Dados e viés: quando as IAs são excludentes.

[ENSINO MÉDIO - 45 MIN]

Objetivo: reconhecer a possibilidade de viés nos dados que alimentam as inteligências artificiais, e como isso impacta a sociedade.

Introdução

Algoritmos são sequências de instruções que permitem que um computador analise determinada quantidade de dados (informações em forma de números, textos ou imagens) e aprenda ou tome decisões a partir deles. Antigamente esses algoritmos aprendiam com quantidades limitadas de dados, alimentados pelos engenheiros. Hoje nossa sociedade digital gera quantidades massivas de dados a cada instante, e as IAs podem aprender a partir deles: por exemplo, bancos de milhões de imagens, ou ainda milhões de páginas da Internet.

Também alimentamos as IAs com os dados que nós mesmos produzimos, mesmo sem perceber, ao utilizar dispositivos como telefones celulares, smartwatches ou mesmo nosso aspirador de pó inteligente! Todos esses dispositivos podem guardar dados relativos às nossas escolhas, preferências, comportamentos e locais, e as IAs, aprendendo com eles, podem nos oferecer sugestões de atividades, anúncios, lembretes etc.

Um dos problemas que enfrentamos atualmente é que, apesar das IAs aprenderem com uma grande quantidade de dados, estes não são necessariamente representativos. Na verdade, muitas vezes eles são tendenciosos: algumas culturas, estilos de vida, idiomas, grupos de pessoas ou comportamentos podem estar sub representados, outros desproporcionalmente representados. Isso pode levar à cristalização de certos estereótipos ou a ausências importantes, gerando discriminação social.

No filme Coded Bias, por exemplo, a engenheira Joy Buolamwini descobre o viés racista e machista por trás das IAs que operam os sistemas de reconhecimento facial. Se os bancos de imagens que treinam esses sistemas possuem majoritariamente imagens de rostos brancos, sem refletir adequadamente a diversidade que existe na sociedade, o resultado é que as máquinas não aprendem a “ler” rostos de pessoas pretas com a mesma precisão, e portanto cometem erros com muito mais frequência quando se trata de identificar essas pessoas.

Observe algumas situações reais e busque entender como elas podem gerar prejuízo para determinados indivíduos ou grupos:

EDUCAMÍDIA + IA | CC BY-NC-ND 4.0

1. Um TikToker pede ao MidJourney que gere imagens fotográficas realistas de "pessoa com autismo" e o que ele devolve são centenas de imagens de jovens homens brancos em atitude solitária e sombria. Assista o vídeo do experimento . Que efeitos você acha que isso pode gerar? Isso é bom ou ruim para pessoas com autismo?

2. Uma prefeitura quer lançar um programa de prevenção de doenças crônicas como pressão alta e diabetes, e, para decidir onde as unidades de atendimento serão colocadas, analisa em que áreas da cidade a população gasta mais com essas doenças. Essa é uma boa maneira de fazer o programa chegar até quem mais precisa? Por que?

3. O ChatGPT é um enorme sucesso ao ser lançado. A base de dados que o treinou é um conjunto de milhões de textos na internet, quase todos na língua inglesa. Como isso pode impactar o conteúdo criado no programa?

4. A Amazon fez uma tentativa de utilizar IA para avaliar currículos de candidatos na área de tecnologia; com pouco tempo de uso, percebe que o algoritmo favorece candidatos homens, e descarta currículos onde são mencionadas participações em times ou clubes femininos. Por que esse viés pode ter aparecido?

Notas para o professor:

1. *A produção do MidJourney pode amplificar um estereótipo e preconceitos existentes na sociedade. Pessoas com autismo não podem ser realizadas, felizes, produtivas e bem inseridas na sociedade; pelo contrário, são identificadas com jovens homens brancos, isolados da sociedade – um grupo que é sempre tido como potencialmente problemático e até perigoso. Fonte: video.*

2. *Não é uma boa maneira. Nos EUA o atendimento médico é muito caro, e a maioria da população não tem um bom plano de saúde. A métrica de gastos só vai mostrar onde as pessoas têm mais recursos para usar os serviços médicos – nas áreas mais pobres pode até haver maior incidência dessas doenças crônicas – mas isso não se reflete em demanda nos hospitais porque essa população, não tendo plano de saúde, procura menos os serviços médicos.*

3. *Ao utilizar somente textos em inglês, e portanto visões de mundo ocidentais hegemônicas, o ChatGPT está favorecendo uma cultura sobre as outras. São deixadas de lado outras tecnologias, outras formas de lidar com a natureza, com o trabalho, com as relações e até com a morte. Pesquisas sobre temas que interessam mais a populações marginalizadas não fazem parte da base de dados, nem sua literatura.*

EDUCAMÍDIA + IA | CC BY-NC-ND 4.0

4. *O algoritmo foi treinado com os currículos recebidos nos dez anos anteriores. Como há uma disparidade de gênero na tecnologia, com presença muito maior de homens do que de mulheres, o algoritmo aprendeu que candidatos homens são melhores, e passou a desvalorizar e até a descartar currículos de mulheres. (Fonte: Amazon scraps secret AI recruiting tool that showed bias against women)*

[PARA O EDUCADOR]

Sugestão de sequência didática

1. Utilize o nosso Glossário Anotado EducaMídia+IA para explorar a definição de algoritmo. Proponha a seguinte pergunta: o software é neutro? Ou os algoritmos podem embutir valores?
2. Peça que os alunos façam pesquisas de imagens na internet utilizando os termos "cabelo feio" e "cabelo bonito". Compare os resultados. O que apareceu? Por que? Como eles se sentiram com esses resultados? Como outras pessoas se sentiriam?
3. Mostre o vídeo do UOL: Como os algoritmos espalham racismo e desigualdade de gênero (a partir de 3m22s) e solicite reflexões.
4. Peça que os alunos leiam o texto introdutório deste material e a seguir distribua os exemplos entre grupos para 10 min de discussão. Passeie entre os grupos, orientando a conversa a partir das notas para o professor.
5. Peça aos grupos que exponham para a turma seus exemplos e algumas das reflexões que fizeram.
6. Reflexão final: como podemos identificar vieses e garantir que a tecnologia não viole direitos?

Para saber mais:

CNN Brasil - **Entre Vozes #01: Preconceito programado? Como operam os algoritmos** (podcast)
UOL - **Como os algoritmos espalham racismo e desigualdade de gênero** (vídeo)
Portal Geledés - **"Racismo algorítmico": pesquisador mostra como os algoritmos podem discriminar** (artigo)
LIS (The London Interdisciplinary School) - **How AI Image Generators Make Bias Worse** (video)
Vox - **Why algorithms can be racist and sexist** (artigo - usar o Google tradutor se necessário)

Criado por Mariana Ochs para o EducaMídia | Instituto Palavra Aberta, 2024.
Disponível em acesso livre sob os termos da licença Creative Commons Atribuição / NãoComercial / SemDerivações 4.0 Internacional (CC BY-NC-ND 4.0).