

**UNIVERSIDADE TUIUTI DO PARANÁ**

**RHIAN FELIPE SILVA DE LARA**

**PAPEL DA IA NO REFORÇO DE ATITUDES RACISTAS.**

**CURITIBA**

**2022**

**RHIAN FELIPE SILVA DE LARA**

**PAPEL DA IA NO REFORÇO DE ATITUDES RACISTAS.**

Relatório apresentado ao curso de Ciência da Computação da Universidade Tuiuti do Paraná, como requisito avaliativo da disciplina de IA.

Orientador: Prof.Chaua

**CURITIBA**

**2022**

SUMÁRIO

1	PROBLEMA ÉTICO DA IA . . . . .	3
	REFERÊNCIAS . . . . .	5

## 1 PROBLEMA ÉTICO DA IA

Inteligência Artificial tornou-se uma palavra familiar. Tornou-se também um manipulador de todas as famílias. A explosão descontrolada da IA em todos os negócios e modelos de negócios tem sido um impulsionador fenomenal de crescimento, mas levanta questões que precisam ser respondidas.

Cientistas de dados e pesquisadores de IA irão cada vez mais conduzir o comportamento humano, impactar a forma como as empresas tomam decisões e até mesmo orientar o governo. Além disso, esses modelos passarão cada vez mais de designs tradicionais compreensíveis para humanos para modelos complexos de Deep-Learning que envolvem uma quantidade incrível de complexidade. Na prática, muito disso vem com pouca ou nenhuma discussão em torno da ética de tais sistemas. A IA tem a capacidade de remover o elemento humano em um processo de tomada de decisão. Talvez o mais importante, ele pode tomar decisões puramente objetivas com base em dados objetivos brutos.

Há um crescente reconhecimento da importância dos efeitos políticos, sociais, econômicos e estratégicos da inteligência artificial (IA). Isso levanta questões éticas importantes sobre a programação, uso e regulamentação da IA. À medida que as máquinas estão se aproximando de adquirir habilidades de linguagem semelhantes às humanas, elas também estão absorvendo os preconceitos profundamente arraigados ocultos nos padrões de uso da linguagem. Mas Bryson alertou que a IA tem o potencial de reforçar os preconceitos existentes porque, ao contrário dos humanos, os algoritmos podem não estar equipados para neutralizar conscientemente os preconceitos aprendidos. “Um perigo seria se você tivesse um sistema de IA que não tivesse uma parte explícita que fosse impulsionada por ideias morais, isso seria ruim”, disse ela.

O artigo mais recente mostra que alguns vieses implícitos mais preocupantes vistos em experimentos de psicologia humana também são prontamente adquiridos por algoritmos. As palavras “feminino” e “mulher” estavam mais associadas às profissões de artes e humanidades e ao lar, enquanto “masculino” e “homem” estavam mais próximas das profissões de matemática e engenharia.

E o sistema de IA era mais propenso a associar nomes europeus americanos a palavras agradáveis, como “presente” ou “feliz”, enquanto nomes afro-americanos eram mais comumente associados a palavras desagradáveis.

As descobertas sugerem que os algoritmos adquiriram os mesmos vieses que levam as pessoas (no Reino Unido e nos EUA, pelo menos) a combinar palavras agradáveis e rostos brancos em testes de associação implícita.

Esses preconceitos podem ter um impacto profundo no comportamento humano.

Um estudo anterior mostrou que um currículo idêntico tem 50% mais chances de resultar em um convite para entrevista se o nome do candidato for europeu-americano do que se for afro-americano. Os resultados mais recentes sugerem que os algoritmos, a menos que sejam explicitamente programados para resolver isso, estarão repletos dos mesmos preconceitos sociais.

Outro caso a ser debatido se relaciona com o comportamento da IA conforme pessoas idosas. As barreiras ao acesso tecnológico podem excluir os idosos do processo de pesquisa, design e desenvolvimento de tecnologias digitais. Sua ausência no design e desenvolvimento de tecnologia também pode ser racionalizada com a crença de que os idosos são incapazes de usar a tecnologia. Como tal, os idosos e suas perspectivas raramente estão envolvidos no desenvolvimento de IA e políticas relacionadas, financiamento e serviços de apoio.

As experiências e necessidades únicas dos adultos mais velhos são negligenciadas, apesar da idade ser um preditor mais poderoso do uso da tecnologia do que outras características demográficas, incluindo raça e gênero.

A IA é treinada por dados, e a ausência de adultos mais velhos pode reproduzir ou até amplificar as suposições egípcias acima em sua produção. Muitas tecnologias de IA estão focadas em uma imagem estereotipada de um idoso com problemas de saúde, um segmento estreito da população que ignora o envelhecimento saudável. Isso cria um ciclo de feedback negativo que não apenas desencoraja os adultos mais velhos a usar a IA, mas também resulta em mais perda de dados desses dados demográficos que melhorariam a precisão da IA. Mesmo quando os idosos são incluídos em grandes conjuntos de dados, eles geralmente são agrupados de acordo com divisões arbitrárias pelos desenvolvedores. Por exemplo, os adultos mais velhos podem ser definidos como todos com 50 anos ou mais, apesar das coortes mais jovens serem divididas em faixas etárias mais restritas. Como resultado, os idosos e suas necessidades podem se tornar invisíveis para os sistemas de IA.

Dessa forma, os sistemas de IA reforçam a desigualdade e ampliam a exclusão social para setores da população, criando uma “subclasse digital” composta principalmente por grupos mais velhos, pobres, raciais e marginalizados.

## REFERÊNCIAS

HALL, Lucy B. Clapton, William. "Programming the machine: Gender, race, sexuality, AI, and the construction of credibility and deceit at the border "; Econstor. Disponível em: <<http://hdl.handle.net/10419/250400>>

MOORE, Jesse. "Can A Machine Be Racist?"; Towards Data Science. Disponível em: <<https://towardsdatascience.com/can-a-machine-be-racist-5809b18e5a91>>

DEVLIN, Hannah. "AI programs exhibit racial and gender biases, research reveals"; the Guardian. Disponível em: <<https://www.theguardian.com/technology/2017/apr/13/ai-programs-exhibit-racist-and-sexist-biases-research-reveals>>