

Como estou a combater o viés nos algoritmos | Joy Buolamwini

Olá, eu sou a Joy, uma poetisa do código, numa missão para deter uma força invisível que está a crescer, uma força a que dei o nome de "**olhar codificado**", o meu termo para o viés algorítmico. O viés algorítmico, tal como o viés humano, resulta em injustiça. No entanto, os algoritmos, como vírus, podem espalhar o viés a uma escala massiva e a um ritmo acelerado. O viés algorítmico também pode levar a experiências excludentes e práticas discriminatórias. Deixem-me mostrar-vos o que quero dizer. (Vídeo)

Joy Buolamwini: Olá, câmara. Eu tenho um rosto. Consegues ver o meu rosto? E sem óculos? Consegues ver o rosto dela. E o meu rosto? Estou a usar uma máscara. Consegues ver a minha máscara?

Então, como é que isto aconteceu? Por que estou sentada à frente de um computador com uma máscara branca, a tentar ser detetada por uma webcam barata? Bem, quando não estou a combater o olhar codificado como poetisa do código, sou estudante de pós-graduação no **MIT Media Lab**. Lá, tenho a oportunidade de trabalhar em todo o tipo de projetos fantásticos, incluindo o **Aspire Mirror**, um projeto que criei para projetar máscaras digitais na minha reflexão. Assim, de manhã, se quiser sentir-me poderosa, posso colocar uma máscara de leão. Se quiser sentir-me inspirada, posso ter uma citação.

Usei um software genérico de reconhecimento facial para construir o sistema, mas descobri que era muito difícil testá-lo, a menos que usasse uma máscara branca. Infelizmente, já me tinha deparado com este problema antes. Quando era estudante de licenciatura na **Georgia Tech**, onde estudava ciências da computação, costumava trabalhar com robôs sociais. Uma das minhas tarefas era fazer um robô brincar ao jogo do "cucu", em que os parceiros cobrem o rosto e depois o descobrem dizendo: "Cucu!". O problema é que o jogo não funciona se não te conseguem ver, e o meu robô não conseguia ver-me.

Acabei por pedir ajuda ao rosto da minha colega de quarto para concluir o projeto, entreguei o trabalho e pensei: "Bem, alguém vai resolver este problema." Não

muito tempo depois, estava em **Hong Kong** para uma competição de empreendedorismo. Os organizadores decidiram levar os participantes num tour pelas startups locais. Uma dessas startups tinha um robô social, e quiseram fazer uma demonstração. A demonstração funcionou com toda a gente, até chegar a mim, e provavelmente já sabem o que aconteceu. Ele não conseguiu detetar o meu rosto. Perguntei aos programadores o que se passava e descobri que tínhamos usado o mesmo software genérico de reconhecimento facial.

A meio do mundo, aprendi que o viés algorítmico pode viajar tão rápido quanto o tempo necessário para descarregar alguns ficheiros da internet. Então, o que está a acontecer? Por que é que o meu rosto não está a ser detetado? Bem, temos de olhar para a forma como damos visão às máquinas. A **visão computacional** usa técnicas de **aprendizagem automática** para fazer reconhecimento facial. Como funciona? Cria-se conjuntos de treino com exemplos de rostos. *Este é um rosto. Este é um rosto. Este não é um rosto.* Com o tempo, ensinas o computador a reconhecer outros rostos. No entanto, se os conjuntos de treino não forem suficientemente diversos, qualquer rosto que se afaste da norma estabelecida será mais difícil de detetar — o que estava a acontecer comigo.

Mas não se preocupem, há boas notícias. Os conjuntos de treino não surgem do nada. Nós podemos realmente criá-los. Há uma oportunidade de criar conjuntos de treino de **espectro completo** que reflitam um retrato mais rico da humanidade.

Agora, viram nos meus exemplos como os robôs sociais me ajudaram a perceber a exclusão causada pelo viés algorítmico. Mas o viés algorítmico também pode levar a práticas discriminatórias. Em todo os **EUA**, os departamentos de polícia estão a começar a usar software de reconhecimento facial para combater o crime. Um relatório da **Georgetown Law** mostrou que **um em cada dois adultos** nos EUA — cerca de **117 milhões de pessoas** — têm os seus rostos incluídos em bases de dados de reconhecimento facial. Atualmente, os departamentos de polícia podem consultar estas bases sem regulamentação, usando algoritmos que não foram auditados quanto à precisão.

Sabemos, no entanto, que o reconhecimento facial não é infalível, e etiquetar rostos corretamente continua a ser um desafio. Vemos isso no Facebook, por exemplo. Os meus amigos e eu rimos quando somos mal identificados em fotografias. Mas identificar erradamente um suspeito de crime não tem graça nenhuma, nem violar as liberdades civis.

A aprendizagem automática também está a ser usada para além do reconhecimento facial. No seu livro "**Weapons of Math Destruction**", a cientista de dados **Cathy O'Neil** fala sobre algoritmos generalizados, misteriosos e destrutivos (ADM) que estão a tomar decisões que impactam cada vez mais aspetos das nossas vidas:

- Quem é contratado ou despedido?
- Consegues obter um empréstimo ou seguro?
- És admitido na faculdade desejada?
- Pagas o mesmo preço pelo mesmo produto que outra pessoa numa plataforma online?

A aplicação da lei também está a usar a aprendizagem automática para **policimento preditivo**. Alguns juízes usam pontuações de risco geradas por máquinas para determinar penas de prisão. Temos de pensar nestas decisões: **são justas?** Já vimos que o viés algorítmico nem sempre conduz a resultados justos.

Então, o que podemos fazer? Podemos adotar práticas de **codificação inclusiva**. Tudo começa com as pessoas:

- **Quem codifica importa.** Temos equipas diversas que verificam os pontos cegos uns dos outros?
- **Como codificamos importa.** Incorporamos justiça nos sistemas que criamos?
- **Por que codificamos importa.** A tecnologia deve servir para desbloquear oportunidades de igualdade.

Para avançar neste sentido, lancei a **Algorithmic Justice League**. Em **codedgaze.com**, qualquer pessoa preocupada com a justiça pode relatar viés, pedir auditorias e participar na conversa com a hashtag **#codedgaze**.

Convido-vos a juntarem-se a mim para criar um mundo onde a tecnologia funciona para todos nós, não apenas para alguns. Um mundo onde a inclusão e a mudança social são centrais. Obrigada. (Aplausos)

Mas deixo-vos uma pergunta: **vão juntar-se a mim nesta luta?**