DeepBCI

Linthe van Rooij

José Bênção Fernandes

Hidde Buiting

Privacy and VSD

Q1 (Privacy): According to the definitions of privacy, which kind is at stake here and why? What
 solution would you propose?

The privacy as the proper flow of information is at stake in here. The lady in question used social media to share an event with her friends and family, not to ask for targeted advertising nor for emails. Privacy as secrecy is not at stake since the information is shared by the user herself. It is suggested that search history and social media posts should not be read and kept by tech companies, or to adjust quickly to a change in one's life situation.

 Q2 (Privacy): Identify three reasons why privacy should be preserved in this case. Give reasons
 why.

The targeted advertising used information she shared with friends and family, and after the stillbirth it harmed her.

It is now unknown who, aside from the persons she shared this with, has information about her pregnancy.

The information that the lady shared was used to advertise to her, which is not something she envisioned.

 Q3 (VSD): Suppose you are responsible for the Value-Sensitive Design section in a company
 that deals with algorithms such as the one just discussed. Which empirical, conceptual, and
 technical investigations would you carry out to avoid cases such as Gillian's?

An empirical investigation that needs to be carried out is to investigate the emotional impact that it could have on people with similar cases. A conceptual investigation would be aimed at privacy and, respect and empathy to make sure that the algorithm is aligned with human values. A technical investigations would improve an algorithms capability to detect and adjust to significant life changes.

Q4 (VSD): Identify the three values to be design in an algorithm such as the one discussed. Why do you think these are the most important? Mention at least two potential conflicts among
them. Give a brief explanation why they are in conflict.

Core values that need to be taken into account for an algorithm like this include empathy, privacy and adjustability. These are in conflict because an algorithm like this shares information with parties without asking explicit consent for that. The adjustability of this algorithm conflicts with privacy since it requires user data. Adjustability of the algorithm conflicts with empathy as well, since an event like the one discussed can not always be picked up by the algorithm. Therefore it can lack empathy while it is designed to adjust to the personal situation.

Bias and fairness

Q1 (Bias): What form(s) of bias can you identify in this case? Give reasons as to why these are the most likely form(s) of bias.

It is a racial bias that is partly caused by an economic bias. People of color are more likely to have a lower income, and a lower income results in a certain way of using healthcare. This way causes a bias that treats people with lower incomes differently than people with higher incomes. Because people of color are often exposed to this kind of bias, they may lose some trust in doctors whom they feel are exhibiting bias. This effect results in a racial bias.

Q2 (Bias): The article stipulates some reasons as to why this algorithm discriminates against black patients. Could you identify which source of bias this is? Of the remaining two sources, could you indicate how such an algorithm could discriminate?

It is a preexisting bias. It is trained on healthcare costs data, which is unrepresentative for realistic healthcare risk prediction. The amount of money spent on healthcare differs for black and white patients due to preexisting societal inequality. Because of this, it shows an existing societal prejudice.

Q3 (Fairness): Evidently, a biased algorithm leads to forms of unfairness. Could you mention at least two forms of (un)fairness perpetrated by the algorithm (TIP: think of the sources of (un)fairness, such as data, outcomes, and the design of the algorithm)? How could these forms of (un)fairness be addressed by the algorithm in order to be reduced (or utterly eliminated)?

There is unfairness in outcome, black patients, who have more chronic conditions, are not identified as high-risk and thus don't receive the same level of care as white people. Second, there's data fairness, where using healthcare spending as a proxy for need inherently disadvantages poorer patients, who are often black, resulting in unfair predictions. This could be reduced by using data that better represents the healthcare risk a patient has.

Q4 (Fairness): Suppose you have been hired by the National Health Agency (NHA) of some country to implement a "risk-care management" algorithms nationwide. Based on the previous experience gained from the algorithm just discussed, which considerations would you recommend NHA to take into account in order to avoid current and potential sources of (un)fairness? Which provisions would you take up for increasing fairness in risk-care management algorithms?

Training data needs to be diverse in order to represent all of society. The decision that is made by the algorithm should focus on clinical aspects rather than the amount of money that is spent. The algorithm should regularly be controlled and checked on bias to correct for potential emerging biases.