Identifying ethical concerns

18 Jan 2022

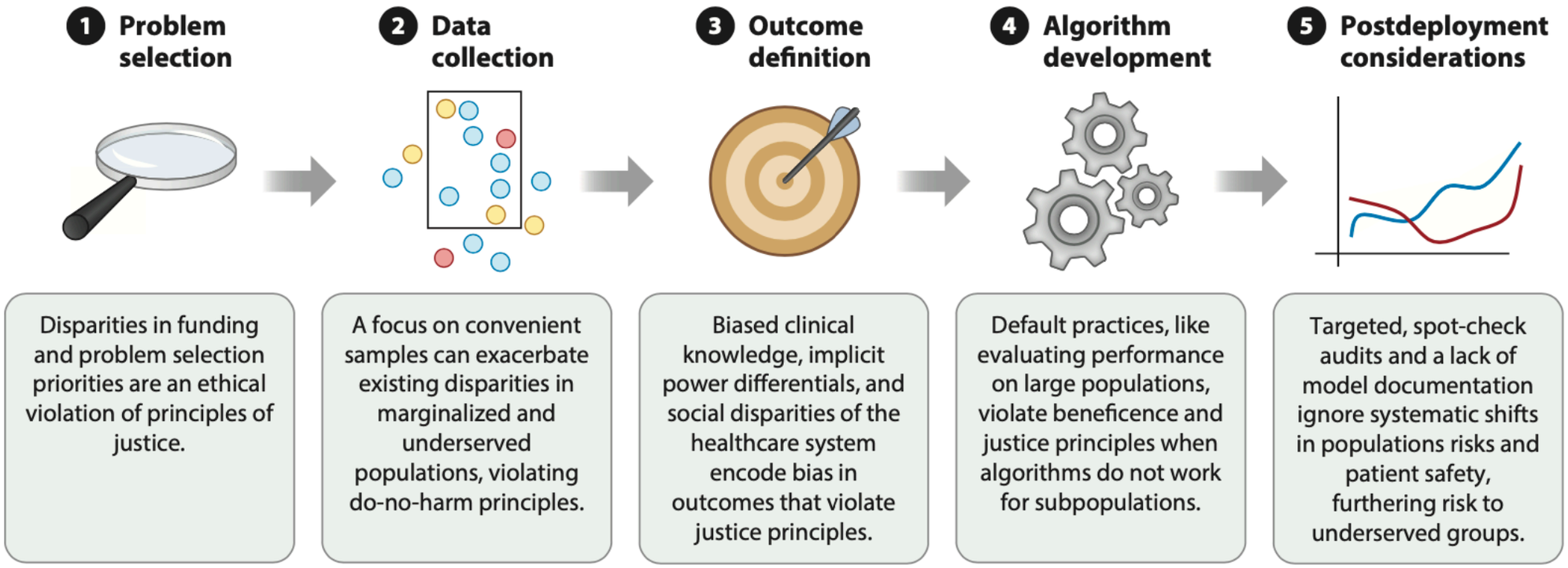

Figure 1

We motivate the five steps in the ethical pipeline for healthcare model development. Each stage contains considerations for machine learning where ignoring technical challenges violates the bioethical principle of justice, either by exacerbating existing social injustices or by creating the potential for new injustices between groups. Although this review's ethical focus is on social justice, the challenges that we highlight may also violate ethical principles such as justice and beneficence. We highlight a few in this illustration.

Chen et al. "Ethical Machine Learning in Healthcare"

ShotSpotter

The company's algorithms initially classified the sound as a firework. That weekend had seen widespread protests in Chicago in response to George Floyd's murder, and some of those protesting lit fireworks.

But after the 11:46 p.m. alert came in, a ShotSpotter analyst manually overrode the algorithms and "reclassified" the sound as a gunshot. Then, months later and after "post-processing," another ShotSpotter analyst changed the alert's coordinates to a location on South Stony Island Drive near where Williams' car was seen on camera.

Todd Feathers, "Police Are Telling ShotSpotter to Alter Evidence From Gunshot-Detecting AI"

The company has not allowed any independent testing of its algorithms, and there's evidence that the claims it makes in marketing materials about accuracy may not be entirely scientific.

Over the years, ShotSpotter's claims about its accuracy have increased, from 80 percent accurate to 90 percent accurate to 97 percent accurate. According to Greene, those numbers aren't actually calculated by engineers, though.

Todd Feathers, "Police Are Telling ShotSpotter to Alter Evidence From Gunshot-Detecting AI"

Discussion

- What are potential harms of the Shotspotter algorithm? Are these potential biases?
- Is this algorithm biased? In what sense?
- Can this algorithm be made fair?

TikTok

The document says watch time isn't the only factor TikTok considers. The document offers a rough equation for how videos are scored, in which a prediction driven by machine learning and actual user behavior are summed up for each of three bits of data: likes, comments and playtime, as well as an indication that the video has been played:

Plike X Vlike + Pcomment X Vcomment + Eplaytime X Vplaytime + Pplay X Vplay

"The recommender system gives scores to all the videos based on this equation, and returns to users videos with the highest scores," the document says. "For brevity, the equation shown in this doc is highly simplified. The actual equation in use is much more complicated, but the logic behind is the same."

Ben Smith, "How TikTok Reads Your Mind"

> Julian McAuley, a professor of computer science at the University of California San Diego, who also reviewed the document, said in an email that the paper was short on detail about how exactly TikTok does its predictions, but that the description of its recommendation engine is "totally reasonable, but traditional stuff." The company's edge, he said, comes from combining machine learning with "fantastic volumes of data, highly engaged users, and a setting where users are amenable to consuming algorithmically recommended content (think how few other settings have all of these characteristics!). Not some algorithmic magic."

Ben Smith, "How TikTok Reads Your Mind"

Discussion

- What are potential harms of the "TikTok algorithm?" Are these potential biases?
- Is the "TikTok algorithm" really reading your mind?
- Is this algorithm biased? In what sense?
- Can this algorithm be made fair?

Discussion

- https://docs.google.com/document/d/1jXvBqjORh6xdbHys24DomFP_j1dYQ33rSuERehiu8jQ/edit?usp=sharing