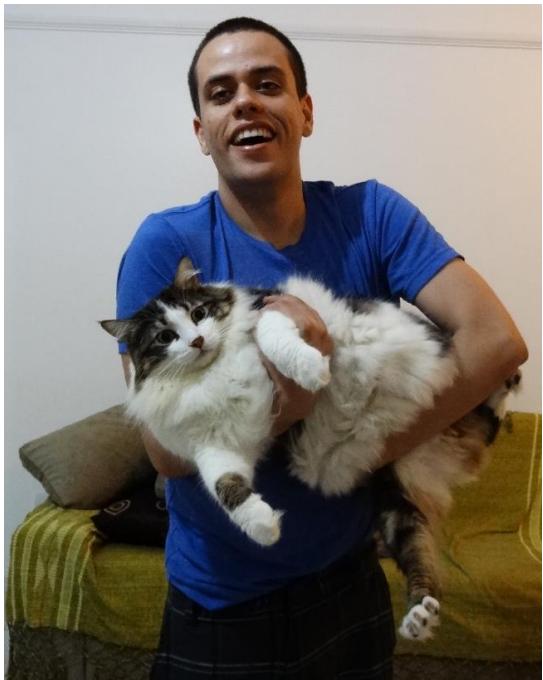


Primeiros Passos para o Sucesso em Machine Learning

Bruno Cabral

Quem sou eu ?



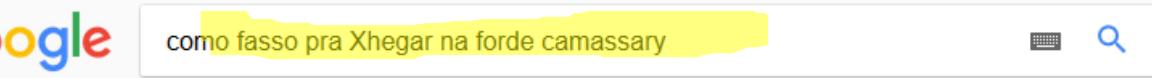
Mestre em Sistemas de
Recomendação
Programador OpenSource
CEO da Potelo



1 - Estamos cercados por IA
e não tem como fugir.

Está em todos os lugares

- Google Translate
- Google Assistant
- Google Calendar/e-mail
- Youtube
- Facebook



Todas Maps Shopping Notícias Imagens Mais Configurações Ferramentas

Aproximadamente 371.000 resultados (0,69 segundos)

Exibindo resultados para **como faço pra Chegar na ford camaçari**

Em vez disso, pesquisar por [como fasso pra Xhegar na forde camassary](#)



Childish Gambino - This Is America (Official Video)
Donald Glover 4:05
250M views



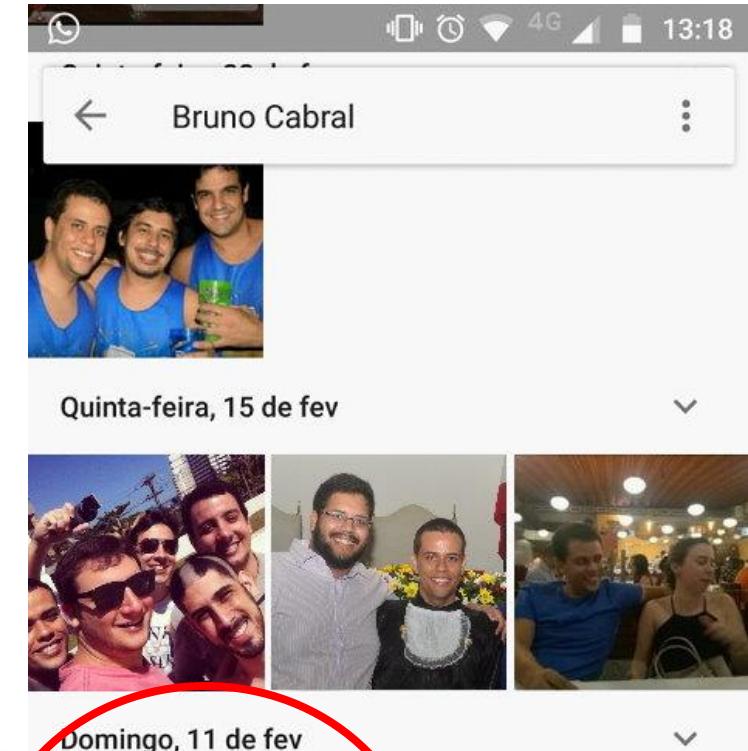
EXPLICANDO THIS IS AMERICA (frame a frame)
Dossiê do Felipe 18:23
707K views



MURILLO COUTO REAGINDO À QUANTO CUSTA O OUTFIT? (O...
Murilo Couto 9:29
636K views
New

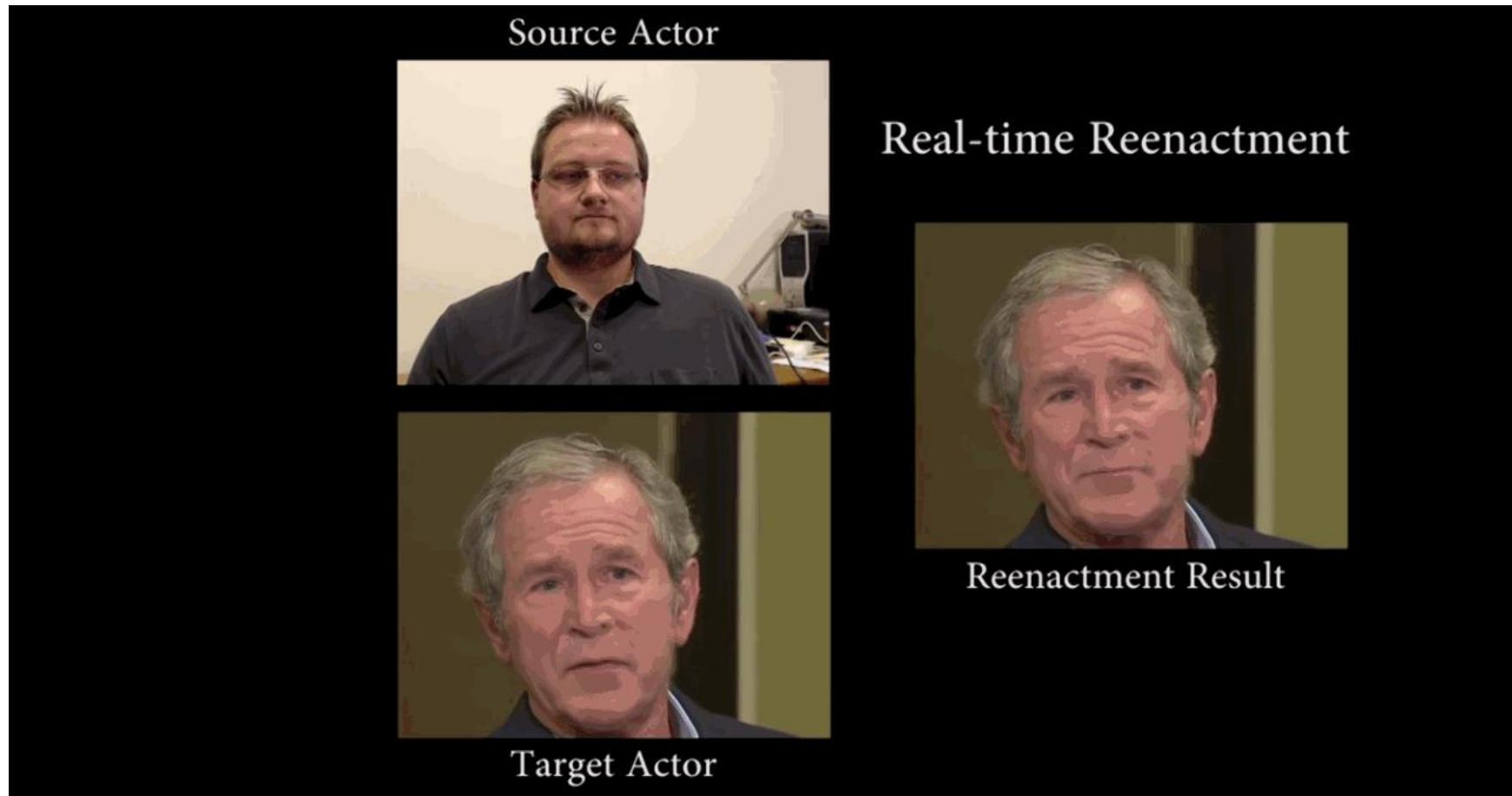


ISSO É BRASIL | Paródia This is America - Childish Gambino
Galo Frito 3:27
246K views
New

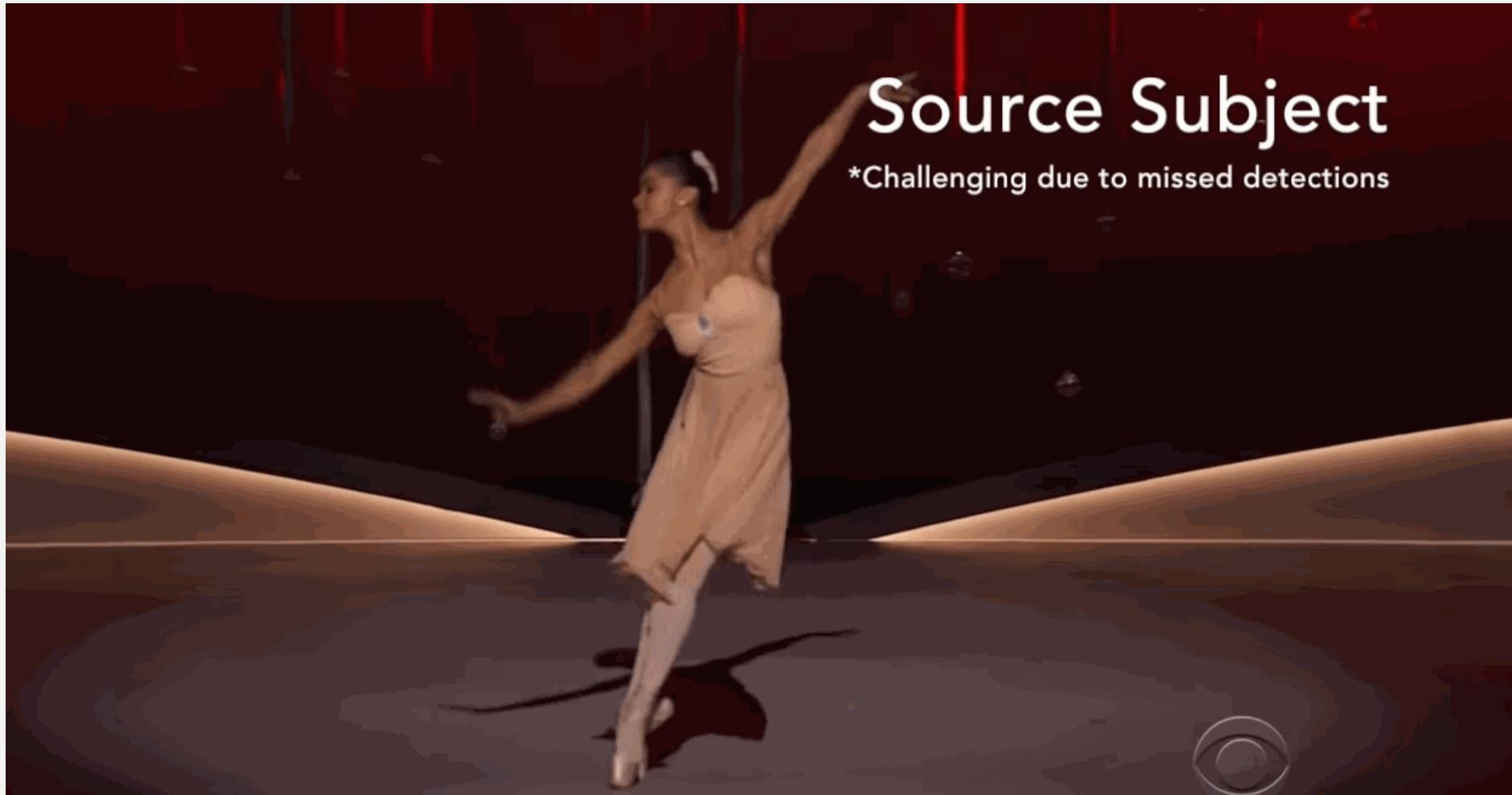


Segunda-feira, 29 de jan

Podem falar pelo os outros



Dançar como um profissional



Segurança Pública



Mas na real, o que é IA?



Algoritmos genéticos

Robô

Redes neurais

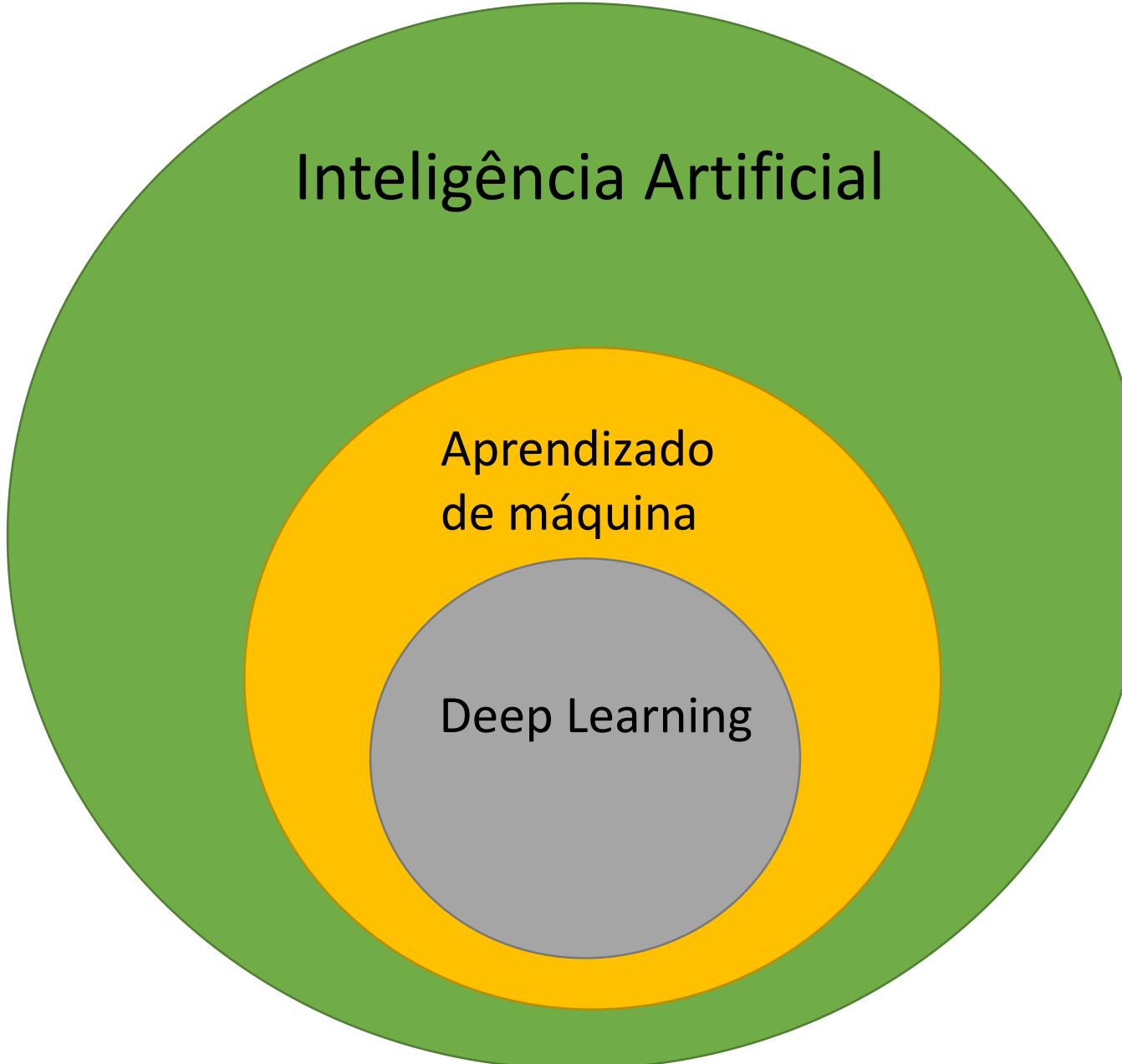
Big data

Chatbot

Natural Language Processing

Redes neurais
profundas

Data Mining



É a inteligência similar à humana exibida por mecanismos ou software (Regras, Elisa)

Algoritmos de aprendizagem que aprendem modelos automaticamente a partir de dados (SVM, Árvores de decisão)

Uma forma de aprendizado de máquina **inspirada** no cérebro humano (redes neurais) com modelos profundos

Tipos de aprendizado



1 -Supervisionado

Data



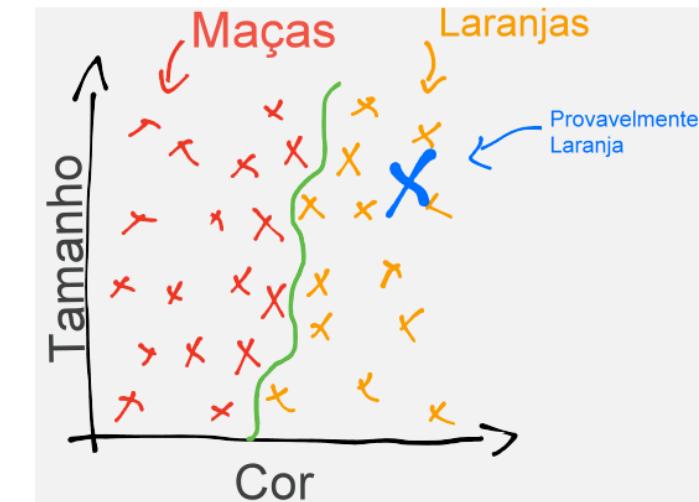
Rótulo

Gato

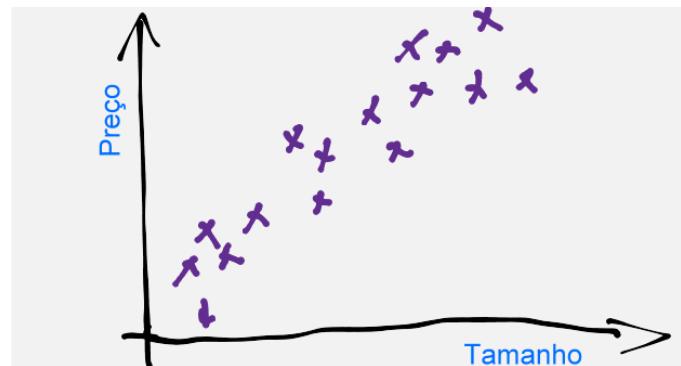


Comida

★ Classificação ★



Regressão



Eu comprei 2 maçãs por 2 reais, quanto custa 10 maçãs ?

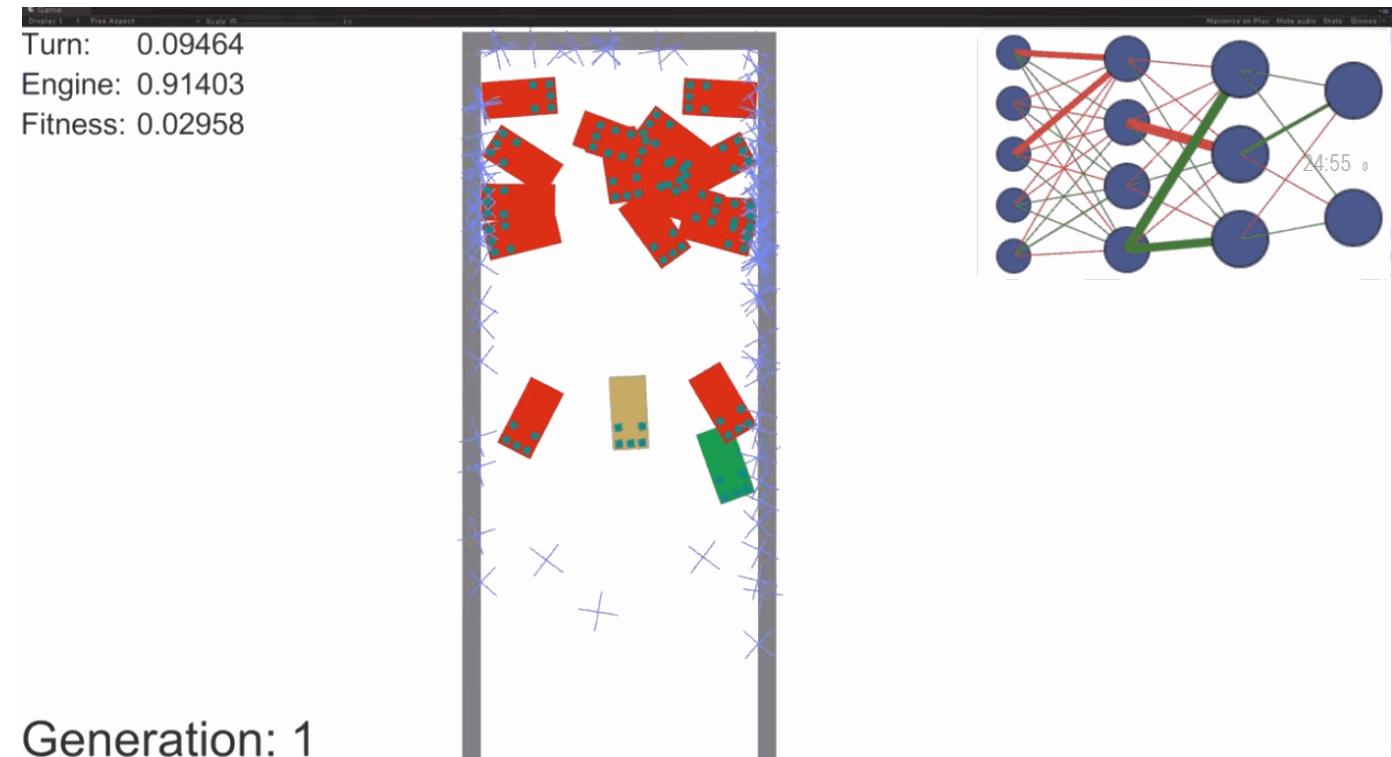
2 - Não supervisionado

Aprende sem ter nada (aprende a estrutura)



3 - Aprendizado por Reforço

Aprende tendo apenas o resultado final/feedback



A programação tradicional consiste em definir os passos de forma bem detalhada



```
def detectar_letras  
...  
def detectar_contornos  
...  
def detectar_cores  
...  
def detectar_estado  
...
```

Caro

Trabalhoso

Machine Learning muda a dinâmica



```
modelo = treinar(exemplos)
```

```
resultado = modelo.predizer(novo_exemplo)
```

MHM-0058

Classificador

Tem vezes que não temos nem ideia de quais seriam os passos

Como detectar uma transação suspeita?
Que fuja do padrão ?

Como você codificaria as regras para esse padrão?



Melhor deixar a máquina achar o suspeito



A screenshot of a credit card statement page. At the top, there's a small image of a credit card. Below it, a table shows the summary of the bill in R\$:

Resumo da fatura em R\$	
Total da fatura anterior	2.851,63
Créditos e Pagamentos	0,01
Débitos e despesas nacionais	690,97
Débitos e despesas internacionais	0,00
Total desta fatura	3.542,59

Below this, there's a section for payment information:

vencimento	pagamento total	pagamento mínimo	parcelamento da fatura
02/03/2014	R\$ 3.542,59	R\$ 2.042,59	Não disponível

Then, there's a table for credit limits and movements:

Límite de crédito R\$	Movimentações
Limite total de crédito 1.500,00	DATA MOVIMENTAÇÃO VALOR EM R\$
De retirada de recursos	16/02 MULTA POR ATRASO 57,03
País (saque)	16/02 JUROS DE MORA 22,81
De retirada de recursos	16/02 JUROS DE MORA 0,82
Exterior (saque)	16/02 JUROS DE MORA 1,87
Exterior (saque)	16/02 JUROS DE MORA 0,12
Exterior (saque)	16/02 JUROS DE MORA 0,24
Exterior (saque)	16/02 JUROS DE MORA 0,02
Exterior (saque)	16/02 JUROS DE MORA 0,05
Exterior (saque)	16/02 JUROS DE MORA 0,10
Exterior (saque)	16/02 JUROS DE MORA 0,04
Exterior (saque)	16/02 JUROS DE MORA 0,17
Exterior (saque)	16/02 DB DIFERENÇA COTACAO US 2,28
Exterior (saque)	16/02 JUROS DE FINANCIAMENTO 401,49
Total de débitos efetuados	Total de débitos efetuados -0,00
Total de débitos efetuados	Total de débitos efetuados 487,04

At the bottom left, there's a world map with several red and blue dots indicating activity or locations.



Temos que saber pois os usuários
desejam o nível de iteração que
estão acostumados

E também para não cair no Hype

BUZZWORD

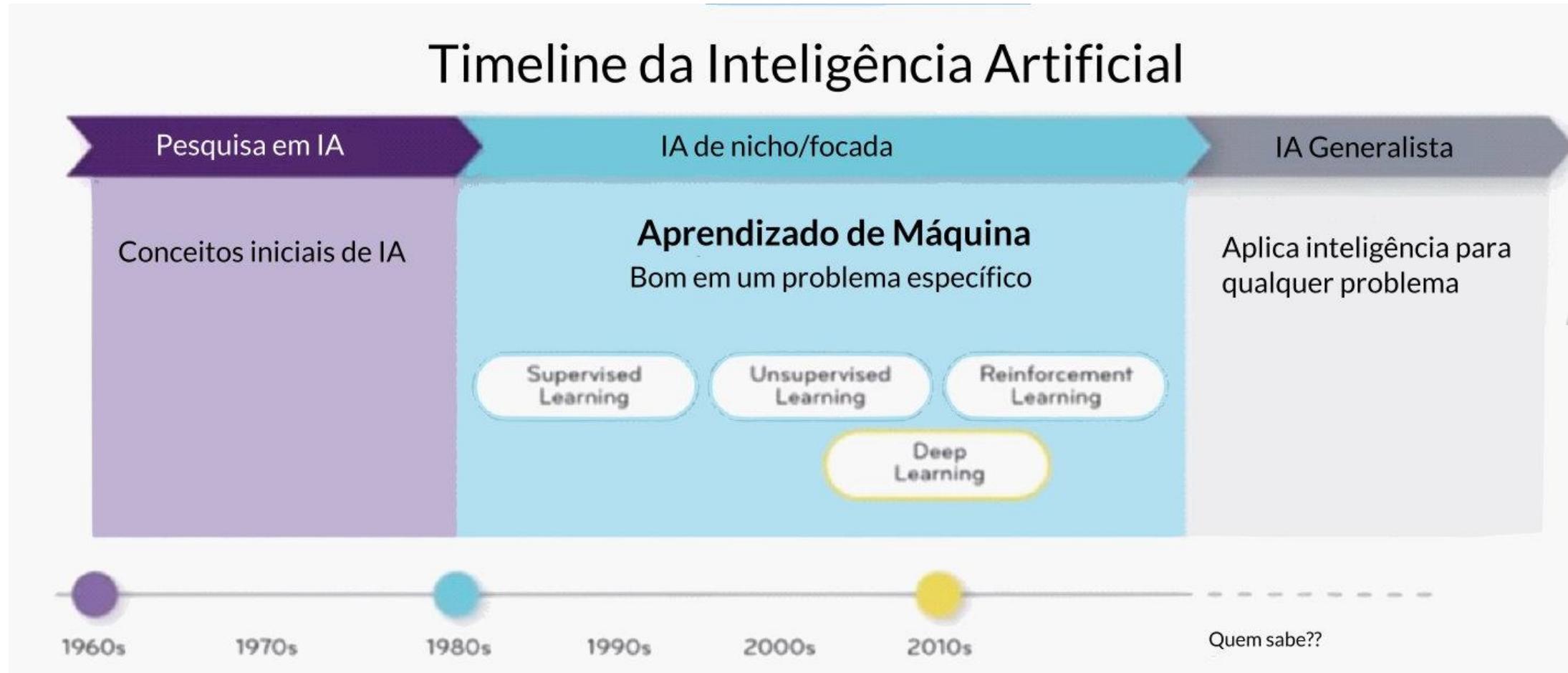
B I N G O

Machine learning	Next-gen	Innovative	Mobile	Revenue sharing
Apps	Resilient	SaaS	Blockchain	Efficiency
Data-mining	IoT	★	Gamification	Sensors
Pilot	Savings	Dashboard	Scalable	Disruption
SoLoMo	Engagement	Cloud	Big data	Smart

Redes neurais e Inteligência artificial

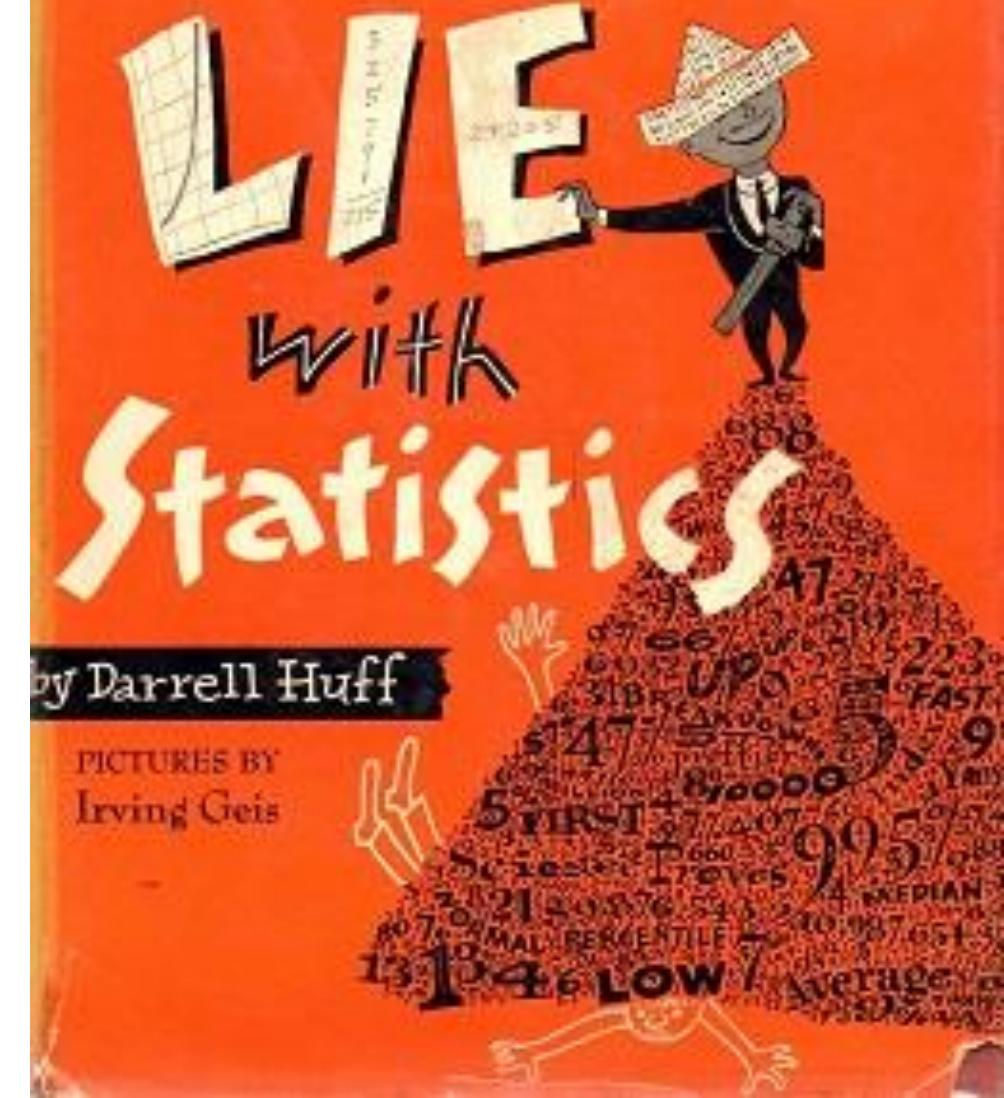


Onde realmente estamos?

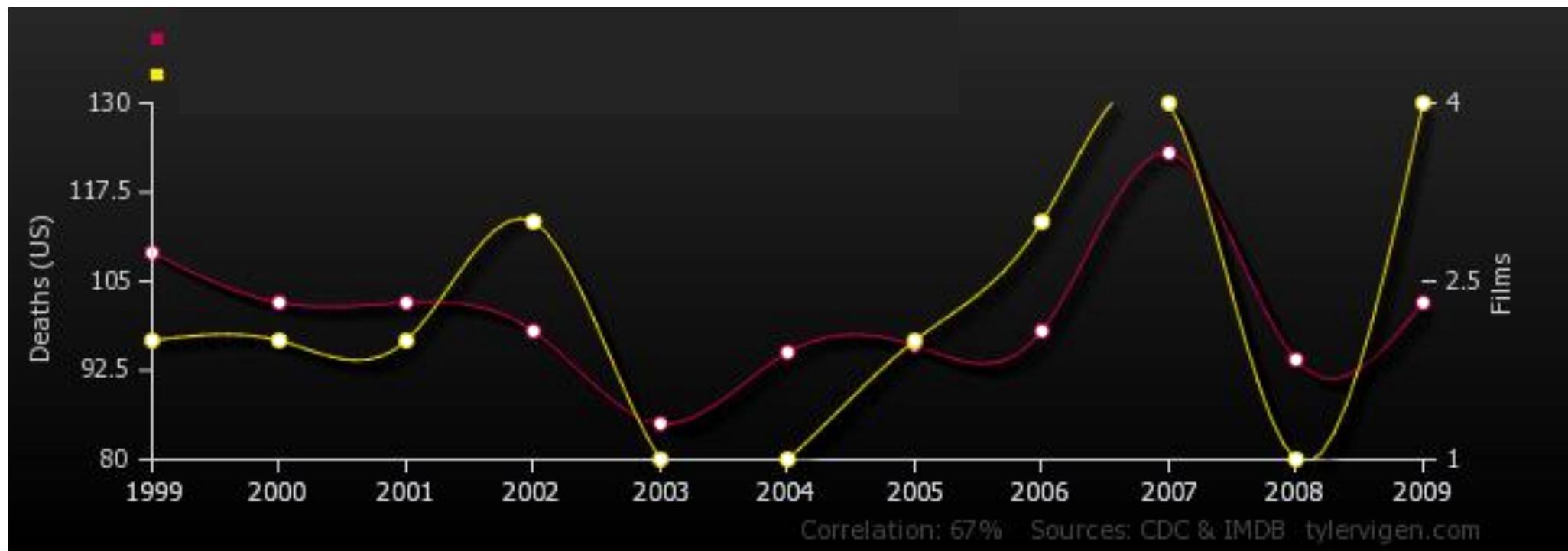


“Há três tipos de mentiras:
mentiras, mentiras terríveis e
estatísticas” - Benjamin Disraeli

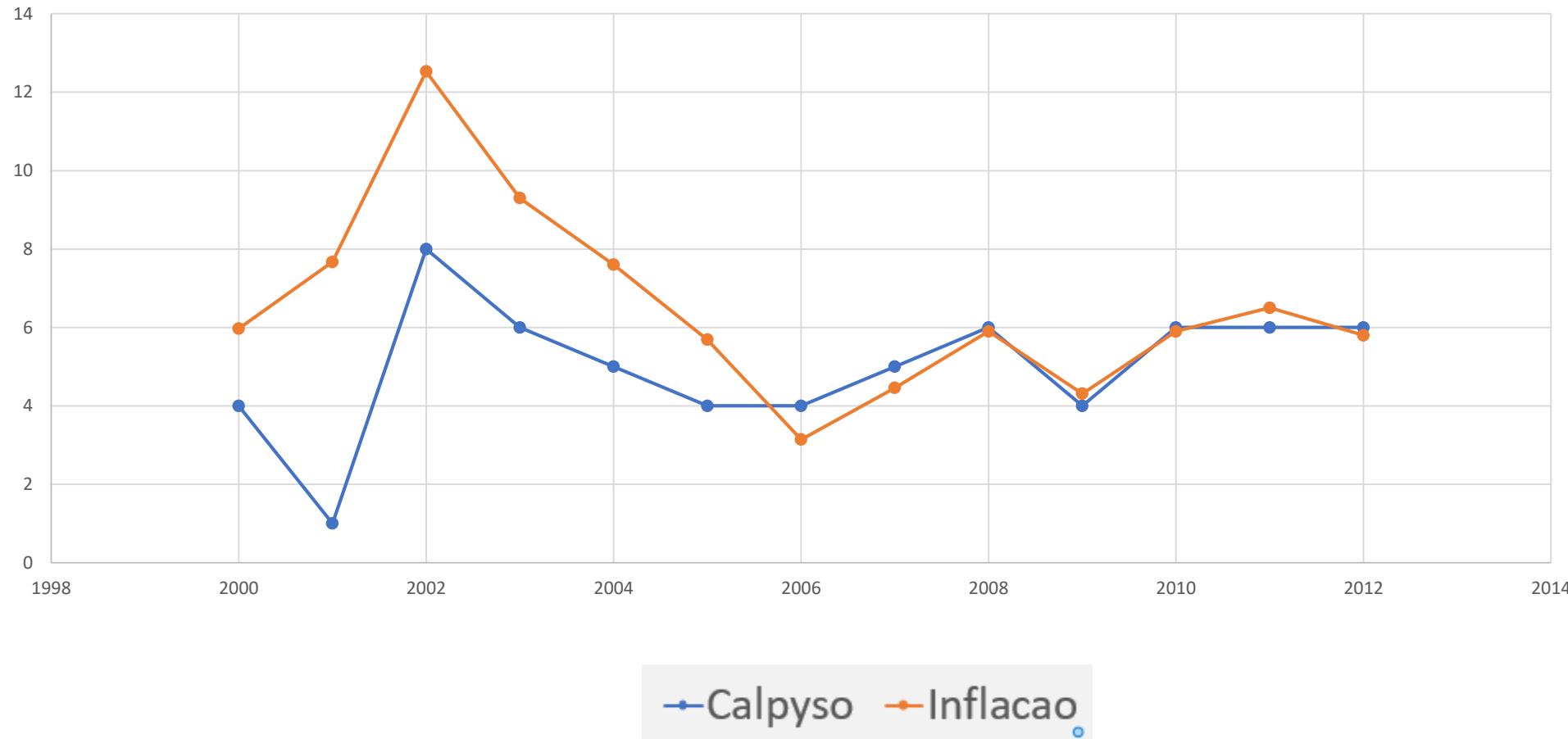
Now you, too, can out-double-talk your accountant;
confuse your political opponent;
How to
prove that your product has secret built-in goodness!



Afogamentos vs Filmes com Nicolas Cage



Taxa de Inflação vs Singles Banda Calypso



Whatsapp é fantástico, mas...



Como Resolvemos isso??

Deep troll (Gemidão do Whatsapp) detector

Vamos treinar uma rede neural para detectar audios com o gemidão.

Vamos colocar a mão na massa!



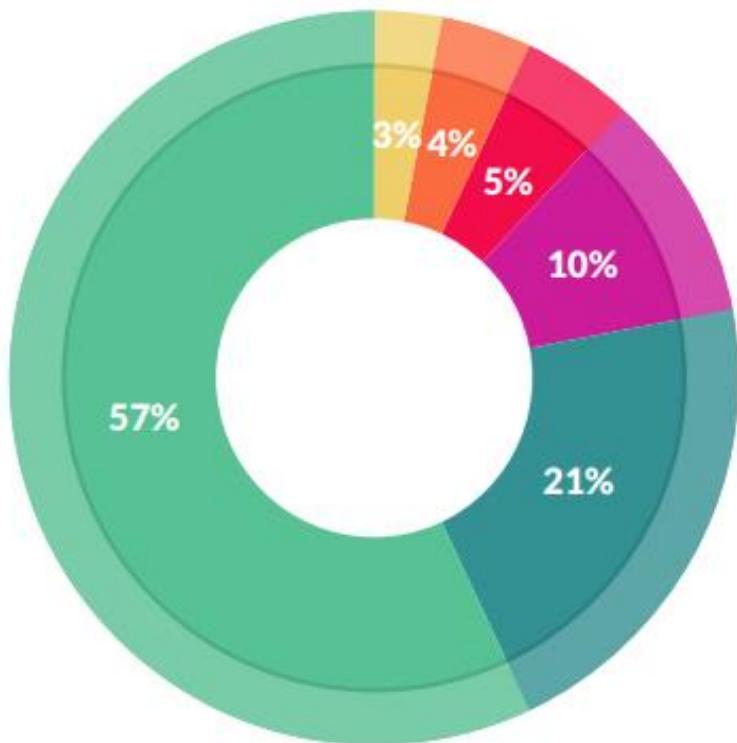
Créditos para **Andrios Robert Silva Pereira**
<https://github.com/andriosr>

Lições aprendidas

- Preparar o conjunto de dados dá muito trabalho
- Ter um pipeline eficiente de coleta e gestão dos dados é muito importante
- Muita coisa é na tentativa e erro
- Redes neurais profundas são lentas (vc precisa de uma GPU)



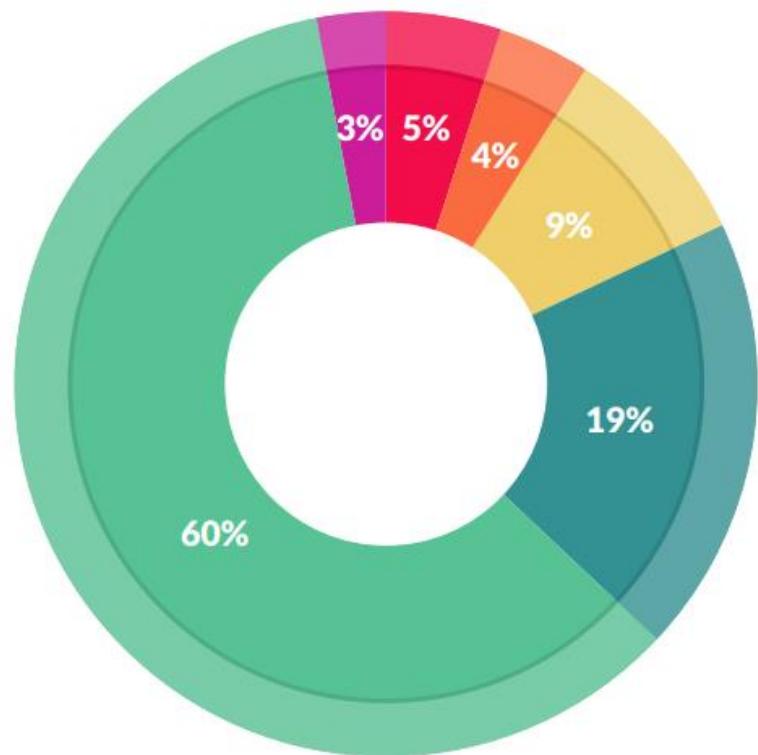
MAIOR TEMPO GASTO É LIMPANDO DADOS



What's the least enjoyable part of data science?

- *Building training sets: 10%*
- *Cleaning and organizing data: 57%*
- *Collecting data sets: 21%*
- *Mining data for patterns: 3%*
- *Refining algorithms: 4%*
- *Other: 5%*

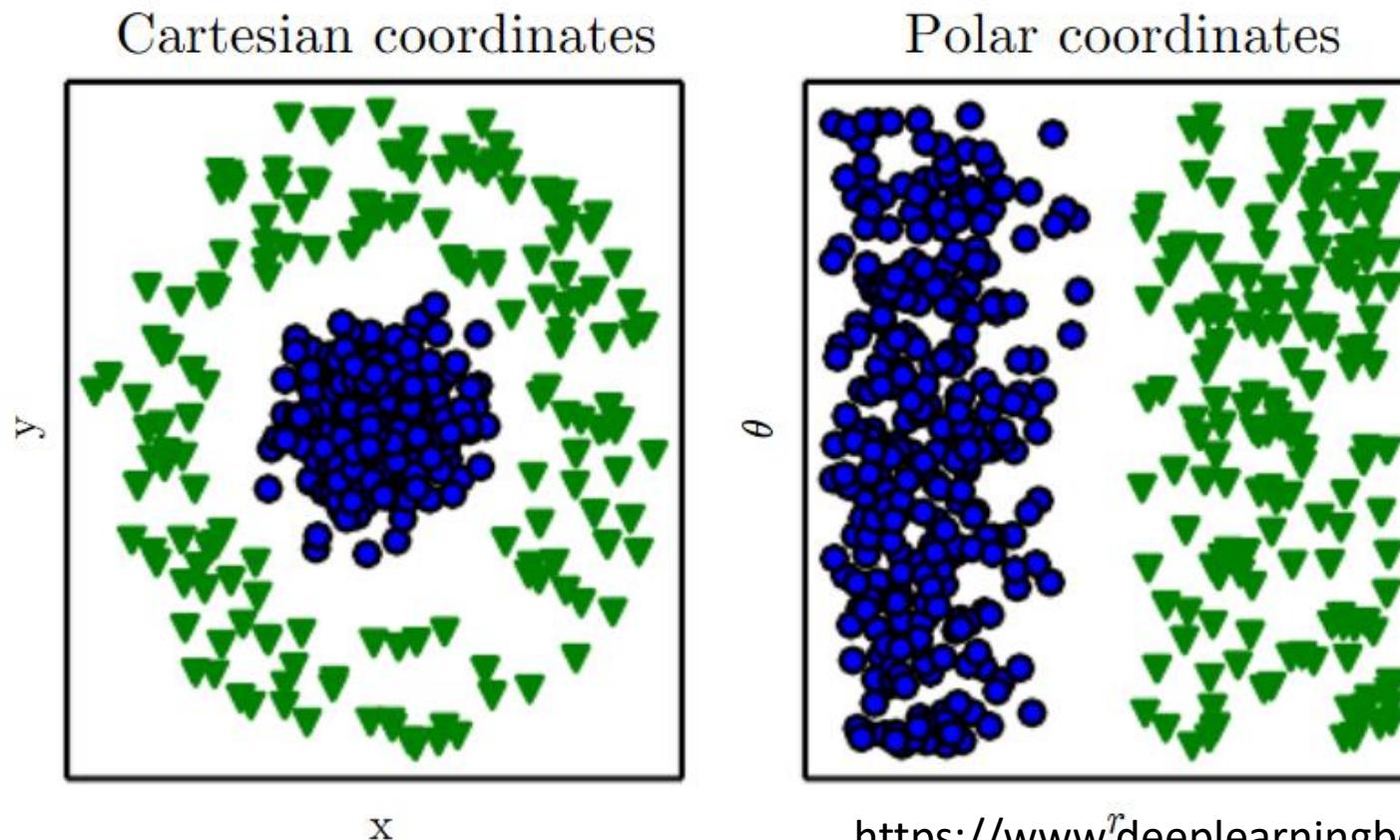
MAIOR TEMPO GASTO É LIMPANDO DADOS



What data scientists spend the most time doing

- *Building training sets: 3%*
- *Cleaning and organizing data: 60%*
- *Collecting data sets; 19%*
- *Mining data for patterns: 9%*
- *Refining algorithms: 4%*
- *Other: 5%*

Importância de representar os dados

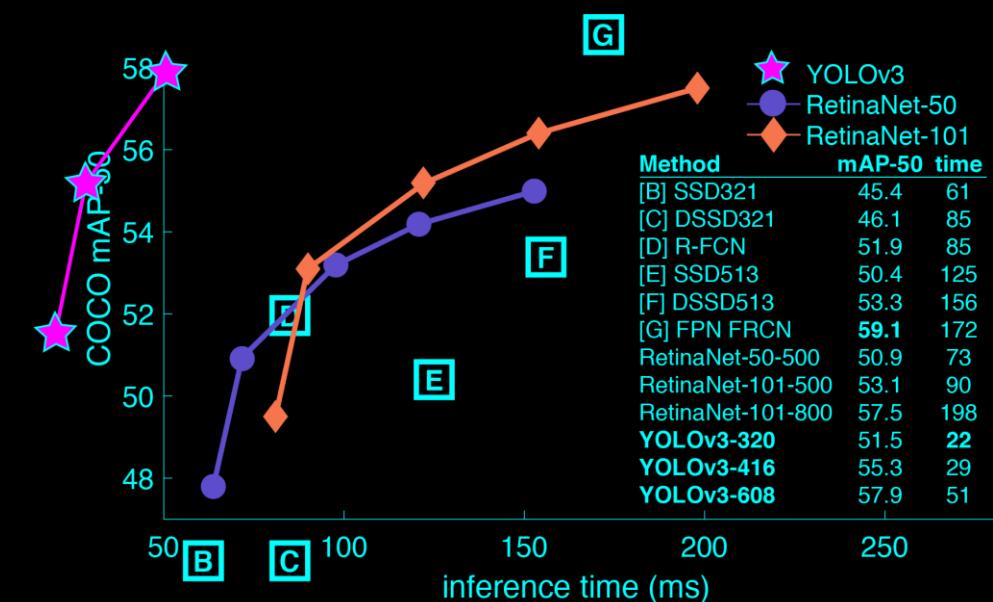
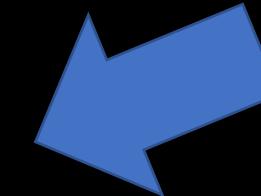


Moreno do Whatsapp

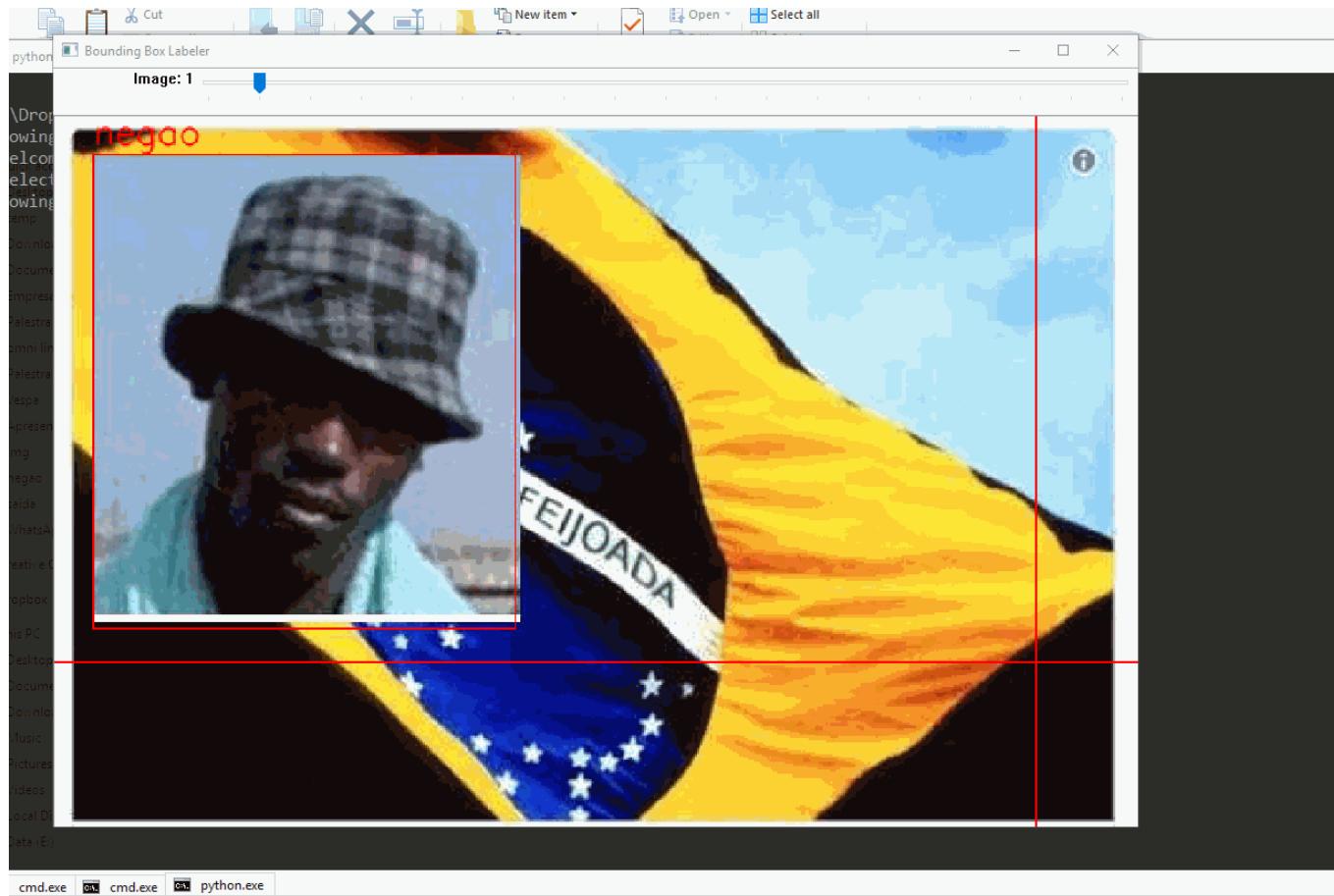


YOLO 3

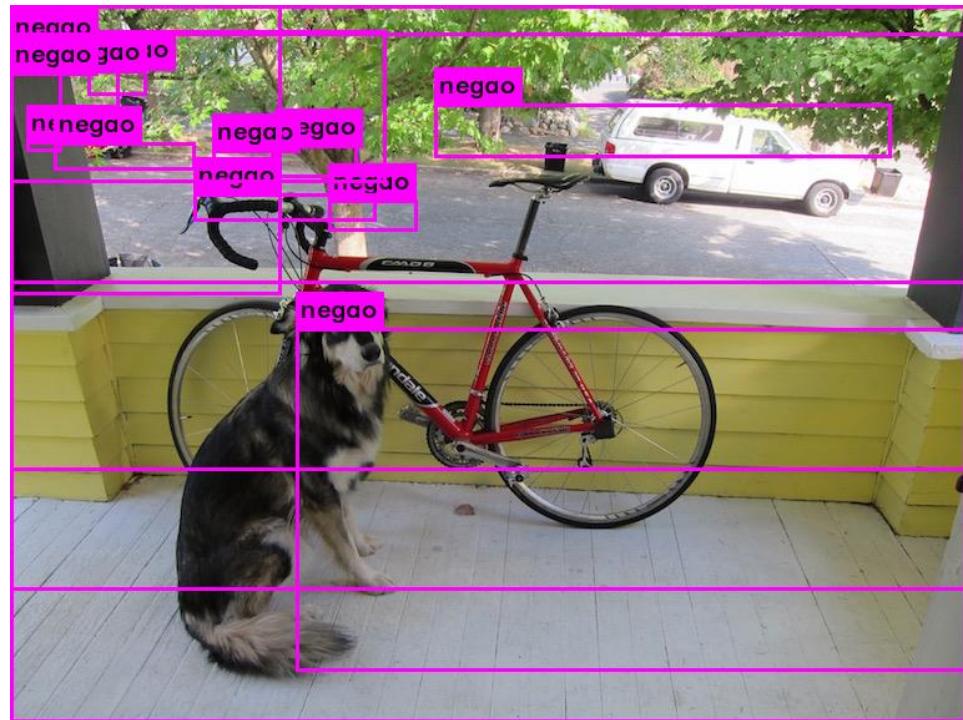
Temos a nossa solução home-made para resolver CAPTCHA (feita em 2017)
Quero usar o BOMZÃO!!



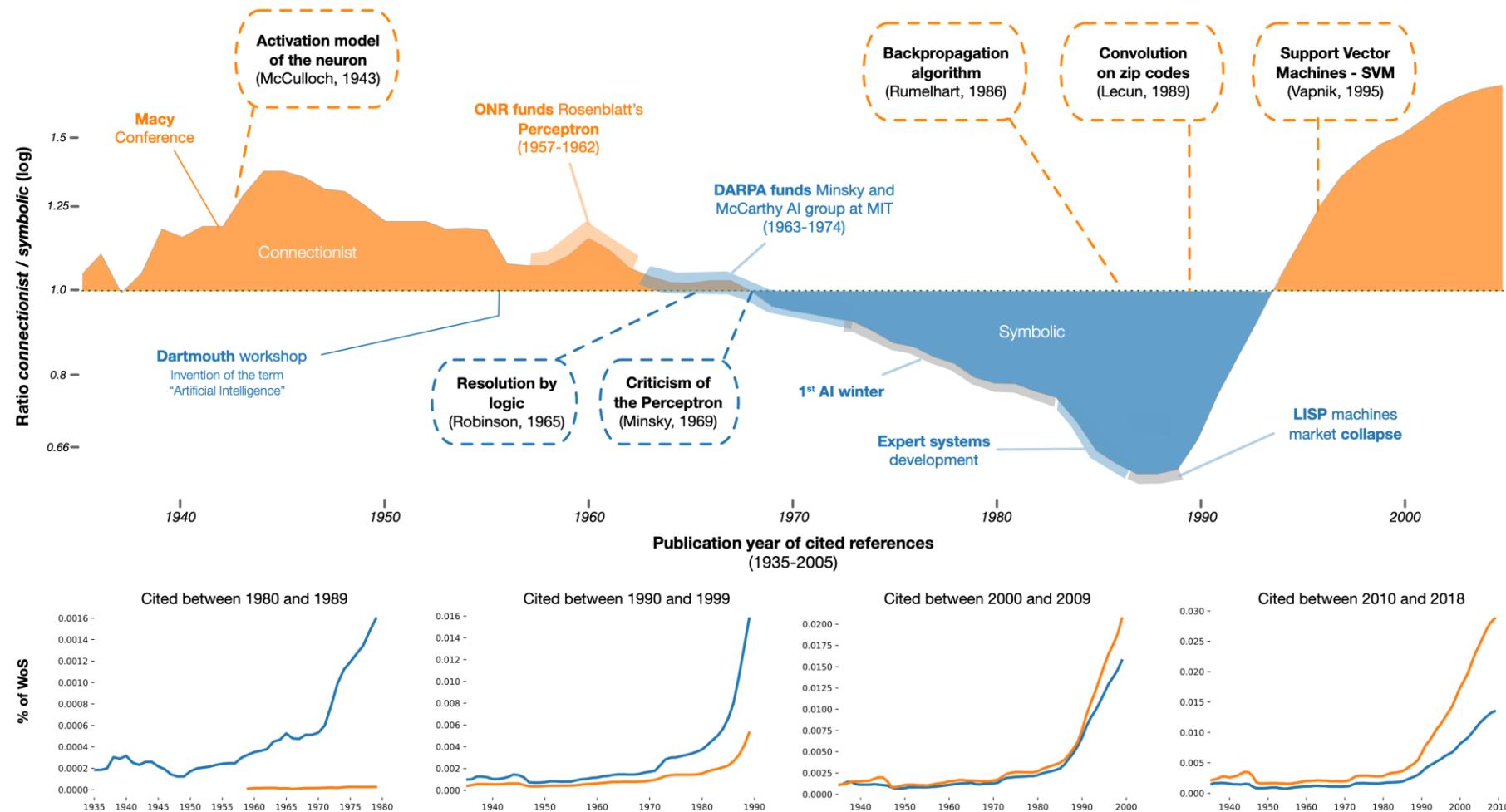
Marcando os dados feliz



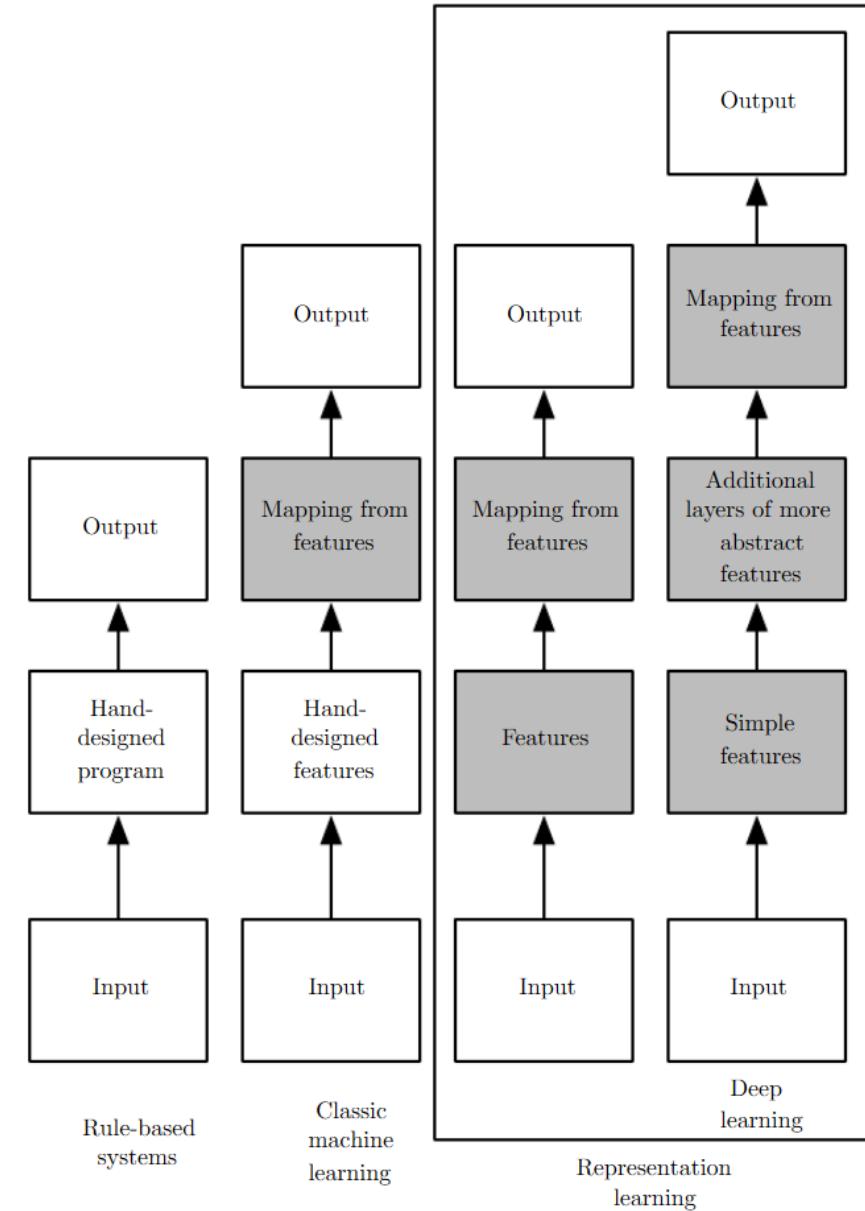
- Depois de 9 horas de treino
- Acordei e vejo a loss de 200 =(

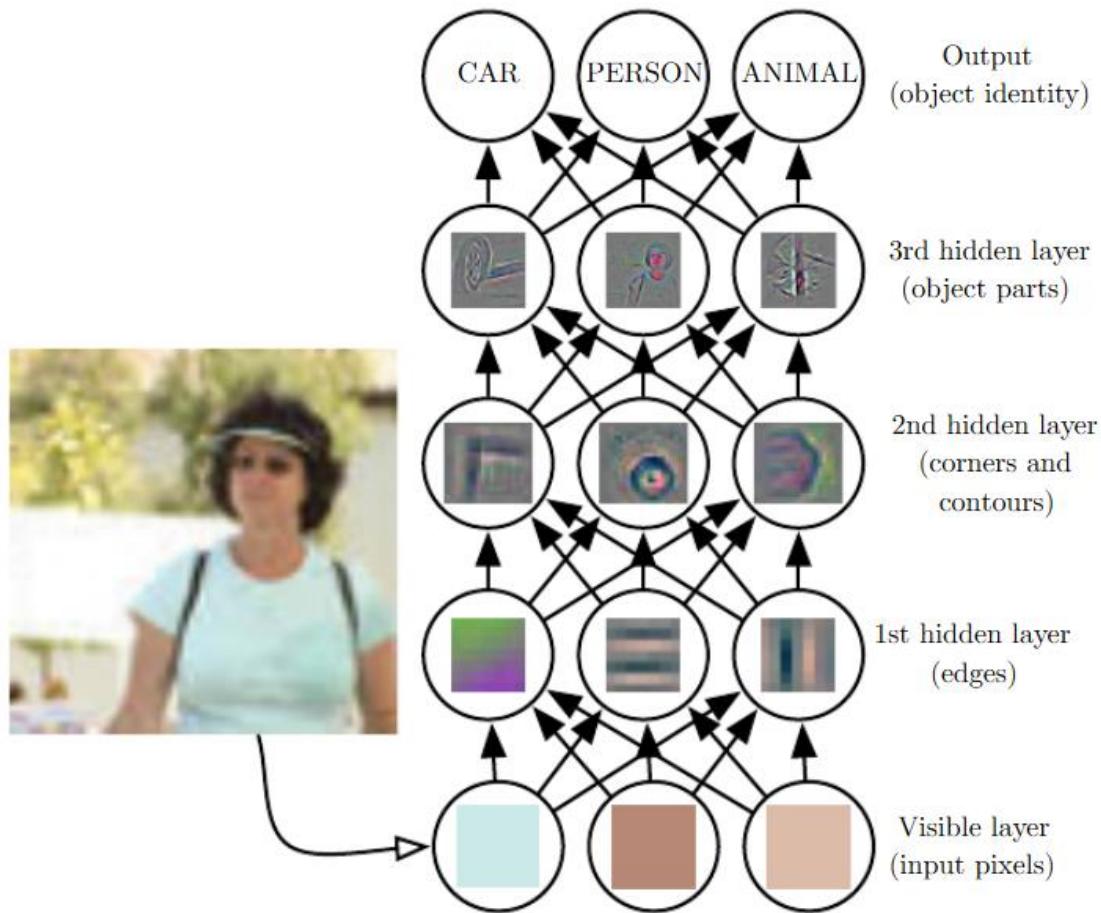


Nem tudo se resolve com Redes Neurais



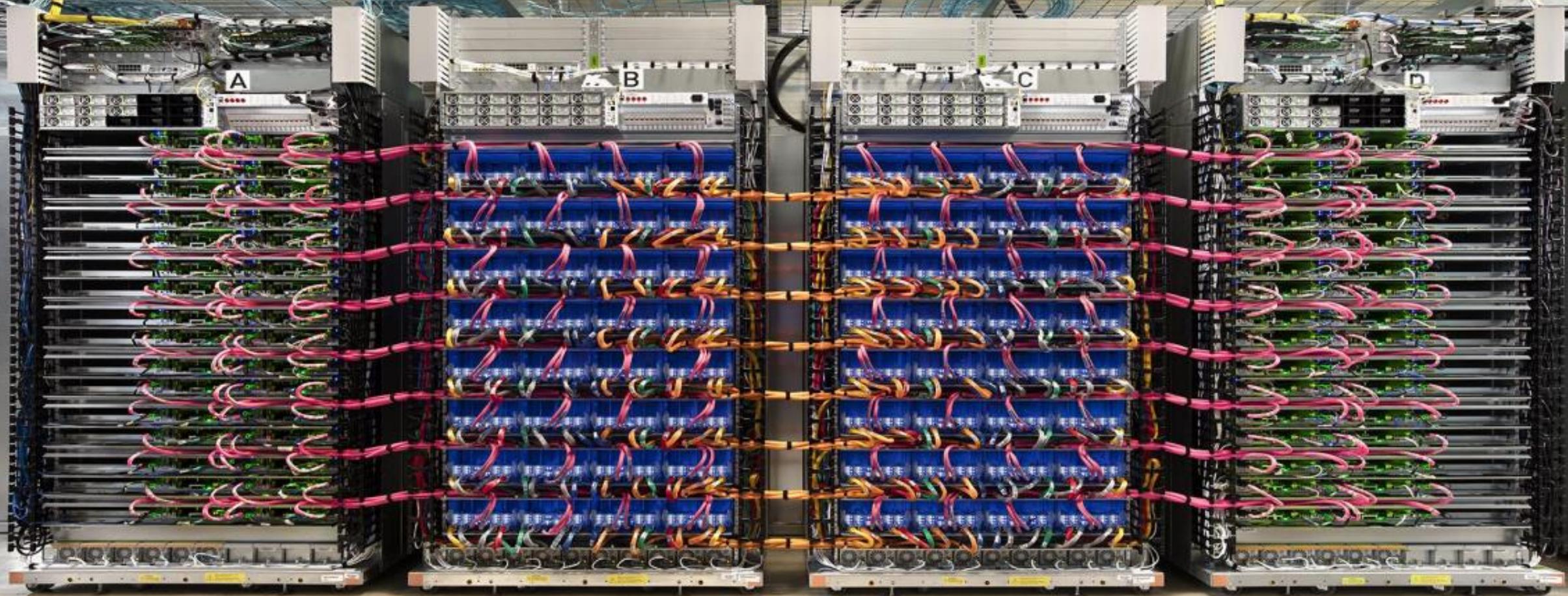
E pq Deep Learning?





Veja como representa bem.

Parece que fez a seleção de features que vc teria que fazer na mão sozinho



TPU Pod
64 2nd-gen TPUs
11.5 petaflops
4 terabytes of memory



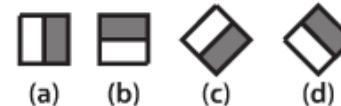
Cascade classifiers with HAAR Features

Usar o OpenCV para criar um detector de Objetos que extraí HAAR-like Feature e alimenta uma Random Forest

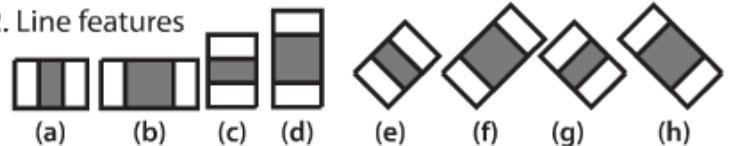
10 Segundos



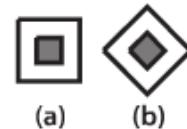
1. Edge features



2. Line features

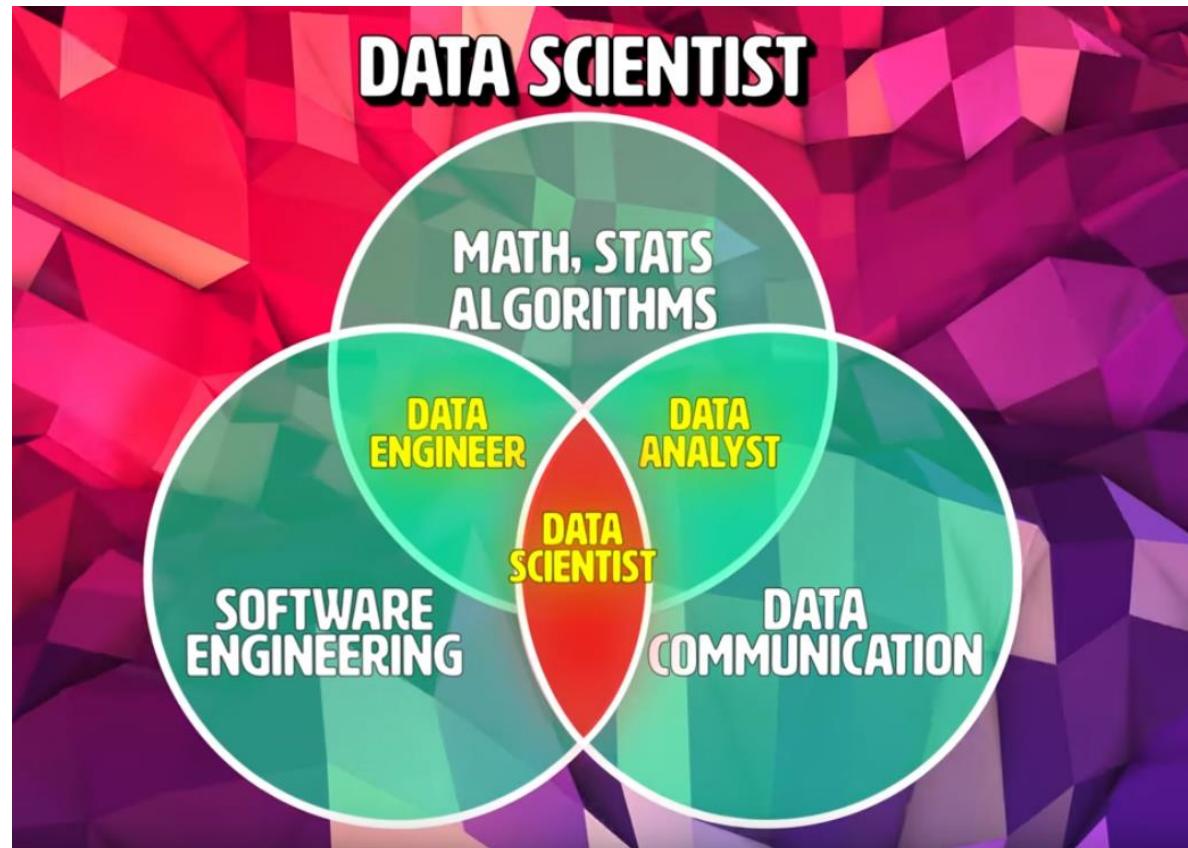


3. Center-surround features

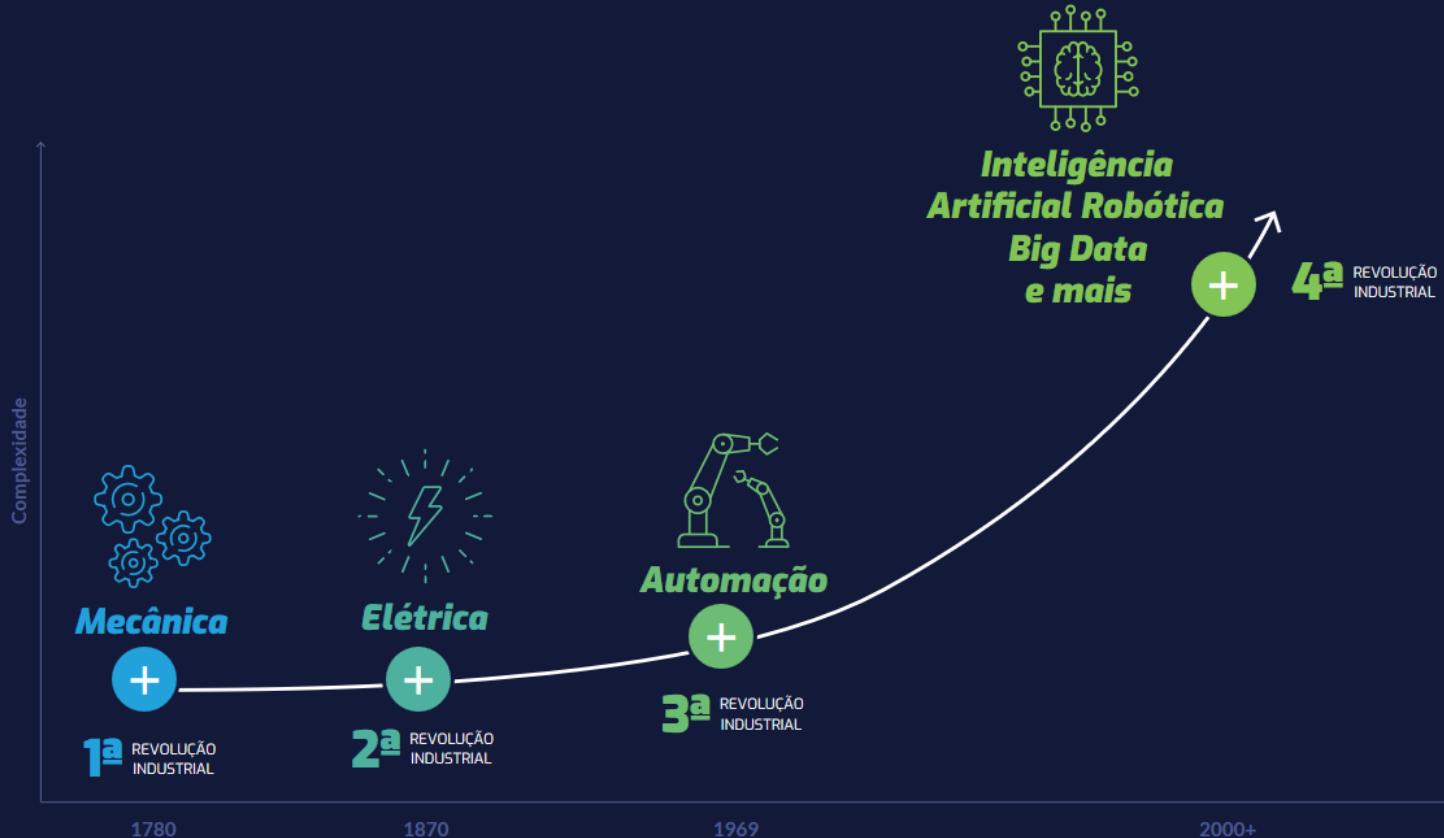


Vamos Mostrar o BOT de Whatsapp

Pq Data Science ?



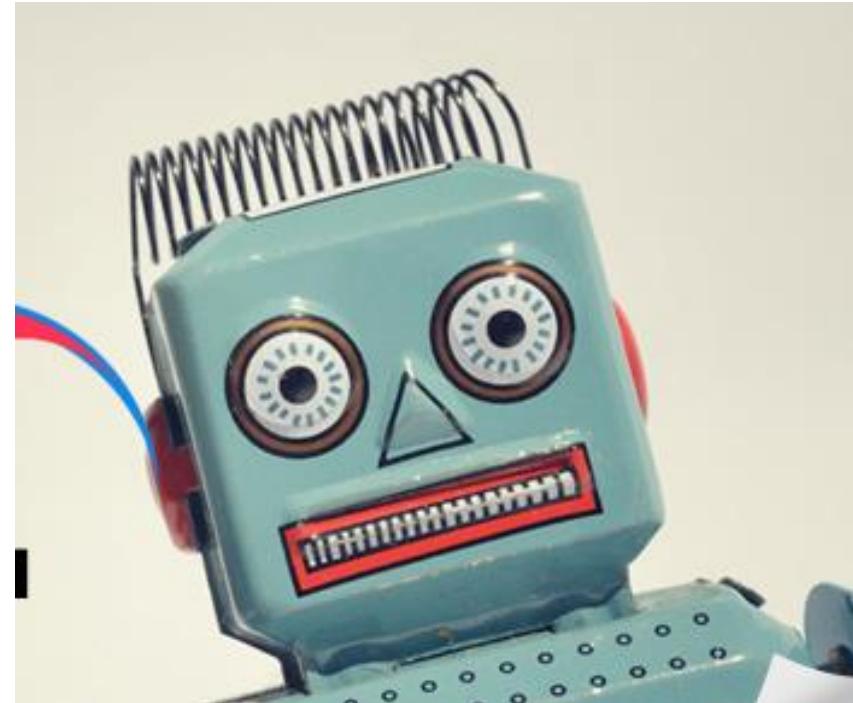
Indústria 4.0



“Completa descentralização do controle dos processos produtivos e uma proliferação de dispositivos inteligentes interconectados, ao longo de toda a cadeia de produção e logística.”

O Profissional 4.0

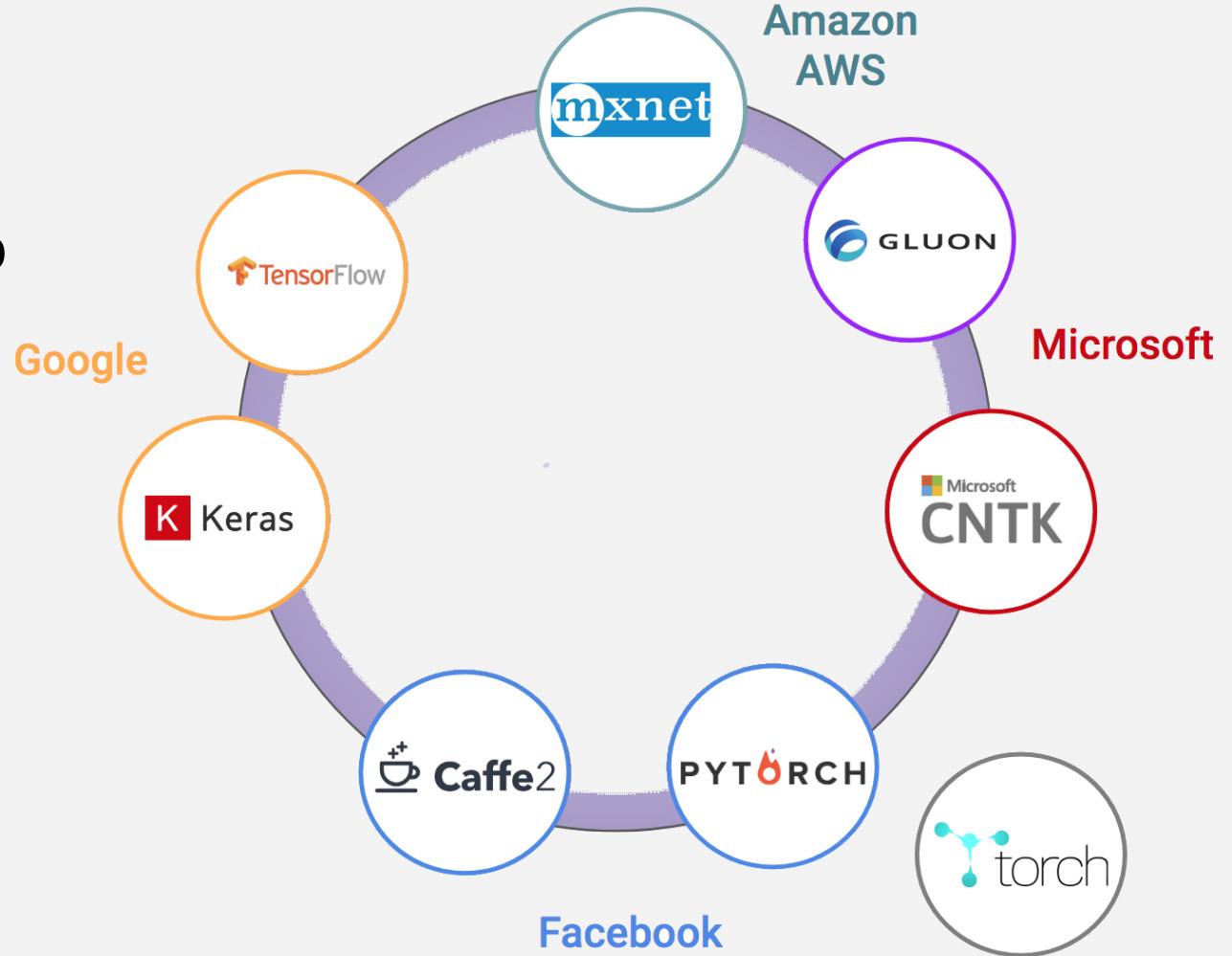
- Formação multidisciplinar e flexível
- Domínio de novas ferramentas, idiomas e competências emocionais.
- Consegue olhar de forma sistêmica para os processos da empresa e seus respectivos reflexos.
- Eles **não desempenham funções repetitivas**, são mestres em automatizar e integrar as soluções tecnológicas
- **Consegue coletar e realizar a análise de um grande volume de dados**



O valor está nos dados

O valor está nas suas Peças,
Contratos, Cláusulas, Modelos,
Contatos, network, conhecimento
de mercado...

**Você já começou a guardar
os seus dados?**





Venha trabalhar em uma
startup de Data Science



Advogado do escritório
Peixoto e Silva (SP)

Processo trabalhista
com grandes chances
de ganho

Acompanhamento Processual

1359785-80.2019.8.26.089

• 97% CHANCES DE GANHO

2 recursos restantes

VER RECURSOS POSSÍVEIS 

ENVOLVIDOS NO PROCESSO

 Gabriel Peixoto 

 Marivaldo Júnior 
OAB 123456/BA

• Trabalhou na empresa
de 2016 a 2018



Algumas referencias legais para começar

- <https://www.deeplearningbook.org/>
- <https://www.khanacademy.org/>
- <https://www.kaggle.com/learn/overview>
- <http://aiplaybook.a16z.com/>
- <https://towardsdatascience.com>
- <https://arxiv.org/>



Siraj Raval

Muito Obrigado



Bruno Cabral
bruno@potelo.com.br