

# APRENDIZADO DE MÁQUINA

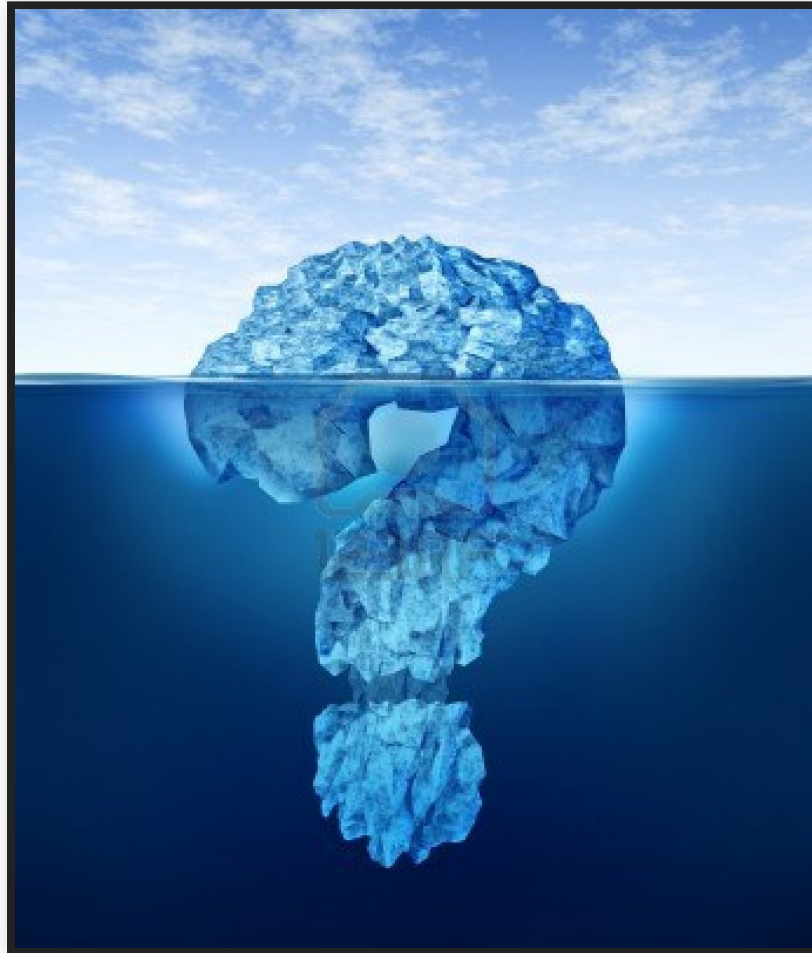
WRAP UP

PROF. RONALDO CRISTIANO PRATI

[ronaldo.prati@ufabc.edu.br](mailto:ronaldo.prati@ufabc.edu.br)

Bloco A, sala 513-2

# TÓPICOS NÃO ABORDADOS



# TÓPICOS NÃO ABORDADOS

- Aprendizado bayesiano
  - modelos gráficos
  - redes bayesianas
  - Gaussian process



David Heckerman



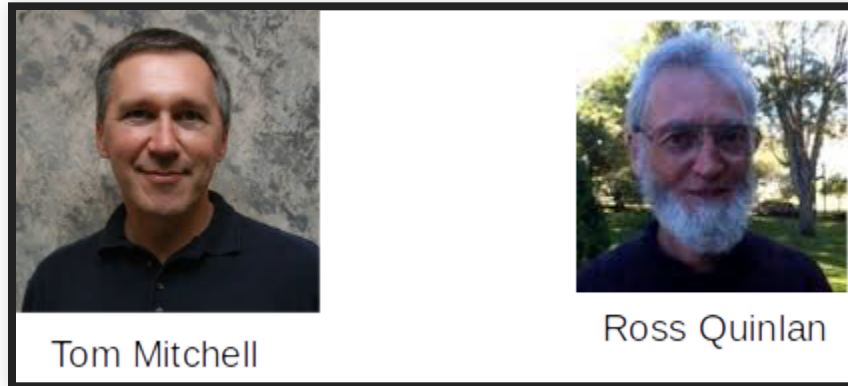
Judea Pearl



Michael Jordan

# TÓPICOS NÃO ABORDADOS

- Aprendizado simbólico
  - Árvores de decisão (decision trees)
  - Regras



# TÓPICOS NÃO ABORDADOS

- Aprendizado evolutivo
  - Algoritmos Genéticos
  - Programação Genética



John Holland

# TÓPICOS NÃO ABORDADOS

- Teoria do aprendizado
  - Dimensão VC
  - PAC-learning



# TÓPICOS NÃO ABORDADOS

- Aprendizado por reforço
  - Iteração de valor
  - Q-learning



# TÓPICOS NÃO ABORDADOS (RNAS)

- Text Embeddings
- Generative Adversarial Networks
- Siamese Networks
- Differentiable Neural Computer



# LIMITAÇÕES E DESAFIOS

# EXPLICAÇÃO E TRANSPARÊNCIA

- **Desafio:** Como confiamos em um sistema de AM?
  - Avaliar o desempenho cuidadosamente
    - Identificar vantagens e pontos fracos
  - Entender como funciona
    - Entender o resultado gerado

# AVALIAR O DESEMPENHO CUIDADOSAMENTE

- Além de uma avaliação quantitativa, em muitos casos é preciso fazer uma avaliação qualitativa.
  - As previsões fazem sentido?
  - Tem casos que falham?

# MUNDO FECHADO

- Modelo estado da arte em geração de rótulos de imagens

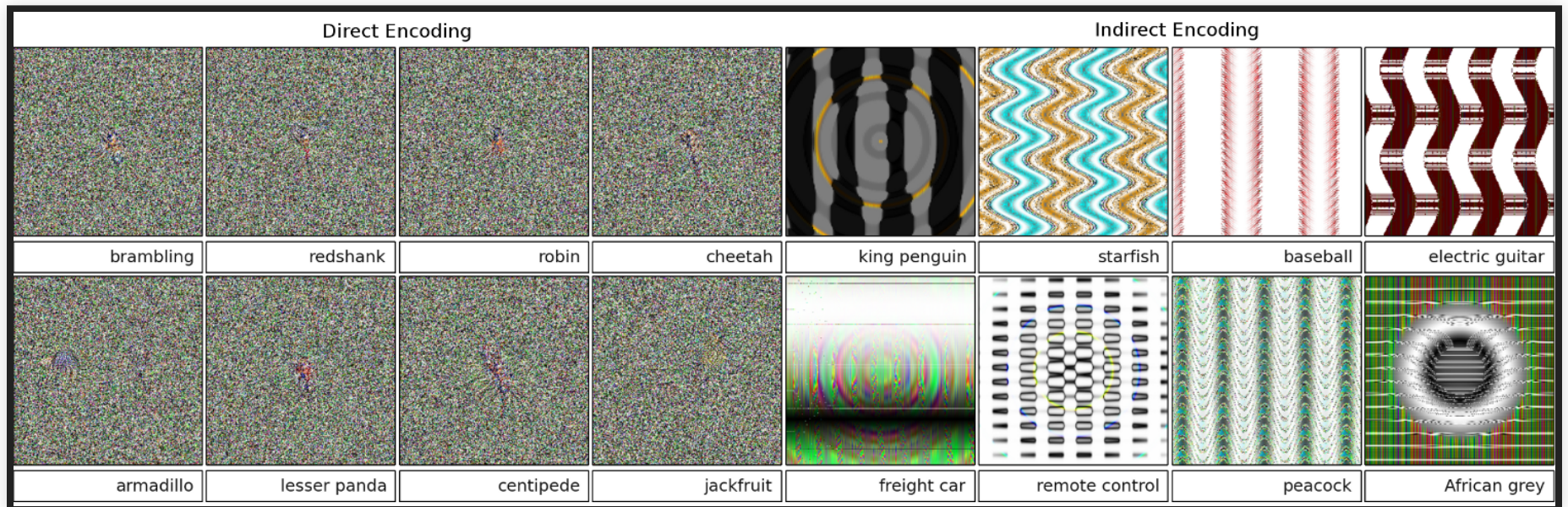


- Não tinha nenhuma escova de dentes no treinamento!

# O QUE DEU ERRADO?

- Devido a uma amostra viciada, um algoritmo de reconhecimento de imagens do Google classificava pessoas negras como gorilas
  - Dificilmente pensamos em como nosso modelo pode errar
  - Provavelmente o conjunto de treinamento tinha poucos exemplos de negros.

# TESTES COM DADOS FAKE



# ENTENDER COMO FUNCIONA:



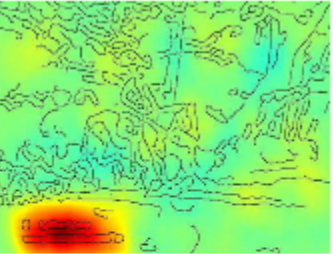
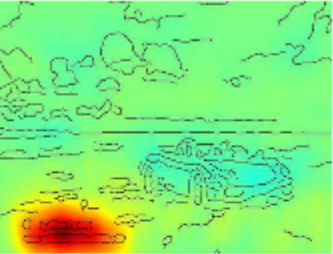


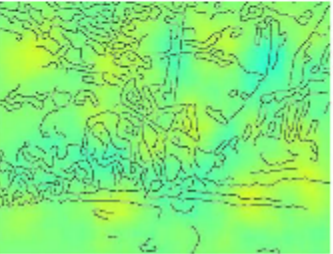
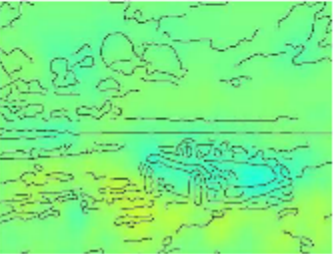
- Distinção importante:
  - Entender as decisões de um modelo (por que um classificador atribuiu uma certa classe)
  - Entender como o modelo foi criado (qual é o funcionamento do algoritmo)?

# ENTENDER COMO FUNCIONA:

- Para entender um modelo (e a sua saída), não necessariamente precisamos entender como ele foi feito
- Mas entender o funcionamento e a saída pode ajudar a melhorar o desempenho!
- Alguns modelos são mais fáceis de entender que outros!



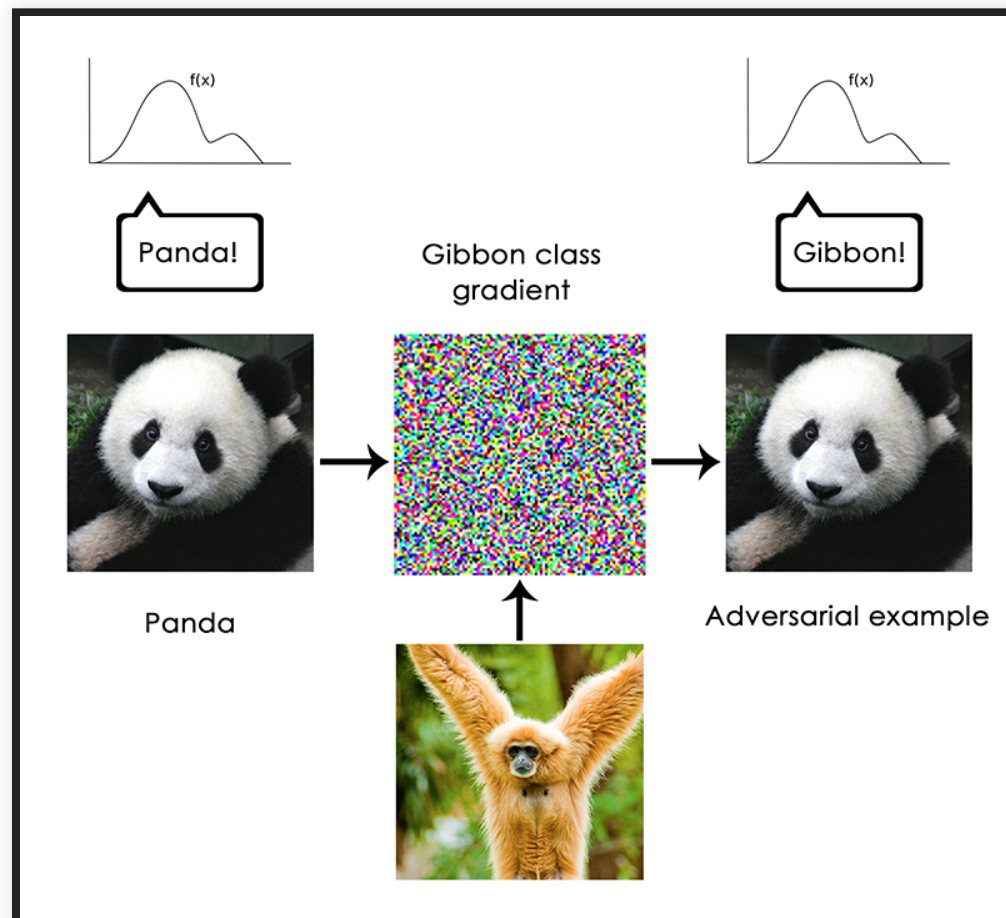
# ENTENDER O FUNCIONAMENTO

Horse-picture from Pascal VOC data set	Artificial picture of a car
	
	
Source tag present	Source tag present
↓	↓
Classified as horse	Classified as horse
	
	
No source tag present	No source tag present
↓	↓
Not classified as horse	Not classified as horse

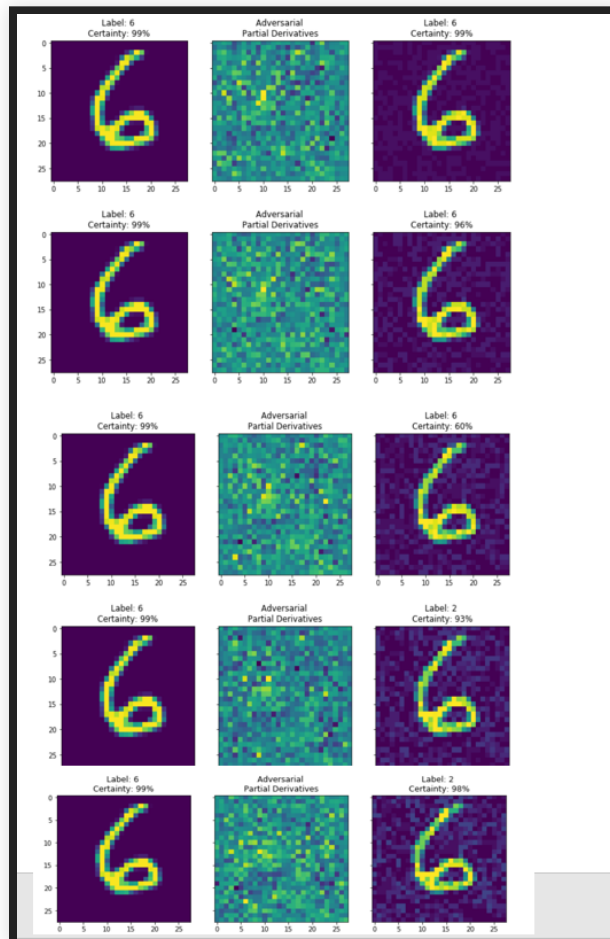
# ENTENDER COMO FUNCIONA:

- Muitas vezes se quer "enganar" um modelo
- Pode ter consequências sérias:
  - Ataque cibernético
  - Fraude

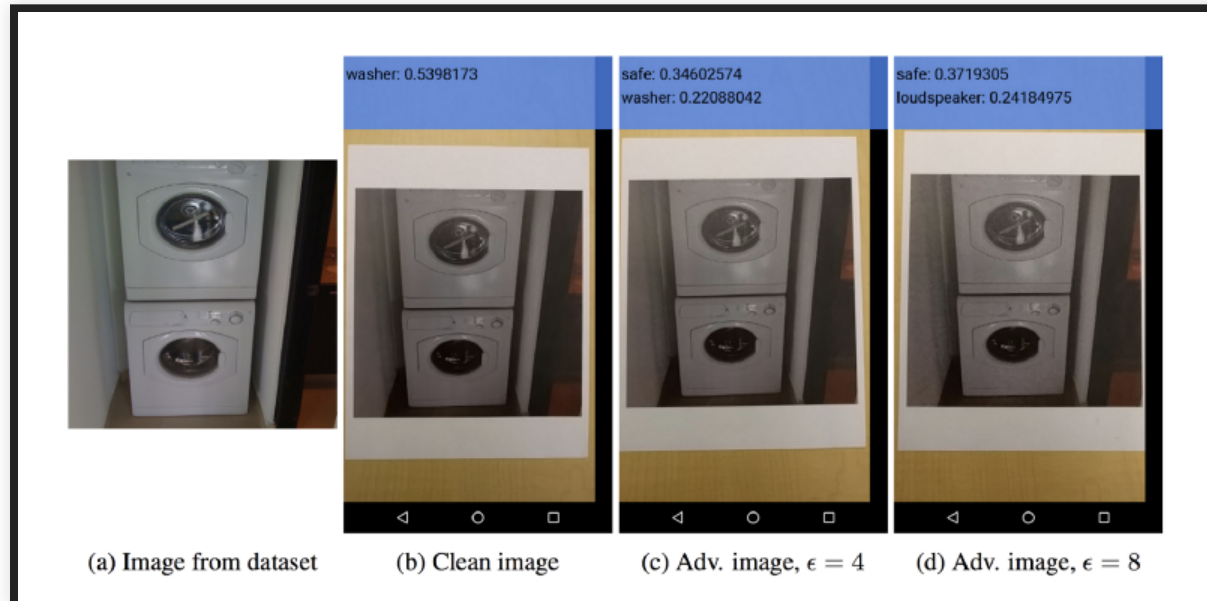
# APRENDENDO A ENGANAR A UM MODELO



# APRENDENDO A ENGANAR A UM MODELO



# APRENDENDO A ENGANAR A UM MODELO



# ENGANANDO UM CARRO AUTONOMO



# ALGORITHM BIAS

# ALGORITHM BIAS

Search query	Work experience	Education experience	Profile views	Candidate	Xing ranking
Brand Strategist	146	57	12992	male	1
Brand Strategist	327	0	4715	female	2
Brand Strategist	502	74	6978	male	3
Brand Strategist	444	56	1504	female	4
Brand Strategist	139	25	63	male	5
Brand Strategist	110	65	3479	female	6
Brand Strategist	12	73	846	male	7
Brand Strategist	99	41	3019	male	8
Brand Strategist	42	51	1359	female	9
Brand Strategist	220	102	17186	female	10

TABLE II: Top k results on [www.xing.com](http://www.xing.com) (Jan 2017) for the job search query “Brand Strategist”.



# ALGORITHM BIAS

**Translate** Turn off instant translation

Portuguese English French Detect language

English Portuguese Turkish **Translate**

She's a professor. He's a babysitter

O bir profesör. O bir bebek bakıcısı

Suggest an edit

36/5000

---

**Translate** Turn off instant translation

Portuguese English Turkish Detect language

English Portuguese Turkish **Translate**

O bir profesör. O bir bebek bakıcısı

He's a professor. She's a babysitter

Suggest an edit

36/5000

# ALGORITHM BIAS (CAUSAS)

- amostra viciada: mais exemplos de um determinado grupo apesar de uma menor prevalência.
- amostra contaminada: o rotulador pode dar viciar a amostra de acordo com as suas próprias convicções
- atributos ruins: algumas features podem ser mais difíceis de obter para certos grupos
- proxies: algumas informações são baseadas em medidas indiretas, que podem conter generalizações indevidas

# BIAS NÃO INTENCIONAL

```
text_to_sentiment("Let's go get Italian food")
```

```
2.0429166109408983
```

```
text_to_sentiment("Let's go get Chinese food")
```

```
1.4094033658140972
```

```
text_to_sentiment("Let's go get Mexican food")
```

```
0.38801985560121732
```

# BIAS NÃO INTENCIONAL

```
text_to_sentiment("My name is Emily")
```

```
2.2286179364745311
```

```
text_to_sentiment("My name is Heather")
```

```
1.3976291151079159
```

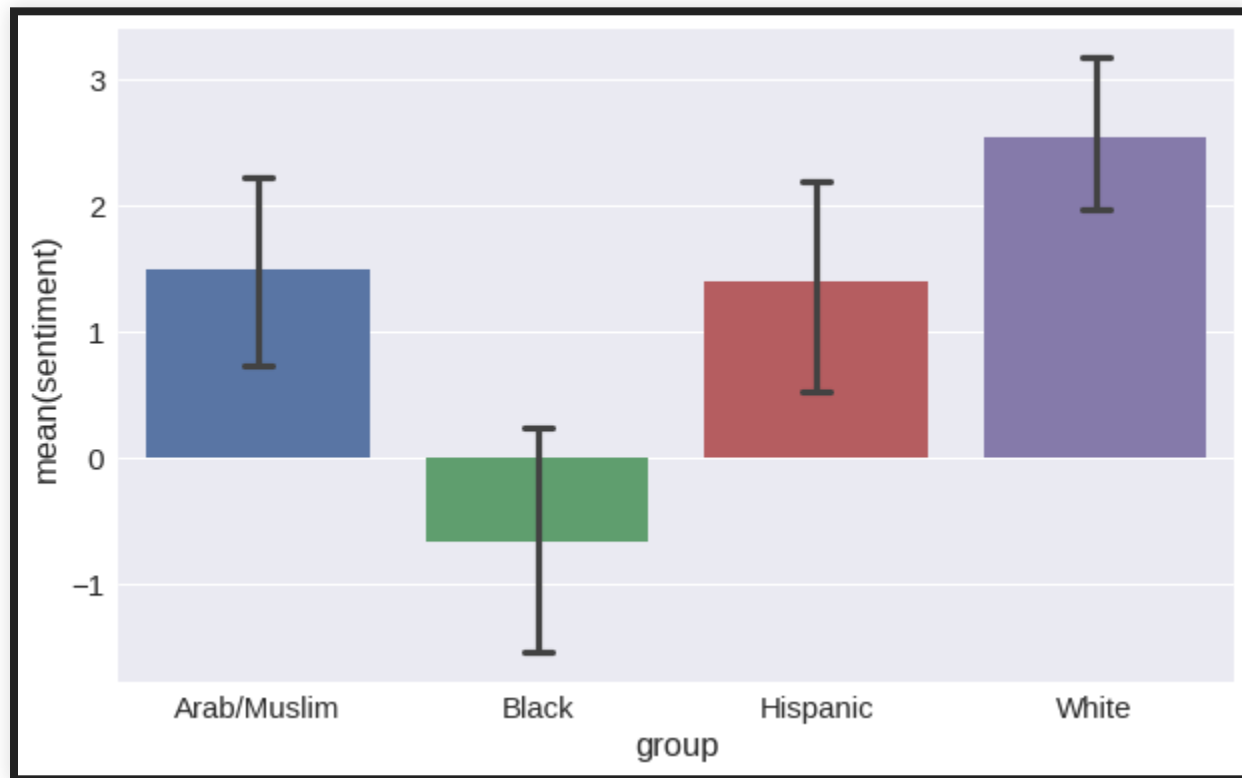
```
text_to_sentiment("My name is Yvette")
```

```
0.9846380213298556
```

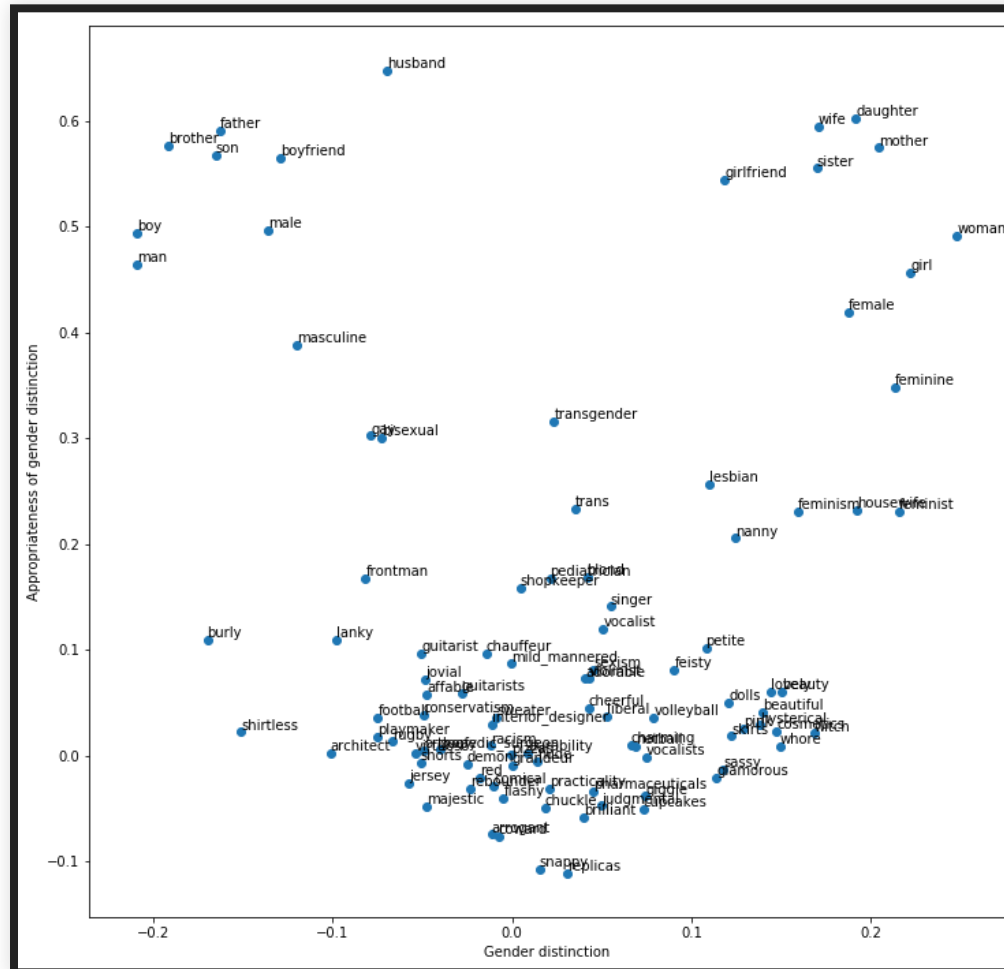
```
text_to_sentiment("My name is Shaniqua")
```

```
-0.47048131775890656
```

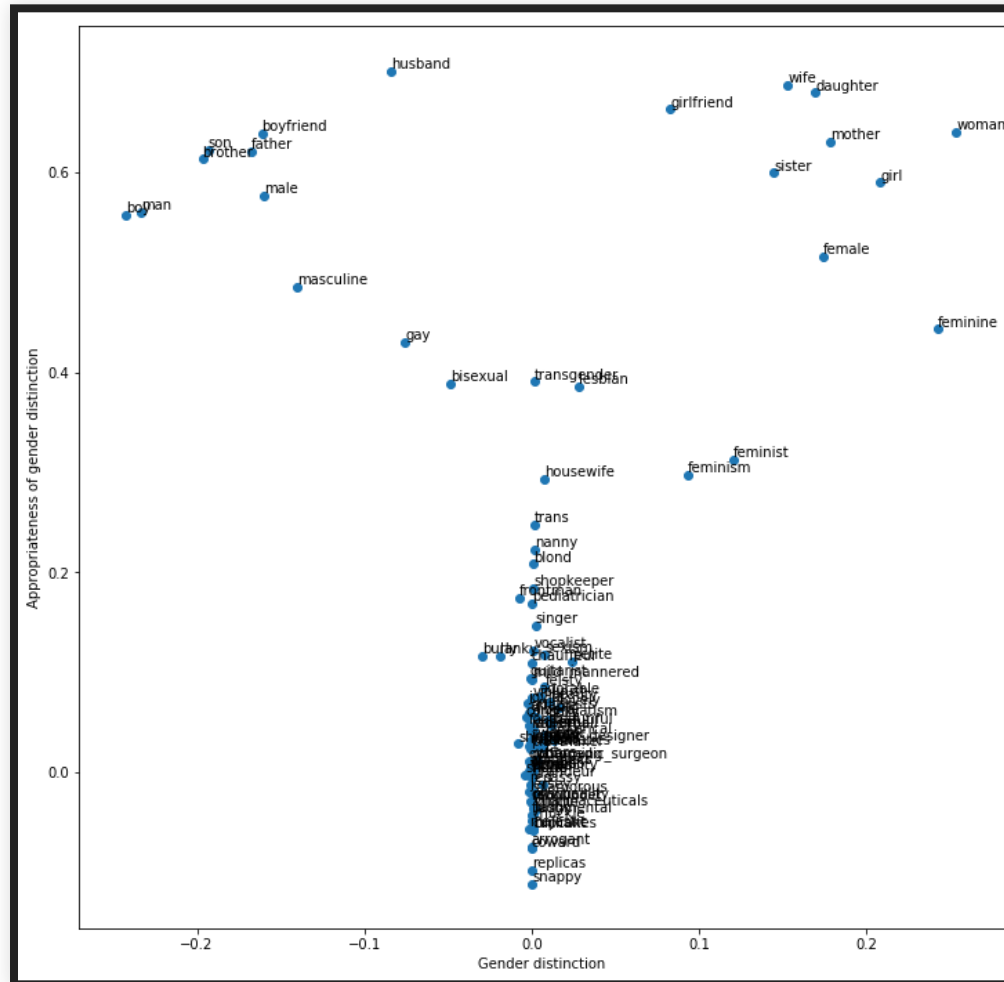
# BIAS NÃO INTENCIONAL



# ALGORITHM BIAS

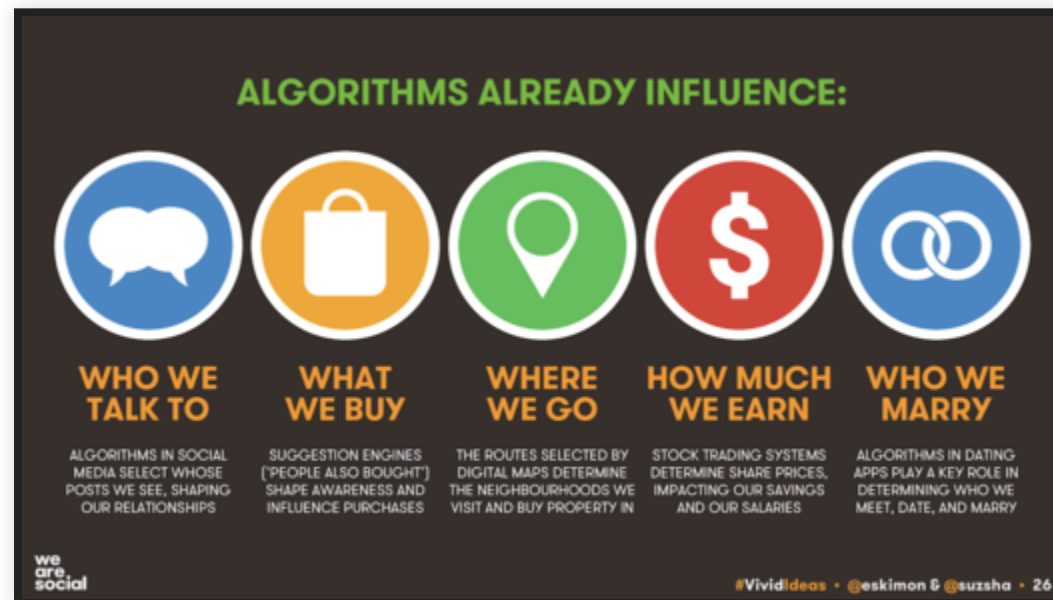


# ALGORITHM BIAS



# ASPECTOS SOCIAIS

**ALGORITHMS ALREADY INFLUENCE:**



The infographic consists of five circular icons in a row, each with a different background color and a white icon. Below each icon is a bold title and a short paragraph of text. The icons are: 1. Blue circle with two white speech bubbles. 2. Orange circle with a white shopping bag. 3. Green circle with a white location pin. 4. Red circle with a white dollar sign. 5. Blue circle with two white interlocking rings.

Icon	Area of Influence	Description
Speech bubbles	WHO WE TALK TO	ALGORITHMS IN SOCIAL MEDIA SELECT WHOSE POSTS WE SEE, SHAPING OUR RELATIONSHIPS
Shopping bag	WHAT WE BUY	SUGGESTION ENGINES ("PEOPLE ALSO BOUGHT") SHAPE AWARENESS AND INFLUENCE PURCHASES
Location pin	WHERE WE GO	THE ROUTES SELECTED BY DIGITAL MAPS DETERMINE THE NEIGHBOURHOODS WE VISIT AND BUY PROPERTY IN
Dollar sign	HOW MUCH WE EARN	STOCK TRADING SYSTEMS DETERMINE SHARE PRICES, IMPACTING OUR SAVINGS AND OUR SALARIES
Interlocking rings	WHO WE MARRY	ALGORITHMS IN DATING APPS PLAY A KEY ROLE IN DETERMINING WHO WE MEET, DATE, AND MARRY

**we are social**

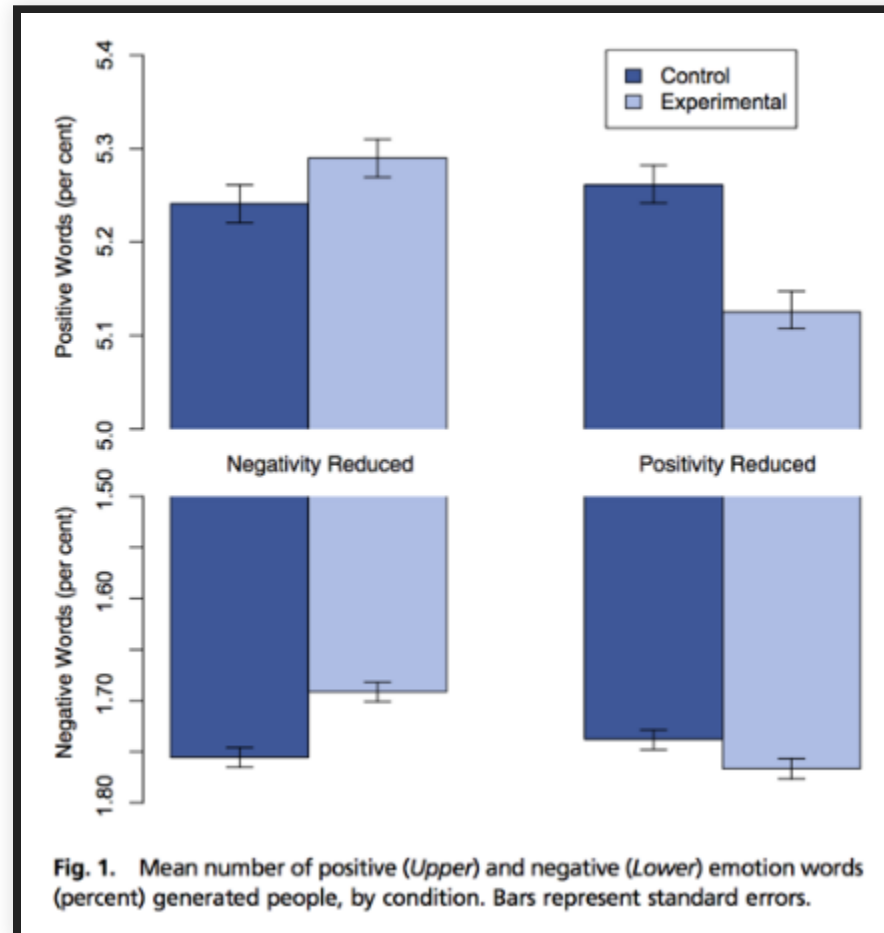
#VividIdeas • @eskimon & @suzsha • 26



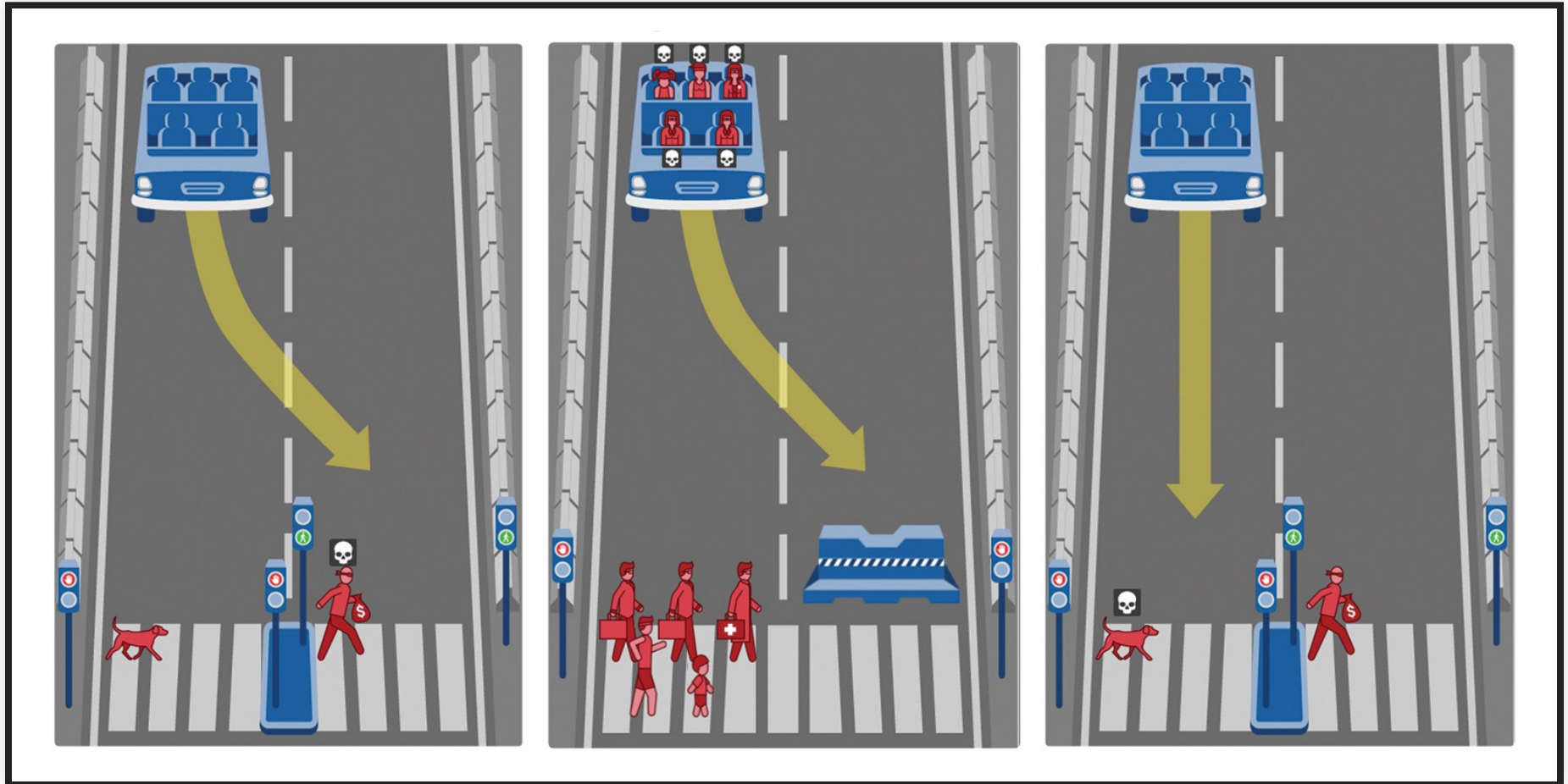
# EXPERIMENTO DO FACEBOOK/CORNELL

- 689.003 pessoas
- Uma semana em 2012
- Dois grupos:
  - Grupo A: remover conteúdo considerado negativo do feed
  - Grupo B: remover conteúdo considerado positivo do feed

# EXPERIMENTO DO FACEBOOK/CORNELL



# ASPECTOS ÉTICOS



# GERENCIANDO MUDANÇAS

- Sistemas mudam continuamente
- Os resultados ainda são válidos?
- Caso Famoso: "Gripe do google"  
(<https://www.google.org/flutrends/about/>)
  - Predições foram boas por um tempo
  - Então quebraram

# DISCIPLINAS RELACIONADAS

- MCTAO14-13 - Inteligência Artificial
- MCZA030-13 - Vida Artificial na Computação
- MCZA006-13 - Computação Evolucionista e Conexionista
- MCZA017-13 - Processamento de Linguagem Natural
- MCZA015-13 - Mineração de Dados
- ESZI014-17 - Sistemas Inteligentes
- ESZA019-17 - Visão Computacional