Stanford eCorner

Ethically Approaching Bias in AI

31-05-2023

URL: https://stvp.stanford.edu/clips/ethically-approaching-bias-in-ai

Rahul Roy-Chowdhury, Grammarly's CEO, explains his company's approach to designing its tool that flags potentially biased or insensitive text. They have a team of linguists to create "rules of the road" based on research and scholarship, and the tool suggests changes rather than requiring them. But, he says, such a tool requires constant evaluation.

Transcript

Woman I'm curious about when you were speaking 00:00:08,103 about responsibility and this sensitive text feature.. I guess kind of to play devil's advocate, where for you is the line between filtering out, like potentially triggering speech and like censorship and these AI sort of like taking a limiting, yeah access to certain ideas? - Yeah, it's a great question 00:00:32,400 and it's a question we ask ourselves every single day.. I don't know that there's any perfect line etched in stone or in paint or whatever the metaphor is.. We have a team of linguists, they are tracking and studying authoritative sources of what constitutes different kinds of bias, hate speech.. And we try to assemble from a set of these sources, sort of rules of the road.. You know, what is the thinking around this? What's the research, what's the scholarship, and how can you apply that in a meaningful way in the product? And that changes, right? So there's not like you do it and you say, "Okay, check the box, I'm done." There's a constant evaluation.. When Covid became prevalent, one of the things that was new was lots of biased ways to refer to Covid.. That became a thing.. And so Grammarly's sensitive text classifier caught those things and was able to suggest to users, I mean, at the end of the day, we can't write for you, but we'll suggest to you that, "Hey, the thing you're writing is actually considered pretty biased as a way to refer to this virus.. And so maybe you should consider this and here's why, and you should consider this other thing." So this is a thing that keeps changing and adapting and we are trying to be as close to current as we can...