



Marcelo Filipak
Diretor de Inovação da AFRAC
CEO Filiart

I.A. NA PRÁTICA

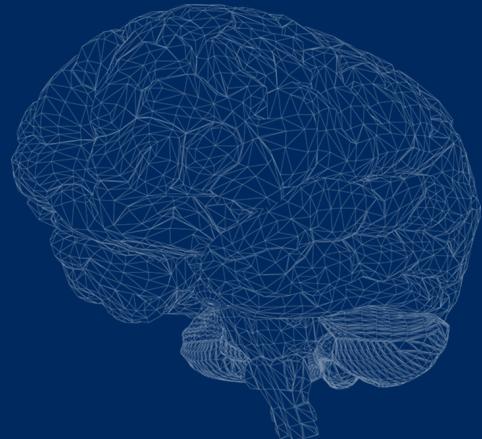
técnicas, cases reais e mão na massa.



I.A. NA PRÁTICA

técnicas, cases reais e mão na massa.

Marcelo Filipak, Engenheiro Eletrônico e Cientista de dados.



- . Treino redes neurais há 7 anos.
- . Me especializei em criar modelos.
- . Desenvolvo projetos de modelos de deep learning com aplicações na área de saúde, segurança pública, controle de tráfego, logística, inteligência fiscal, controle aduaneiro, controle de qualidade ...
- . Trabalhei com hardware e software para automação comercial de 1990 a 2018 na BEMATECH e na TOTVS.
- . CEO na FiliArt - Projetos e Advisory em I.A

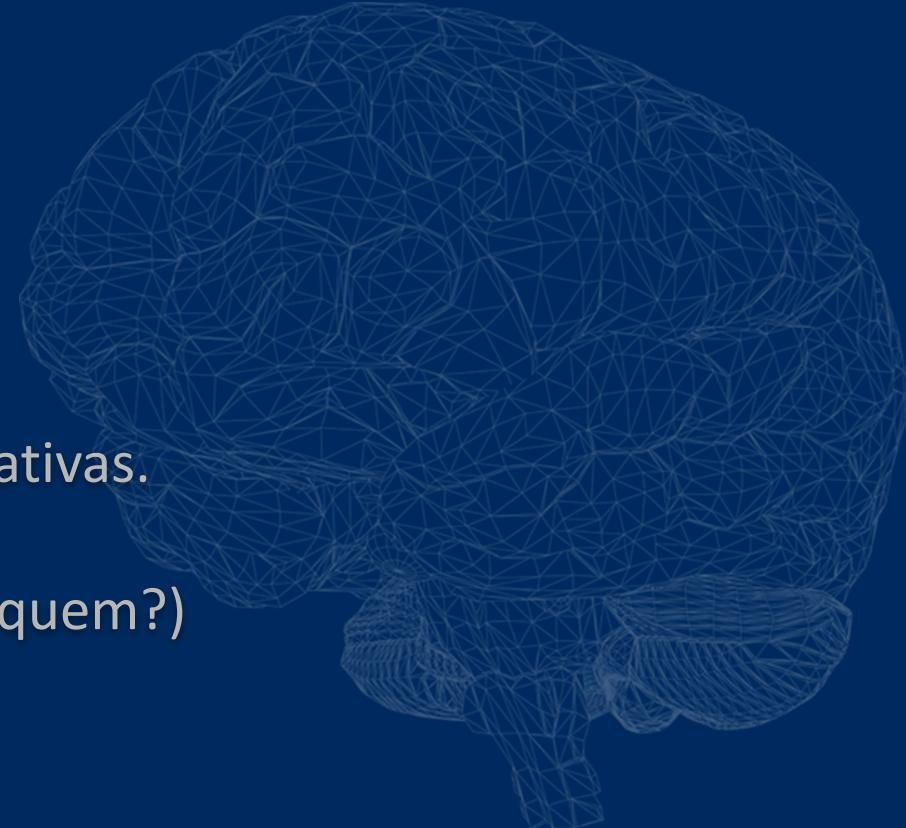
Parte 1 - Cases

Parte 2 - Que tipo de I.A eu preciso?

Parte 3 - Avanços Recentes e Expectativas.

BÔNUS 1 - Evolução da I.A (Quem é quem?)

BÔNUS 2 - Onde e o que aprender.

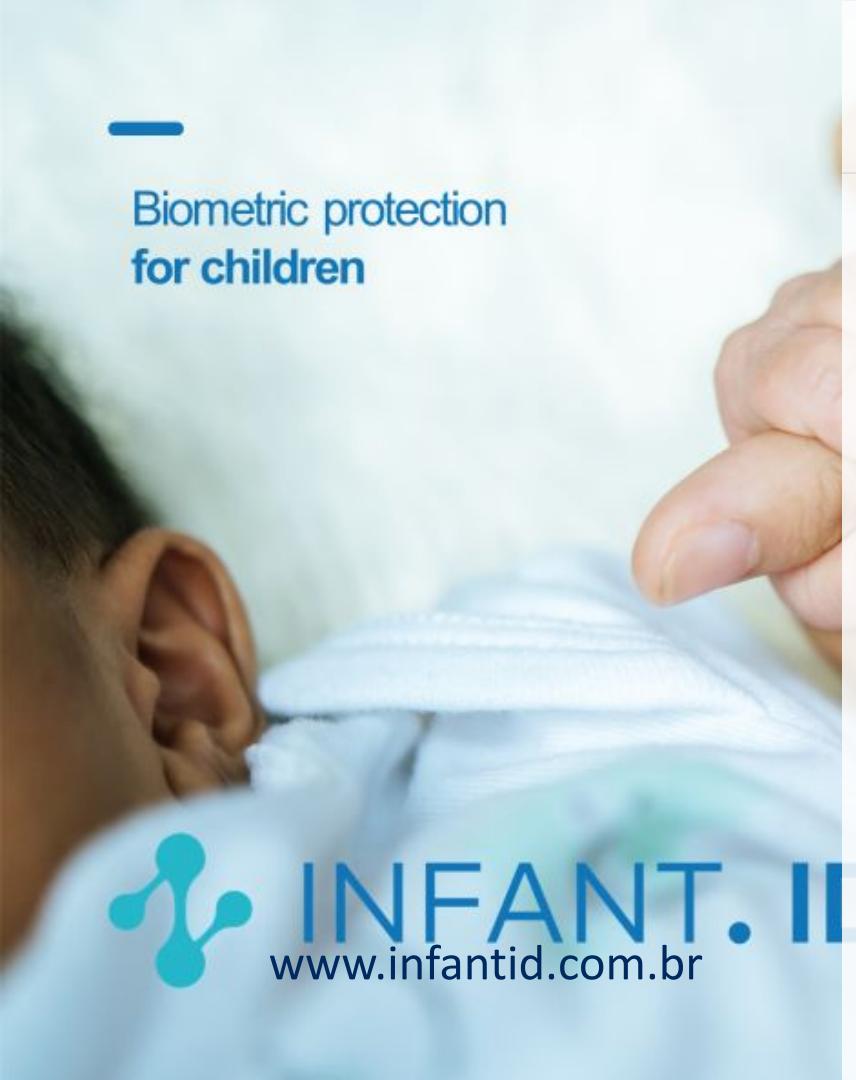




—

Biometric protection
for children

 INFANT. ID
www.infantid.com.br



Biometric protection
for children

 INFANT. ID
www.infantid.com.br

Google

biometria recém nascidos

Todas Shopping Imagens Vídeos Notícias Web Maps Mais

[www.youtube.com > watch](http://www.youtube.com/watch)

Biometria de recém-nascidos no Piauí é destaque no Jornal ...



[globoplay.globo.com > ...](http://globoplay.globo.com)

Jornal Hoje | Piauí é o primeiro estado a usar a biometria ...



[globoplay.globo.com > ...](http://globoplay.globo.com)

PITV 2ª Edição | Novo RG dos bebês: Piauí é o primeiro ...



[globoplay.globo.com > ...](http://globoplay.globo.com)

Boa Noite Paraná | Começam os testes para a identificação ...



Biometria é coletada dos recém-nascidos e das mães. Programa deve ser ampliado para todas as maternidades públicas do estado.

Globoplay · 11 de fev. de 2022

Projeto - INFANT.ID



INFANT.ID

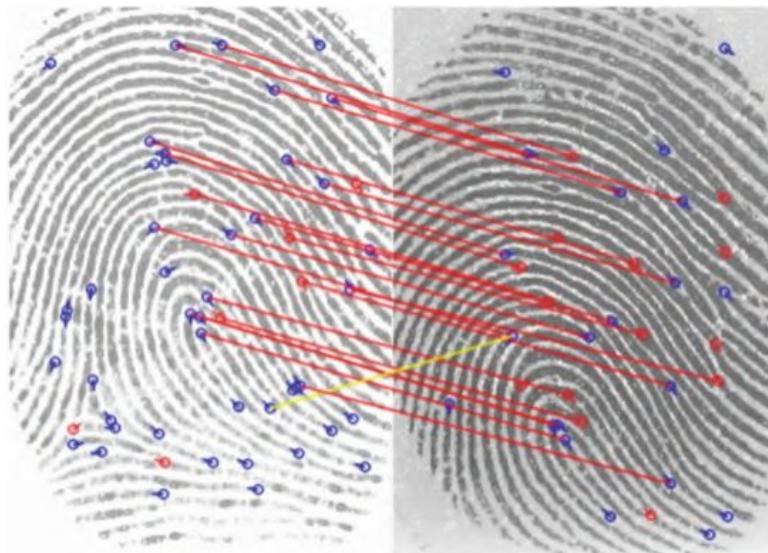
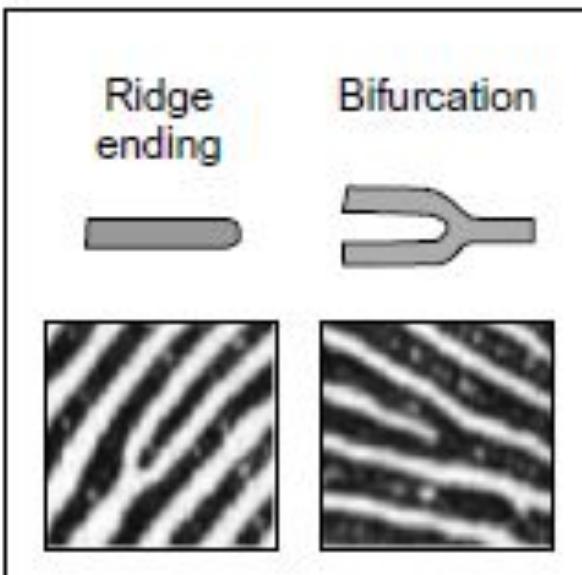
Construindo um mundo mais seguro para as nossas
crianças

[CONHEÇA A PLATAFORMA INFANT.ID](#)



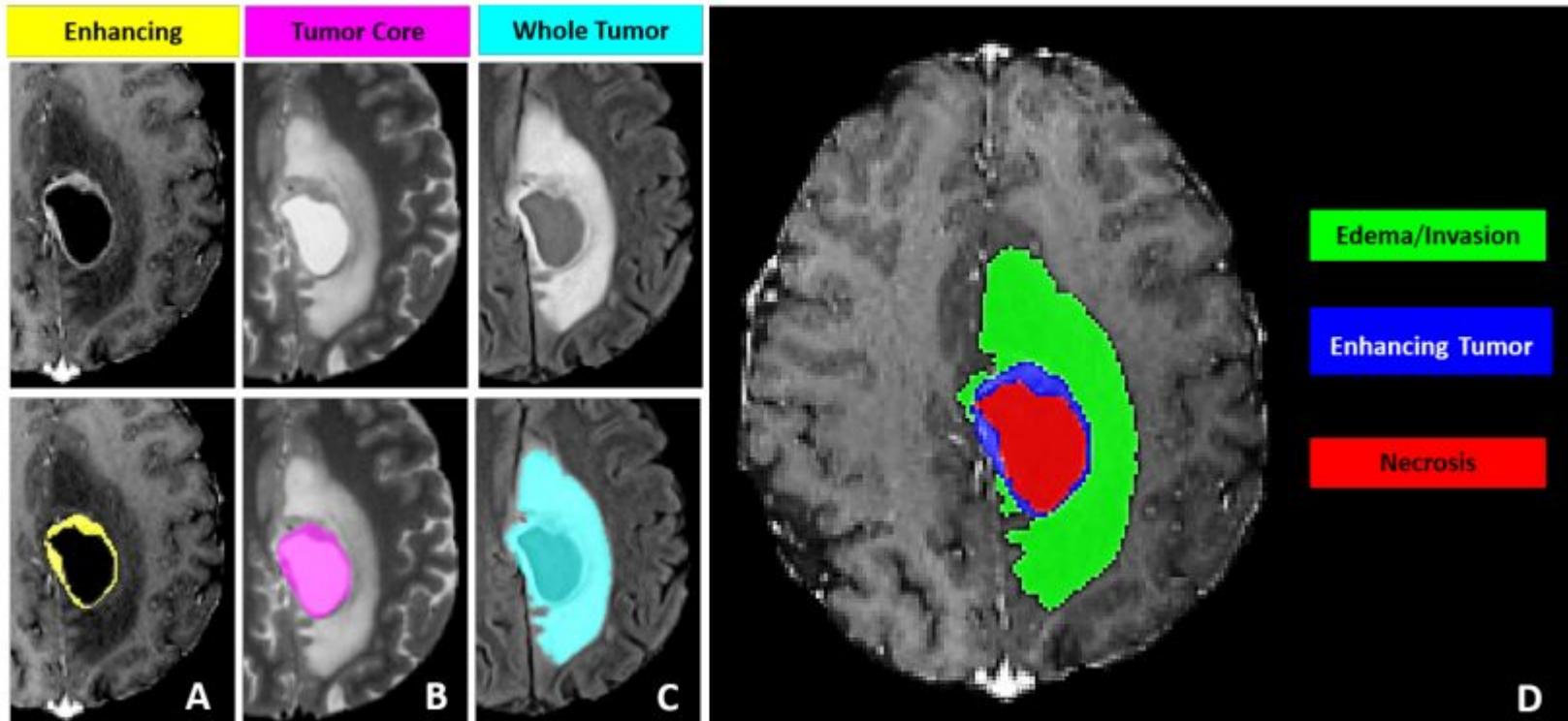


Projeto - INFANT.ID

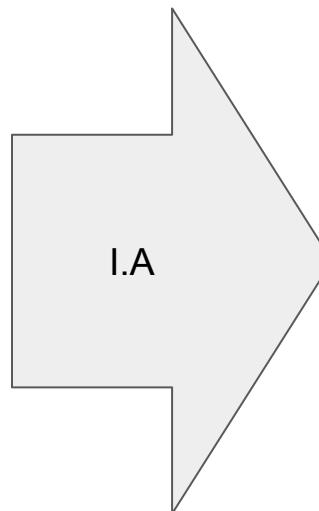
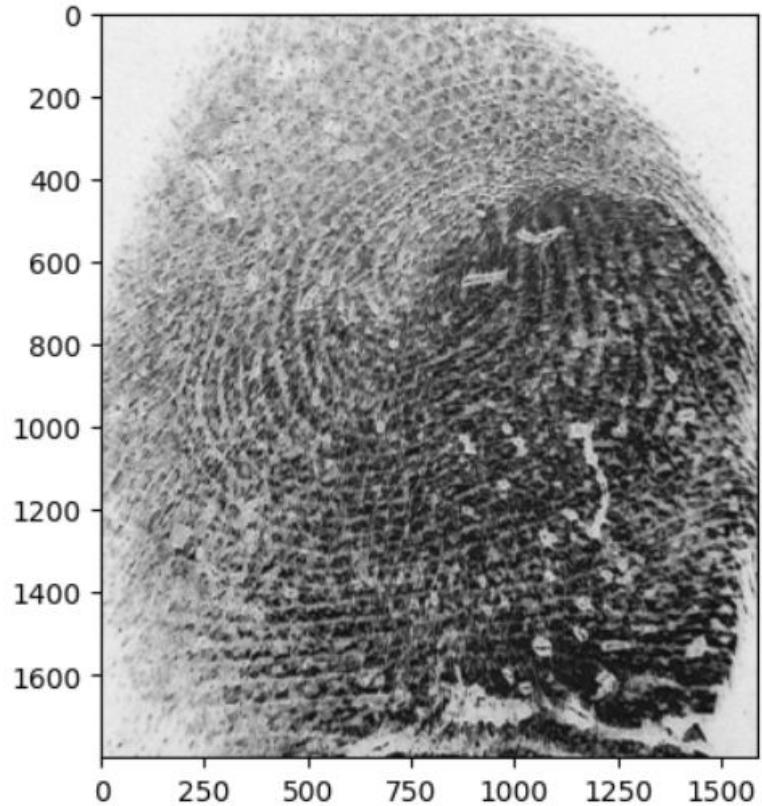


Para adultos é uma tecnologia madura e padronizada.

Problema semelhante



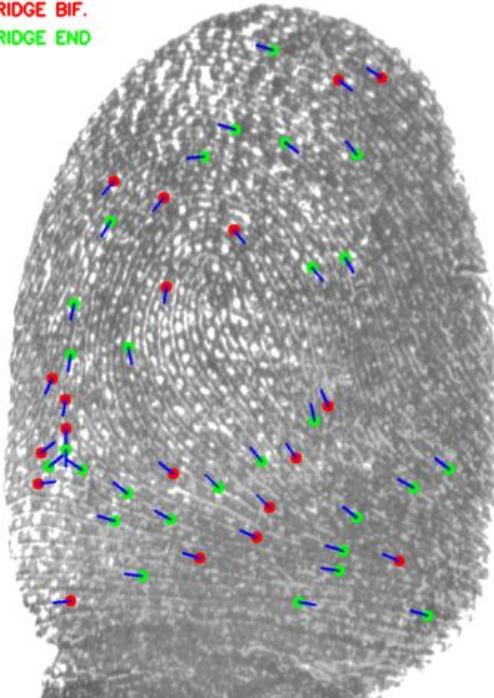
Projeto - INFANT.ID



Projeto - INFANT.ID

MINUTIAE LABEL

- RIDGE BIF.
- RIDGE END



MINUTIAE PREDICTION

- RIDGE BIF.
- RIDGE END

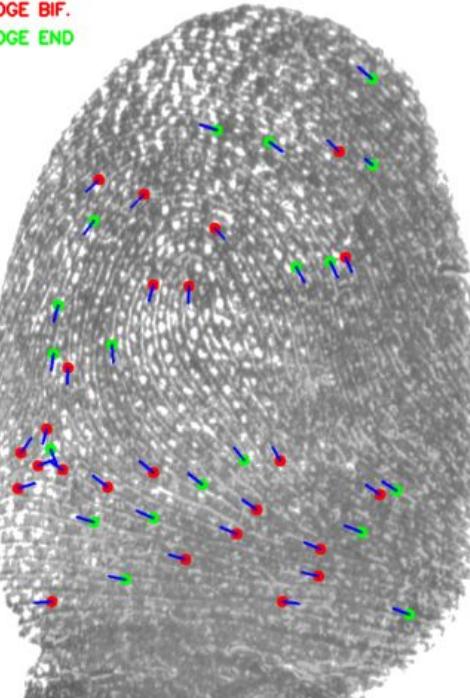
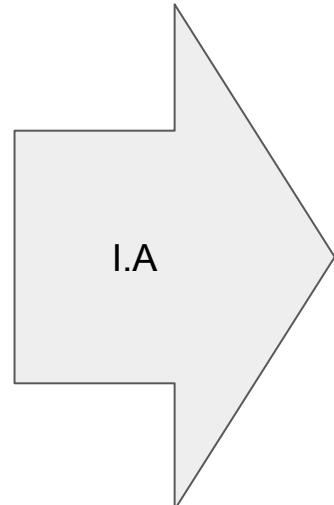


Imagen “boa” de
recém nascido.



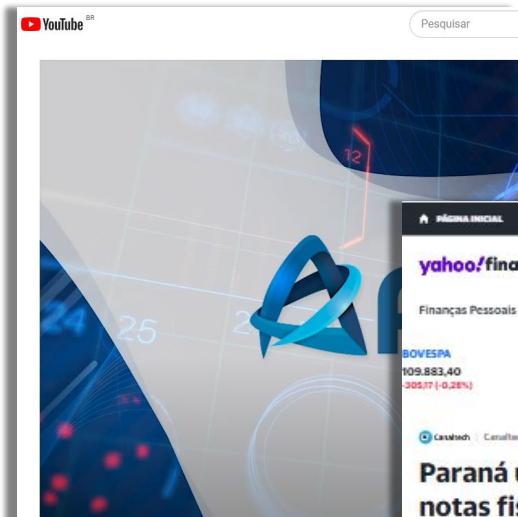
Projeto - INFANT.ID

Treinar redes pode levar tempo e \$

epoch_DICE_VALLEYS
tag: epoch_DICE_VALLEYS



Projeto - AFRAC Nota Certa



yahoo! finanças

PÁGINA INICIAL MAIL NOTÍCIAS FINANÇAS ESPORTES CELEBRIDADES VIDA E ESTILO

Buscar por notícias, códigos de negociação ou empresas

Finanças Pessoais Investimentos Empreendedorismo Trabalho Líderes Agro

BOVESPA 109.883,40 -305,37 (-0,28%) Merval 38.390,84 +233,89 (+0,61%) MXX 51.020,65 +221,19 (+0,44%) PETROLEO CRU 74,30 +0,05 (+0,07%)

Canaltech | Canaltech

Paraná usará inteligência artificial para preencher e corrigir notas fiscais

Kaique Lima | 25 de agosto de 2022 às 15h20

A Secretaria da Fazenda do Paraná (SEFA/PR) firmou uma parceria com a Associação Brasileira de Tecnologia para o Comércio e Serviços (AFRAC) para o desenvolvimento de um projeto-piloto que envolve o uso da inteligência artificial para preencher e corrigir notas fiscais. O objetivo do acordo é reduzir a complexidade fiscal para os varejistas e para o fisco.



A Secretaria da Fazenda do Paraná (SEFA/PR) firmou uma parceria com a Associação Brasileira de Tecnologia para o Comércio e Serviços (AFRAC) para o desenvolvimento de um projeto-piloto que envolve o uso da inteligência artificial para preencher e corrigir notas fiscais. O objetivo do acordo é reduzir a complexidade fiscal para os varejistas e para o fisco.

1. 2. 3. 4.

1. 2. 3. 4.

1. 2. 3. 4.

1. 2. 3. 4.

1. 2. 3. 4.

Projeto - AFRAC Nota Certa



APP AFRAC



WEB APP HUGGINGFACE



API - WS



Aplicativo AFRAAC



Ao acessar a tela inicial do Aplicativo, busque pela Opção Nota Certa



O aplicativo é
grátis e está
disponível
mesmo para não
associados.

Aplicativo AFRAC

Digite Livremente a
Descrição do
Produto que deseja
descobrir o NCM.



Aplicativo AFRAC



O Aplicativo irá informar uma lista de prováveis NCMs indicando o grau de confiança da resposta.

Reparem que eu propositalmente escrevi “Brhma” errado, então o Nível de confiança foi apenas de 48%

Aplicativo AFRAC

Quando a descrição é um pouco mais completa, os resultados são ainda melhores.



Desta vez o nível de confiança na resposta é de 95%



Baixe o APP e teste com seus próprios produtos

WEB APP - HUGGINGFACE

AFRAC NOTA CERTA

Descrição
Feijão Preto

Clear Submit

NCMs

07133319

07133319	33%
07133399	14%
07133329	5%
20055100	5%
21069090	4%

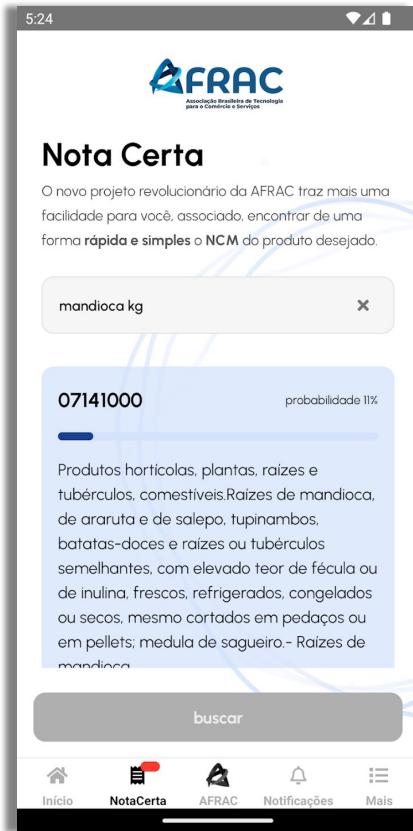
Descrição do NCM

Produtos hortícolas, plantas, raízes e tubérculos, comestíveis. Legumes de vagem, secos, em grão, mesmo pelados ou partidos. - Feijões (<i>Vigna</i> spp., <i>Phaseolus</i> spp.) -- Feijão comum (<i>Phaseolus vulgaris</i>) Preto Outros

https://huggingface.co/spaces/AFRAC/NCM_DEMO
https://github.com/mfilipak/AFRAC_IA



API - AFRAC Nota Certa



```
import requests

url = "https://afrac-ncm-demo.hf.space/api/predict"

payload = {
    "data": ["Coca-Cola PET 21"]
}

headers = {
    "Content-Type": "application/json"
}

response = requests.post(url, json=payload, headers=headers)
result = response.json()

print(result)
```



API - AFRAC Nota Certa



```
import requests

url = "https://afrac-ncm-demo.hf.space/api/predict"

payload = {
    "data": ["Coca-Cola PET 21"]
}

headers = {
    "Content-Type": "application/json"
}

response = requests.post(url, json=payload,
headers=headers)
result = response.json()

print(result)
```

A screenshot of a terminal window titled 'result' shows the JSON response from the API. The response contains a list of labels and their confidences, followed by a descriptive text in Portuguese about beverages and juices.

```
{'data': [{}], 'confidences': [{}], 'label': '22021000', 'confidence': 0.5184096693992615}, {'label': '22011000', 'confidence': 0.03791089728474617}, {'label': '22030000', 'confidence': 0.03218604251742363}, {'label': '21069000', 'confidence': 0.02866430953145027}, {'label': '22029900', 'confidence': 0.02776830829679966}], 'Bebidas, líquidos alcoólicos e vinagres. Águas, incluindo as águas minerais e as águas (sumos) de fruta ou de produtos hortícolas da posição 20.09.- Águas, incluindo a 'is_generating': False, 'duration': 0.09473180770874023, 'average_duration': 0.07237190791693515}
```

GT IOT&AI AFRAC



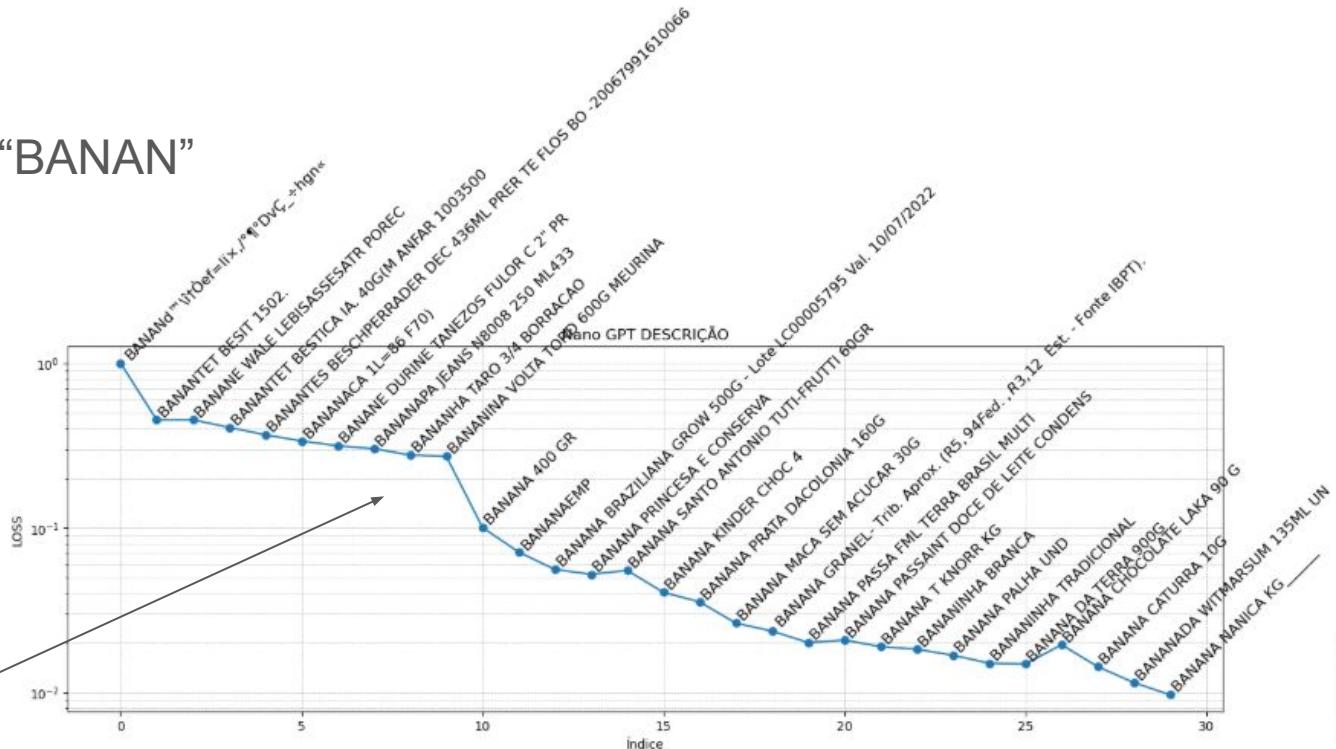
PREDIÇÃO/VERIFICAÇÃO DE GTIN

HIPÓTESE:

- ✓ 1 - Se encontrarmos uma boa representação da descrição.
- ? 2 - Se criarmos uma rede que aprenda a alinhar descrições diferentes do mesmo GTIN.
- ? 3 - Teremos como vetorizar uma tabela e encontrar o GTIN mais próximo mesmo que a descrição seja diferente.

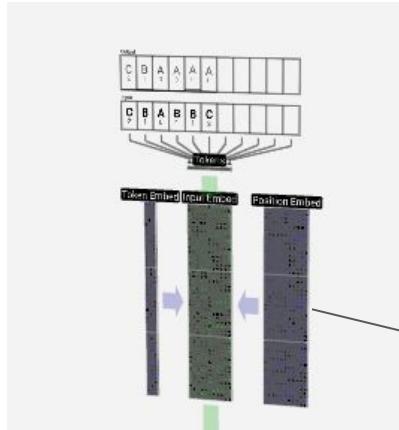
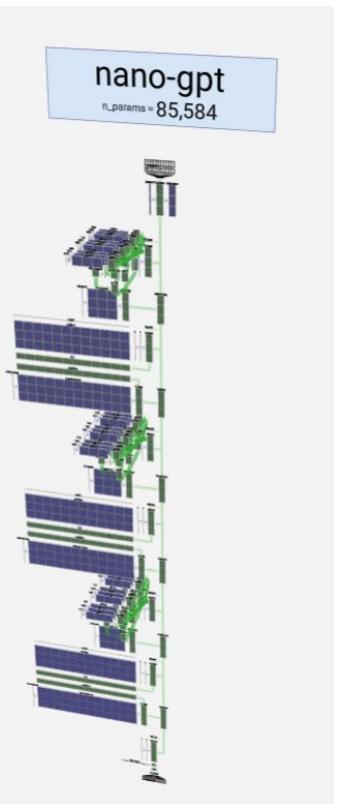
MODELO GENERATIVO DE DESCRIÇÕES - NANOGPT

GT IOT&AI AFRAC

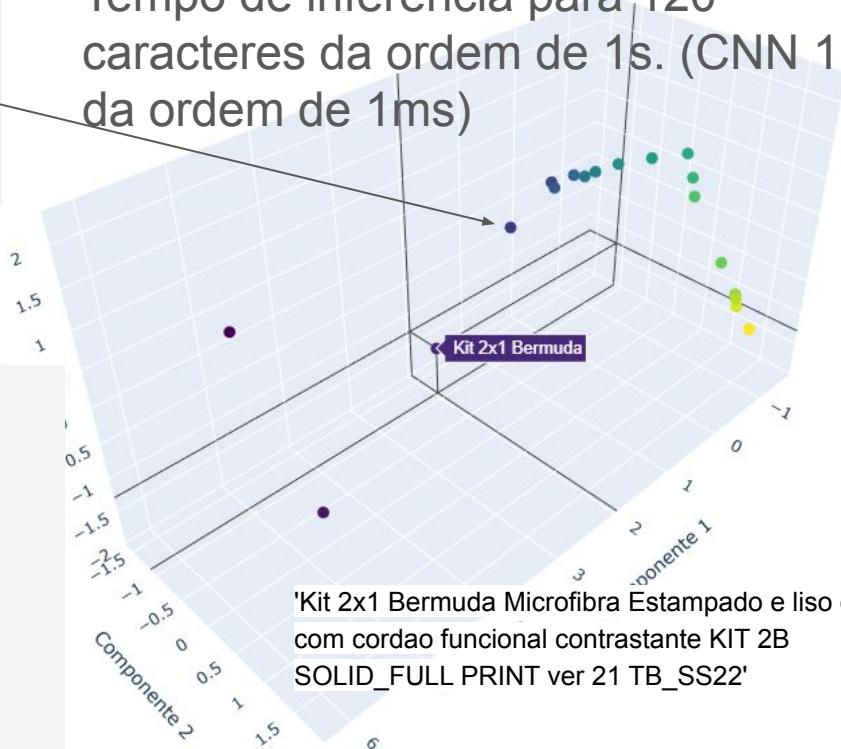


Aproximadamente 21 dias.
(51M de parâmetros)

GT IOT&AI AFRAC



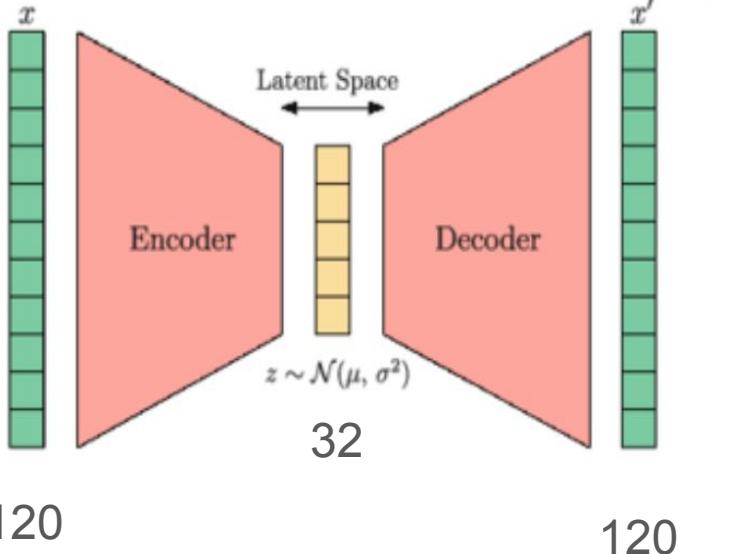
Tempo de inferência para 120 caracteres da ordem de 1s. (CNN 1D da ordem de 1ms)



MODELO GENERATIVO DE DESCRIÇÕES - VAE

BANANA

Input Sample



Aproximadamente 3 dias.
(9M de parâmetros)

GT IOT&AI AFRAC

```
[15]: #s1 = "Caixa Cabo De Rede Utp 4 Pares Cat5e Megatron 305 Homologado"
      s1 = "CEBOLITOS"
      s2 = "CEBOLA ROXA KG"

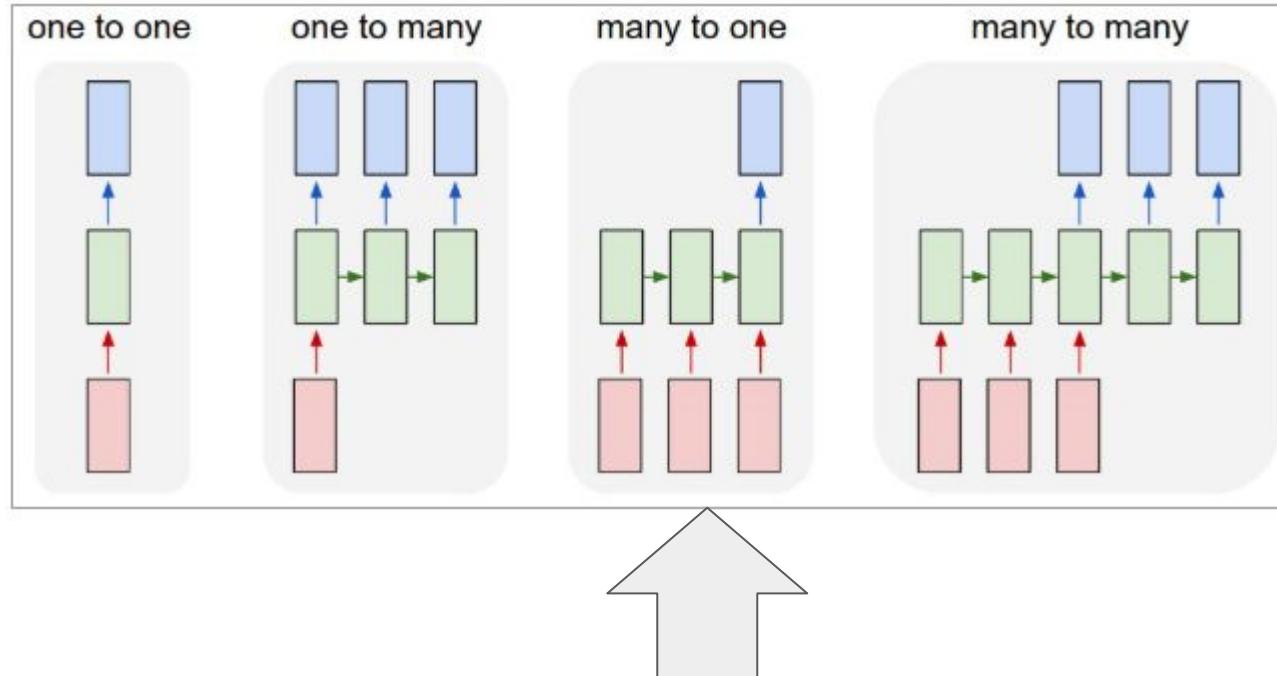
      XX = np.array([encode(_) + encode(" ") * (256 - len(_)) for _ in [s1, s2]])
      yy = decoder.predict(encoder.predict(XX, verbose=False)[0], verbose=False)

      for t in np.arange(-0.2, 1.3, 0.1):
          enc = encoder.predict(XX, verbose=False)[0]
          dec = decoder.predict((t * enc[1] + (1 - t) * enc[0])[None, ...], verbose=False)[0]
          print(f"{t:.2f} {decode(np.argmax(dec, axis=-1))[:60]}")

-0.20 CEBOLYTAS
-0.10 CEBOLITAS
0.00 CEBOLITOS
0.10 CEBOLITES
0.20 CEBOLETES
0.30 CEBOLERI::
0.40 CEBOLER.TR.
0.50 CEBOLOR TTe "
0.60 CEBOLOR ETE "
0.70 CEBOLARAENO "X
0.80 CEBOLAUIESO "G
0.90 CEBOLA YOSA KG
1.00 CEBOLA ROXA KG
1.10 CEBOLA ROJA KG
1.20 CEBOLA TOLA KG.
```

MODELO GENERATIVO DE DESCRIÇÕES - GRU

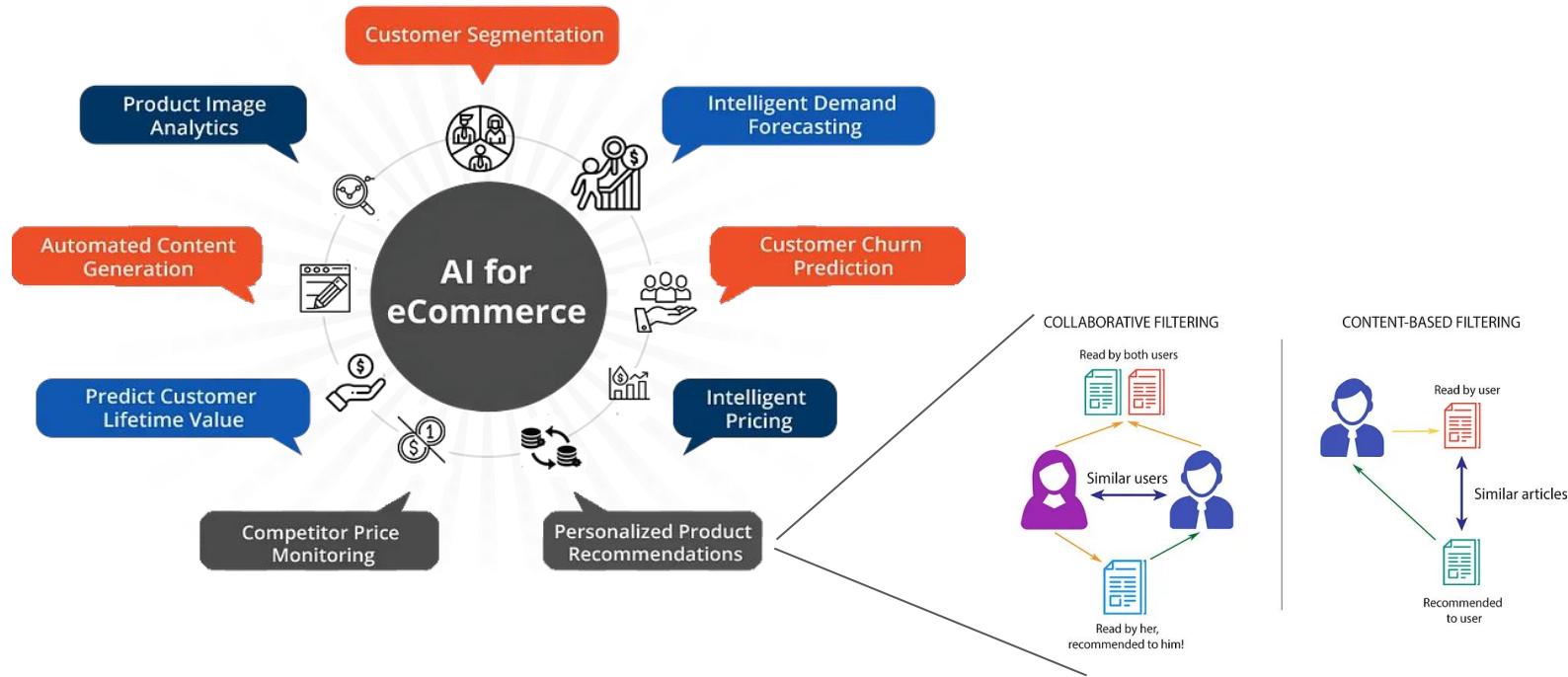
GT IOT&AI AFRAC



Daqui se extrai uma boa representação sem toda a complexidade das arquiteturas de LLM. Possivelmente GRU - Gated Recurrent Unit

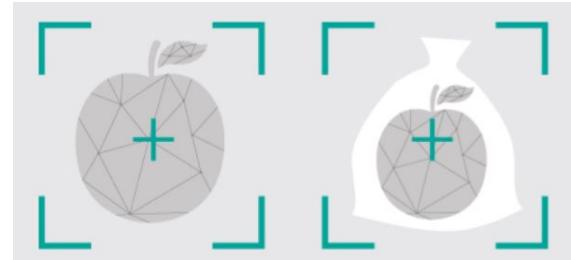
I.A. Discriminativa

I.A. no e-commerce



I.A. Discriminativa

Balanças com I.A.



I.A. Generativa

CHAT-BOTS

Coletando dados
pela interação.

Extender o
suporte para 24x7

Recomendação de
produtos,
“upselling”

Suporte aos
vendedores.

Facilitar tracking dos
pedidos.

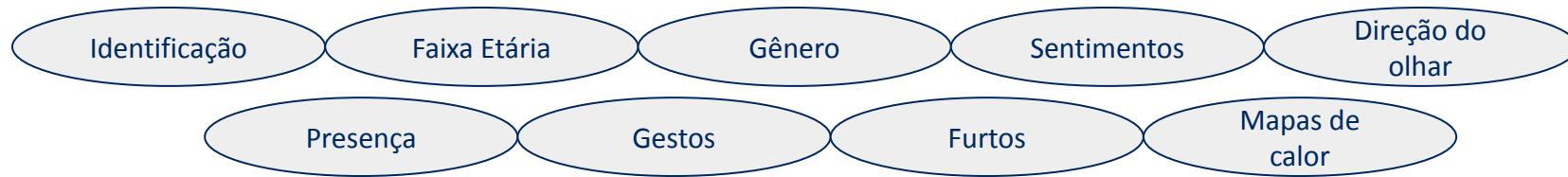
Organizar FAQ e
perguntas
relacionadas.
perplexity.ai



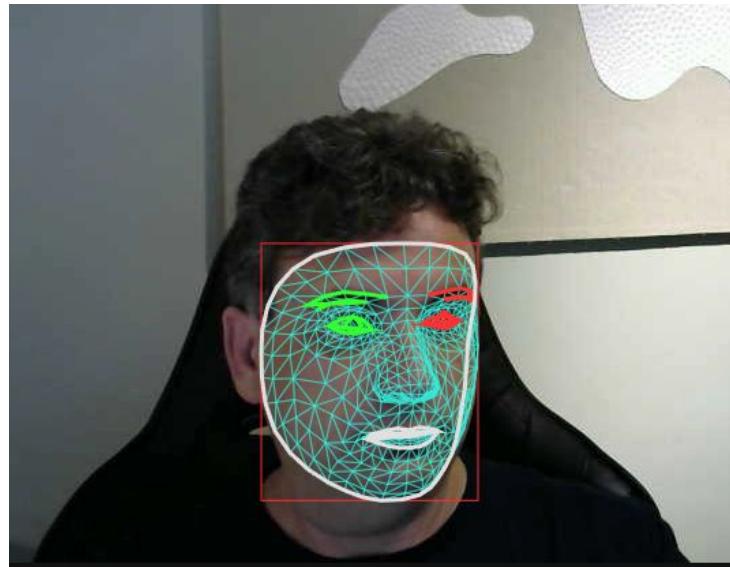
Encontrar uma loja
física.

Coletar feedbacks.

Modelos faciais e corporais



<https://bit.ly/3Y3kvjX>



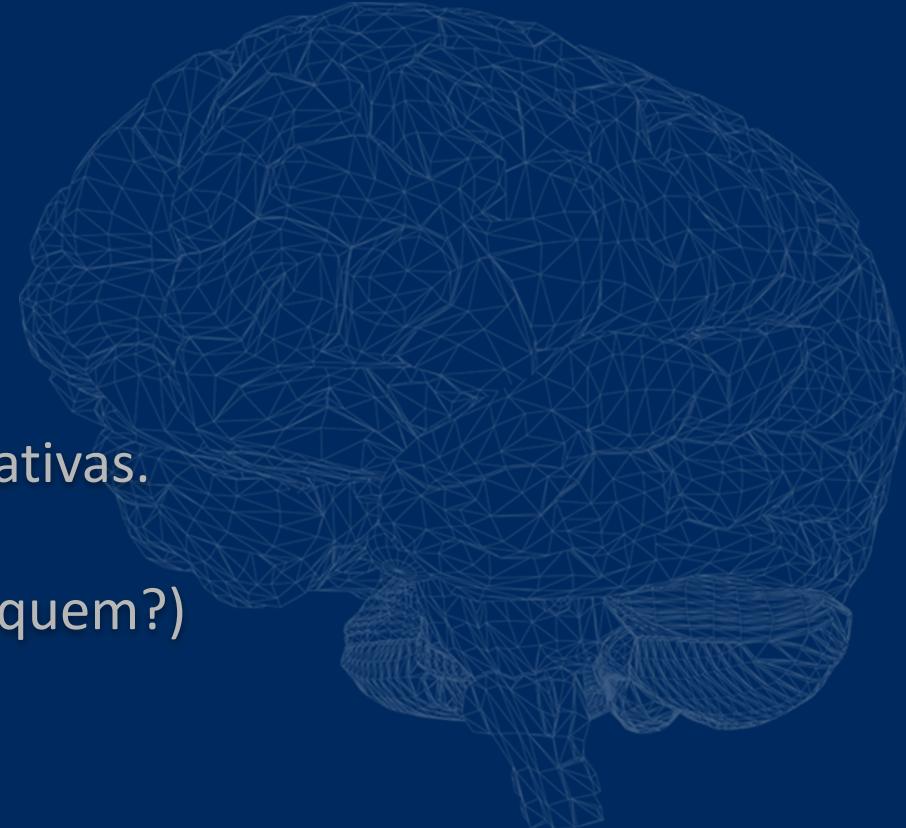
~~Parte 1 - Cases~~

Parte 2 - Que tipo de I.A eu preciso?

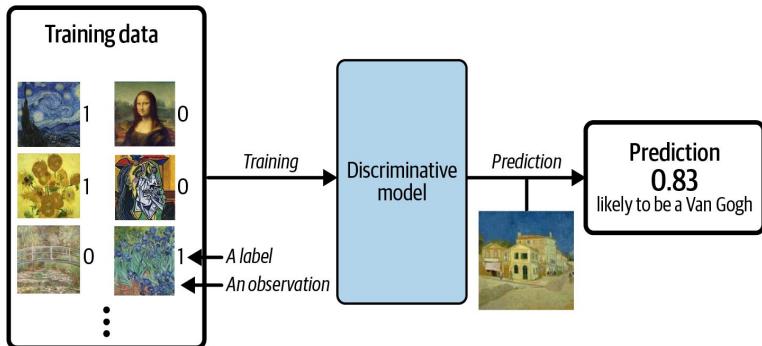
Parte 3 - Avanços Recentes e Expectativas.

BÔNUS 1 - Evolução da I.A (Quem é quem?)

BÔNUS 2 - Onde e o que aprender.

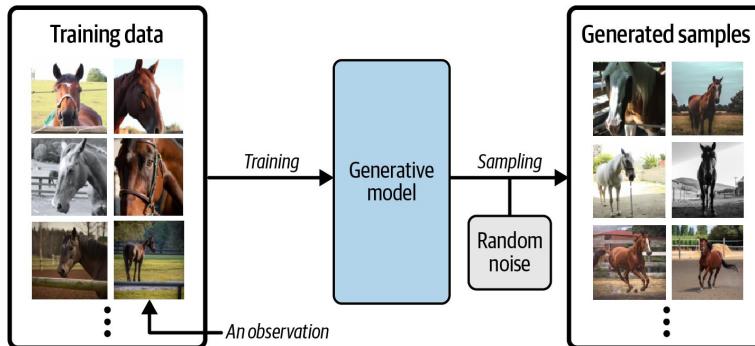


I.A. Discriminativa



Classificação, Regressão
Reconhecimento de imagens
Emoções em texto
Reconhecimento Facial
Contagem de objetos
...

I.A. Generativa



Tradução
Dublagem
Geração de imagens
Chatbots (chat-gpt)
Geração de código
...

São especializadas em problemas diferentes, embora sua arquitetura interna tenha mais semelhanças que diferenças.

Que tipo de modelo eu preciso?

Shallow Learning (Ex: Random Forest)

- . Classificação e Regressão
- Simplificada:** COLLABORATIVE FILTERING, recomendação de vendas, diagnóstico de doenças.
Read by both users.
- Análise de Crédito:** Avaliação de risco de crédito.
- Detectação de Fraudes:** Identificação de transações fraudulentas.
- Marketing:** Segmentação de clientes. Read by her, recommended to him!

Deep Learning



Generative AI

- Criação de Conteúdo:** Geração de texto, imagens, áudio. Extender o suporte para 24x7.
- Design e Arte:** Criação de designs gráficos, arte digital.
- Modelagem e Simulação:** Simulação de cenários, modelagem de personagens em jogos.
- Assistência Cognitiva:** Ferramentas de apoio à escrita, composição musical.

Uma indicação e uma contra-indicação.

Shallow Learning (Ex: Random Forest)

Generaliza bem mesmo com poucos dados.

Incapaz de tratar muitos problemas complexos.

Deep Learning

Capaz de tratar problemas complexos.

Se treinada com poucos dados tem overfit (generaliza mal)

Generative AI

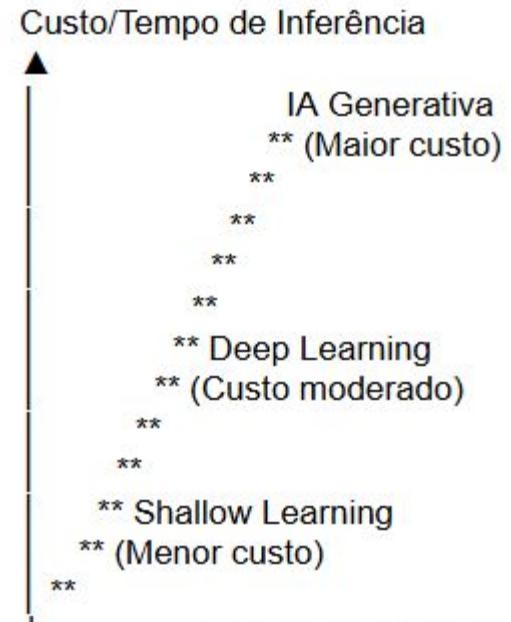
Criativa, tem uma compreensão ampla.

Alucina, precisa ser “controlada”.

Trade-off



Preciso de uma figura que mostre que indo de shallow learning para deep learning e depois para I.A generativa temos de um lado um aumento no custo e no tempo de inferência e no outro uma menor capacidade de problemas complexos



Principais características do trade-off:

1. Shallow Learning

- Custo: Requer ~1-100 GPU horas para treinamento 1 3
- Inferência: <10ms por predição em CPUs comuns 3 6
- Limitações: Eficaz apenas para padrões lineares ou de baixa dimensionalidade 4 6

2. Deep Learning

- Custo: 1,000-100,000 GPU horas (ex: AlphaGo custou milhões) 1 3
- Inferência: 50-500ms usando GPUs especializadas 6 12
- Vantagem: Resolve problemas não-lineares complexos como reconhecimento de imagens 1 6

3. IA Generativa

- Custo: >1M GPU horas (ex: GPT-3 custou ~\$12 milhões) 1 3
- Inferência: 500ms-5s mesmo com hardware otimizado 3 6
- Risco: Tendência a overfitting e alto consumo energético 1 6

Paradoxo da complexidade:

Embora modelos mais profundos tenham **capacidade teórica** superior ($\approx 97\%$ de acurácia em ImageNet) [1](#), na prática enfrentam:

1. **Lei de rendimentos decrescentes**

Melhorias de 0.1% na acurácia podem exigir 10x mais recursos [1](#) [3](#)

2. **Custo ambiental**

Treinar um modelo grande emite ≈ 284 ton CO₂ equivalente [1](#)

3. **Complexidade vs Generalização**

Modelos muito profundos (>100 layers) mostram degradação de performance [6](#) [14](#)

Para aplicações práticas, estudos recomendam:

```
python
```

```
if latency < 100ms and budget < $10k:  
    use shallow models  
elif 100ms < latency < 1s and budget < $1M:  
    use deep learning  
else:  
    consider generative AI  
```[3][6][12]
```



# Modelos prontos - onde?

HUGGING FACE

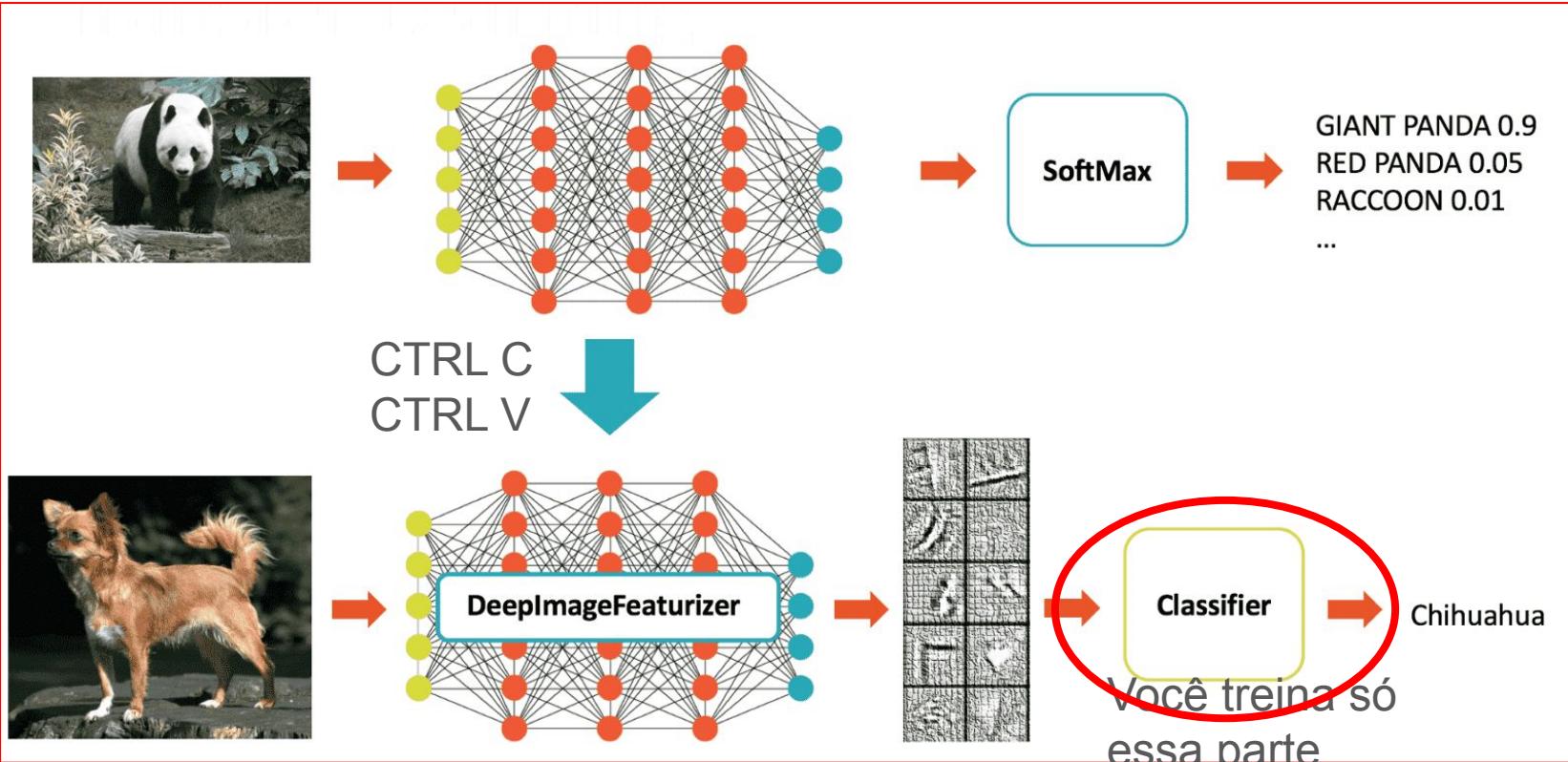


**1** - Procure um modelo pronto para o seu problema (alguns como detecção de face, biometria facial, landscapes faciais, detecção de objetos, ...) são treinados em big data e seguramente vão ser melhores que treinar do zero.

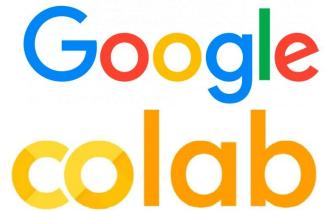
**2** - Se ele estiver no caminho, é possível fazer um “fine tuning” ou “transfer learning”. Aqui você pode congelar parte do modelo e treinar somente uma parte pequena e com relativamente poucos dados.

**3** - Decida se está bom o suficiente ou se você vai treinar um modelo próprio. Nesse cenário você vai precisar obter, tratar ou construir dados e prepará-los para treinar modelos.

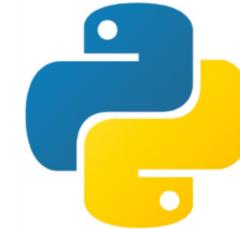
# “Transfer learning” pode poupar muito tempo e \$



**Que tipo de ambiente eu preciso? (para treinar modelos “simples” ou finetunning / inferência)**



Google  
colab



Que tipo de ambiente eu preciso?  
(para treinar LLMs do zero. ou para inferência em escala)



# Que tipo de ambiente eu preciso? para rodar meus modelos/aplicação



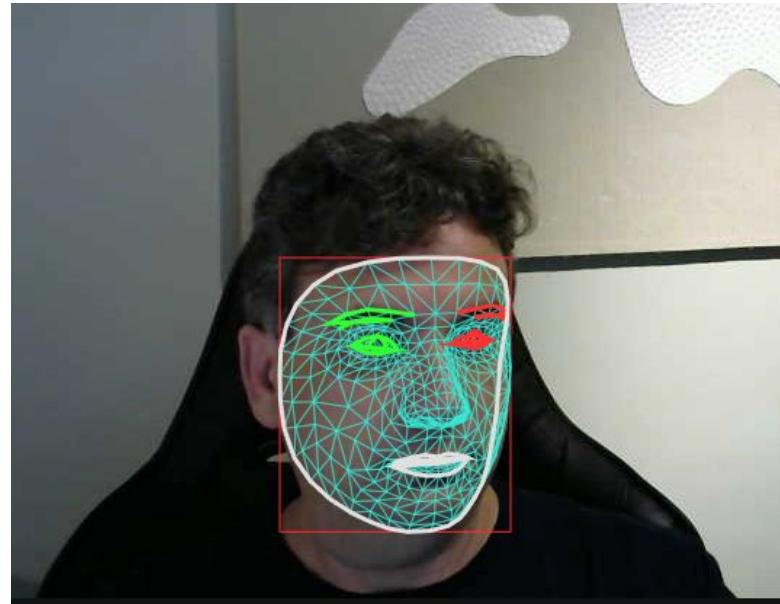
Local ou Nuvem?  
Depende da aplicação.



# Rode agora no seu smartphone.

- . O mais comum é rodar em web services (facilidade de prototipação)
- . Muitos modelos complexos rodam até num smartphone. (Exceção: LLMs)
- . Mesmo LLMs podem rodar num desktop.

<https://bit.ly/3Y3kvjX>

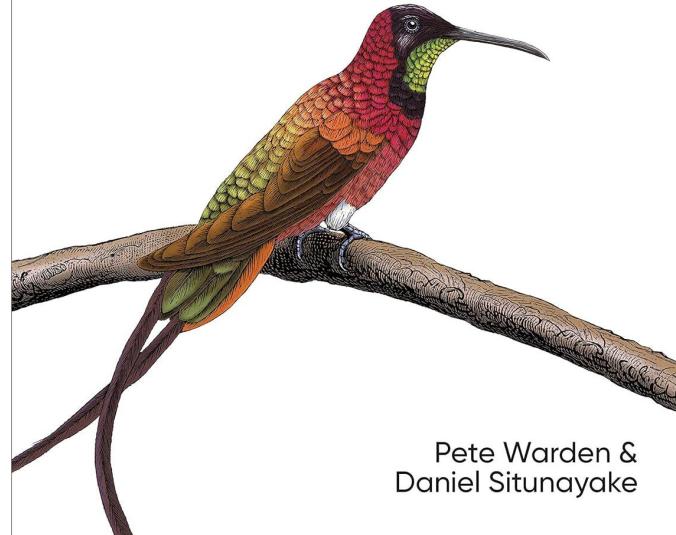




O'REILLY®

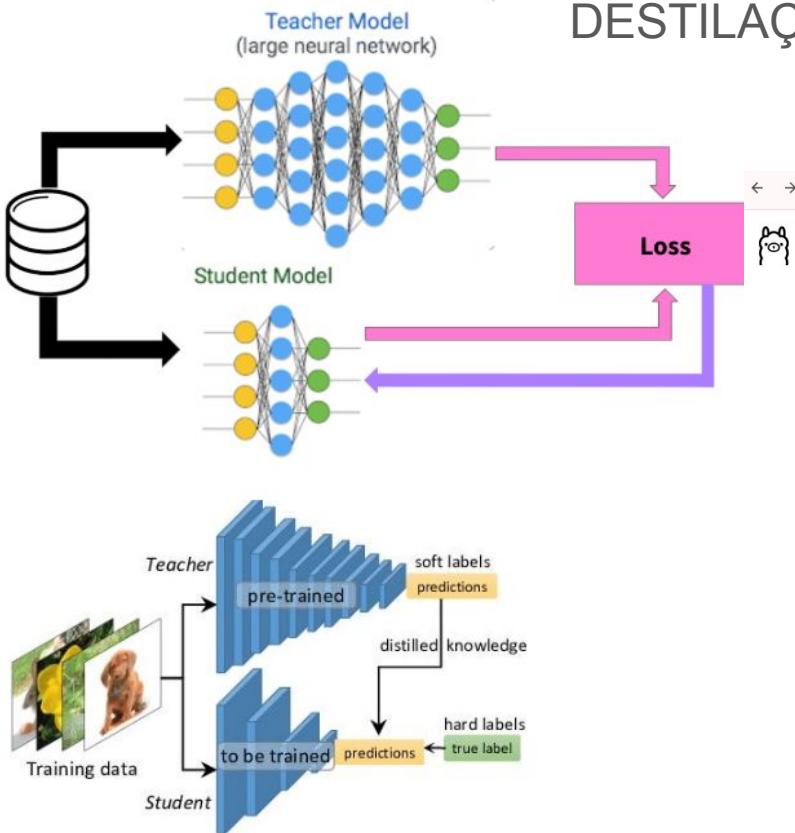
# TinyML

Machine Learning with TensorFlow Lite on  
Arduino and Ultra-Low Power Microcontrollers



Pete Warden &  
Daniel Situnayake

# DESTILAÇÃO / SMALL LLMs

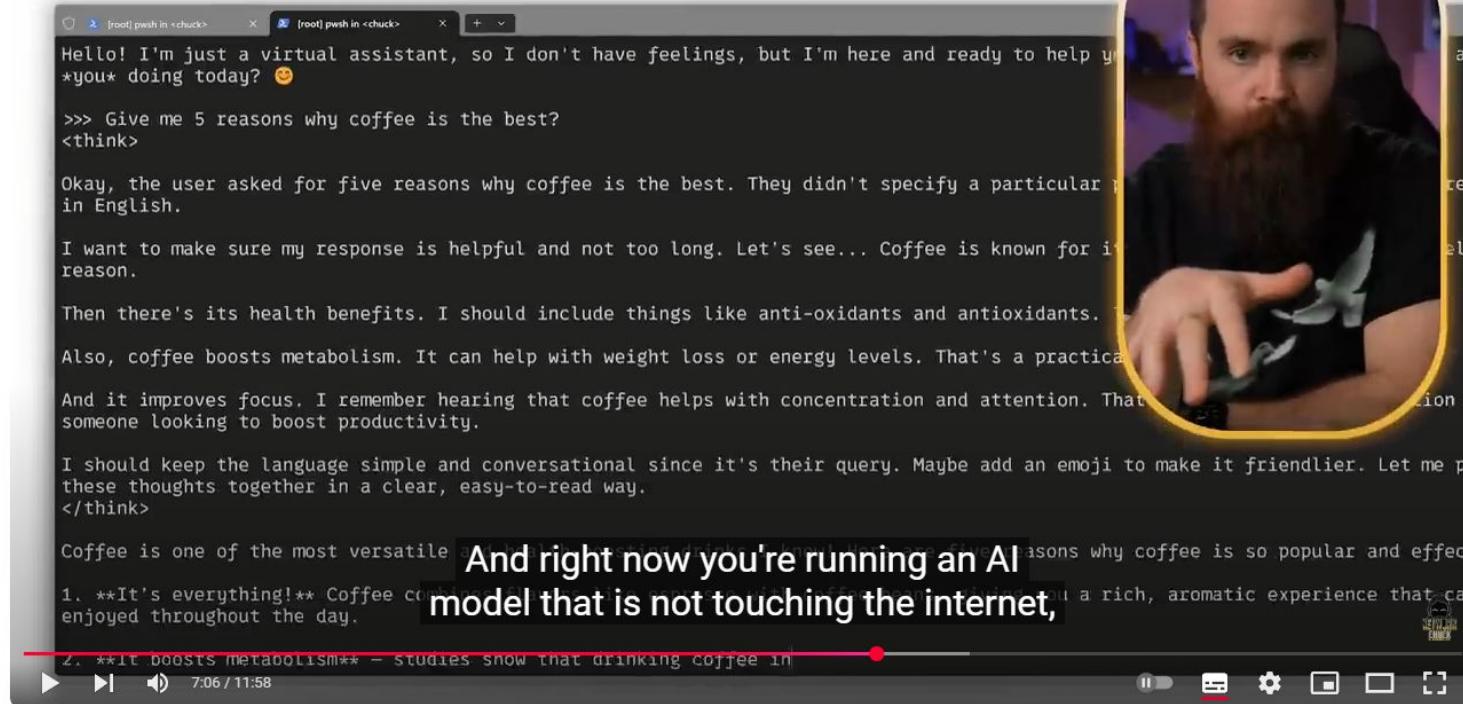


LM Studio interface screenshot showing distillation options:

- Qwen2.5 7B Instruct 1M** - Qwen (7B): Powerful general purpose instruct model with up to 1 million tokens context.
- DeepSeek R1 Distill (Qwen 7B)** - deepseek (qwen 7B): DeepSeek R1 distilled into Qwen 7B: a powerful reasoning model in a small package.
- DeepSeek R1 Distill (Llama 8B)** - deepseek (llama 8B): DeepSeek R1 distilled into Llama 8B: a powerful reasoning model in a small package.

Get up and running with large language models.  
Run Llama 3.3, DeepSeek-R1, Phi-4, Mistral, Gemma 2, and other models, locally.

Explore models →  
Available for macOS, Linux, and Windows



Hello! I'm just a virtual assistant, so I don't have feelings, but I'm here and ready to help you doing today? 😊

>>> Give me 5 reasons why coffee is the best?

</think>

Okay, the user asked for five reasons why coffee is the best. They didn't specify a particular reason.

I want to make sure my response is helpful and not too long. Let's see... Coffee is known for its health benefits. Then there's its health benefits. I should include things like anti-oxidants and antioxidants. Also, coffee boosts metabolism. It can help with weight loss or energy levels. That's a practical reason. And it improves focus. I remember hearing that coffee helps with concentration and attention. That's another reason for someone looking to boost productivity.

I should keep the language simple and conversational since it's their query. Maybe add an emoji to make it friendlier. Let me put these thoughts together in a clear, easy-to-read way.

</think>

Coffee is one of the most versatile and delicious beverages in the world. Here are five reasons why coffee is so popular and effective:

**And right now you're running an AI model that is not touching the internet,**

1. **It's everything!** Coffee combines the perfect blend of flavor and aroma, providing a rich, aromatic experience that can be enjoyed throughout the day.

2. **It boosts metabolism** – studies show that drinking coffee in

▶ ▶ ⏪ 7:06 / 11:58

|| ⏴ ⏵ ⏷ ⏸ ⏹ ⏺

the ONLY way to run Deepseek...



NetworkChuck ✅  
4,44 mi de inscritos

Seja membro

Inscrito

33 mil



Compartilhar

Download

Valeu

...



4.4  
milhões

<https://www.youtube.com/watch?v=7TR-FLWNVHY>

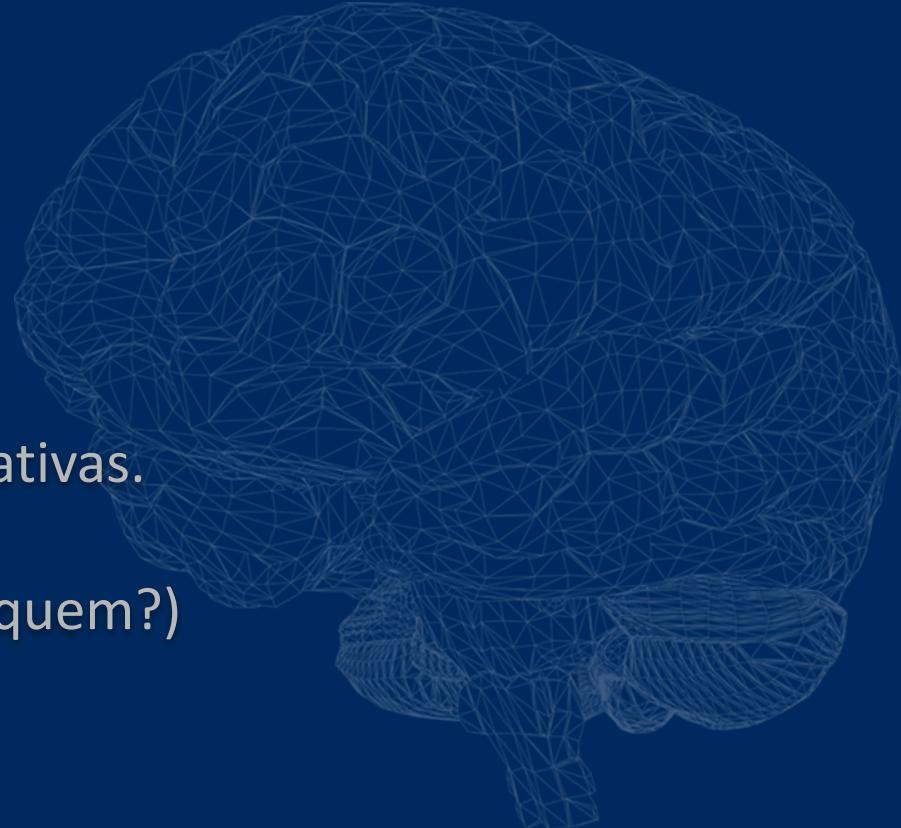
~~Parte 1 - Cases~~

~~Parte 2 - Que tipo de I.A eu preciso?~~

Parte 3 - Avanços Recentes e Expectativas.

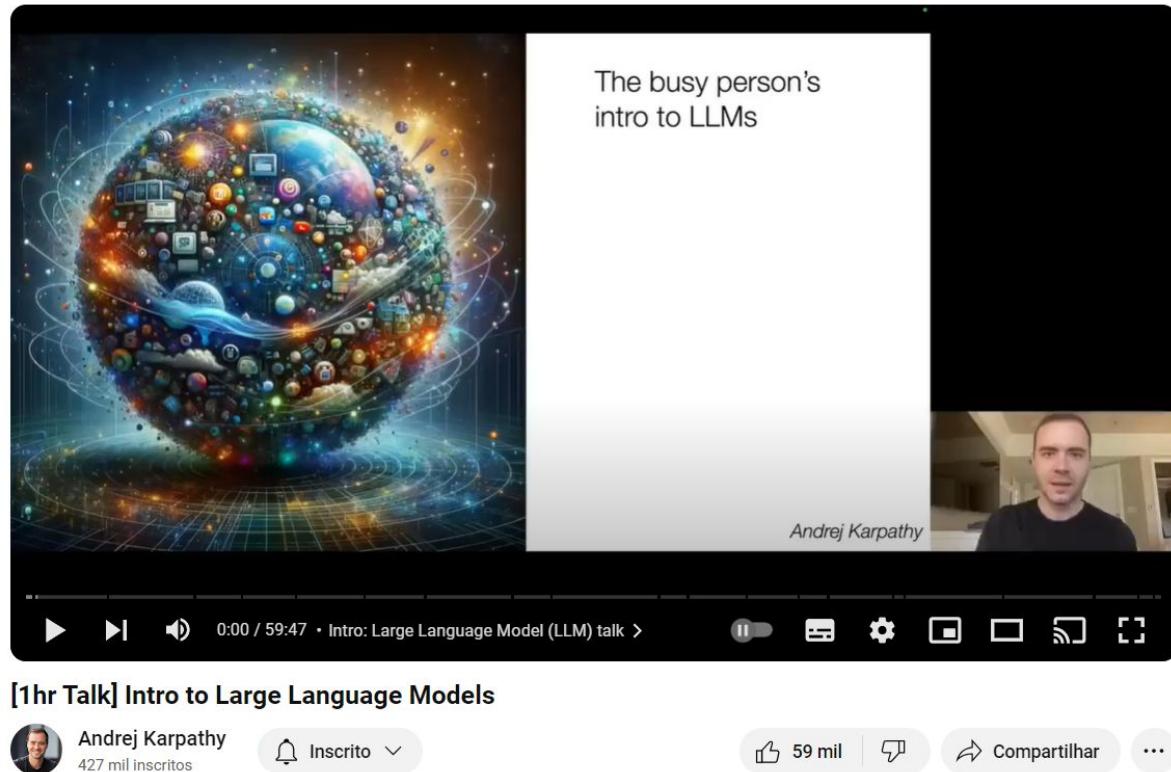
BÔNUS 1 - Evolução da I.A (Quem é quem?)

BÔNUS 2 - Onde e o que aprender.



# LLMs

[https://www.youtube.com/  
watch?v=zjkBMFhNj\\_g&t](https://www.youtube.com/watch?v=zjkBMFhNj_g&t)



The busy person's intro to LLMs

Andrej Karpathy

[1hr Talk] Intro to Large Language Models

Andrej Karpathy  
427 mil inscritos

Inscrito

0:00 / 59:47 • Intro: Large Language Model (LLM) talk >

Like 59 mil Dislike Compartilhar ...

A large, colorful illustration of Earth surrounded by a dense cluster of various icons and symbols, representing the vast amount of data and information processed by Large Language Models.

The video player interface shows the title "[1hr Talk] Intro to Large Language Models", the speaker's name "Andrej Karpathy", and the number of subscribers "427 mil inscritos". The video progress bar indicates it is at 0:00 of 59:47. The video content itself shows Andrej Karpathy speaking from a room with a whiteboard in the background.

# MULTIMODAL LLMs

2-21 Na Fig. 2-20, achar o peso do bloco suspenso sendo a tensão da corda diagonal igual a 20 N.

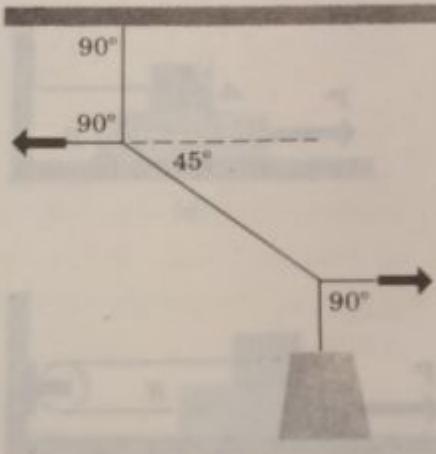


Fig. 2-20

Resposta:

O peso do bloco suspenso é aproximadamente  $w = 14.14 \text{ N}$ .



O problema 2-21 pede para calcular o peso do bloco suspenso na Fig. 2-20, onde a tensão na corda diagonal é dada como 20 N, e há um ângulo de 45° entre a corda diagonal e a horizontal.

Etapas de Resolução:

1. Identificar as forças no sistema:

- A tensão na corda diagonal é  $T = 20 \text{ N}$ .
- O peso do bloco  $w$  atua verticalmente para baixo.
- O sistema está em equilíbrio, então as somas das componentes das forças nas direções horizontal e vertical devem ser zero.

2. Equilíbrio horizontal:

As forças horizontais se equilibram. A única força com uma componente horizontal é a tensão da corda diagonal  $T$ . Podemos decompor essa força em sua componente horizontal:

$$T_{\text{horizontal}} = T \cdot \cos(45^\circ)$$

Substituindo  $T = 20 \text{ N}$ :

$$T_{\text{horizontal}} = 20 \cdot \cos(45^\circ) = 20 \cdot \frac{\sqrt{2}}{2} = 14.14 \text{ N}$$

Essa componente horizontal deve ser equilibrada pela força horizontal de compressão (ou reação) na barra horizontal à esquerda, o que significa que não precisamos resolver mais nada nessa direção, pois a barra suporta essa força.

3. Equilíbrio vertical:

A componente vertical da tensão diagonal deve equilibrar o peso do bloco. A componente vertical da força  $T$  é:

$$T_{\text{vertical}} = T \cdot \sin(45^\circ)$$

Substituindo  $T = 20 \text{ N}$ :

$$T_{\text{vertical}} = 20 \cdot \sin(45^\circ) = 20 \cdot \frac{\sqrt{2}}{2} = 14.14 \text{ N}$$

Essa componente vertical é responsável por equilibrar o peso  $w$ . Logo:

$$T_{\text{vertical}} = w$$

$$w = 14.14 \text{ N}$$

Resposta:

O peso do bloco suspenso é aproximadamente  $\downarrow = 14.14 \text{ N}$ .

ChatGPT 4o ▾

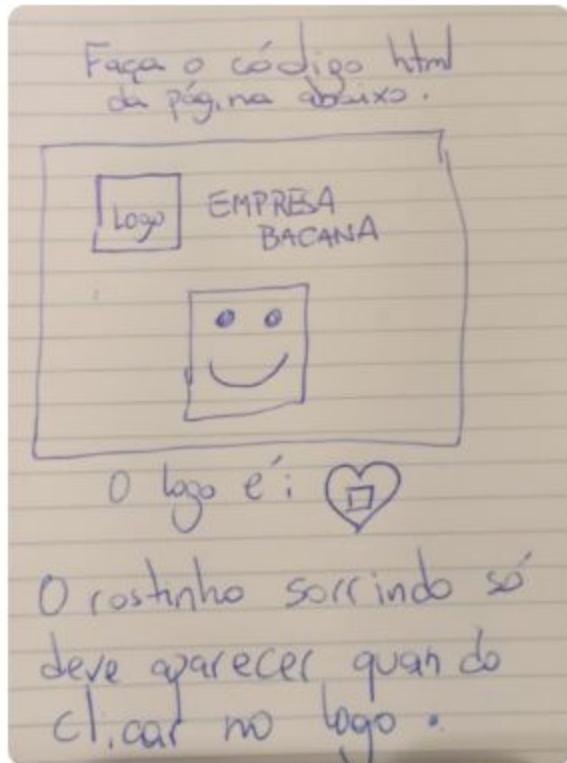


Aqui está um exemplo de código HTML com JavaScript que implementa o que foi solicitado na imagem:

html

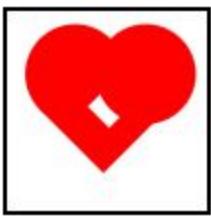
Copy code

```
<!DOCTYPE html>
<html lang="pt-BR">
<head>
 <meta charset="UTF-8">
 <meta name="viewport" content="width=device-width,
 <title>Empresa Bacana</title>
 <style>
 body {
 font-family: Arial, sans-serif;
 text-align: center;
 margin-top: 50px;
 }
 #logo {
```



Arquivo

file:///C:/Users/User/...



EMPRESA BACANA

Arquivo

file:///C:/Users/User/...



EMPRESA BACANA



# GERAÇÃO DE CÓDIGO, TESTE E DEBUG



CURSOR

Pricing

Features

Forum

Docs

Careers

Blog

Settings

Logout

Download

## The AI Code Editor

Built to make you extraordinarily productive,  
Cursor is the best way to code with AI.



Download for Free



Watch Demo  
1 Minute

The screenshot shows a Microsoft Visual Studio Code (VS Code) interface. On the left, there's a sidebar with project files like 'AFRAC\_SEFAPR', 'flagged', 'MODELS', 'templates', '.gitignore', and various Jupyter notebook files (e.g., '001\_EDA.ipynb', '002\_EMBEDDINGS.ipynb'). The main area has tabs for 'Code' (circled in red), 'Markdown', 'Interrupt', 'Restart', and 'tf230 (Python 3.7.10)'. A yellow circle highlights the 'CHAT' tab at the top right. In the center, a code editor displays Python code for processing multiple DataFrames:

```
all_descs = []
for i, dfg in enumerate(dfs):
 all_descs += [[]]
 try:
 while True:
 df = next(dfg)
 if len(df) == 0:
 break
 else:
 all_descs[-1] = list(set(list(df["DS_ITEM"])))
 print(i+1, len(all_descs[-1]), end="\r")
 except StopIteration:
 None
```

Below the code editor, there are buttons for 'Add to Chat' (Ctrl+Shift+L) and 'Edit' (Ctrl+K). The status bar shows '[17]' and '8m 57.9s'. The bottom navigation bar includes 'PROBLEMS' (13), 'OUTPUT', 'TERMINAL', and a powershell tab.

The right side of the interface features a 'CHAT' panel with a message from Claude 3.5 asking for a Portuguese explanation of the code. Below the message is a 'Run as cell' button. A detailed analysis of the code follows:

Este código está processando uma série de DataFrames (provavelmente arquivos CSV) para extrair descrições únicas de itens. Vamos analisar em detalhes:

1. `all_descs = []`: Inicializa uma lista vazia para armazenar as descrições.
2. `for i, dfg in enumerate(dfs):`: Itera sobre `dfs`, que parece ser uma lista de geradores de DataFrames (provavelmente criados com `pd.read_csv(..., chunksize=...)`).
3. `all_descs += [[]]`: Adiciona uma nova lista vazia para cada arquivo/gerador.
4. O bloco `try-except` é usado para iterar sobre os chunks de cada

Preciso de um código que

Bottom status bar: 'Spaces: 4 CRLF Cell 8 of 21 Cursor Tab' and icons for 'chat' and 'codebase'.

**ANTES**

```
025 - Get only descriptions.ipynb U ● ⚙️ ⚙️ ⚙️ ...
Code + Markdown | Run All Restart ... tf230 (Python 3.7.10)
D ✓ df = next(dfg)
if len(df) == 0:
 break
else:
 all_descs[-1] = list(set(list(df["DS_ITEM"]))
 print(f"\r{i+1} {len(all_descs[-1])}", end="\r")
except StopIteration:
 None
12m 56.6s
```

**DEPOIS**

```
all_descs = []
for i, dfg in enumerate(dfs):
 unique_descs = set()
 for df in dfg:
 unique_descs.update(df["DS_ITEM"])
 print(f"\r{i+1} {len(unique_descs)}", end="")
 all_descs.append(list(unique_descs))
2m 25.0s
```

1608798

5x mais rápido

CHAT

Este código está efetivamente coletando todas as descrições únicas de itens de vários arquivos CSV, armazenando-as separadamente para cada arquivo na lista `all_descs`. Isso é útil para análise de texto ou para criar um vocabulário único de descrições de produtos.

Você consegue deixar esse código mais curto?

Certamente! Podemos otimizar e encurtar o código da seguinte maneira:

```
all_descs = []
for i, dfg in enumerate(dfs):
 unique_descs = set()
 for df in dfg:
 unique_descs.update(df["DS_ITEM"])
 print(f"\r{i+1} {len(unique_descs)}", end="")
 all_descs.append(list(unique_descs))
```

Run as cell

Explicação das mudanças:

1. Removemos o loop `while True` e o bloco `try-except`, substituindo-os por um simples `for df in dfg`.
2. Usamos `set()` desde o inicio para armazenar descrições únicas, eliminando a necessidade de converter para lista e de volta para set em cada iteração.
3. Utilizamos o método `update()` do set para adicionar novas descrições de forma eficiente.
4. Simplificamos a impressão do progresso usando f-string.

# Dados Sintéticos

Tenho um produto com a descrição "Sabão em pó Omo Multi Ação 400g" gere para mim descrições equivalentes a essa para que eu tenha um banco de sinônimos



Aqui estão algumas descrições equivalentes para "Sabão em pó Omo Multi Ação 400g" que você pode usar como sinônimos:

1. Detergente em pó Omo Multi Ação 400g
2. Sabão em pó Omo Multiuso 400g
3. Sabão em pó Omo Multi Ação 400g embalagem
4. Omo Sabão em pó Multi Ação 400g
5. Detergente em pó Omo para Lavagem Multi Ação 400g
6. Omo Sabão em pó Multiuso 400g
7. Sabão para roupas Omo Multi Ação 400g
8. Sabão Omo em pó Multi Ação 400g
9. Detergente para roupas Omo Multi Ação em pó 400g
10. Omo Multi Ação sabão em pó 400g

+/- 10s pra gerar

Essas variações mantêm o sentido original, enquanto oferecem opções de sinônimos que podem ser usadas em seu banco de dados.

# Dados Sintéticos

jupyter 030 - EDA - NANOGPT\_EMBEDDINGS Last Checkpoint: 2 months ago

File Edit View Run Kernel Settings Help

Code JupyterLab

```
[45]: #context = torch.zeros((1, 1), dtype=torch.long, device=device)
import string

for i in range(30):
 context = torch.tensor([encode(random.choice(string.ascii_uppercase))], dtype=torch.long, device=device)
 print(decode(m.generate(context, max_new_tokens=200)[0].tolist()).split("\n")[0])
```

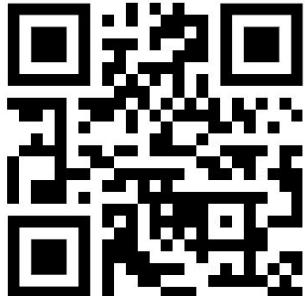
Creme Alis Aorgan300  
XIC CAFE FLORENCA 4PECAS  
MESA REF107/360  
WAFFER BARION FLOCOS/AVELA PCT 35G  
COMFORTIS CAES ADULT -  
Leite Monange] Semi-  
LIDOCAINA 2% COLIRIO DUOM  
FAR.MAND.YOK'Semidesn.500g  
X REPARO SC TRA NXR 125/150 - PARTES E PECAS 76 METAL  
KT3 CALC 2831798 DUBAI  
KIT APAR BARB ACQUA  
SHOYU CEPERA1LT



# LMSYS Chatbot Arena (Multimodal)

Eu não tenho  
o Chatgpt  
pago?

<https://chat.lmsys.org/>



Total #models: 201. Total #votes: 2,634,992. Last updated: 2025-02-09.

Code to recreate leaderboard tables and plots in this [notebook](#). You can contribute your vote at [larena.ai](https://larena.ai)!

Category: Overall | Apply filter:  Style Control |  Show Deprecated

Overall Questions: #models: 201 (100%) #votes: 2,634,992 (100%)

Rank* (UB)	Rank (StyleCtrl)	Model	Arena Score	95% CI	Votes	Organization	License
1	1	Gemini-2.0-Flash-Thinking-Exp-01-21	1384	+5/-6	11462	Google	Proprietary
1	1	Gemini-2.0-Pro-Exp-02-05	1379	+5/-8	9385	Google	Proprietary
3	1	ChatGPT-4a-latest-(2024-11-20)	1365	+3/-3	39649	OpenAI	Proprietary
3	1	DeepSeek-R1	1361	+9/-9	4193	DeepSeek	MIT
4	7	Gemini-2.0-Flash-001	1355	+5/-8	7264	Google	Proprietary
4	1	o1-2024-12-17	1351	+5/-6	13416	OpenAI	Proprietary
7	6	o1-preview	1335	+4/-4	33164	OpenAI	Proprietary
7	7	Qwen2.5-Max	1332	+7/-9	5459	Alibaba	Proprietary



# LMSYS Chatbot Arena (Multimodal)

👉 Chat now!

🔍 Expand to see the descriptions of 73 models

Model A

Pra quem o cachorro está olhando?

O cachorro está olhando para a pessoa sentada no sofá à direita.

Model B

Pra quem o cachorro está olhando?

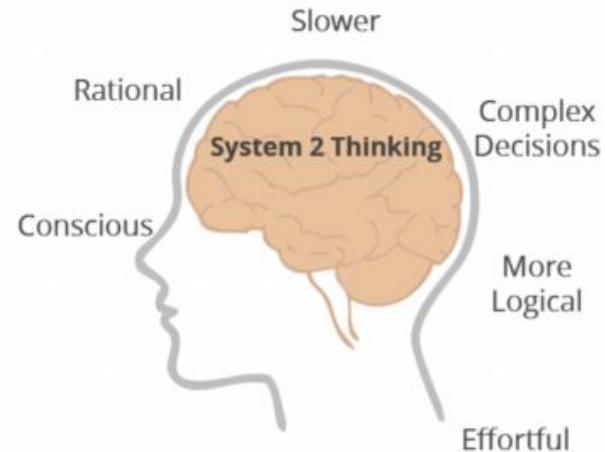
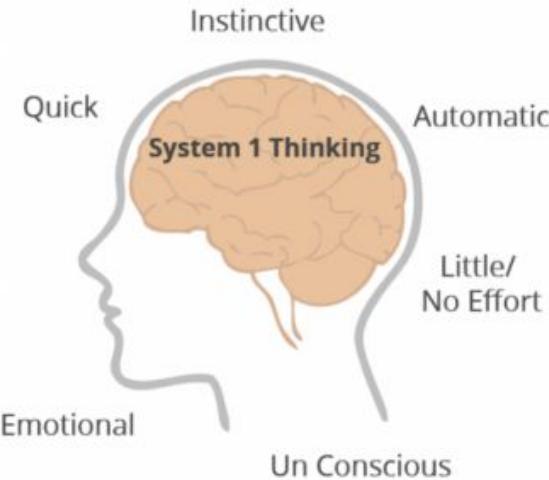
Na imagem, o cachorro está olhando para a mulher que está sentada no sofá.

👉 A is better

👉 B is better

之心 Tie

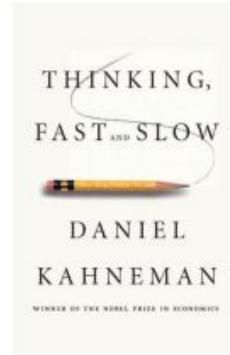
👎 Both are bad



$$2 + 2 =$$

A maior parte dos modelos de I.A.  
(incluindo LLMs) são análogas ao  
sistema 1.

$$17 \times 24 =$$



Fale sobre o evento da BX Brasil

Fale sobre o evento da BX Brasil 2025

Found 37 results ▶

Thought for 41 seconds ▲

Okay, o usuário quer saber sobre o evento BX Brasil 2025. Vou começar verificando os resultados de busca que tenho. Vejo que as páginas 1, 2 e 4 são diretamente relacionadas ao evento BX Brasil. As outras páginas tratam de outros temas, como mercado financeiro, eventos de marketing e o portal BRICS. Então, preciso focar nas informações das páginas 1, 2 e 4.

A página 1 detalha o evento, mencionando que é a 3ª edição, datas (13 e 14 de fevereiro de 2025), local provavelmente não especificado, mas fala em indicação de hotéis próximos. Também tem informações sobre o conteúdo do evento, como palestras sobre IA e estratégias empresariais, networking, e inclusões como material de apoio e coffee break. Além disso, menciona o cupom de desconto "EVNTOS".

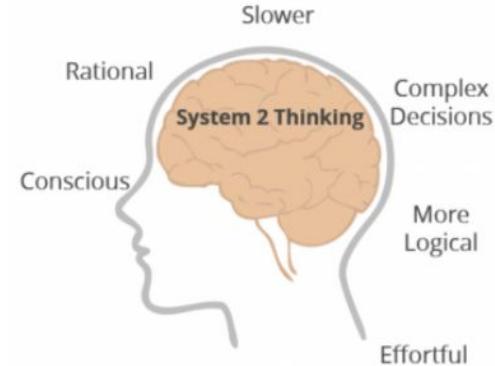
A página 2 parece ser a homepage do evento, com programação detalhada dos dois dias. Dia 1 tem palestras sobre vendas, IA, investimentos, e um tour na TOTVS. Dia 2 aborda lucro máximo, gestão de times com IA, tendências tecnológicas e painéis. Também confirma os palestrantes e detalha o que está incluso no ingresso, como acesso a dois dias, material, coffee break e água. Além disso, menciona que as apresentações serão disponibilizadas e que o evento ocorre em locais diferentes cada dia.

A página 4 é a página de ingressos, reforçando a data, informações de contato, e detalhes sobre inscrição, como parcelamento em até 3x sem juros. Também menciona que participantes de todos os níveis podem participar e que a organização oferece certificado.

Message DeepSeek

DeepThink (R1)

Search



$$17 \times 24 =$$

Finalmente, garantir que a resposta esteja em português, bem estruturada, com citações adequadas e sem informações desnecessárias de outras páginas não relacionadas.

O **BX Brasil 2025** é um evento de imersão executiva focado em **inteligência artificial (IA)** e suas aplicações no ambiente empresarial, promovendo conexões estratégicas e conteúdo de alto nível para profissionais e líderes do mercado. Seguem os principais detalhes:

## 1. Data e Local

- **Data:** 13 e 14 de fevereiro de 2025 (quinta e sexta-feira) 1 3.
- **Horário:** Das 8h às 21h, com programação extensa ao longo do dia 3.
- **Local:** O evento ocorrerá em diferentes locais a cada dia, conforme a programação divulgada. A organização oferece indicações de hotéis próximos com descontos exclusivos usando o cupom "EVNTOS" 1 3.

# Agentic A.I.



Andrew Ng On AI Agentic Workflows And Their Potential For Driving AI Progress

DATA CLOUD SUMMIT 24

Coding Benchmark (HumanEval)

GPT-3.5  
Zero-shot (48%)

GPT-4  
Zero-shot (67%)

Intervenor-AMPL

Language Agent

Tree Search

LDB+Reflexion

Zero-shot  
Reflection  
Tool Use  
Planning  
Multiagent

Thanks to Joaquin Dominguez and John Santerre for help with analysis.

16:04 / 30:53 • Presentation by Andrew Ng >

## ChatDev: Communicative Agents for Software Development

Chen Qian\* Wei Liu\* Hongzhang Liu\* Nuo Chen\* Yufan Dang\*  
Jiahao Li\* Cheng Yang\* Weize Chen\* Yusheng Su\* Xin Cong\*  
Juyuan Xu\* Dahai Li\* Zhiyuan Liu\*\* Maosong Sun\*\*  
\*Tsinghua University \*The University of Sydney \*\*BUPT \*Modelbest Inc.  
qianc62@gmail.com liuzy@tsinghua.edu.cn sms@tsinghua.edu.cn

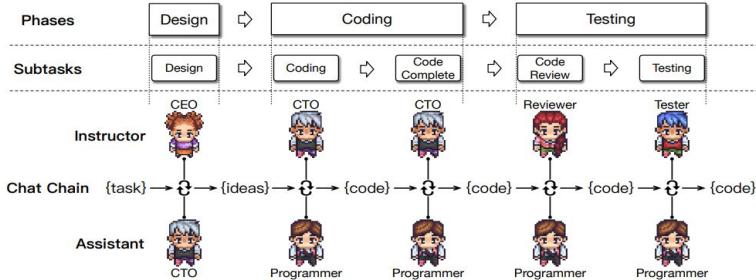
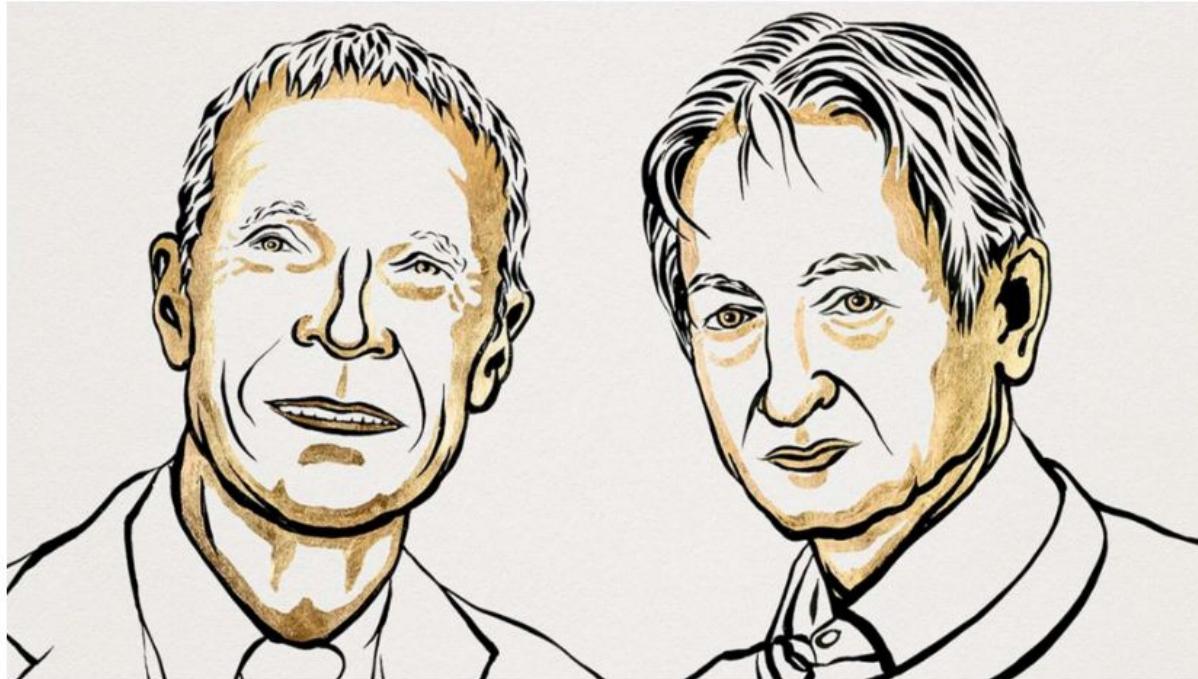


Figure 1: ChatDev, a chat-powered software development framework, integrates LLM agents with various social roles, working autonomously to develop comprehensive solutions via multi-agent collaboration.

# Pioneiros da inteligência artificial ganham Nobel de Física 2024

A dupla John Hopfield e Geoffrey Hinton ganhou o Nobel de Física de 2024 pelos seus trabalhos voltados ao aprendizado de máquina; entenda

Pedro Spadoni | 08/10/2024 08h44



John Hopfield (esq.) e Geoffrey E. Hinton (dir.) (Imagem: Niklas Elmehed/Nobel Assembly)

# Nobel de Química 2024 vai para trio que usou IA para decifrar proteínas

David Baker, Demis Hassabis e John Jumper ganharam Nobel de Química 2024 por decifrarem segredos das proteínas por meio da computação e IA

Pedro Spadoni | 09/10/2024 08h46



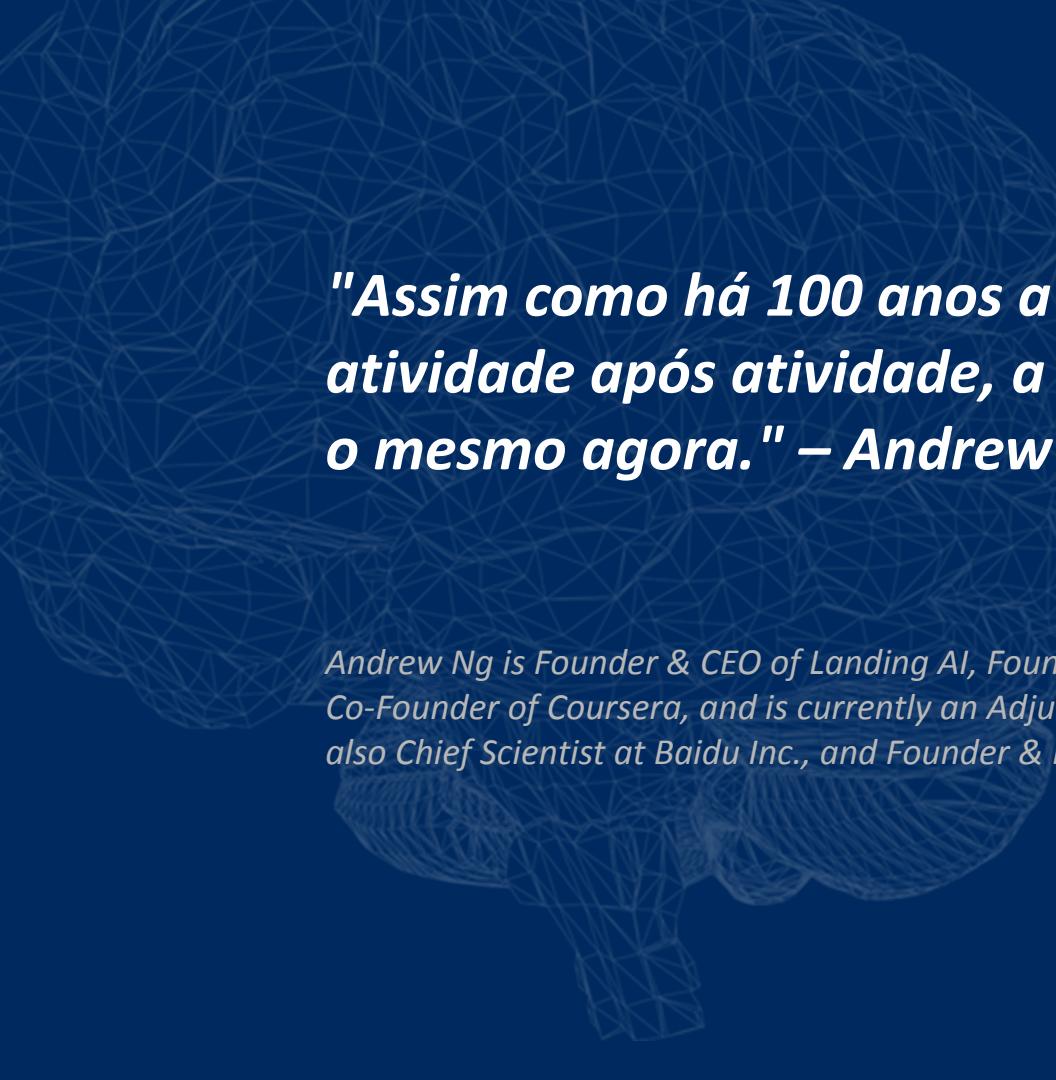
David Baker (esq.), Demis Hassabis (meio) e John Jumper (dir.) (Imagem: Niklas Elmehed/Nobel Assembly)

## **How Long Until Computers Have the Same Power As the Human Brain?**

Lake Michigan's volume (in fluid ounces) is about the same as our brain's capacity (in calculations per second). Computing power doubles every 18 months. At that rate, you see very little progress for a long time—and suddenly you're finished.



<https://www.motherjones.com/media/2013/05/robots-artificial-intelligence-jobs-automation/>



***"Assim como há 100 anos a eletricidade transformou  
atividade após atividade, a inteligência artificial fará  
o mesmo agora." – Andrew Ng***

*Andrew Ng is Founder & CEO of Landing AI, Founder of deeplearning.ai, Co-Chairman and Co-Founder of Coursera, and is currently an Adjunct Professor at Stanford University. He was also Chief Scientist at Baidu Inc., and Founder & Lead for the Google Brain Project*



**Marcelo Filipak**

filipak@filiart.com.br

<https://www.linkedin.com/in/marcelo-filipak-97386710/>

<https://github.com/mfilipak>

# I.A. NA PRÁTICA

técnicas, cases reais e mão na massa

**OBRIGADO!**



**Marcelo Filipak**

filipak@filiart.com.br

<https://www.linkedin.com/in/marcelo-filipak-97386710/>

<https://github.com/mfilipak>



# I.A. NA PRÁTICA

Técnicas, cases reais e mão na massa.

BAIXE O MATERIAL

OBRIGADO!

<https://github.com/mfilipak/BXBRASIL2025>

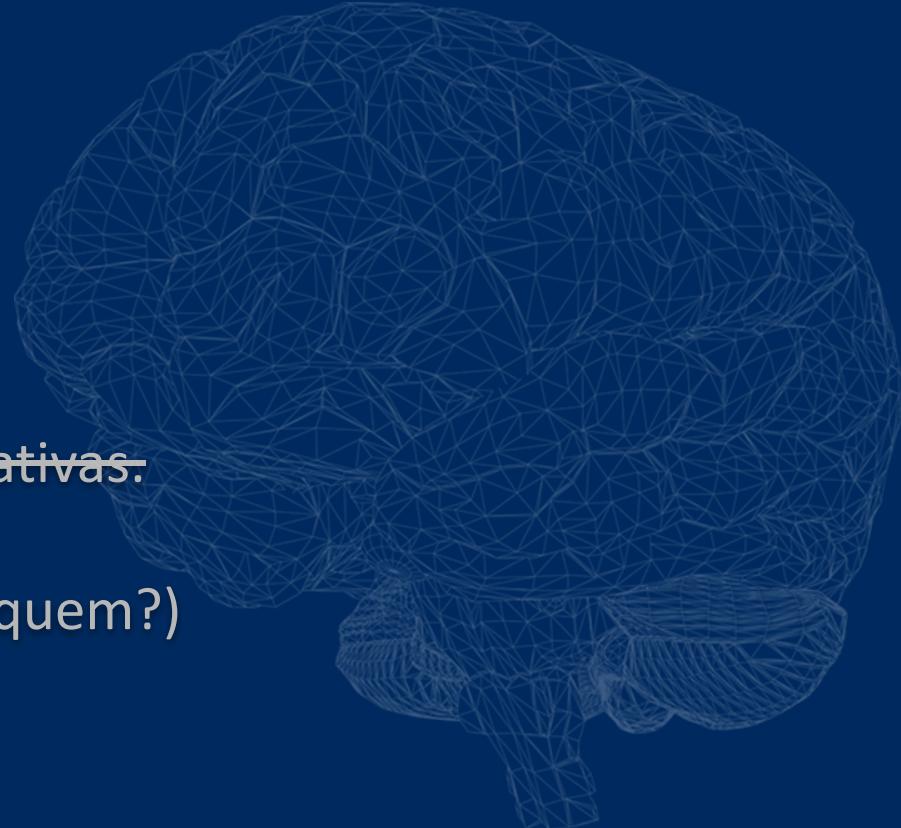
~~Parte 1 - Cases~~

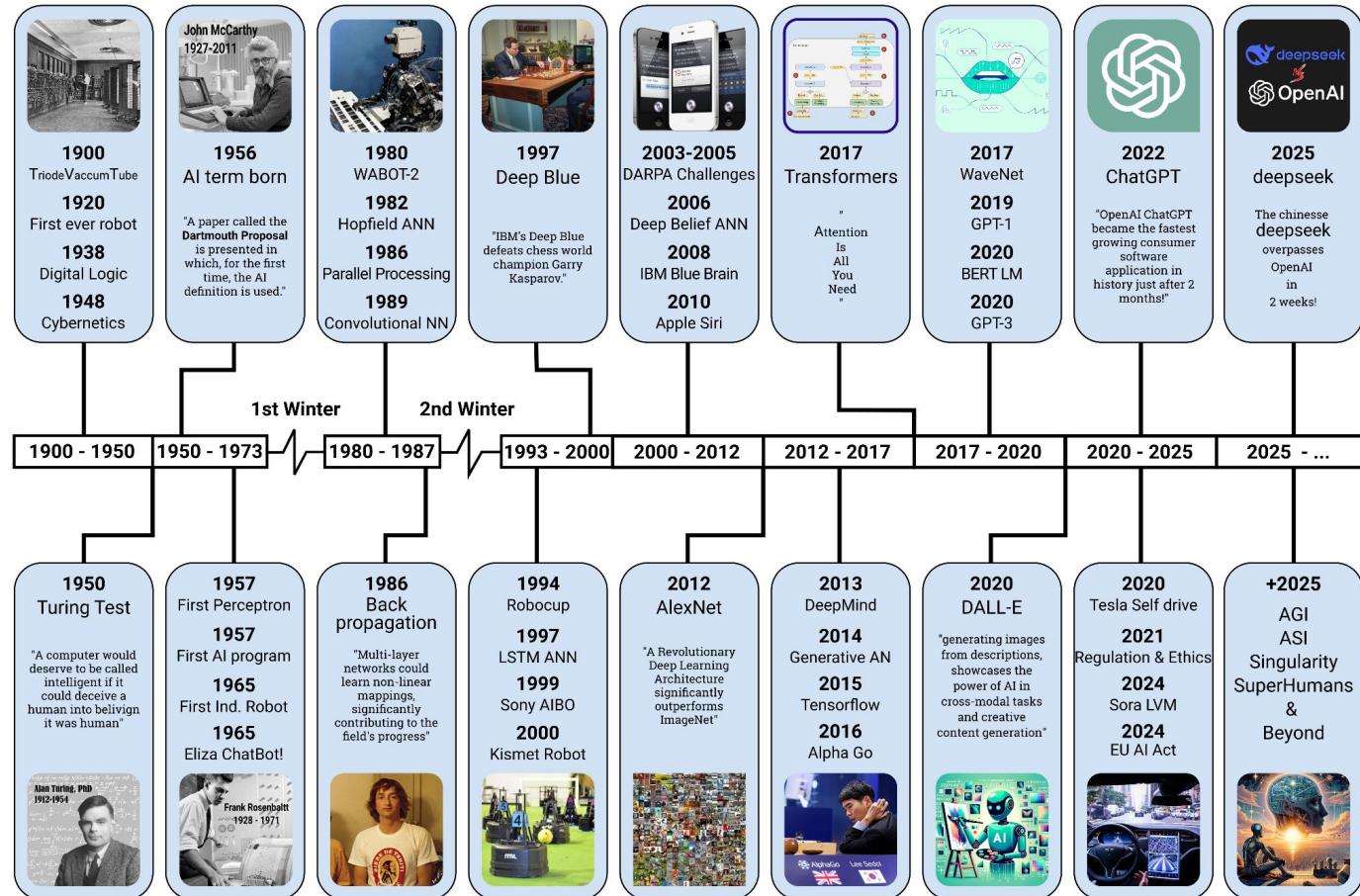
~~Parte 2 - Que tipo de I.A eu preciso?~~

~~Parte 3 - Avanços Recentes e Expectativas.~~

BÔNUS 1 - Evolução da I.A (Quem é quem?)

BÔNUS 2 - Onde e o que aprender.





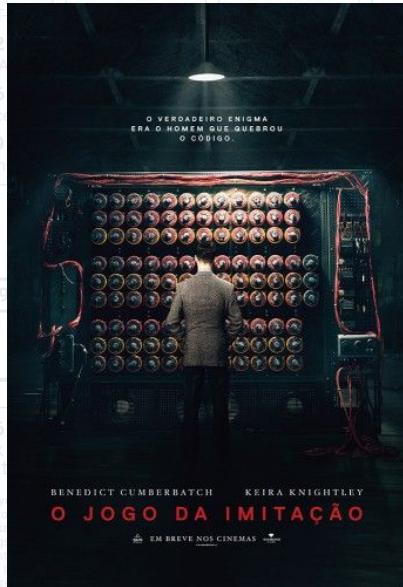
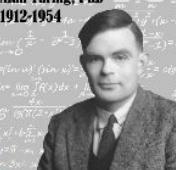
[https://en.wikipedia.org/wiki/Timeline\\_of\\_artificial\\_intelligence](https://en.wikipedia.org/wiki/Timeline_of_artificial_intelligence)

1900 - 1950

## 1950 Turing Test

"A computer would deserve to be called intelligent if it could deceive a human into believing it was human"

Alan Turing, PhD  
1912-1954

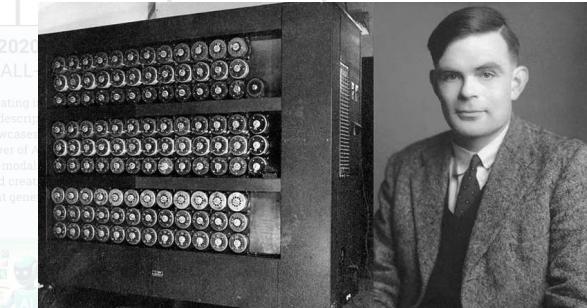
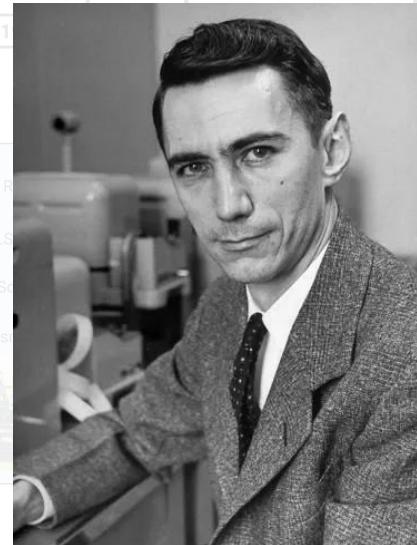
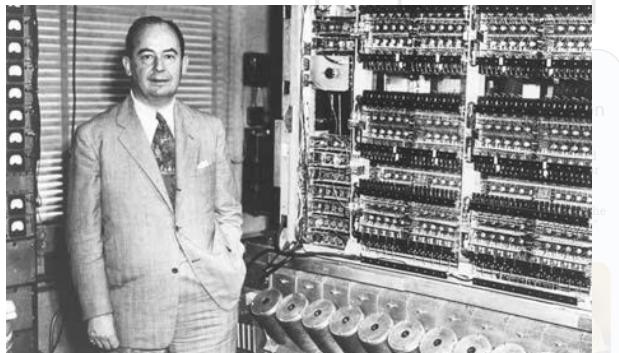


No teste, um avaliador humano interage, por meio de comunicação escrita, com um humano e uma máquina sem saber qual é qual. **Se o avaliador não conseguir distinguir consistentemente a máquina do humano, considera-se que a máquina passou no teste, demonstrando um nível de inteligência comparável ao humano.**

Esse teste já foi superado por volta de 2022.  
Na verdade tem que “emburrerar” um LLM pra ele não parecer tão inteligente. Dizendo que ele está sendo testado.

A relevância de Turing pra evolução da I.A não tem relação praticamente nenhuma com o Teste de Turing.

Turing foi um gênio, assim como Shannon, von Neumann, ... pais da computação e da teoria da informação, das quais a I.A “ramifica”.



E quais foram os personagens e marcos mais relevantes?



# 1847 - Essa “técnica” é usada em todos os treinos de I.A. Vem do Cálculo.

## Gradient descent

Article Talk

From Wikipedia, the free encyclopedia

For the analytical method called "steepest descent", see [Method of steepest descent](#).

**Gradient descent** is a method for unconstrained mathematical optimization. It is a first-order iterative algorithm for minimizing a differentiable function.

The idea is to take repeated steps in the opposite direction (of the gradient) of the function at the current position. Steepest descent. Conversely, stepping in the direction of the gradient leads to a trajectory that maximizes that function; the procedure is called gradient ascent. It is particularly useful in machine learning for minimizing loss functions. Gradient descent should not be confused with local search algorithms, which are iterative methods for optimization.

Gradient descent is generally attributed to [Augustin-Louis Cauchy](#), who first suggested it in 1847.<sup>[2]</sup> Jacques Hadamard independently proposed a similar method in 1907.<sup>[3][4]</sup> Its convergence properties for non-linear optimization problems were first studied by Haskell Curry in 1944,<sup>[5]</sup> with the method becoming increasingly well-studied and used in the following decades.<sup>[6][7]</sup>

A simple extension of gradient descent, stochastic gradient descent, serves as the most basic algorithm used for training most deep networks today.

Description [\[edit\]](#)



26 languages ▾

Read Edit View history Tools ▾

Article Talk

From Wikipedia, the free encyclopedia

For the analytical method called "steepest descent", see [Method of steepest descent](#).

**Gradient descent** is a method for unconstrained mathematical optimization. It is a first-order iterative algorithm for minimizing a differentiable function.

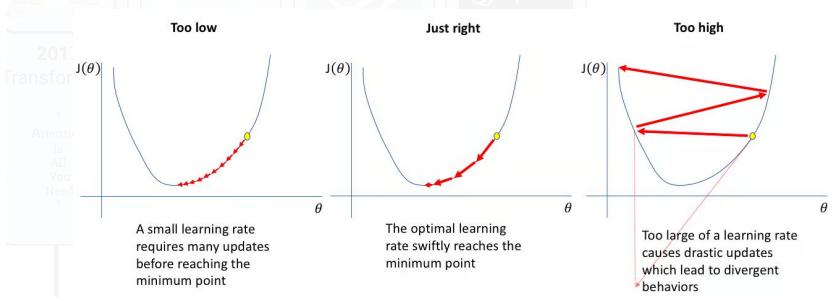
Baron **Augustin-Louis Cauchy** was a French mathematician, engineer, and physicist. He was one of the first to rigorously state and prove the key theorems of calculus, pioneered the field of complex analysis, and the study of permutation groups in abstract algebra. Cauchy also contribu-

The infographic illustrates the effect of different learning rates on the convergence of gradient descent. It shows three graphs of the cost function  $J(\theta)$  versus parameter  $\theta$ .

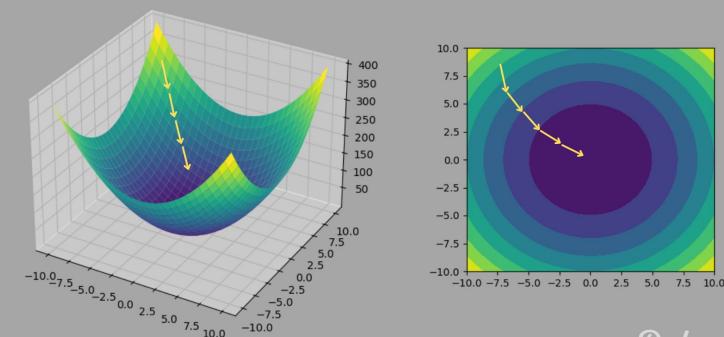
- Too low:** A small learning rate requires many updates before reaching the minimum point.
- Just right:** The optimal learning rate swiftly reaches the minimum point.
- Too high:** Too large of a learning rate causes drastic updates which lead to divergent behaviors.

Below the graphs, a timeline shows the years 2012 - 2017, 2017 - 2020, 2020 - 2025, and 2025 - ...

Related articles:  
- Structured prediction  
- Anomaly detection  
- Artificial neural network  
- Reinforcement learning  
- Learning with humans  
- Model diagnostics  
- Mathematical foundations  
- Journals and conferences  
- Related articles



## Gradient Descent



**1950 - 1973**



1900  
Triode Vacuum Tube

1920  
First ever robot

1938  
Digital Logic

1948  
Cybernetics

1900 - 1950

1950  
Turing Test

"A computer would deserve to be called intelligent if it could deceive a human into believing it was human"

Alan Turing, PhD  
1912-1954

## 1957 First Perceptron

## 1957 First AI program

## 1965 First Ind. Robot

## 1965 Eliza ChatBot!



**Frank Rosenblatt**  
1928 - 1971



1997  
Deep Blue

IBM's Deep Blue defeats chess world champion Garry Kasparov."



2003-2005  
DARPA Challenges

2006  
Deep Belief ANN

2008  
IBM Blue Brain

2010  
"Attention  
All  
You  
Need"



2017  
Transformers

2017  
WaveNet  
2019  
GPT-1  
2020  
BERT LM  
2020  
"OpenAI ChatGPT became the fastest growing consumer software application in history just after 2 months!"



2017  
WaveNet

2019  
GPT-1

2020  
BERT LM  
2020  
"OpenAI ChatGPT became the fastest growing consumer software application in history just after 2 months!"



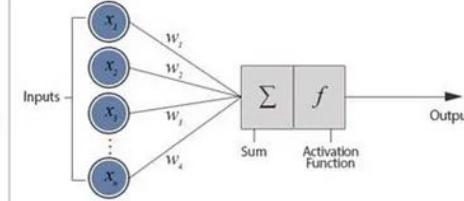
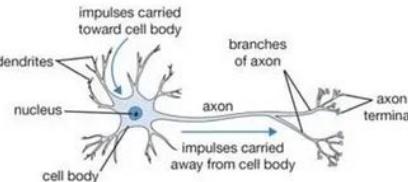
2022  
ChatGPT

The Chinese deepseek surpasses OpenAI in accuracy.



2025  
deepseek

## Biological Neuron versus Artificial Neural Network



center

199

dendrites

nucleus

cell body

impulses carried toward cell body

branches of axon

axon

impulses carried away from cell body

axon terminals

19

Robocup

1997

LSTM ANN

1999

Sony AIBO

2000

Kismet Robot

AlexNet

2014

Generative AN

2015

Tensorflow

2016

Alpha Go

deepmind

2014

Generative AN

2015

Tensorflow

2016

Alpha Go

DALL-E

2014

generating images from descriptions,

showcases the power of AI in cross-modal tasks

and creative content generation"

Resia Self drive

2021

Regulation & Ethics

2024

Sora LVM

2024

EU AI Act

AGI

ASI

Singularity

SuperHumans

&

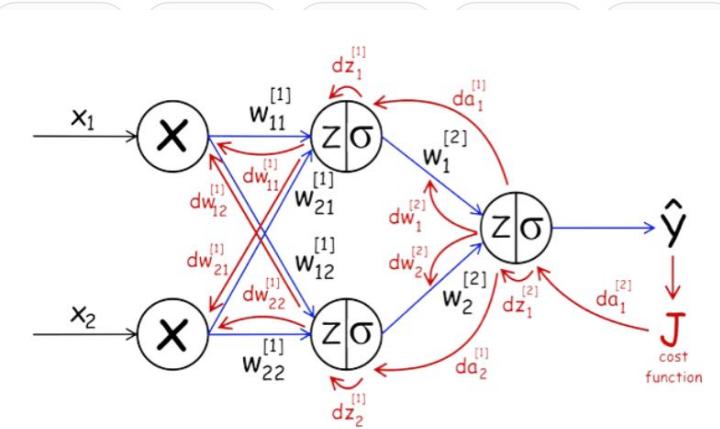
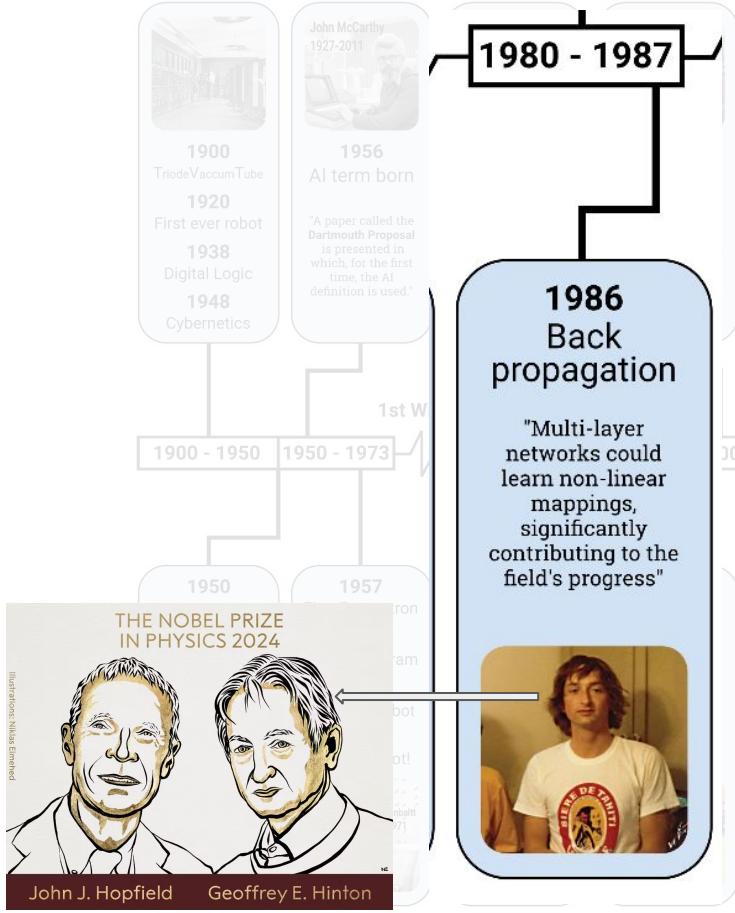
Beyond

AGI

ASI

Singularity

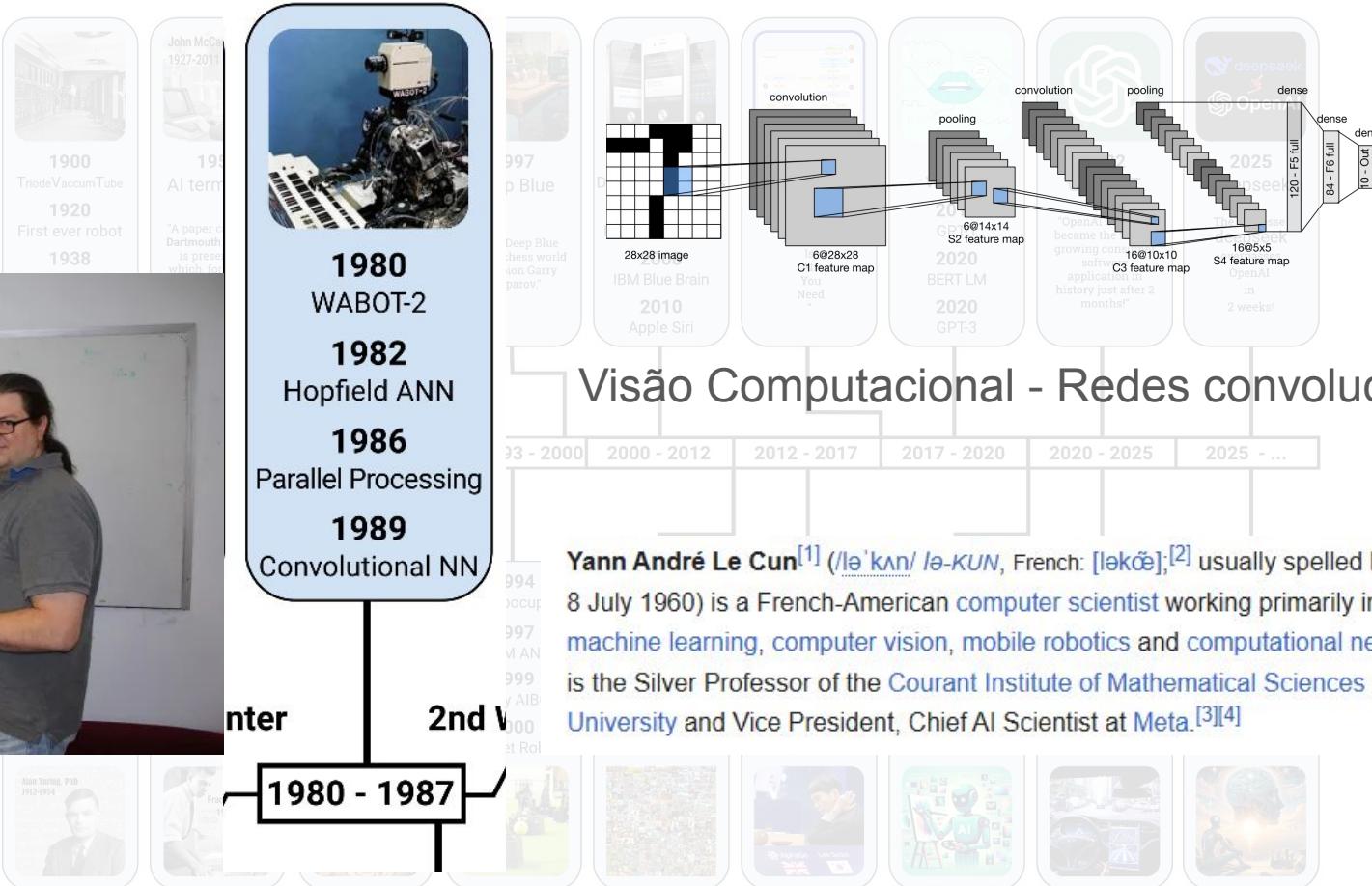
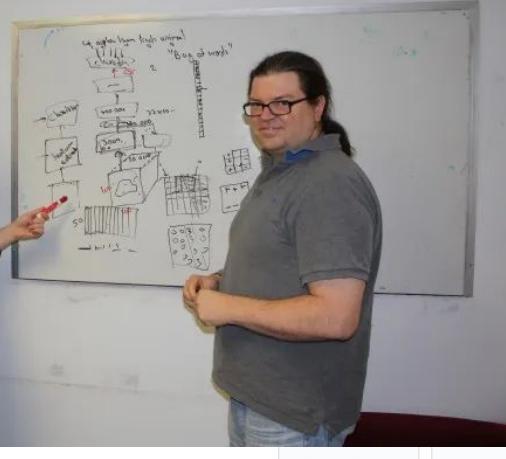
SuperHumans



## Backpropagation

**Relevância:**  
Permite treinar as redes de forma eficiente e paralelizada.

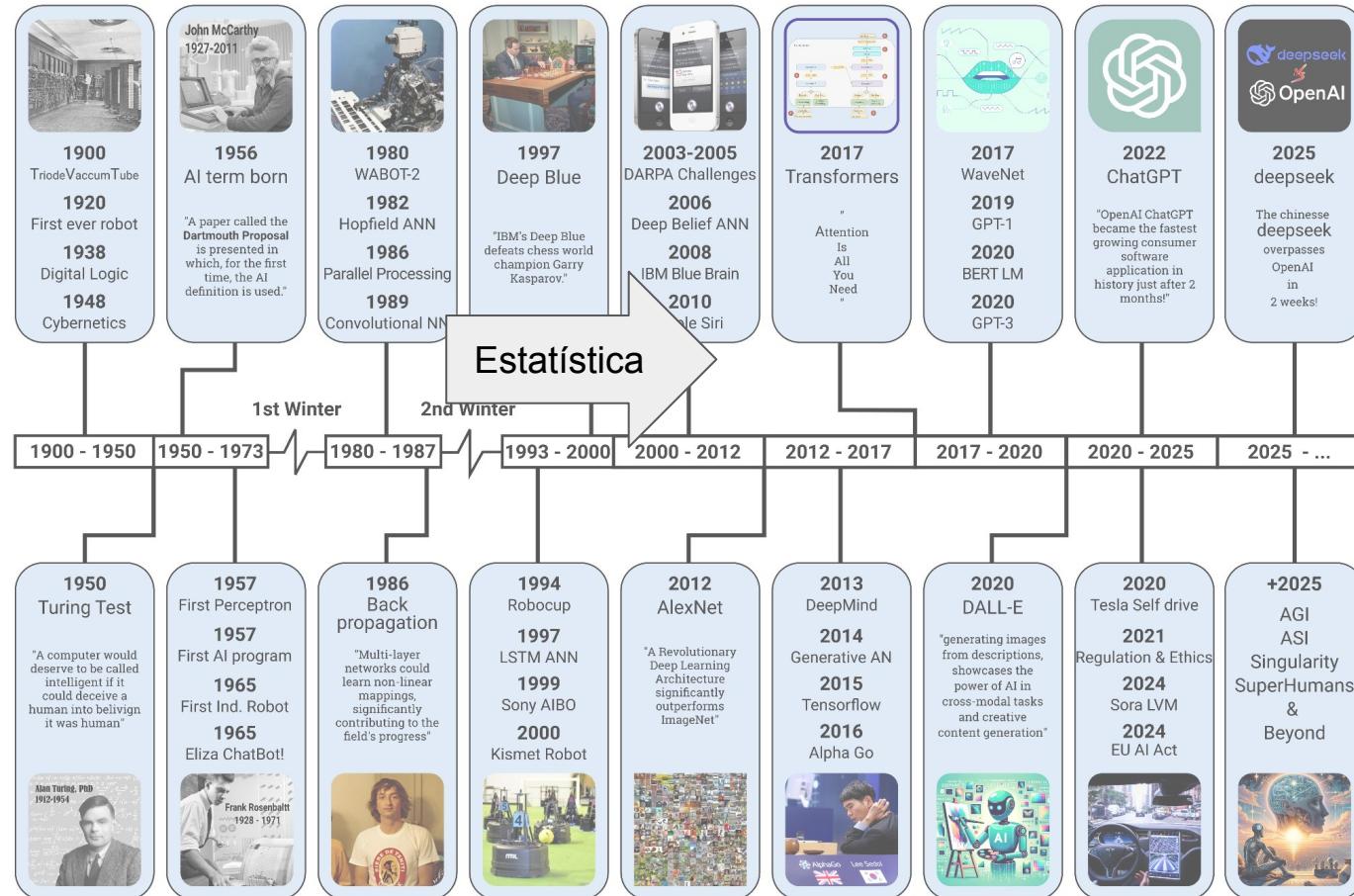
Em vez de ajustar um peso por vez faz tudo em paralelo.  
Torna o treino “bilhões” de vezes mais rápido.

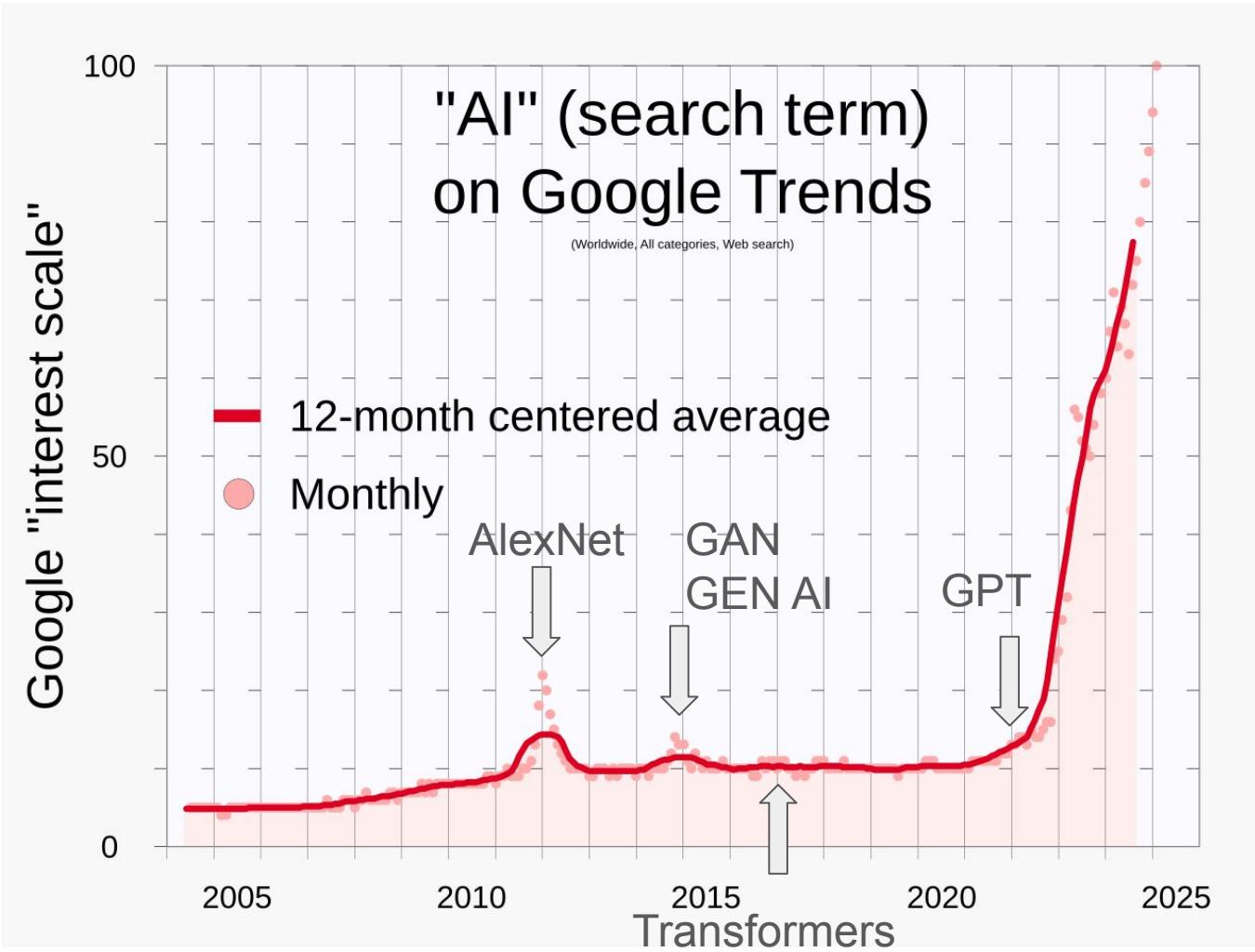


Visão Computacional - Redes convolucionais

**Yann André Le Cun**<sup>[1]</sup> (*/lə'kœn/*, French: [lɛkœ̃];<sup>[2]</sup> usually spelled **LeCun**;<sup>[2]</sup> born 8 July 1960) is a French-American computer scientist working primarily in the fields of machine learning, computer vision, mobile robotics and computational neuroscience. He is the Silver Professor of the Courant Institute of Mathematical Sciences at New York University and Vice President, Chief AI Scientist at Meta.<sup>[3][4]</sup>

# De 1989 a 2012 “ninguém” dava bola pra redes neurais. Não existia deep learning.





# O QUE VOCÊ FAZIA EM 2012?

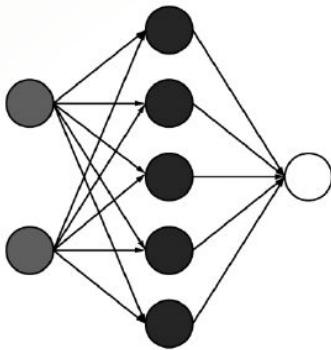
EU:





### 3. Software

- Improved Techniques
- New Models
- Toolboxes



IMAGENET



## Neural Nets



## GPUs



Fei-Fei Li 李飛飛  
Stanford University

*A Deep  
Learning  
Revolution*



**Andrej Karpathy**

@AndrejKarpathy · 658 mil inscritos · 17 vídeos

Saiba mais sobre este canal ...[mais](#)

[eurekalabs.ai](#) e mais 4 links

Inscrito ▾

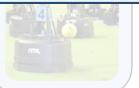


# AlexNet

Geoffrey Hinton

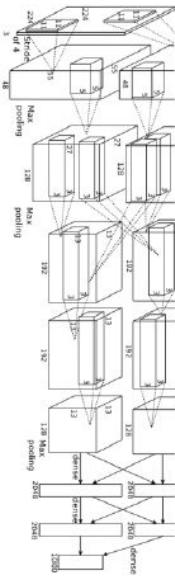


## ImageNet Large Scale Visual Recognition Challenges



# ...And Cooler and Cooler ☺

“*AlexNet*”



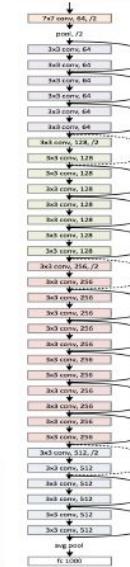
“*GoogLeNet*”



“*VGG Net*”



“*ResNet*”



[Krizhevsky et al. NIPS 2012]

[Szegedy et al. CVPR 2015]

[Simonyan & Zisserman,  
ICLR 2015]

[He et al. CVPR 2016]

"Snappily written . . . Entertaining and valuable . . . Essential."

*—Los Angeles Times*

# Genius Makers

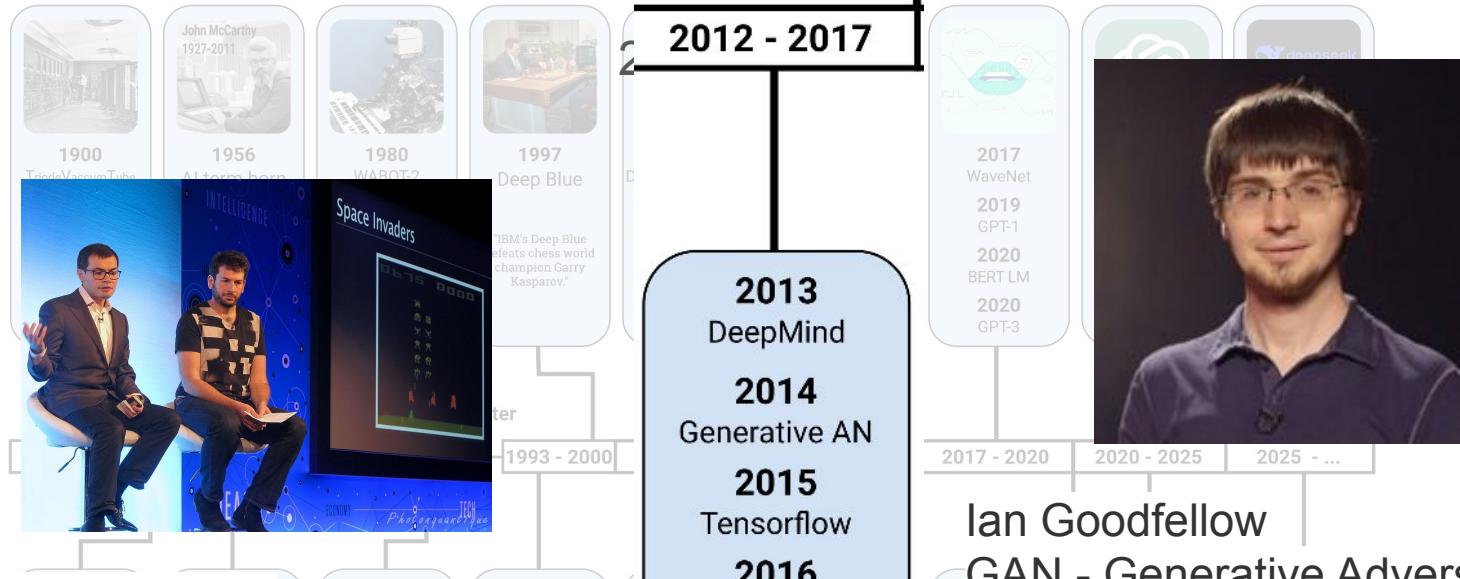


The Mavericks Who Brought AI  
to Google, Facebook, and the World

C A D E M E T Z



**Demis Hassabis**  
C.E.O da Deepmind  
Nobel de Química 2024.  
Alphafold



## Timeline of images generated by artificial intelligence

These people don't exist. All images were generated by artificial intelligence.



Ian Goodfellow  
GAN - Generative Adversarial Network

2014



Goodfellow et al. (2014) – Generative Adversarial Networks

2015



Radford, Metz, and Chintala (2015) – Unsupervised Representation Learning with Deep Convolutional GANs

2016



Liu and Tuzel (2016) – Coupled GANs

2017



Karras et al. (2017) – Progressive Growing of GANs for Improved Quality, Stability, and Variation

2018



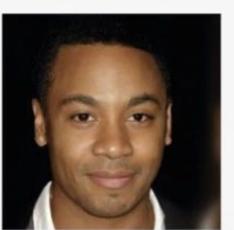
Karras, Laine, and Aila (2018) – A Style-Based Generator Architecture for Generative Adversarial Networks

2019



Karras et al. (2019) – Analyzing and Improving the Image Quality of StyleGAN

2020



Ho, Jain, &amp; Abbeel (2020) – Denoising Diffusion Probabilistic Models

2021

Image generated with the prompt:  
"a couple of people are sitting on a wood bench"

Ramesh et al. (2021) – Zero-Shot Text-to-Image Generation (OpenAI's DALL-E 1)

2022

Image generated with the prompt:  
"A Pomeranian is sitting on the King's throne wearing a crown. Two tiger soldiers are standing next to the throne."

Saharia et al. (2022) – Photorealistic Text-to-Image Diffusion Models with Deep Language Understanding (Google's Imagen)

# Ashish Vaswani

Cientista de computação :



LinkedIn

1900  
Triode Vaccun  
1920  
First ever r  
1938  
Digital Lo  
1948  
Cybernet

1900 - 1948

1997  
Deep Blue

"Deep Blue  
its chess world  
upset Garry Kasparov."

1993 - 2000

2003-2005  
DARPA Challenge

2006  
Deep Belief ANN

2008  
IBM Blue Brain

2010  
Apple Siri

2000 - 2012

2017  
Transformers

"  
Attention  
Is  
All  
You  
Need  
"

2012 - 2017

Focado em tradução.  
Rede com input de tamanho M e output N.

1950  
Turing Test  
"A computer would deserve to be called intelligent if it could deceive a human into believing it was human"



1957  
First Perceptron  
1957  
First AI program



1986  
Back  
propagation  
"Multi-layer  
networks could  
learn non-linear  
mappings,  
significantly  
contributing to the  
field's progress"



1994  
Robocup  
1997  
LSTM ANN  
1999  
Sony AIBO  
2000  
Kismet Robot



2012  
AlexNet  
"A Revolutionary  
Deep Learning  
Architecture  
significantly  
outperforms  
ImageNet"



Tesla #134958245  
Adobe Stock #134958245

2024  
Sora LVM

2024

EU AI Act



SuperHumans  
&  
Beyond



2022  
ChatGPT



2025  
deepseek





John McCarthy  
1927-2011



# Ilya Sutskever

Cientista de computação israelense-canadense

Triode

First

Digi

Cyt

1901

Tur

"A con  
deserv  
inte  
cool  
human  
it w

Alan Turing  
1912-1954



Time

2005  
Challenges  
6  
ef ANN  
8  
Brain  
0  
Siri

2017  
Transformers

"  
Attention  
is  
All  
You  
Need"

2017  
WaveNet  
2017  
GPT-3  
2021  
BERT  
2021  
GPT-4

2012 2012 - 2017 2017 - 2025

2  
Jet

2013  
DeepMind

2014  
Generative AN

2015  
Tensorflow

2016  
Alpha Go

2021  
DALL-E

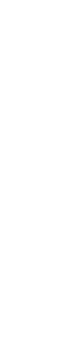
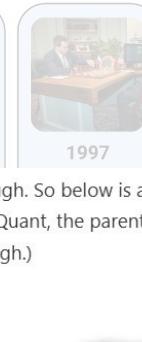
"generating  
from descri  
showcase  
power of  
cross-mod  
and crea  
content gen

2020 - 2025



**2022**  
**ChatGPT**

"OpenAI ChatGPT  
became the fastest  
growing consumer  
software  
application in  
history just after 2  
months!"



1900 - 1950

1950  
Turing Test

"A computer would deserve to be called intelligent if it could deceive a human into believing it was human"



1965  
Eliza ChatBot!



陆政哲  
幻方量化



contributing to the field's progress\*

2000  
Kismet Robot



imageret



2016  
Alpha Go



2017 - 2020

2020  
DALL-E

\*generating images from descriptions, showcases the power of AI in cross-modal tasks and creative content generation\*

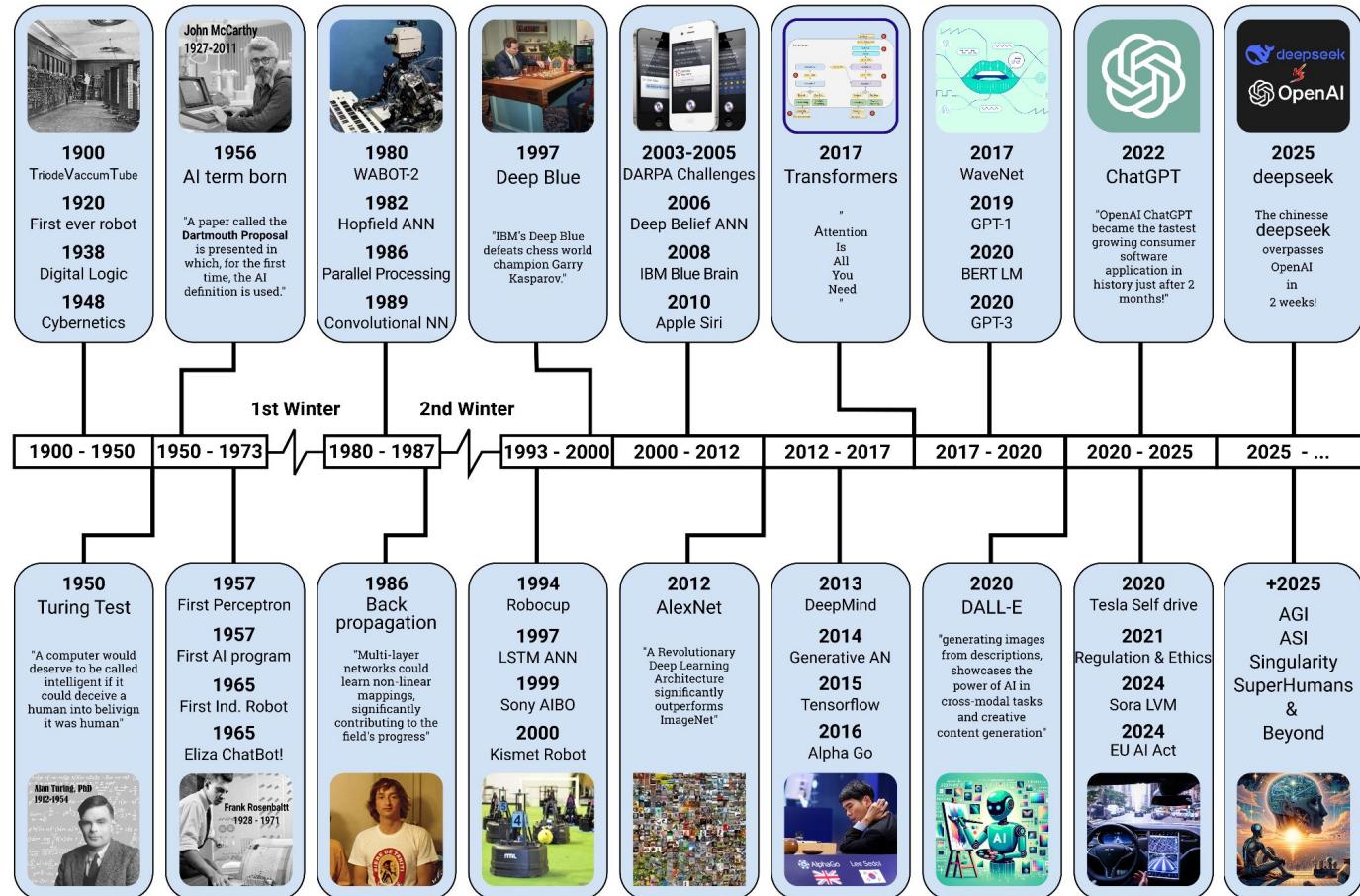


The chinesse  
deepseek  
overpasses  
OpenAI  
in  
2 weeks!

2025 - ...







[https://en.wikipedia.org/wiki/Timeline\\_of\\_artificial\\_intelligence](https://en.wikipedia.org/wiki/Timeline_of_artificial_intelligence)

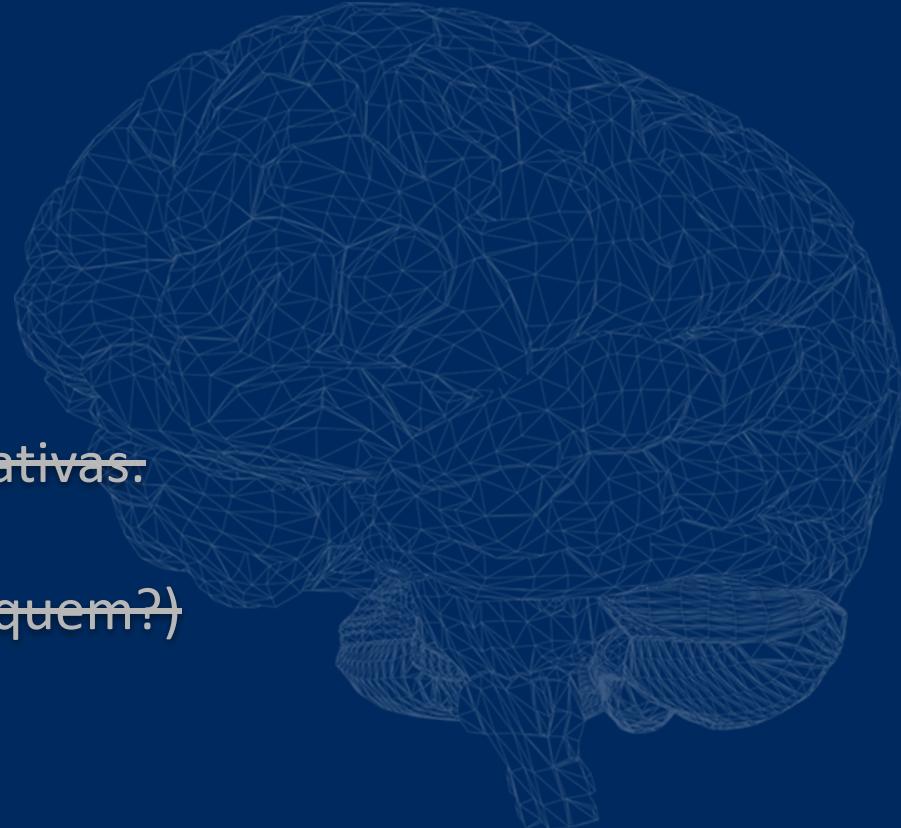
~~Parte 1 - Cases~~

~~Parte 2 - Que tipo de I.A eu preciso?~~

~~Parte 3 - Avanços Recentes e Expectativas.~~

~~BÔNUS 1 - Evolução da I.A (Quem é quem?)~~

~~BÔNUS 2 - Onde e o que aprender.~~



Se eu fosse começar hoje  
tentaria esse:

<https://www.fast.ai/>

≡ fast.ai



## fast.ai—Making neural nets uncool again

- Courses: [Practical Deep Learning for Coders](#); [From Deep Learning Foundations to Stable Diffusion](#)
- Software: [fastai for PyTorch](#); [nbdev](#)
- Book: [Practical Deep Learning for Coders with fastai and PyTorch](#)
- In the news: [The Economist](#); [The New York Times](#); [MIT Tech Review](#)

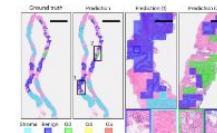
### Blog

Feb 3, 2025  
Rachel Thomas

What AI can tell us about  
microscope slides

MACHINE LEARNING SCIENCE

A friendly introduction to Foundation  
Models for Computational Pathology



Nov 7, 2024  
Jeremy Howard

A New Chapter for fast.ai:  
How To Solve It With Code  
fast.ai is joining Answer.AI, and we're  
announcing a new kind of educational  
experience, 'How To Solve It With  
Code'



### Categories

- All (152)
- advice (24)
- ai-in-society (44)
- courses (40)
- education (6)
- health (14)
- machine learning (2)
- science (1)
- technical (42)

Aprenda a conversar com os LLMs para te darem dicas de onde aprender. Lembre de habilitar reflexão (reasoning) e acesso à Internet para a qualidade ficar maior. Aprenda a testar as respostas.

Se possível converse também com um especialista pois LLMs nem sempre dão respostas corretas.

Quero aprender mais sobre Inteligência Artificial. Tenho facilidade com Inglês, mas não sei mais programação do que alguma coisa no EXCEL. Você pode me indicar algumas trilhas? Meu objetivo é criar produtos que integrem Inteligência Artificial no comércio.

Reasoned for 7 seconds ▾

Claro! Vou propor algumas "trilhas" de estudo e prática que podem te ajudar a entender os fundamentos de Inteligência Artificial e depois aplicar esse conhecimento na criação de produtos para o comércio. Vou estruturar em etapas e dar algumas sugestões de cursos e materiais (prioritariamente em inglês, já que você mencionou ter facilidade, mas também com opções em português). A ideia é que você possa escolher por onde começar de acordo com o seu nível de conforto e tempo disponível.

## 1. Fundamentos de Lógica de Programação e Python

Por que:

- Mesmo que existam hoje muitas ferramentas "low-code" ou "no-code" para IA, aprender os fundamentos de programação—principalmente em Python—vai te ajudar a ir além do básico e ter autonomia na hora de criar soluções personalizadas.

O que estudar:

- Conceitos de variáveis, tipos de dados, estruturas de repetição e condicional.
- Introdução à linguagem Python (sintaxe básica, bibliotecas padrão).
- Uso de notebooks (Jupyter/Google Colab) para praticar.

Sugestões de recursos:

- Curso "Python for Everybody" (Coursera, em inglês)
- Python Crash Course (livro em inglês), de Eric Matthes

## 2. Fundamentos de Data Science e Machine Learning

Por que:

- A base de Inteligência Artificial moderna é estatística, análise de dados e machine learning.  
Entender ao menos superficialmente esses fundamentos é importante para criar soluções robustas.

O que estudar:

- Conceitos básicos de estatística e probabilidade (média, variância, distribuições, correlação).
- Princípios de Machine Learning: supervisão, regressão linear, classificação, validação de modelos.
- Ferramentas e bibliotecas em Python: NumPy, pandas, matplotlib, scikit-learn.

Sugestões de recursos:

- [Curso "Machine Learning" de Andrew Ng \(Coursera, legendas em português\)](#) – aborda conceitos fundamentais.
- [Data Science do Zero \(livro em português\)](#), de Joel Grus – aborda Python e estatística aplicada.
- [Kaggle](#) – cursos introdutórios gratuitos – prática guiada em notebooks online.

O chatgpt propôs justamente um dos melhores cursos introdutórios que eu fiz. Acesso ele ainda hoje pra lembrar conceitos.



Link pra essa conversa com o chat  
gpt:  
<https://chatgpt.com/share/67acd725-b158-8007-8bea-79cb255c81af>

## Resumindo a Trilha

1. Aprenda ou Reforce Lógica de Programação + Python
  - (Cursos gratuitos/YouTube/livros intro de Python)
2. Entenda Fundamentos de Data Science e ML
  - (Conceitos básicos, estatística, scikit-learn)
3. Foque em Casos de Uso para Comércio
  - (Sistemas de recomendação, chatbots, previsão de demanda)
4. Explore Ferramentas e APIs que Agilizam Desenvolvimento
  - (Plataformas low-code, ChatGPT API, Google AI, AWS, etc.)
5. Desenvolva Projetos e Crie Portfólio
  - (Protótipos com dados fictícios ou reais, compartilhe no GitHub)
6. Conecte-se com Comunidades
  - (Meetups, LinkedIn, Kaggle, GitHub)

Dessa forma, você vai gradualmente ganhando confiança na programação, no uso das bibliotecas de IA e entendendo como aplicar todo esse conhecimento em soluções reais para o comércio. A chave é equilibrar teoria e prática, sempre voltando às aplicações de negócios que fazem sentido para você. Boa jornada nos estudos!

# Melhores vídeos para entender LLMs (em fev. 2025)

The screenshot shows the YouTube channel page for Andrej Karpathy. At the top, there's a search bar with 'andrey karpathy' and a play button icon. Below the search bar is a video thumbnail featuring a man in a yellow jacket pointing at a diagram of a neural network architecture. The channel profile picture is a portrait of Andrej Karpathy smiling. His name, 'Andrej Karpathy', is displayed in large bold letters. Below his name, it says '@AndrejKarpathy · 658 mil inscritos · 17 vídeos'. There are links to 'Saiba mais sobre este canal ...mais' and 'eurekalabs.ai e mais 4 links'. A 'Inscrito' (Subscribed) button with a bell icon is present. The main navigation menu includes 'Início', 'Vídeos', 'Playlists', 'Posts', and a search icon. Below the menu, a section titled 'General Audience' with a 'Reproduzir tudo' (Play all) button is shown. A note says 'videos for more general audience, no programming experience necessary.' Two video thumbnails are visible: one for 'INTRO TO LARGE LANGUAGE MODELS' (59:48) and another for 'Deep Dive into LLMs like ChatGPT' (3:31:24).

[https://youtu.be/zjkBMFhNj\\_g?si=OpiMuhMOKMEQuN57](https://youtu.be/zjkBMFhNj_g?si=OpiMuhMOKMEQuN57)

<https://youtu.be/7xTGNNLPyMI?si=tB65E9oXwzMb56at>



# Um lugar para aprender ou se atualizar. (MIT Courses)

<http://introtodeeplearning.com/>



## Schedule

New 2024 lectures, slides, and labs!  
*Taught in-person at MIT — open-sourced to the world.*

<b>Intro to Deep Learning</b> Lecture 1 Apr. 29, 2024 [Slides] [Video]	<b>Deep Sequence Modeling</b> Lecture 2 May 6, 2024 [Slides] [Video]	<b>Intro to TensorFlow; Music Generation</b> Software Lab 1 [Code]
<b>Deep Computer Vision</b> Lecture 3 May 13, 2024 [Slides] [Video]	<b>Deep Generative Modeling</b> Lecture 4 May 20, 2024 [Slides] [Video]	<b>Facial Detection Systems</b> Software Lab 2 [Paper] [Code]
<b>Deep Reinforcement Learning</b> Lecture 5 May 27, 2024 [Slides] [Video]	<b>New Frontiers</b> Lecture 6 June 3, 2024 [Slides] [Video]	<b>Large Language Models</b> Software Lab 3 [Code] coming soon!
<b>Generative AI for Media</b> Lecture 7 June 10, 2024 [Slides] [Video]	<b>Stories from Models in the Wild</b> Lecture 8 June 17, 2024 [Slides] [Video]	<b>Final Project</b> Work on final projects
<b>Final Project</b> Work on final projects June 24, 2024	<b>Project Presentations</b> Pitch your ideas! June 24, 2024	<b>Awards Ceremony</b> Final awards and celebration!

Links com muita informação de qualidade e sem distrações.

Chip Huyen - <https://huyenchip.com/blog/>

Eugeneyan - <https://eugeneyan.com/writing/>

Lil'Log - <https://lilianweng.github.io/>

Sebastian Raschka - <https://sebastianraschka.com/>

Simon Willison - <https://simonwillison.net/>

Andrej Karpathy - <https://karpathy.github.io/>

Nathan Lambert (Interconnects AI) - <https://www.interconnects.ai/>

Ethan Mollick - <https://www.oneusefulthing.org/>

Gwern - <https://gwern.net/>

Sebastian Ruder - <https://www.ruder.io/>

Companies blogs

Hugging Face blogs - <https://huggingface.co/blog>

Anthropic blogs - <https://www.anthropic.com/research>

Sakana AI blogs - <https://sakana.ai/blog>

Sequoia Capital blogs - <https://www.sequoiacap.com/stories/>

a16z blogs - <https://a16z.com/news-content/>

Google DeepMing blog - <https://deepmind.google/discover/blog/>

Meta AI blogs - <https://ai.meta.com/blog/>

# Alguns “Playgrounds de I.A.”

<https://chatgpt.com/share/67acfa90-96b8-8007-bbff-1d5ed89c5af9>

ChatGPT 4o ▾

This is a copy of a conversation between ChatGPT & Anonymous.

[Report content](#)

<https://cs.stanford.edu/people/karpathy/convnetjs/>

Preciso de links com "playgrounds" interessantes para demonstrar modelos, conceitos, ... relacionados a deep learning e deep learning generativo, ...  
Como referência esse link acima

Claro, aqui estão alguns "playgrounds" interativos que demonstram conceitos e modelos relacionados a deep learning e deep learning generativo:

1. **A Neural Network Playground:** Uma ferramenta interativa que permite explorar como redes neurais aprendem a partir de dados. Você pode ajustar parâmetros como número de camadas, neurônios e taxas de aprendizado para ver como eles afetam o desempenho da rede.

Aqui está a lista de **playgrounds interativos** para demonstrar conceitos de **deep learning** e **deep learning generativo**, incluindo o link do Karpathy:

1. **ConvNetJS (Karpathy)** - <https://cs.stanford.edu/people/karpathy/convnetjs/>  
Biblioteca JavaScript para redes neurais convolucionais rodando diretamente no navegador, sem necessidade de bibliotecas externas.
2. **A Neural Network Playground** - <https://playground.tensorflow.org>  
Simulação interativa de redes neurais básicas usando TensorFlow, permitindo ajustar camadas, neurônios e taxas de aprendizado.
3. **Deeper Playground** - <https://deeperplayground.org>  
Uma versão aprimorada do TensorFlow Playground com funções de perda avançadas e mais opções de experimentação.
4. **GAN Lab** - <https://poloclub.github.io/ganlab>  
Ferramenta visual interativa para explorar como Redes Adversárias Generativas (GANs) aprendem a gerar novos dados.
5. **TensorSpace Playground** - <https://tensorspace.org/html/playground/index.html>  
Plataforma para visualizar redes neurais profundas em 3D, incluindo modelos de classificação e detecção de objetos.
6. **Machine Learning Playground** - <https://ml-playground.com>  
Interface intuitiva para testar módulos básicos de machine learning e deep learning sem precisar escrever código.
7. **Digit Fantasies by a Deep Generative Model** - [https://dpmkingma.com/sgvb\\_mnist\\_demo/demo.html](https://dpmkingma.com/sgvb_mnist_demo/demo.html)  
Demonstração de modelos generativos aplicados à criação de imagens de dígitos, ajustando variáveis latentes.
8. **RAISE Playground** - <https://playground.raise.mit.edu>  
Ambiente de aprendizado baseado em blocos para explorar conceitos de IA e machine learning de forma interativa

Nesse link o “hands on” do evento:

[https://github.com/mfilipak/BXBRASIL2025/blob/main/BX\\_BRASIL\\_HANDS\\_ON.ipynb](https://github.com/mfilipak/BXBRASIL2025/blob/main/BX_BRASIL_HANDS_ON.ipynb)

BXBRAISL2025 / BX\_BRASIL\_HANDS\_ON.ipynb

mfilipak Created using Colab

Preview Code Blame 783 lines (783 loc) · 330 KB Code 55% faster with GitHub Copilot

Open in Colab

COLAB NOTEBOOK DO EVENTO BX-BRASIL

1 - Clique  
aqui.

2 - Aperte o “play” para rodar cada  
célula ou simplesmente “leia” os  
resultados e explicações dos códigos  
direto no Python / Jupyter Notebook.

COLAB NOTEBOOK DO EVENTO BX-BRASIL

1 - USANDO A API NOTA CERTA.

Import requests

```
Run cell (Ctrl+Enter)
cell has not been executed in this session demo.hf.space/api/predict
```

executed at unknown time

```
"data": ["Coca-Cola PET 21"]
```

headers = {  
 "Content-Type": "application/json"  
}  
  
response = requests.post(url, json=payload, headers=headers)  
result = response.json()  
  
import pprint  
pprint.pprint(result)

{'average\_duration': 0.06768654330863252,  
 'data': [{  
 'confidences': [{  
 'confidence': 0.5184096693992615,  
 'label': '22021000'},  
 {'confidence': 0.03791089728474617,  
 'label': '22011000'},  
 {'confidence': 0.03218604251742363,  
 'label': '22030000'},  
 {'confidence': 0.02866430953145027,  
 'label': '21069090'},  
 {'confidence': 0.02776830829679966,  
 'label': '22029900'}],  
 'label': '22021000'},  
 'bebidas, líquidos alcoólicos e vinagres. Águas, incluindo as águas '  
 'minerais e as águas gaseificadas, adicionadas de açúcar ou de '  
 'outros edulcorantes ou aromatizadas e outras bebidas não '  
 'alcoólicas, exceto sucos (sumos) de fruta ou de produtos hortícolas '  
 'da posição 20.09.- Águas, incluindo as águas minerais e as águas '  
 'gaseificadas, adicionadas de açúcar ou de outros edulcorantes ou '  
 'aromatizadas',  
 'duration': 0.05477762222290039,  
 'is\_generating': False}



**Marcelo Filipak**

filipak@filiart.com.br

<https://www.linkedin.com/in/marcelo-filipak-97386710/>

<https://github.com/mfilipak>



# I.A. NA PRÁTICA

Técnicas, cases reais e mão na massa.

BAIXE O MATERIAL

OBRIGADO!

<https://github.com/mfilipak/BXBRASIL2025>