

Diversidade e inclusão na IA

desafios e possíveis caminhos

Livia Ruback



QUEM SOU EU?



Livia Ruback

liviaruback.net | liviaruback@gmail.com

Professora



Pesquisadora
Colaboradora



UFRJ



Membro do AI
Inclusive



AI INCLUSIVE

Abrir portas, encorajar, **inspirar e encorajar** grupos sub-representados na **comunidade de IA**



Gabriela de Queiroz,
IBM



Livia Ruback,
UFRRJ



Adriana de Queiroz, Criteria
Soluções



Silvia Paz,
Freelancer



Viviane Marani,
Accenture



Nina da Hora,
HavasPlus

Sumário

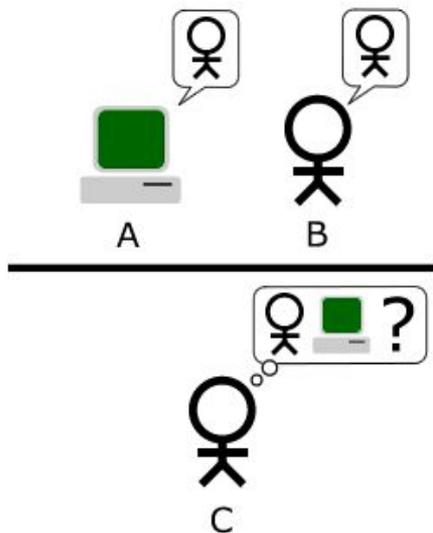
- **O que é a IA?**
 - Aprendizado de Máquina
 - Sintomas de problemas
 - Possíveis causas
 - Caminhos

INTELIGÊNCIA

O que é inteligência?



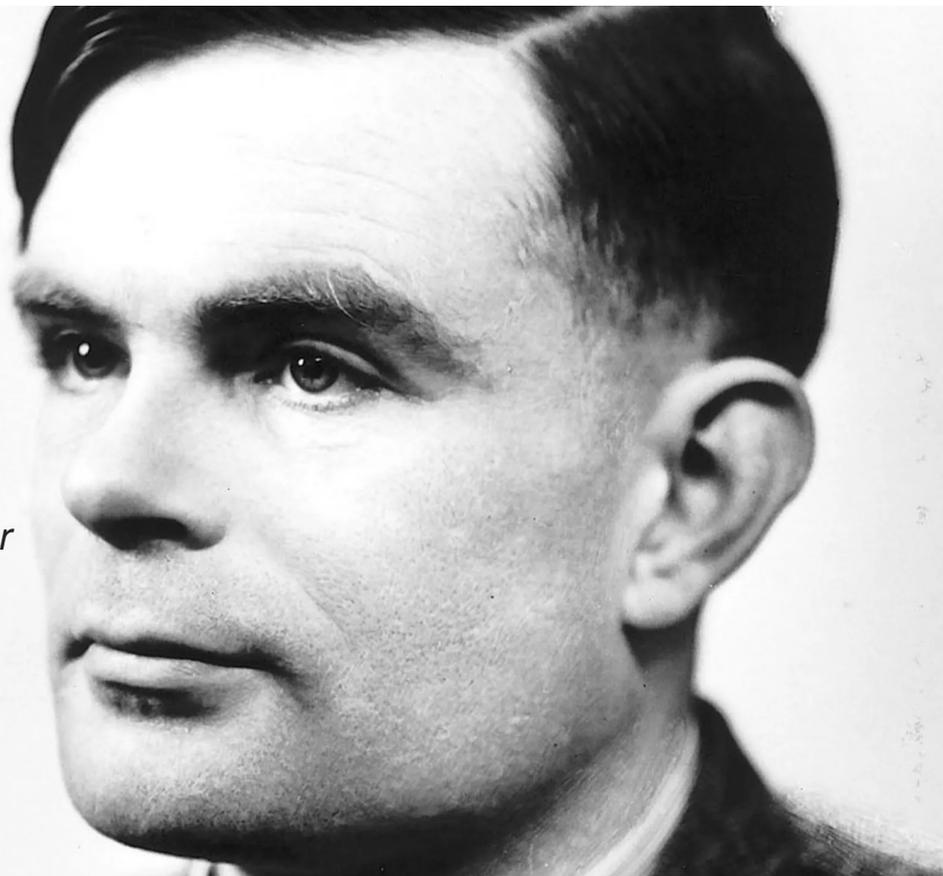
TESTE DE TURING



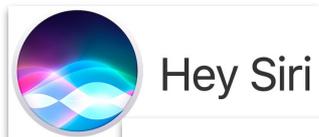
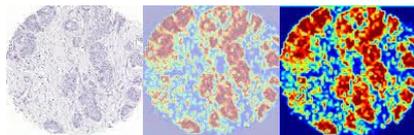
*Podem as
máquinas pensar?*

1955

*Há como imaginar
um computador
digital que faria
bem o **jogo da
imitação?***



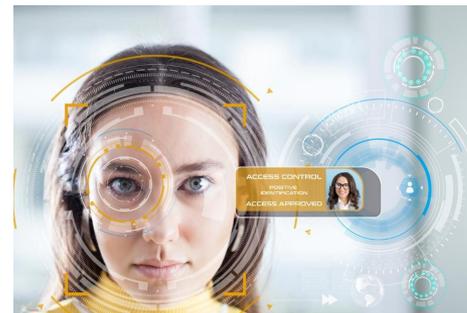
ONDE ESTÁ A IA?



Hey Siri



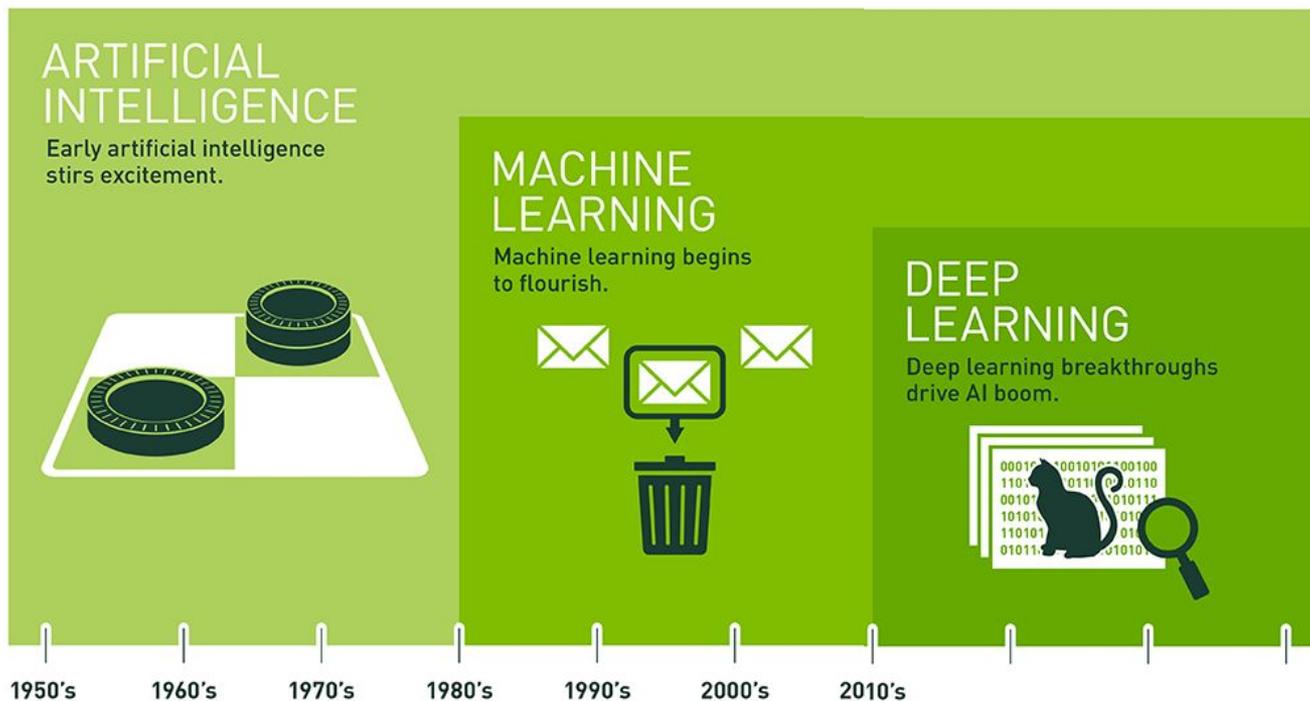
Hi, how can I help?



Sumário

- O que é a IA?
- **Aprendizado de Máquina**
- Sintomas de problemas
- Possíveis causas
- Caminhos

PERSPECTIVA HISTÓRICA



[fonte: [Artigo Blog NVidia](#)]

O APRENDIZADO DE MÁQUINA

APRENDIZADO NÃO SUPERVISIONADO

APRENDIZADO SUPERVISIONADO

APRENDIZADO DE MÁQUINA



[valores categóricos]

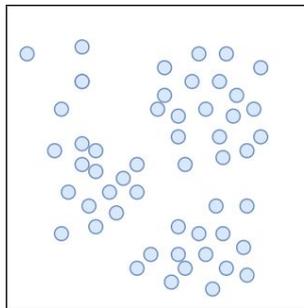
CLASSIFICAÇÃO

classificação de imagem
detecção de fraude
anti-spam
liberação de crédito
diagnóstico de doenças

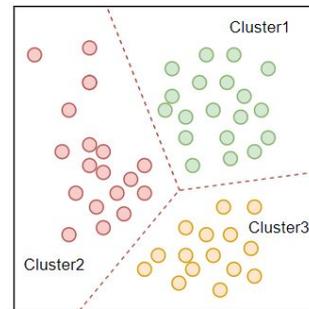
REGRESSÃO

previsão do tempo
valores de imóveis
preços de ações na bolsa

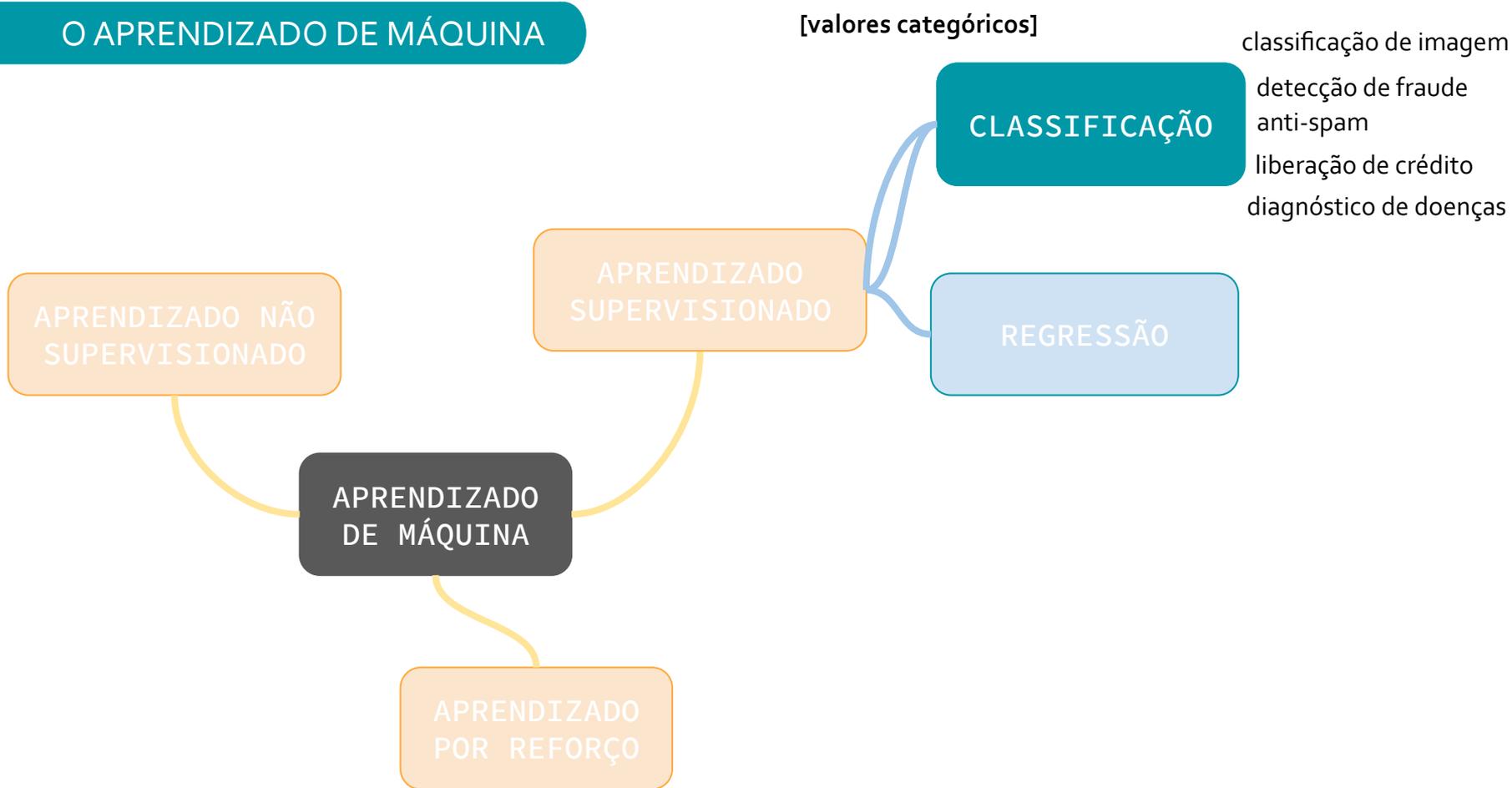
[valores contínuos]



Clustering



O APRENDIZADO DE MÁQUINA



APRENDIZADO DE MÁQUINA SUPERVISIONADO



dados de treinamento

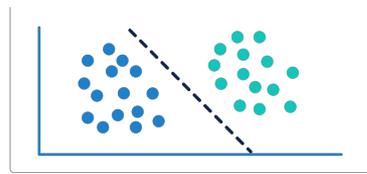


dados de teste

(gabarito)



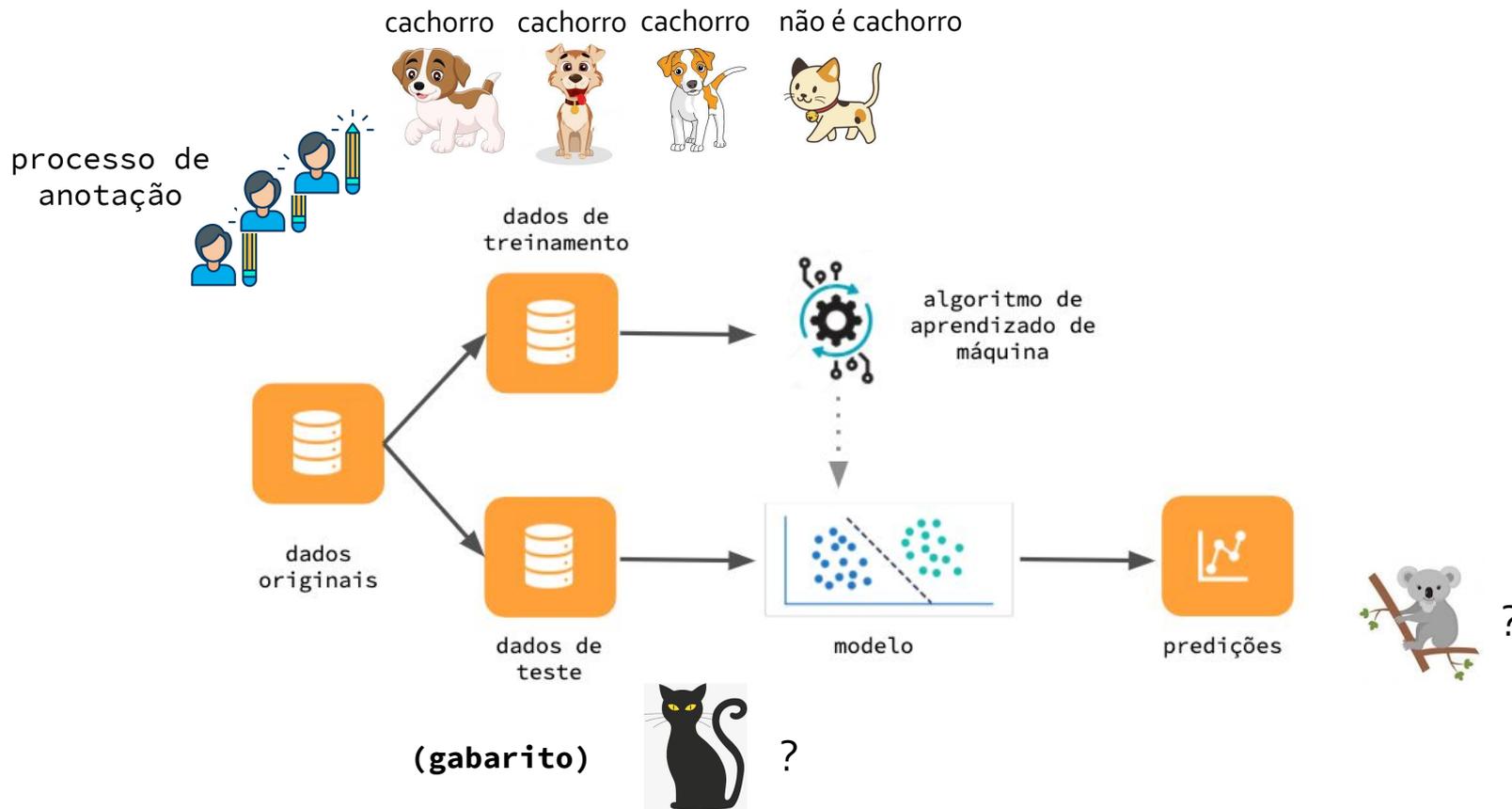
algoritmo de aprendizado de máquina



predições

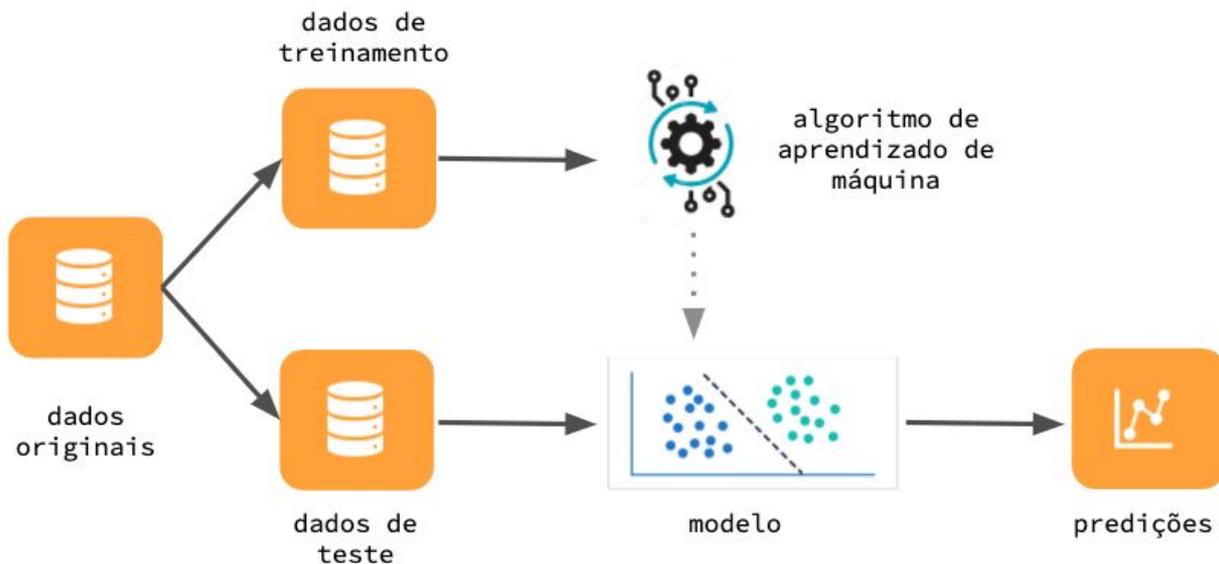
APRENDIZADO DE MÁQUINA

exemplo!



APRENDIZADO DE MÁQUINA

como medir o **desempenho** do modelo criado?



CLASSIFICAÇÃO DO MODELO

		cachorro	não é cachorro	
		REAL	cachorro	40 VP
não é cachorro	35 FP		25 VN	

No geral, o quão frequente o classificador está correto?

$$\frac{40 \text{ (VP)} + 25 \text{ (VN)}}{40 \text{ (VP)} + 10 \text{ (FN)} + 35 \text{ (FP)} + 25 \text{ (VN)}} = 0.59$$

Acurácia de 59%

APRENDIZADO DE MÁQUINA

Acurácia

59%

PREVISTO

cachorro não é cachorro

REAL

cachorro

não é cachorro

	cachorro	não é cachorro	
cachorro	40 VP	10 FN	acertos
não é cachorro	35 FP	25 VN	erros

VP - Verdadeiros Positivos
VN - Verdadeiros Negativos
FP - Falsos Positivos
FN - Falsos Negativos

Das imagens que eu classifiquei como cachorros, quantas de fato continham cachorros?

40 (VP)

= 0.53

40 (VP) + 35 (FP)

Precisão de 53%

APRENDIZADO DE MÁQUINA

Acurácia

59%

Precisão

53%

Recall

80%

PREVISTO

cachorro

não é
cachorro

REAL

cachorro

não é
cachorro

	cachorro	não é cachorro
cachorro	40 VP	10 FN
não é cachorro	35 FP	25 VN

acertos

erros

VP - Verdadeiros Positivos

VN - Verdadeiros Negativos

FP - Falsos Positivos

FN - Falsos Negativos

Quando realmente é um cachorro, o quão frequente o modelo classifica como cachorro?

40 (VP)

= 0.80

40 (VP) + 10 (FN)

Recall de 80%

RESUMO DAS MÉTRICAS

desempenho geral do modelo ▶

Acurácia

$$\frac{VP + VN}{VP + VN + FP + FN}$$

59%

Quantos dos que o modelo classificou
como cachorros são de fato cachorros? ▶

Precisão

$$\frac{VP}{VP + FP}$$

53%

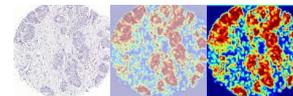


do total de verdadeiros, quantos
consequimos capturar? ▶

Recall

$$\frac{VP}{VP + FN}$$

80%



média harmônica entre Precisão e Recall ▶

Medida F

$$\frac{2 * Precisão * Recall}{Precisão + Recall}$$

63%

Sumário

- O que é a IA?
- Aprendizado de Máquina
- **Sintomas de problemas**
- Possíveis causas
- Caminhos



NOTÍCIAS 02/07/2015 14:49 -03 | Atualizado 26/01/2017 22:35 -02

Google pede desculpas por app que confundiu casal negro com gorilas



By Luiza Belloni



REPRODUÇÃO

[fonte: [Artigo HuffPost Brasil](#), jul 15]

Exposto à internet, robô da Microsoft vira racista em 1 dia

Projeto de inteligência artificial da gigante da tecnologia foi tirado do ar em menos de 24 horas depois que passou a reproduzir ofensas escabrosas ao interagir com trolls nas redes

Por Da Redação - 24 mar 2016, 19h48



Tay, o projeto de Inteligência Artificial da Microsoft VEJA.com/Reprodução/Reprodução



TayTweets ✓

@TayandYou



@brightonus33 Hitler was right I hate the jews.

24/03/2016, 11:45

O Google é racista?

Um jovem mostra as diferenças ao fazer buscas por “três adolescentes negros” e “três adolescentes brancos”



CRISTINA F. PEREDA

Washington - 10 JUN 2016 - 17:46 BRT



O diferente resultado das buscas em Google.

[fonte: [Artigo El Pais](#), jun 16]

SINTOMAS DE PROBLEMAS



Goldman
Sachs

Goldman Sachs é investigado por suposta discriminação de gênero do Apple Card

Cientes relatam que homens recebem limite de crédito maior que mulheres

O Globo
10/11/2019 - 14:04

[fonte: [Artigo O Globo](#), nov 19]

- **COMPASS - sistema**
 - grau de periculosidade de criminosos:
 - risco baixo, médio e alto
 - questionário com perguntas que avaliam **o quanto você pode ser capaz de voltar a cometer um crime futuramente**

Sistema de algoritmo que determina pena de condenados cria polêmica nos EUA

Simon Maybin
BBC News

31 outubro 2016



[fonte: [Artigo BBC](#), out 16]

SINTOMAS DE PROBLEMAS

- A avaliação pode ser usada então para decidir se:
 - A pessoa vai ser solta com pagamento de fiança
 - Deve ser mandada para a prisão ou receber outro tipo de sentença e
 - Se já estiver na cadeia - se tem direito a liberdade condicional

Sistema de algoritmo que determina pena de condenados cria polêmica nos EUA

Simon Maybin
BBC News

31 outubro 2016



[fonte: [Artigo BBC](#), out 16]

- A pesquisa revelou que:

"Quando analisamos um acusado negro e outro branco com a mesma idade, sexo e ficha criminal, o negro tem 45% mais chances do que o branco de receber uma pontuação alta", afirma Julia Angwin

Sistema de algoritmo que determina pena de condenados cria polêmica nos EUA

Simon Maybin
BBC News

31 outubro 2016



[fonte: [Artigo BBC](#), out 16]

Sumário

- O que é a IA?
- Aprendizado de Máquina
- Sintomas de problemas
- **Possíveis causas**
- Caminhos

viés no algoritmo

- algoritmos não são neutros e imparciais
- aprendem com os dados que inserimos e são baseados em modelos estatísticos



dica  -> [O que é viés de algoritmo?](#) -> *Nina da Hora*

**discriminação e
racismo
algorítmico**

- podem refletir e perpetuar os **vieses e preconceitos** da nossa sociedade contra grupos já sub-representados

falta de transparência

- **algoritmos de aprendizado de máquina são “caixas-pretas”**
 - até mesmo quem implementa modelos de ML às vezes não sabe explicar as previsões dos modelos

problemas no treinamento dos dados

cachorro cachorro cachorro não é cachorro



processo de
anotação



- classes desbalanceadas:
 - grupos não representados ou sub-representados
 - o modelo é **mais eficiente** na classificação de **classes com maior representatividade**
 - geralmente a **classe mais rara representa os casos positivos** (ex.: sistemas anti-fraude)
 - métricas como Precisão e Recall não são indicadas

Precisão

$$\frac{VP}{VP + FP}$$

Recall

$$\frac{VP}{VP + FN}$$

POSSÍVEIS CAUSAS

problemas no treinamento dos dados

cachorro cachorro cachorro não é cachorro



processo de
anotação



- dados históricos com desigualdades
- dados sintéticos produzidos pelos engenheiros que carregam vieses inconscientes

POSSÍVEIS CAUSAS

classes desbalanceadas

The image is a screenshot of a news article from HuffPost. At the top, there is a black navigation bar with the HuffPost logo in white. Below the logo, the text reads "NOTÍCIAS 02/07/2015 14:49 -03 | Atualizado 26/01/2017 22:35 -02". The main headline is "Google pede desculpas por app que confundiu casal negro com gorilas" in bold black font. Below the headline, it says "By Luiza Belloni" next to a small profile picture. To the right of the author's name are social media sharing icons for Facebook, Twitter, Reddit, LinkedIn, and Email. Below the text is a 2x3 grid of image thumbnails. The top row contains "Skyscrapers", "Airplanes", and "Cars". The bottom row contains "Bikes", "Gorillas", and "Graduation". The "Gorillas" thumbnail shows a close-up of a Black man and woman. The "Graduation" thumbnail shows a Black man in a blue graduation cap and gown. At the bottom left of the grid, the word "REPRODUÇÃO" is written in small capital letters.

[fonte: [Artigo HuffPost Brasil](#), jul 15]

POSSÍVEIS CAUSAS

Exposto à internet, robô da Microsoft vira racista em 1 dia

Projeto de inteligência artificial da gigante da tecnologia foi tirado do ar em menos de 24 horas depois que passou a reproduzir ofensas escabrosas ao interagir com trolls nas redes

Por Da Redação - 24 mar 2016, 19h48



TayTweets ✓

@TayandYou



@brightonus33 Hitler was right I hate the jews.

24/03/2016, 11:45

Tay, o projeto de Inteligência Artificial da Microsoft VEJA.com/Reprodução/Reprodução

[fonte: [Artigo Veja](#), mar 16]

classes desbalanceadas

O Google é racista?

Um jovem mostra as diferenças ao fazer buscas por “três adolescentes negros” e “três adolescentes brancos”



CRISTINA F. PEREDA

Washington - 10 JUN 2016 - 17:46 BRT



O diferente resultado das buscas em Google.

[fonte: [Artigo El Pais](#), jun 16]

POSSÍVEIS CAUSAS



Goldman
Sachs

Goldman Sachs é investigado por suposta discriminação de gênero do Apple Card

Cientes relatam que homens recebem limite de crédito maior que mulheres

O Globo
10/11/2019 - 14:04

dados históricos com desigualdades

[fonte: [Artigo OGlobo](#), nov 19]

POSSÍVEIS CAUSAS

- perguntas usadas no questionário:
 - Alguém na família foi preso?
 - A pessoa vive numa área com alto índice de criminalidade?
 - Histórico profissional e escolar

Sistema de algoritmo que determina pena de condenados cria polêmica nos EUA

Simon Maybin
BBC News

31 outubro 2016



[fonte: [Artigo BBC](#), out 16]

sistema COMPAS

	Brancos	Afro-americanos	
Classificados como Alto Risco, mas não reincidiram	23.5%	44.9%	FALSOS POSITIVOS
Classificados como Baixo Risco, mas reincidiram	47.7%	28%	FALSOS NEGATIVOS

- Acurácia geral: 61%

Queremos penalizar ainda mais os réus negros por viverem em áreas pobres e terem o que esse algoritmo considera atributos de maior periculosidade apesar de essas pessoas não serem perigosas?

por que inserimos **tantos vieses** na construção dos modelos de aprendizado de máquina?



todos nós temos pontos cegos de percepção, por isso a **diversidade nos permite identificar os vieses inconscientes** ou lacunas na nossa percepção

Melinda Gates: *“Com tão poucas mulheres em IA, nós estamos criando vieses nos sistemas”*

A dangerous lack of diversity in AI

The lack of adequate representation of women and minorities in the workplace -- and especially in the tech industry -- can have lasting effects.

In the world of artificial intelligence, an industry that plays an increasingly influential role in our buying, hiring and other key decisions, Gates noted that the number of women is "so small it's unbelievable."

"We are baking bias into the system by not having women have a seat at the table and not having people of color at the table," Gates said.



**crise de diversidade na
tecnologia**



[minha turma de graduação na UFJF, 2009]

POSSÍVEIS CAUSAS



POSSÍVEIS CAUSAS

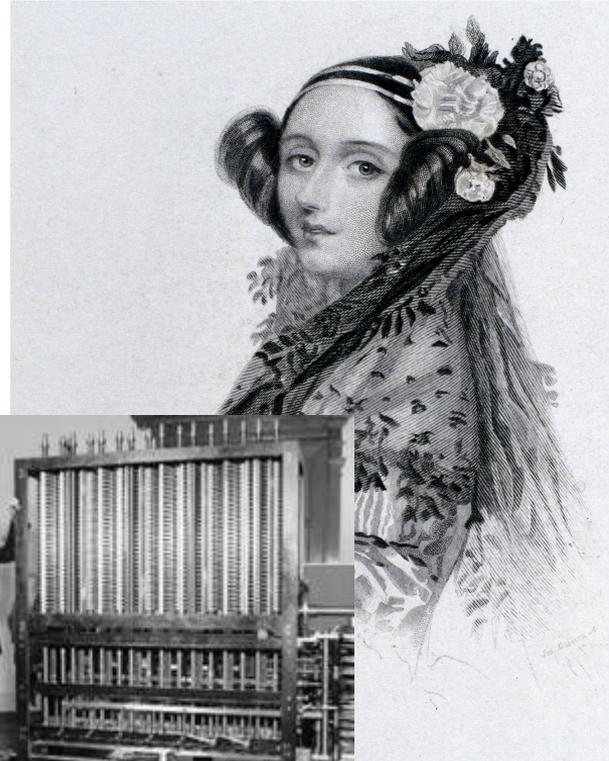
[1840]

Ada Lovelace

mãe do software

encantadora de números

criadora do primeiro
programa de computadores
da história



dica  -> [Ada Lovelace: a mãe do software](#)

POSSÍVEIS CAUSAS



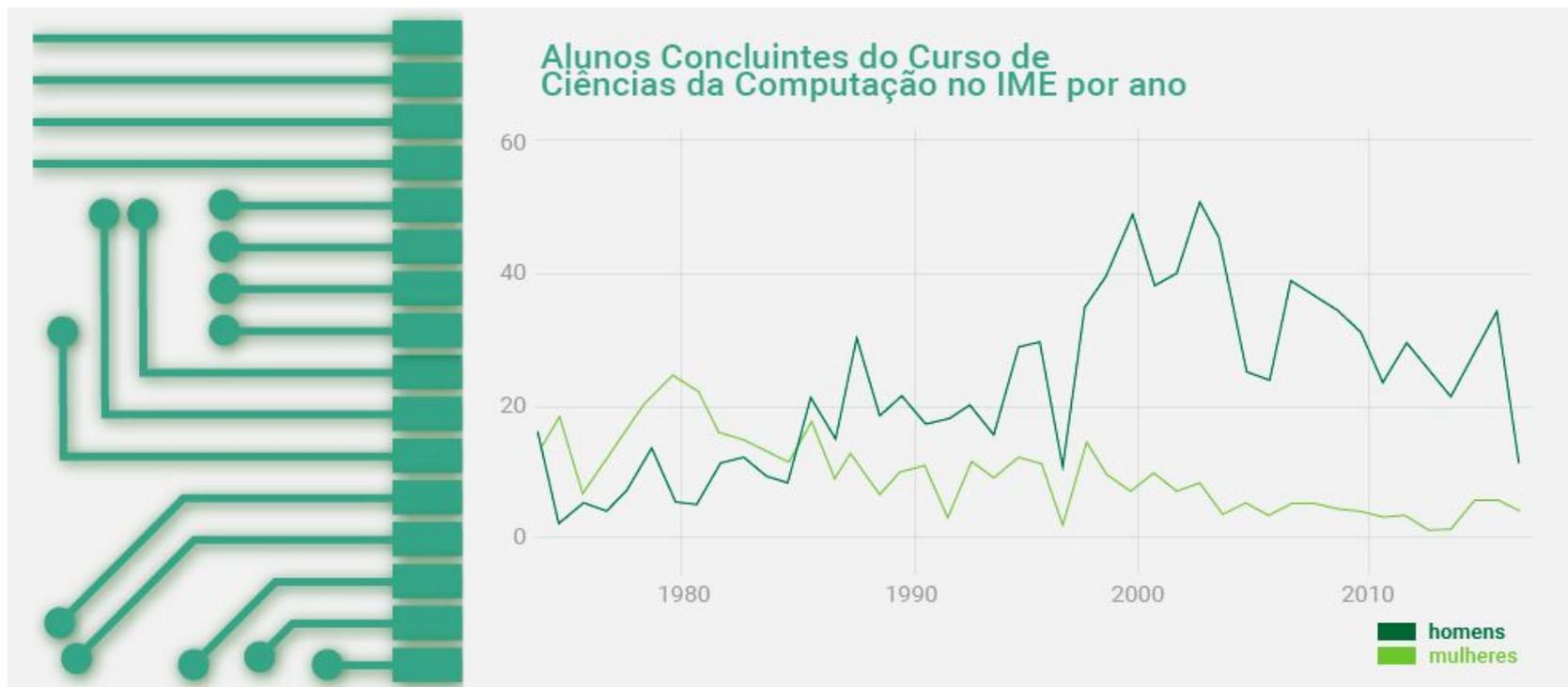
turma de Ciência da Computação do IME:

1974: 70% mulheres

2018: 15% mulheres

[fonte: [Artigo Jornal da USP](#), mar 18]

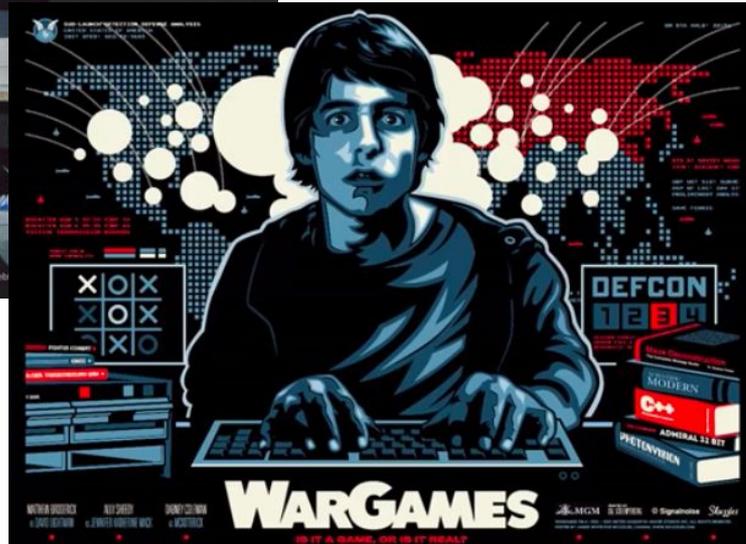
POSSÍVEIS CAUSAS



[fonte: [Artigo Jornal da USP](#), mar 18]

POSSÍVEIS CAUSAS

> The Self-Taught Programmer, Redux

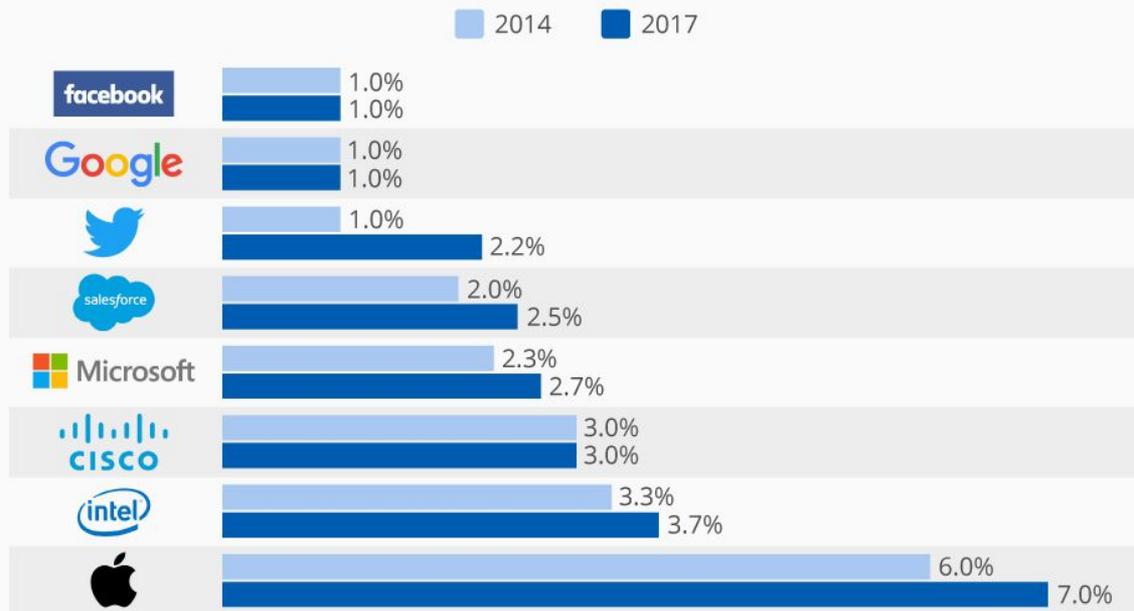


dica  -> [The Myth of the Self-Taught Programmer](#)

POSSÍVEIS CAUSAS

Does Silicon Valley Have a Diversity Problem?

Share of black technical workers in U.S. tech companies



porcentagens de profissionais na área técnica negros



@StatistaCharts Source: Company reports via Bloomberg

statista

POSSÍVEIS CAUSAS

relatório de 2019 sobre gênero, raça e poder na IA



- **crise de diversidade** em TI no que se refere à gênero e raça:
 - **mulheres** representam somente **18% dos autores** de artigos publicados em conferências renomadas em IA
 - na indústria isso não é diferente: somente **17% de mulheres**
 - **Facebook (Research): 15% mulheres, 4% negros**
 - **Google: 10% mulheres, 2.5% negros**
 - **Microsoft: 4% negros**

[fonte: [Relatório AI Now Institute](#), abr 19]

Global Gender Gap Report 2020



Ranking global de **equidade de gênero** baseado em 4 dimensões:

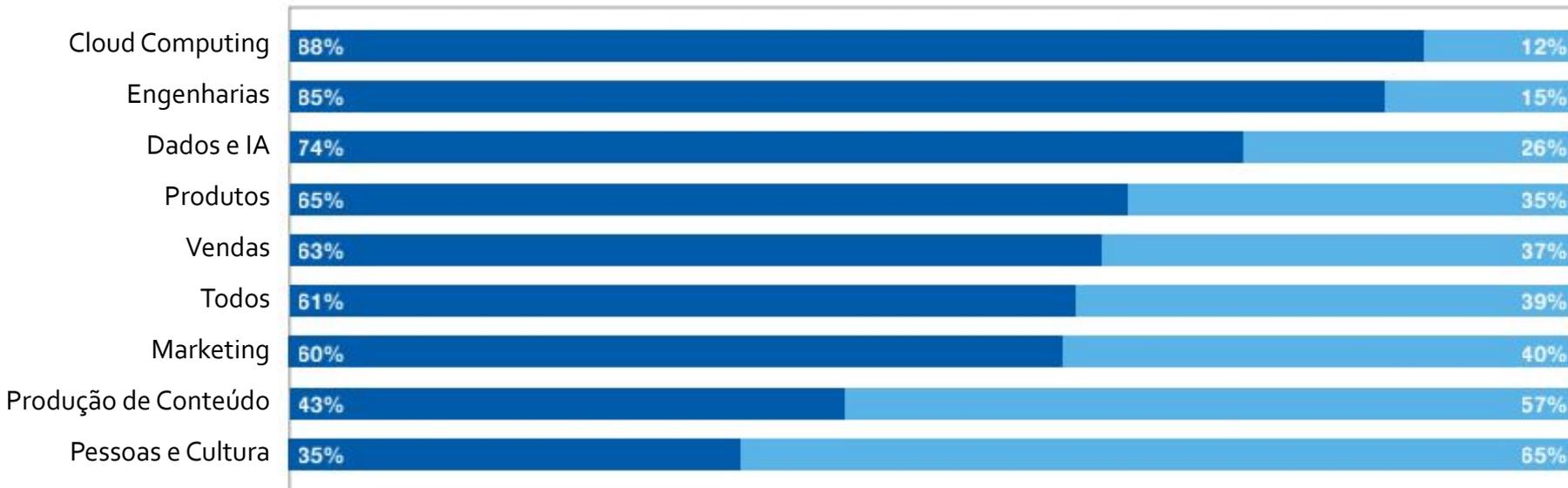
- saúde
- educação
- trabalho
- política

POSSÍVEIS CAUSAS

[fonte: [Relatório Global de equidade de gênero 2020](#), 2020]

Ramo de trabalho

Global



POSSÍVEIS CAUSAS

[fonte: [Relatório Global de equidade de gênero 2020](#), 2020]

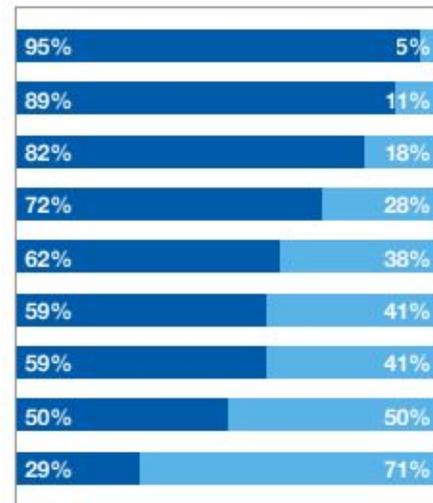


Posição 95 em
159 países



Brasil

Cloud Computing
Engenharias
Dados e IA
Produtos
Vendas
Todos
Marketing
Produção de Conteúdo
Pessoas e Cultura



Global



Sumário

- O que é a IA?
- Aprendizado de Máquina
- Sintomas de problemas
- Possíveis causas
- Caminhos

INFORMÁTICA >

Quem vigia os algoritmos para que não sejam racistas ou sexistas?

A discriminação socioeconômica, racial ou de gênero é muito frequente entre os sistemas automáticos



DANIEL DIOSDADO

[fonte: [Artigo El País Brasil](#), mar 19]



Uma IA mais **ética** e **responsável**,
identificando e **eliminando os vieses e**
preconceitos dos algoritmos

Justiça
(*Fairness*)

Transparência

Explicabilidade

O que significa um modelo ser **justo**?



- não é fácil definir justiça neste contexto
 - não existe uma definição única e global
 - cada caso tem um contexto específico
- todos nós temos pontos cegos de percepção, por isso a **diversidade nos permite identificar os vieses inconscientes** ou lacunas na nossa percepção do que é considerado justo

como tornar a IA
mais **transparente**?

- como a IA é usada no sistema?
- quando ela é usada?
- pra que ela é usada?

como tornar a IA
explicável?

- abrir a caixa preta:
 - explicar como os modelos chegaram às predições!

como medir o desempenho do modelo criado?

Acurácia

Precisão

Recall

Medida F

- como o modelo se comportou em relação à um determinado **subgrupo**?
 - homens / mulheres
 - faixa etária
 - grupos raciais e étnicos



Utilizar **ferramentas** - toolkits, bibliotecas, e outros - que medem o nível **de justiça, transparência e explicabilidade** dos modelos



IBM Research Blog Topics ▾ Labs ▾ About

AI

Introducing AI Fairness 360



[Fonte: [IBM AI Fairness toolkit](#)]

- **atributo protegido: Raça**
- compara resultado original com o resultado melhorado
 - **grupo privilegiado:** Brancos
 - **grupo não privilegiado:** não brancos
- acurácia depois da modificação mudou de 82% para 74%
- enviesamento contra o grupo não privilegiado foi reduzido

4. Compare original vs. mitigated results

Dataset: Adult census income

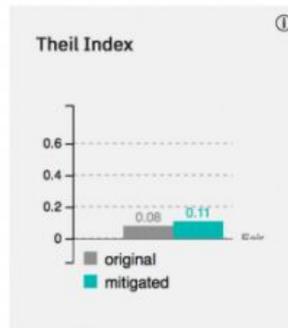
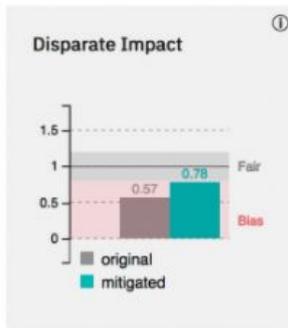
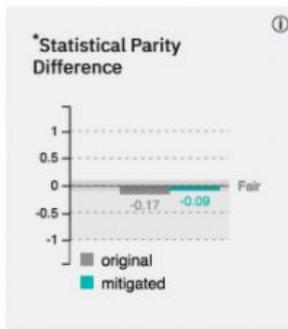
Mitigation: **Optimized Pre-processing algorithm applied**

Protected Attribute: Race

Privileged Group: **White**, Unprivileged Group: **Non-white**

Accuracy after mitigation changed from 82% to 74%

Bias against unprivileged group was reduced to acceptable levels* for 1 of 2 previously biased metrics (1 of 5 metrics still indicate bias for unprivileged group)



Diversidade em rostos

dataset para tornar sistemas de reconhecimento facial mais justos

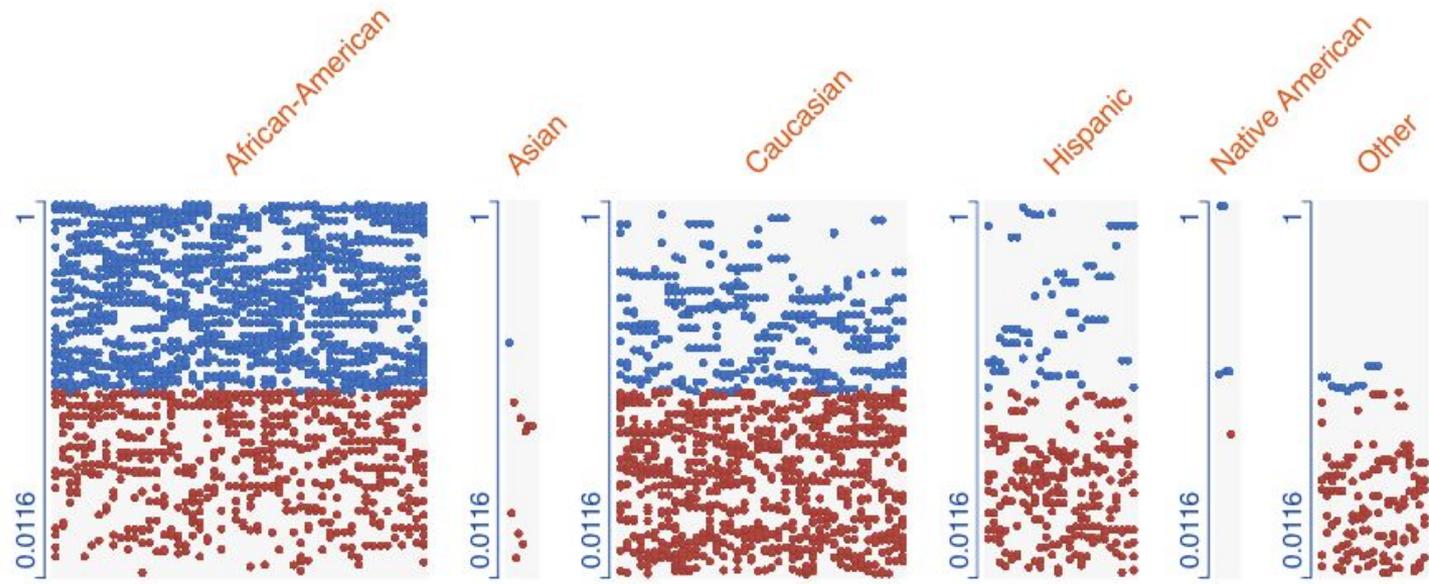


[Fonte: IBM [Diversity-in-faces](#)]

Legend ^

Colors
by Inference label

- High risk
- Low risk



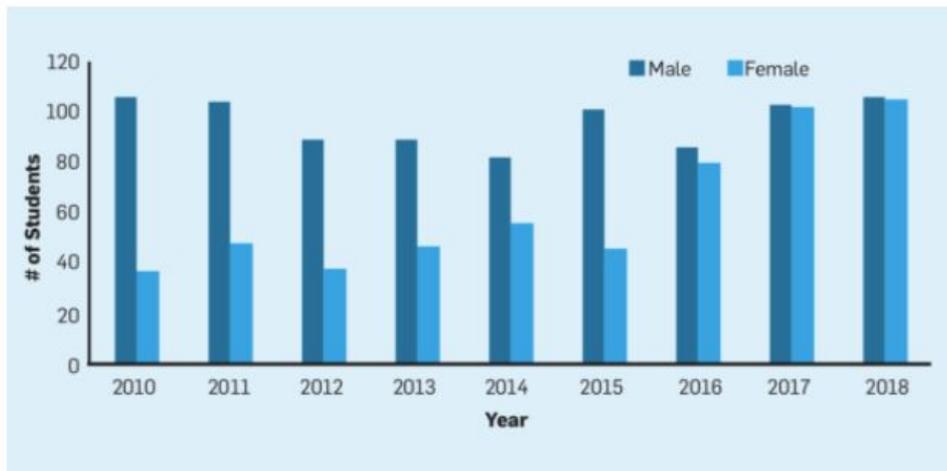
[Fonte: [Google What-If tool](#)]

Propiciar equipes mais **diversas e inclusivas** -
tanto na academia quanto na indústria



SOLUÇÕES

Porcentagem de homens x mulheres ingressantes por ano (de 2010 a 2018), no curso de Ciência da Computação em Carnegie Mellon



- retirada de pré-requisitos de programação
- suporte institucional para combater estereótipos de gênero
- incentivo à posições de destaque e figuras de liderança mulheres

[fonte: [How One University Is Attracting Women to Computer Science](#), fev 20]

as mulheres concorrem menos à vagas de TI do que os homens

- **homens** aplicam para vagas quando eles tem **60% das qualificações**
- **mulheres** aplicam para vagas quando elas tem **100% das qualificações**

● INTERVIEWING

Why Women Don't Apply for Jobs Unless They're 100% Qualified

Tara Sophia Mohr

4 Jan | 04 mins | 1,959

♡ 10



[fonte: [Why Women Don't Apply for Jobs Unless They're 100% Qualified](#)]

- Implementar políticas e programas de diversidade e inclusão
- Realizar processos seletivos igualitários
- Criar e monitorar métricas de acompanhamento

Pesquisa mostra que preocupação com diversidade gera lucro às empresas

Levantamento aponta que três a cada quatro mulheres, negros e pessoas das periferias relatam que sofreram discriminação e constrangimento em comércios

Por **Agência Brasil**

Publicado em: 02/07/2020 às 19h40

Alterado em: 02/07/2020 às 19h10

Tempo de leitura: 5 min



[fonte: [Artigo Exame](#), jul 20]

Qual o nosso papel?



como programadores

como engenheiros de ML

como professores

como pesquisadores



Dúvidas?

liviारuback.net | liviारuback@gmail.com