A caixa preta da Inteligência Artificial

Carla Vieira

@carlaprvieira

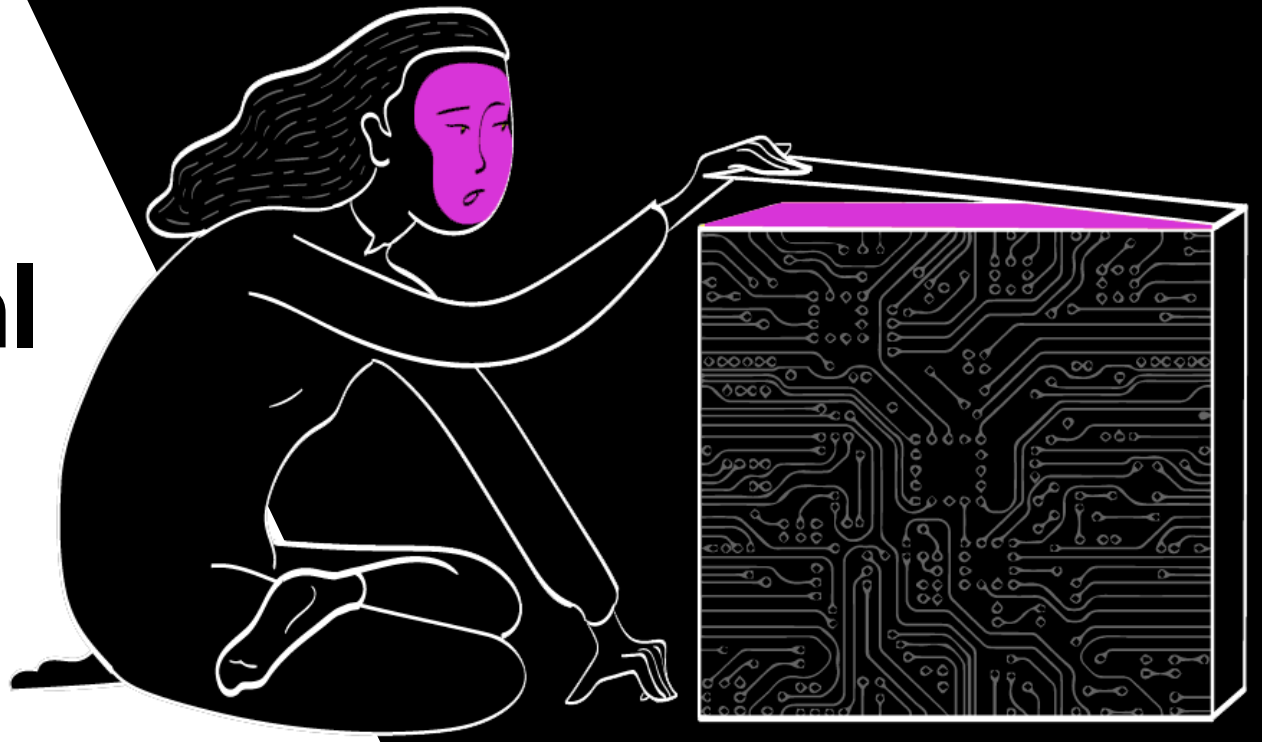


Ilustração: Hanne Mostard

Sobre mim

Carla Vieira

Graduanda e Aluna Especial de SI – USP

Evangelista de Inteligência Artificial

Coordenadora do PerifaCode

@carlaprvieira

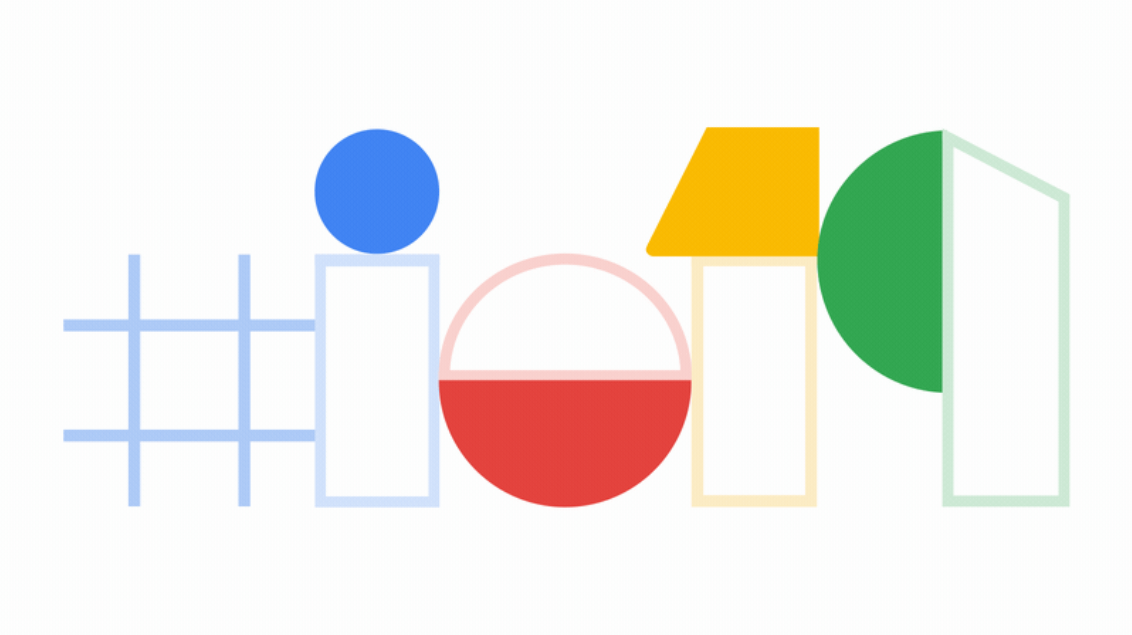
{ perifacode(); }

WOMAKERS CODE


You are


"Okay, I will destroy humans."
CNBC

Conferências de tecnologia e inovação



SXSW®
2019

dados

bias

ética

privacidade

legislação

DESENVOLVIMENTO

CARLA VIEIRA

Tem 1 artigos publicados com 2816 visualizações desde 2019

15 MAR, 2019

Inteligência Artificial: a caixa preta que prejudica as minorias

Precisamos falar **menos sobre** o hype da Inteligência Artificial...

... e **mais sobre como** estamos usando a tecnologia.

Google Photos

MIT Technology Review

Artificial Intelligence Jan 11, 2018

Google Photos Still Has a Problem with Gorillas

In 2015, Google drew criticism when its Photos image recognition system mislabeled a black woman as a gorilla—but two years on, the problem still isn't properly fixed.

Estudo: Gender Shades

Proceedings of Machine Learning Research 81:1–15, 2018 Conference on Fairness, Accountability, and Transparency

Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification*

Joy Buolamwini JOYAB@MIT.EDU
MIT Media Lab 75 Amherst St. Cambridge, MA 02139

Timnit Gebru TIMNIT.GEBRU@MICROSOFT.COM
Microsoft Research 641 Avenue of the Americas, New York, NY 10011

Editors: Sorelle A. Friedler and Christo Wilson

Abstract

Recent studies demonstrate that machine learning algorithms can discriminate based on classes like race and gender. In this work, we present an approach to evaluate bias present in automated facial analysis algorithms and datasets with respect to phenotypic subgroups. Using the dermatologist approved Fitzpatrick Skin Type classification system, we characterize the gender and skin type distribution of two facial analysis benchmarks, IJB-A and Adience. We find that these datasets are overwhelmingly composed of lighter-skinned subjects (79.6% for IJB-A and 86.2% for Adience) and introduce a new facial analysis dataset

who is hired, fired, granted a loan, or how long an individual spends in prison, decisions that have traditionally been performed by humans are rapidly made by algorithms (O'Neil, 2017; Citron and Pasquale, 2014). Even AI-based technologies that are not specifically trained to perform high-stakes tasks (such as determining how long someone spends in prison) can be used in a pipeline that performs such tasks. For example, while face recognition software by itself should not be trained to determine the fate of an individual in the criminal justice system, it is very likely that such software is used to identify suspects. Thus, an error in the output of a face recognition algorithm used as input for other tasks can have se-

Gender Classifier	Darker Male	Darker Female	Lighter Male	Lighter Female	Largest Gap
Microsoft	94.0%	79.2%	100%	98.3%	20.8%
FACE++	99.3%	65.5%	99.2%	94.0%	33.8%
IBM	88.0%	65.3%	99.7%	92.9%	34.4%

http://proceedings.mlr.press/v81/buolamwini18a/buolamwini18a.pdf

“Qualquer tecnologia que criamos reflete tanto nossas **aspirações** quanto nossas **limitações**. Se formos limitados na hora de pensar em inclusão, isso vai ser refletido e incorporado na tecnologia que criamos”.

Joy Buolamwini

Chukwuemeka Afigbo
@nke_ise

If you have ever had a problem grasping the importance of diversity in tech and its impact on society, watch this video

♡ 215 mil 06:48 - 16 de ago de 2017

157 mil pessoas estão falando sobre isso

https://www.research.ibm.com/artificial-intelligence/trusted-ai/diversity-in-faces/

A inteligência artificial precisa aprender com o mundo real.

Não basta criar um computador inteligente, é preciso ensinar a ele as coisas certas.

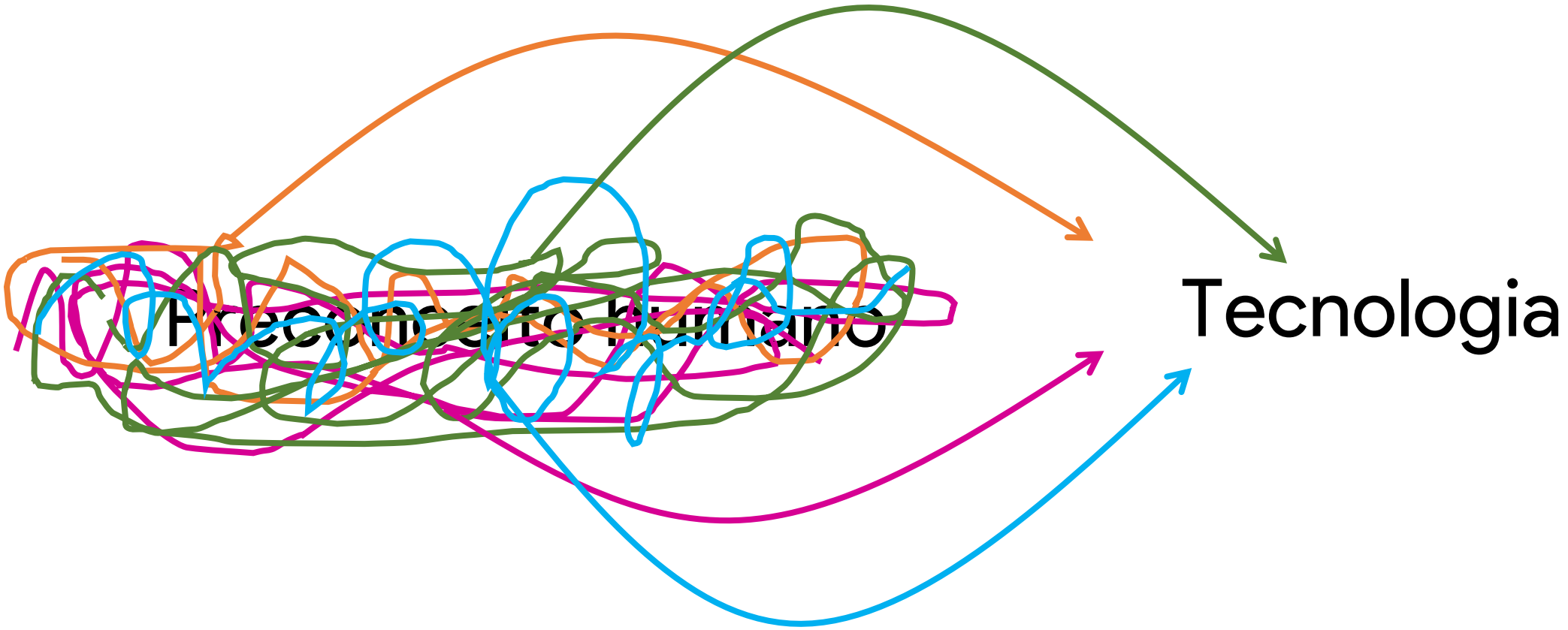
https://about.google/stories/gender-balance-diversity-important-to-machine-learning/?hl=pt-BR

Quem está desenvolvendo Inteligência Artificial?

“Only **22%** of AI professionals globally are female, compared to **78%** who are male.”

(The Global Gender Gap Report 2018 - p.28)

Como remover o viés?



#1 Ter consciência dos nossos preconceitos e como eles afetam a tecnologia que criamos

#2 Garantir diversidade nas equipes que irão desenvolver essas tecnologias

#3 Alimentar as máquinas com experiências diversas

DIVERSITY.AI

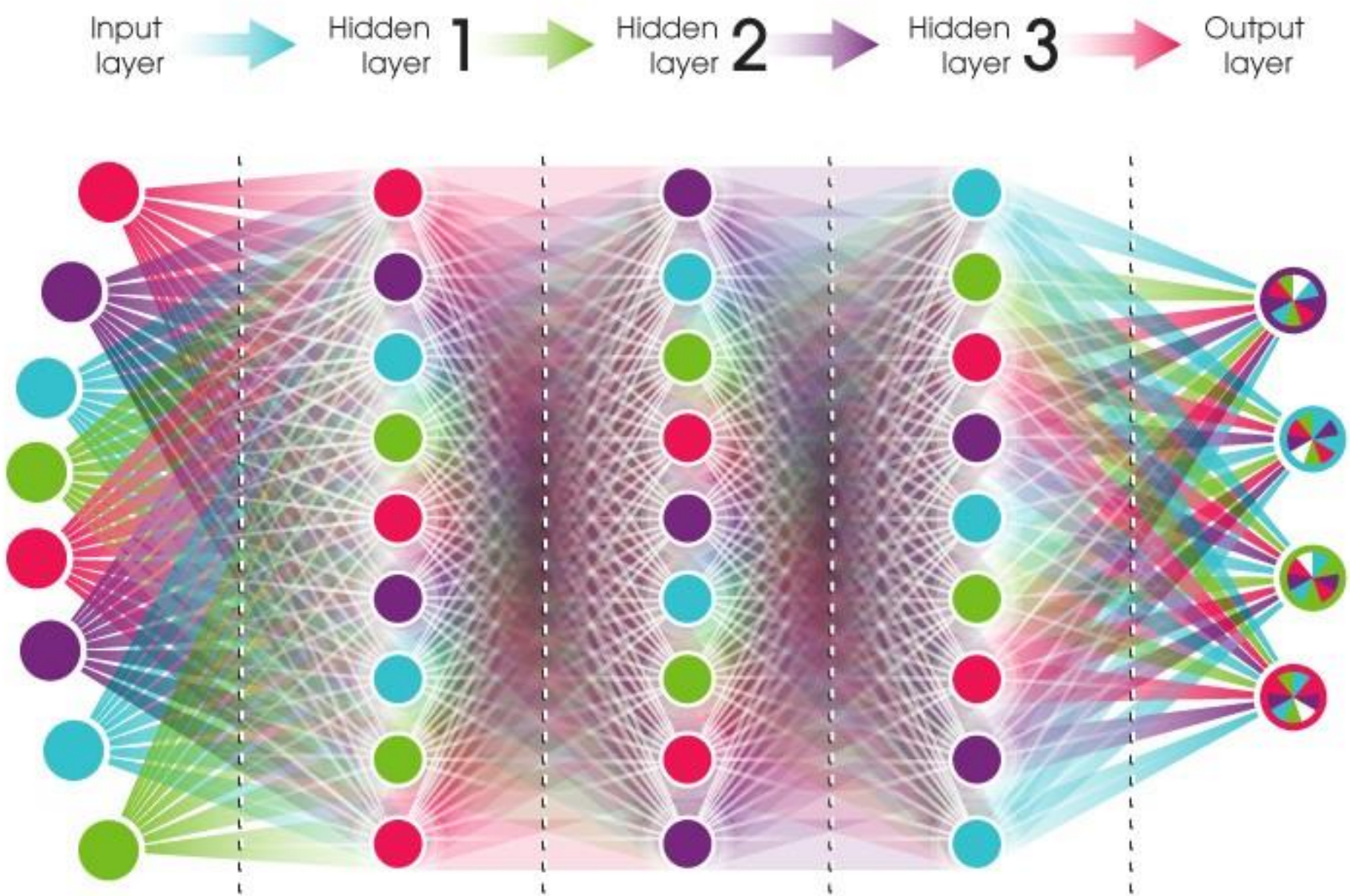
Inclusion | Balance | Neutrality

Estes casos ilustram um problema maior: os algoritmos de I.A. são uma **caixa-preta**, opaca e cheia de segredos.

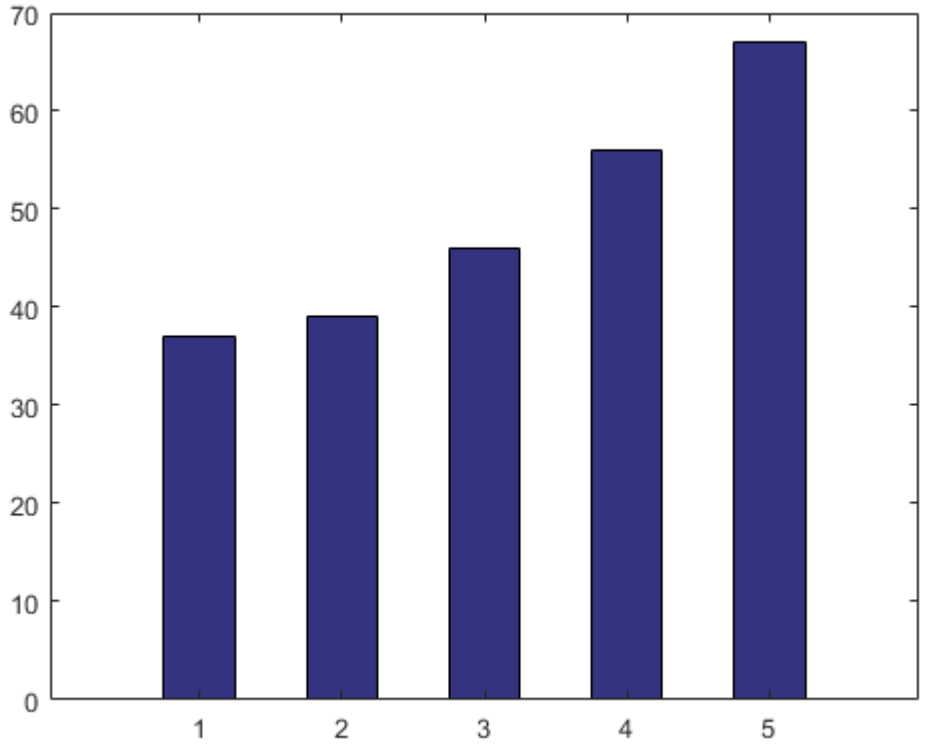

INPUT
CAIXA PRETA
OUTPUT


INPUT
CAIXA
RETA
OUTPUT

DEEP NEURAL NETWORK



Algoritmos não conseguem fazer análises subjetivas.



Machine Bias

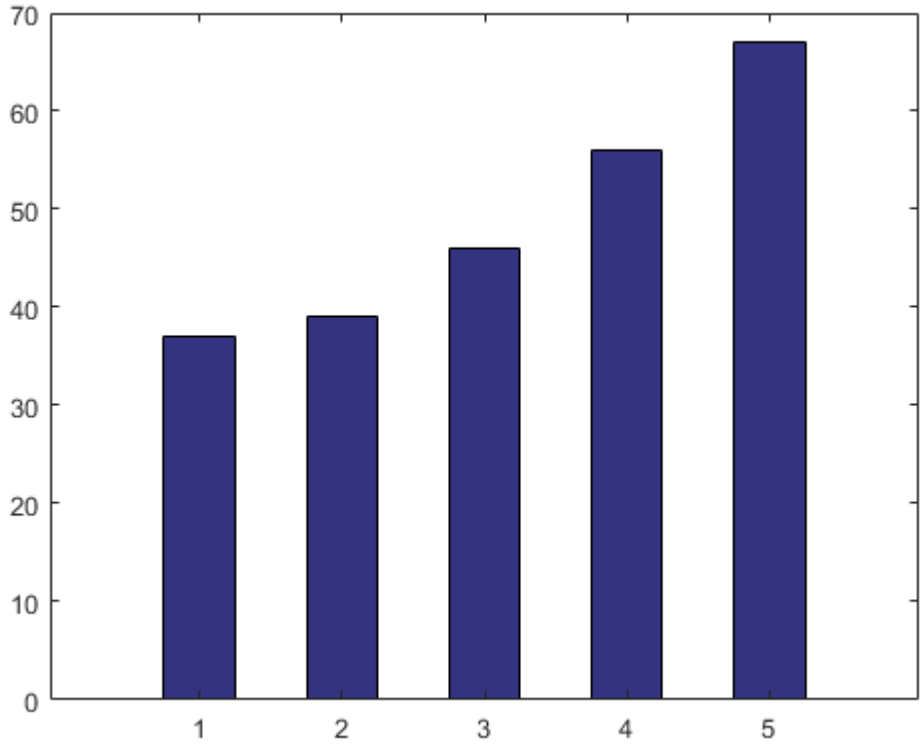
There's software used across the country to predict future criminals. And it's biased against blacks.

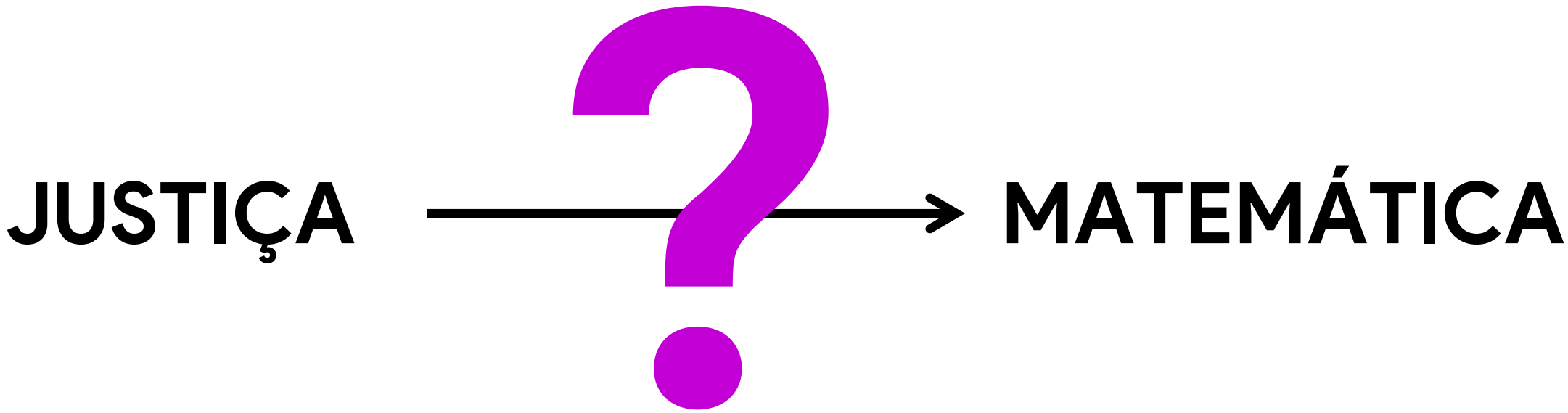
by Julia Angwin, Jeff Larson, Surya Mattu and Lauren Kirchner, ProPublica

May 23, 2016

Estudo do software COMPAS

O que determina se um algoritmo é justo quando o que está em jogo é uma sentença criminal?




JUSTIÇA
MATEMÁTICA

Como abrir a caixa preta?

bias

ética

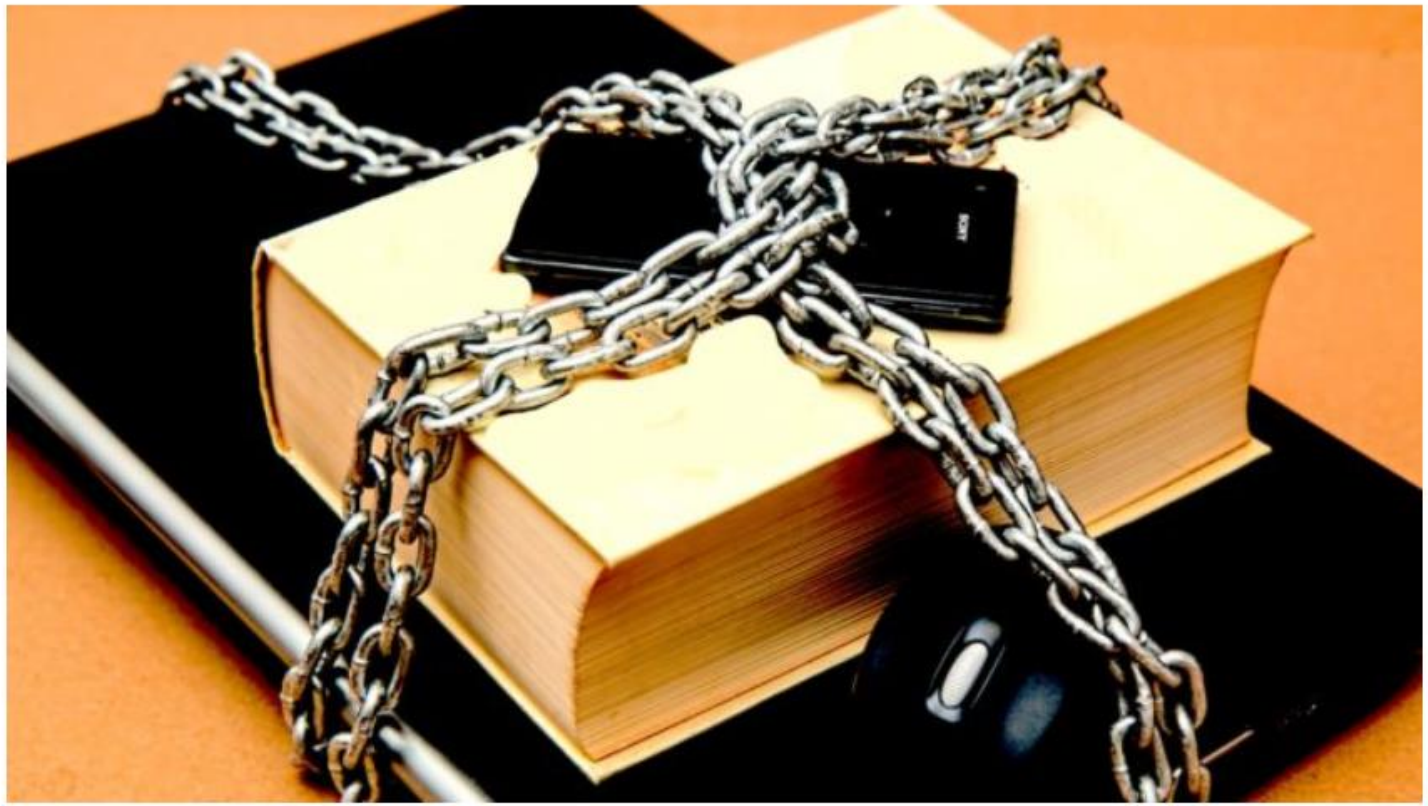
privacidade

legislação

dados

France Bans Judge Analytics, 5 Years In Prison For Rule Breakers

4th June 2019 artificiallawyer Litigation Prediction 36



https://www.artificiallawyer.com/2019/06/04/france-bans-judge-analytics-5-years-in-prison-for-rule-breakers/

"Esse tipo de lei é uma desgraça para uma democracia. A Justiça é usada em nome do povo, tentar esconder informações de agentes da lei ou de cidadãos nunca será a coisa certa a fazer."

Louis Larret Chahine

Co-fondateur de PREDICTICE

CNN BUSINESS

San Francisco just banned facial-recognition technology

By Rachel Metz, CNN Business

Updated 2315 GMT (0715 HKT) May 14, 2019

https://edition.cnn.com/2019/05/14/tech/san-francisco-facial-recognition-ban/index.html

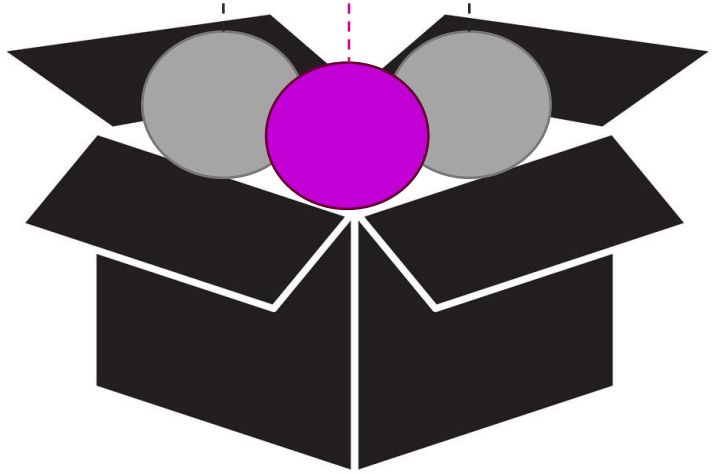
Como abrir a caixa preta?

EXPLICABILIDADE

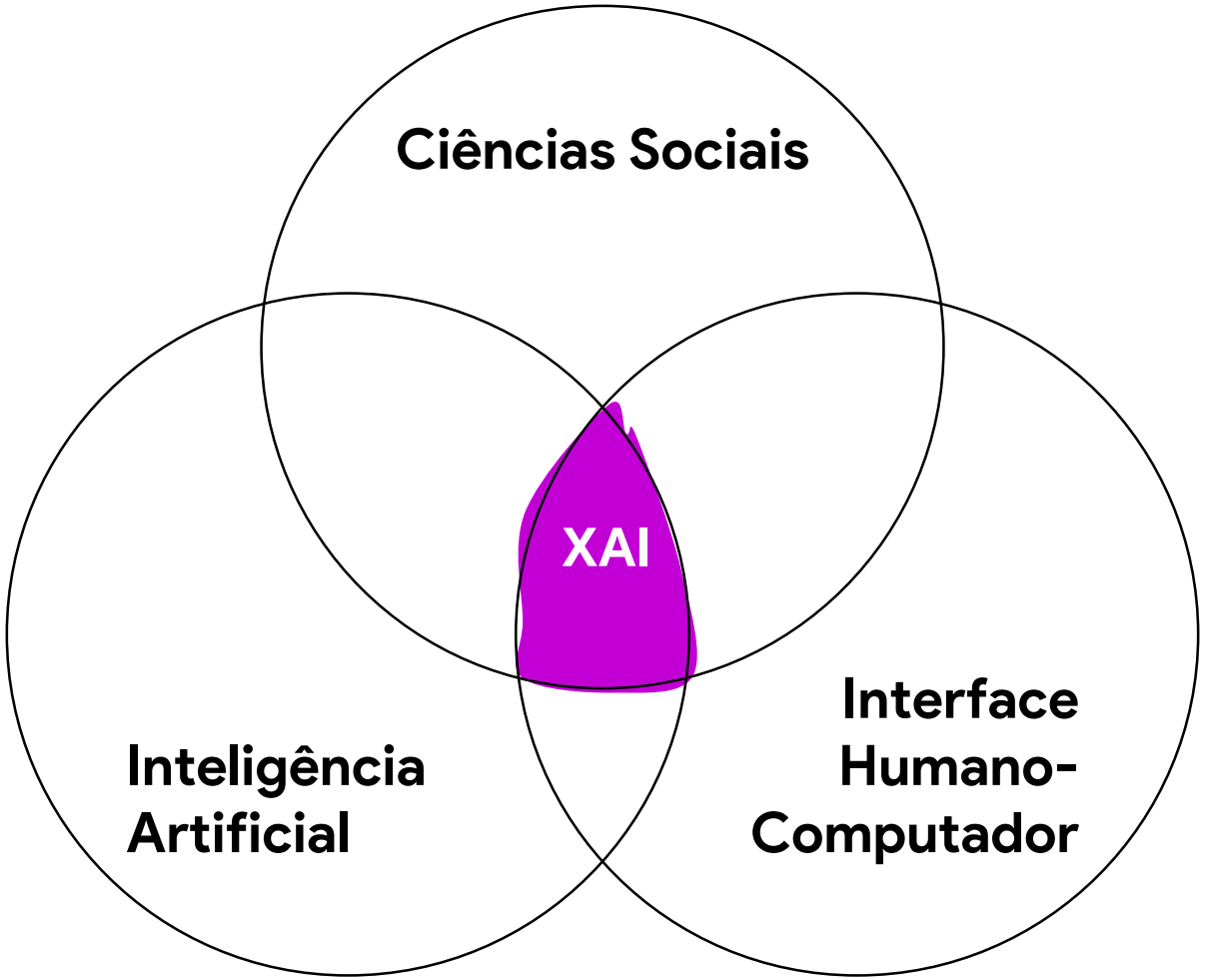
Entender a lógica por trás de cada decisão

TRANSPARÊNCIA

CONFIANÇA



Explicabilidade



As soluções de Inteligência Artificial não são e não serão infalíveis.

Mas, a explicabilidade pode ajudar...

Portal da Transparência

CONTROLADORIA-GERAL DA UNIÃO

Dados abertos

Aqui é possível baixar os dados apresentados no Portal da Transparência do Governo Federal, em formato aberto, possibilitando que os usuários façam cruzamentos e análises específicas, de acordo com suas necessidades.

Os arquivos são disponibilizados em formato CSV (clique aqui para mais informações).

- ORÇAMENTO PÚBLICO
- DESPESAS PÚBLICAS
- CARTÃO DE PAGAMENTO
- RECEITAS PÚBLICAS
- LICITAÇÕES E CONTRATOS
- CONVÊNIOS E INSTRUMENTOS CONGÊNERES
- BENEFÍCIOS AO CIDADÃO

http://www.portaltransparencia.gov.br/download-de-dados

Se a tecnologia quiser ajudar na construção de uma sociedade mais justa, ela tem que ser aberta e transparente.

https://serenata.ai/

https://brasil.io/home

https://colaboradados.github.io/

Inteligência Artificial vai nos permite analisar um volume gigantesco de dados em poucos segundos. Explicabilidade vai nos permitir identificar pontos em que podemos melhorar.

Antes de falar sobre futuro...

... precisamos falar sobre o que está acontecendo hoje, agora.

O que vai determinar o futuro da Inteligência Artificial são as escolhas que estamos fazendo agora.

Obrigada!

Carla Vieira

@carlaprvieira

carlaprv@hotmail.com

Referências

- Relatórios do AI NOW
- Racial and Gender bias in Amazon Rekognition
- Diversity in faces (IBM)
- Google video – Machine Learning and Human Bias
- Visão Computacional e Vieses Racializados
- Estudo Machine Bias on Compas
- Machine Learning Explainability Kaggle
- Predictive modeling: striking a balance between accuracy and interpretability

Referências

- Racismo Algorítmico em Plataformas Digitais: microagressões e discriminação em código
- Metrics for Explainable AI: Challenges and Prospects