underestimated the health care needs of Blacks as compared to Whites with the same risk score. Fortunately, such bias can be tested for, at least for variables for which you suspect underlying bias exists such as race or gender. In this example you would be looking at whether the relationship between risk scores generated by the model and the variable of interest, health needs, differed by race.

Similarly, risk stratification on the basis of cost as a proxy for health needs tends to be biased towards older people with complex chronic conditions at the end of life, when most health care costs are incurred, and against children with acute, potentially fatal but treatable conditions. If the question to be answered is "who needs the most care" it is important to remember that this question is not merely quantitative but also a values question that demands nuance. How is "care" defined? Does the question distinguish between acute and chronic care? How is "need" defined? If patients have untreatable conditions and therefore typically do not receive costly care, does that mean that a model should assign them low risk scores, or that they do not need care?

Questions to ask in problem formulation

- Is the purpose of the AI system to enhance the health and well-being of patients?
- Could there be unintended consequences if misused?
- What is the intent of the system – to identify, classify, predict?
- What is being identified, classified or predicted?
    - E.g. risk, disease, prognosis, costs, health care utilization, admission or readmission, treatment efficacy, decompensation
- How are these terms defined?

Closely related to problem formulation is the choice of **data**. Of course, data that are already available are the easiest to use. But that doesn't mean that these are the right data. For example, electronic health records might be plentiful but they lack a lot of information that could be important to predictive models, such as environmental exposures, diet, or socio- economic factors, all of which we know are highly determinative of health and disease. EHRs from one hospital or health system, or insurance claims data also often do not provide a longitudinal timeline of data points for a patient over time because patients often move between health providers and insurers.

And relying on data from a single time point could be misleading. On the other hand, grouping data together can mask important patterns. In our example of risk stratification, a model trained on data from people of a mix of age ranges could be masking patterns in data from a subpopulation, such as younger people. Of course, detailed knowledge of differences between subpopulations is necessary to know whether customized models are required. How does one know this in advance?

Questions to ask about data

- How well are the variables in the model represented by available data?
- Are proxy measures being used?
- Are there important variables that are likely to be associated with main outcome measures that are not represented in available data?
- Are there likely to be differences between subpopulations in main outcomes, especially by legally protected characteristics such as age, race, ethnicity, or gender, or socially important characteristics such as income and education?

Best practices point to the need to include practicing clinicians who are intimately familiar with the relevant patient populations and conditions, and the data that are necessary for modeling They need to know what data are actually available, and the limitations of the data, especially in terms of likely biases. It is important to have team members with an understanding of the source of systematic error, such as whether error is introduced because of lack of data from specific subpopulations, physician biases (such as the lower likelihood of women with heart disease to get a diagnosis of heart disease as compared to men), or broader social inequities such as differential access to care. That is because understanding the source of systematic error indicates whether and how models can account for bias. Clinician team members can be critical at the problem formulation stage in particular, because deep clinical knowledge is so important to understanding the implications of asking questions in a particular way.