# REFLECTIONS ON ETHICS PRESENTATIONS

*Spring, 2024*

AYSENUR OKAN

Watching the presentations on ethical considerations, I had a chance to reflect on conversations I have had over the years on how biases affect science. Even though science relies on "facts," the way we conduct our research and interpret our findings are not independent from our biases. When we test our hypotheses, the results are largely affected by our interpretations. When we ask questions to participants, interpret and code their answers, and analyze data, our preconceived opinions influence the course of our studies. When we train an algorithm, we add variables that we think are relevant. With all these biases, no scientific product is unbiased.

Watching the "for you page" algorithm video, I had a reaction of discomfort. These were all concepts I had thought of before, but it was difficult to see them all at once. Over the past few months, I have heard of influencers getting punished by their sponsors for talking about political issues on their platforms. A lot of content sharers lost significant income as a result of these practices since their content was limited, the algorithms de-prioritized their content and stopped sharing it with non-followers. A therapist who shares content on Instagram and TikTok shared that he stopped gaining followers after speaking up about genocide, leading to a significant decline in ad revenues. Given that social media platforms were promising alternatives to government-ruled, biased news agencies, these developments show that the advantage of free speech we once thought they offered is no longer available.

It is also deeply concerning that these algorithms increasingly tailor content based on even a single item consumed. Someone who watches a reel about self-harm might actually get

hundreds of videos related to it, potentially risking lives. Additionally, content around racial and gender discrimination is extremely dangerous. Reports show that there is an increase in anti-feminist beliefs in Gen-Z. Podcasters, such as Andrew Tate, have gained popularity and many more people keep sharing similar content, degrading women and spreading false information about feminist values. These practices incite hatred and create a loop of false information cycle that is perpetuated by the lack of control mechanisms. Currently, there are no regulations on these dangerous creations. There is no fact checking, and once a teenager has seen a related video, they will almost certainly keep getting more of that content.

Another concern is that users are profiled based on race, ethnicity, immigration status, nationality, age, sex, and sexual orientation. While these are protected classes, social media platforms are not regulated to ensure the safety of such information. The content is tailored towards users who share identities with either the creators or the modal consumer of the said content. This does not only limit exposure to a diverse set of perspectives, but also raises concern about safety. It is scary that social media companies have access to all this information, in addition to our location, passwords and contacts.

The video on the legal use of machine learning algorithms felt like a dystopian nightmare. To my original point about biases scientific process, it concerns me that the likelihood of re-offense after parole is determined by an algorithm. First, those algorithms rely on research findings that are only capable of providing summary scores. While they are useful and helpful in many ways, the average scores do not reflect an individuals' unique trajectory. Second, the variables that are input into these systems have intricate interactions with each other. Since no single study would be able to determine all possible interactions, it is difficult to have high accuracy. Third, the statistics on which those estimates are based are biased. The judicial system

has long discriminated against racial and ethnic minorities as well as individuals from low socioeconomic backgrounds. Using those values will exacerbate an already crumbling system.

Another concern comes from the issues related to the consumers or users of these systems. The police officers getting a gunshot detection signal and ignoring it based on the fact that it came from a rich neighborhood are not a rare species. A lot of users who use these systems will have their biases and might lean toward ignoring a high likelihood and focusing on the small error rate. Finally, most people are not educated consumers of statistical information. For example, when an algorithm says there is .80 chance that the person on parole will commit another crime, the parole officer may not understand what that would mean for evaluating the potential of a single individual. In this case, the .20 likelihood of having a successful parole is too large to ignore given that a false alarm has the potential to ruin someone's life for no good reason.

I recently participated in a study that was testing a machine learning algorithm that detects depression diagnosis based on pitch during a clinical interview. This study uses AI technology whereby an AI assistant interviews different individuals and records the interview. My job as the participant was to 1) listen to these recordings on a website as the algorithm continuously updated the likelihood of depression based on the information gathered, 2) give a diagnosis myself, and 3) rate the trustability of the algorithm. It was quite disheartening to see how the algorithm would up the likelihood after someone said they had been sleeping poorly because they are studying for an exam. While sleep disturbances could be considered depression symptoms, the periodic disturbance in this case would not be considered within that criterion. On the other than, the AI system did not ask any relevant follow-up questions and was very rigid in its wording. The algorithm also increased the likelihood after people said they had been

diagnosed with depression, without considering how far ago that was. All this shows me that, while they are in earlier stages of development and they will likely improve, the rigidity of machine learning algorithms should always be kept in mind. They do not possess nearly as much cognitive and computational flexibility as humans. Their estimates will always be approximations. Even at their best, their suggestions should only be treated as additional information rather than a final diagnosis.

There is fear in the society that AI will replace humans. I have always thought of the examples of kitchen gadgets and other technological advances (e.g., cars and computers) that provided convenience and ease to human life. Following these examples, AI can improve our quality of life rather than coming for our jobs. However, I also see the other side. Customer service employees have been replaced by chat boxes and we get automated responses that are disguised as the support team member of a big company. I still believe that AI and other technological advances will improve our lives. However, the amount of power and control they have given to big tech companies is terrifying.