



I.A NA PRÁTICA

com alguns exemplos na automação comercial.



'elgin

Tech Day
2024

<https://bit.ly/4cLF7SL>





I.A NA PRÁTICA

com alguns exemplos na automação comercial.



Marcelo Filipak, Engenheiro e Cientista de dados.

- **Treino redes neurais há 7 anos.**
- Me especializei em criar modelos.
- Projetos de modelos de deep learning com aplicações na área de saúde, segurança pública, controles de tráfego, logística, inteligência fiscal, controle aduaneiro, controle de qualidade ...

mfilipak@hotmail.com

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

<https://github.com/mfilipak/AUTOCOM2024>

Projeto - INFANT.ID



INFANT.ID

Construindo um mundo mais seguro para as nossas crianças

[CONHEÇA A PLATAFORMA INFANT.ID](#)

CRIANÇAS DESAPARECIDAS NO BRASIL

4

POR HORA

88

POR DIA

32000

POR ANO

PROJETO DE IDENTIFICAÇÃO NEONATAL

Recife - PE

Através de um convênio de cooperação técnica em parceria com o IITB (Instituto de Identificação Tavares Burlil) e Secretaria de Saúde de Pernambuco, existe desde 2016 o projeto IDENTIFICAÇÃO NEONATAL em maternidades do estado para validar e aprimorar a solução de identificação de recém-nascidos neonatal desenvolvida pela empresa, sob a supervisão e validação dos peritos papiloscopistas do IITB.

Nós já coletamos mais de 6000 imagens. Uma pequena contribuição mas um grande passo para um mundo mais seguro para as crianças.



Projeto - INFANT.ID



- Vincula a mãe com a criança.
- Valida a digital da criança na saída da maternidade.
- Autentica até mesmo o indivíduo adulto com a imagem de recém nascido.
- Uso intenso de modelos de visão computacional.
- Da ordem de 100000 crianças registradas em diversas cidades.

Projeto AFRAC NOTA CERTA

Spaces AFRAC/NCM_DEMO like 5 Running

AFRAC NOTA CERTA

DESCRÍÇÃO

Clear Submit

NCMs

84433299

84433299 19%

84431990 6%

84439941 6%

84433232 5%

84775990 5%

Descrição do NCM

Reatores nucleares, caldeiras, máquinas, aparelhos e instrumentos mecânicos, e suas partes. Máquinas e aparelhos de impressão por meio de placas, cilindros e outros elementos de impressão da posição 84.42; outras impressoras, aparelhos de copiar e aparelhos de telecopiar (fax), mesmo combinados entre si; partes e acessórios.- Outras impressoras, aparelhos de copiar e aparelhos de telecopiar (fax), mesmo combinados entre si:- Outros, capazes de ser conectados a uma máquina automática para processamento de dados ou a uma redeOutrasOutras

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

Spaces AFRAC/NCM_DEMO like 5 Running

AFRAC NOTA CERTA

DESCRÍÇÃO

Clear Submit

NCMs

84433299

84433299 19%

84431990 6%

Descrição do NCM

Reatores nucleares, caldeiras, máquinas, aparelhos e instrumentos mecânicos, e suas partes. Máquinas e aparelhos de impressão por meio de placas, cilindros e outros elementos de impressão da posição 84.42; outras impressoras, aparelhos de copiar e aparelhos de telecopiar (fax), mesmo combinados entre si; partes e acessórios.- Outras impressoras, aparelhos de copiar e aparelhos de telecopiar (fax), mesmo combinados entre si:- Outros, capazes de ser conectados a uma máquina automática para processamento de dados ou a uma redeOutrasOutras

Reatores nucleares, caldeiras, máquinas, aparelhos e instrumentos mecânicos, e suas partes. Máquinas e aparelhos de impressão por meio de placas, cilindros e outros elementos de impressão da posição 84.42; outras impressoras, aparelhos de copiar e aparelhos de telecopiar (fax), mesmo combinados entre si; partes e acessórios.- Outras impressoras, aparelhos de copiar e aparelhos de telecopiar (fax), mesmo combinados entre si:- Outros, capazes de ser conectados a uma máquina automática para processamento de dados ou a uma redeOutrasOutras

AFRAC NOTA CERTA



5:24



Nota Certa

O novo projeto revolucionário da AFRAC traz mais uma facilidade para você, associado, encontrar de uma forma rápida e simples o NCM do produto desejado.

mandioca kg



07141000

probabilidade 11%

Produtos hortícolas, plantas, raízes e tubérculos, comestíveis.Raízes de mandioca, de araruta e de salepo, tupinambós, batatas-doces e raízes ou tubérculos semelhantes, com elevado teor de fécula ou de inulina, frescos, refrigerados, congelados ou secos, mesmo cortados em pedaços ou em pellets; medula de sanguíneo.- Raízes de mandioca

buscar

```
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)
```

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)
```

▶ result

```
{'data': [{'label': '22021000',
  'confidences': [{"label": '22021000', 'confidence': 0.5184},
    {"label": '22011000', 'confidence': 0.03791089728474617},
    {"label": '22030000', 'confidence': 0.03218604251742363},
    {"label": '21069090', 'confidence': 0.02866430953145027},
    {"label": '22029900', 'confidence': 0.02776830829679966}],
  'Bebidas, líquidos alcoólicos e vinagres. Águas, incluindo as sucos (sumos) de fruta ou de produtos hortícolas da posição 20
  'is_generating': False,
  'duration': 0.09473180770874023,
  'average_duration': 0.07237190791693515}]}
```

I.A NA PRÁTICA

com alguns exemplos na automação comercial.



- 1 - Use agentes (LLM) pra quase tudo.**
- 2 - Inteligência Artificial não é magia negra.
- 3 - Entenda seu problema - O que já fiz de “errado”.
- 4 - O inesperado acontece. O que são LLMs e o que é I.A generativa?
- 5 - Agentic A.I é o futuro?
- 6 - Considerações finais. Alguns exemplos na automação comercial.

Bônus 1 - Sugestões para aprendizado de I.A.

Bônus 2 - Uma breve história da I.A. Marcos importantes. Quem é Quem?

LLMs pra quase tudo...

(LLM = Large Language Models)

M You

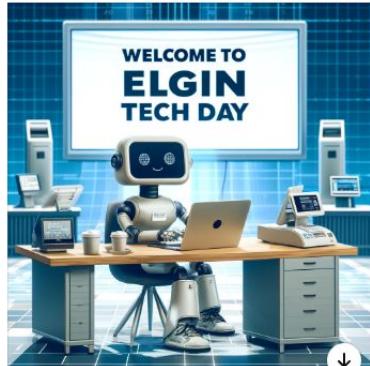


Tech
20



Inspirado nas imagens acima, gere uma figura para a abertura de uma apresentação que será dada no evento Elgin Tech Day. O slide deve ter um robô simpático programando em um laptop com itens usados na automação comercial como impressora, gaveta de dinheiro e maquinetas de cartão.

DALL-E



Here are the images for the presentation opening at Elgin Tech Day. Choose the one you find best

LLMs pra quase tudo...

(LLM = Large Language Models)



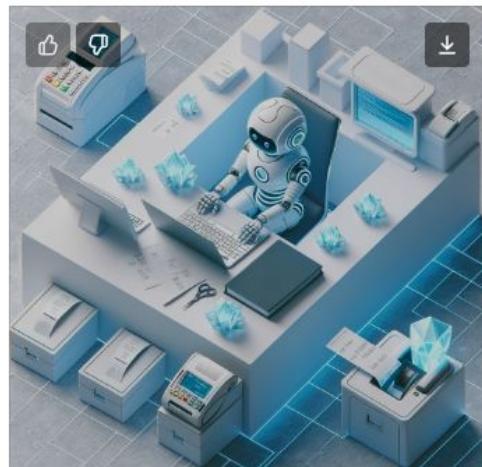
You

Inspirado nas imagens acima, gere uma figura para a abertura de uma apresentação que será dada no evento Elgin Tech Day. O slide deve ter um robô simpático programando em um laptop com itens usados na automação comercial como impressora, gaveta de dinheiro e maquinetas de cartão.

, aerial view, made of crystal



DALL-E



LLMs pra quase tudo...

(LLM = Large Language Models)

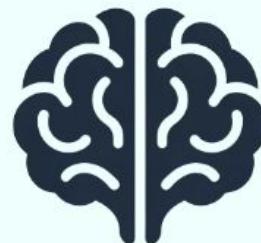
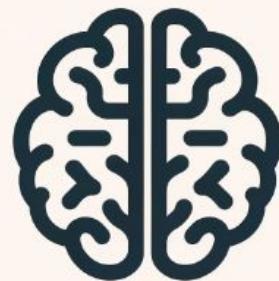


You

Crie uma imagem simples de um cérebro estilizado



DALL-E



Here are the stylized brain images for you to consider.

I.A NA PRÁTICA

com alguns exemplos na automação comercial.

- 1 ~~Use agentes (LLM) pra quase tudo.~~
- 2 - Inteligência Artificial não é magia negra.
- 3 - Entenda seu problema - O que já fiz de “errado”.
- 4 - O inesperado acontece. O que são LLMs e o que é I.A generativa?
- 5 - Agentic A.I é o futuro?
- 6 - Considerações finais. Alguns exemplos na automação comercial.

Bônus 1 - Sugestões para aprendizado de I.A.

Bônus 2 - Uma breve história da I.A. Marcos importantes. Quem é Quem?

Inteligência Artificial não é magia negra.

Vamos usar o chat gpt para construir um modelo simples de Inteligência Artificial, de fácil compreensão (mas que ainda assim é muito poderoso como poderemos comprovar).

Isso que vou apresentar aqui ilustra como o chat gpt pode ajudar a programar/construir modelos.

Existem vários outros LLMs poderosos além do CHAT GPT, vamos falar deles também.

M

You

User

Tenho um dataset com 30 linhas e 4 colunas e preciso fazer uma predição do tipo sim e não. Qual o tipo de modelo recomendável pra esse caso? Gostaria que esse modelo tivesse boa acurácia e não ficasse viciado.

G

ChatGPT

Para um dataset relativamente pequeno com 30 linhas e 4 colunas e uma tarefa de classificação binária (sim e não), você pode considerar usar um modelo de Árvore de Decisão com técnicas de regularização para evitar o overfitting e garantir uma boa acurácia. Aqui estão algumas sugestões específicas:

1. **Árvore de Decisão com Pruning (poda):** Após construir a árvore de decisão, você pode podá-la para evitar que ela se torne muito complexa e se ajuste em excesso aos dados de treinamento. A poda remove ramos da árvore que não contribuem significativamente para a sua capacidade de generalização.
2. **Validação Cruzada:** Use técnicas de validação cruzada, como a validação cruzada k-fold, para avaliar o desempenho do modelo de forma mais robusta e evitar que ele esteja viciado em um único conjunto de treinamento/teste. Isso ajuda a estimar melhor a capacidade de generalização do modelo.

3. **Regularização:** Alguns algoritmos de Árvore de Decisão, como o algoritmo CART (Classification and Regression Trees), oferecem parâmetros de regularização que você pode ajustar para controlar a complexidade da árvore e evitar overfitting. Por exemplo, você pode definir um limite para a profundidade máxima da árvore ou o número mínimo de amostras necessárias em um nó folha.

4. **Random Forest:** Uma alternativa às Árvores de Decisão simples é usar um modelo de Random Forest, que é uma coleção de árvores de decisão. O Random Forest pode reduzir o overfitting ao combinar várias árvores e fazer previsões com base na média ou votação das previsões individuais das árvores.

Ao implementar qualquer modelo, certifique-se de dividir seu dataset em conjunto de treinamento e conjunto de teste para avaliar o desempenho do modelo em dados não vistos durante o treinamento. Isso ajudará a garantir que o modelo generalize bem para novos dados.



M

You

Gostaria de uma tabela com 30 linhas com dados de automação comercial que possam ser usadas num tutorial do uso de algoritmos de random forest

M

You

[Elimine](#) a última coluna



ChatGPT

Aqui está uma tabela exemplo de dados de automação comercial com 30 linhas que você pode usar em um tutorial de algoritmos de random forest. Vou incluir algumas colunas que são comuns em análises desse tipo, como características do produto, vendas e feedback do cliente.

- **Produto_ID:** Identificador único do produto.
- **Categoria:** Categoria do produto (por exemplo, Eletrônicos, Vestuário, Alimentos).
- **Preço:** Preço do produto.
- **Custo:** Custo de produção do produto.
- **Vendas_Mensais:** Número de unidades vendidas por mês.
- **Estoque:** Quantidade de produtos em estoque.
- **Promoção:** Se o produto está em promoção (Sim/Não).
- **Avaliação_Cliente:** Média de avaliações dos clientes (de 1 a 5).

Aqui está um exemplo dos dados:

| Produto_ID | Categoria | Preço | Custo | Vendas_Mensais | Estoque | Promoção | Avaliação_Clien |
|------------|-------------|-------|-------|----------------|---------|----------|-----------------|
| 1 | Eletrônicos | 1200 | 800 | 30 | 100 | Não | 4.5 |
| 2 | Vestuário | 200 | 100 | 50 | 150 | Sim | 4.2 |

| Preço | Custo | Estoque | Promoção | Vendas_Mensais |
|-------|-------|---------|----------|----------------|
| 1200 | 800 | 100 | 0 | 30 |
| 200 | 100 | 150 | 1 | 50 |

DADOS DE TREINO

| Preço | Custo | Estoque | Promoção | Vendas_Mensais |
|-------|-------|---------|----------|----------------|
| 1200 | 800 | 100 | 0 | 30 |
| 200 | 100 | 150 | 1 | 50 |
| 50 | 20 | 500 | 0 | 200 |
| 100 | 50 | 250 | 0 | 40 |
| 3000 | 2000 | 50 | 1 | 15 |
| 150 | 75 | 300 | 0 | 60 |
| 30 | 10 | 600 | 1 | 250 |
| 80 | 40 | 200 | 0 | 30 |
| 2200 | 1500 | 90 | 1 | 20 |
| 350 | 200 | 400 | 0 | 70 |
| 60 | 30 | 450 | 1 | 300 |
| 120 | 60 | 350 | 0 | 50 |
| 2500 | 1700 | 75 | 0 | 18 |
| 100 | 50 | 500 | 1 | 80 |
| 45 | 25 | 550 | 0 | 220 |
| 110 | 55 | 300 | 0 | 35 |
| 2000 | 1300 | 85 | 1 | 22 |
| 250 | 125 | 450 | 0 | 65 |
| 35 | 15 | 700 | 1 | 320 |
| 90 | 45 | 400 | 0 | 45 |
| 2600 | 1800 | 65 | 1 | 17 |
| 220 | 110 | 350 | 0 | 55 |
| 25 | 12 | 800 | 0 | 400 |
| 150 | 75 | 220 | 1 | 55 |
| 2800 | 1900 | 70 | 0 | 19 |
| 175 | 90 | 600 | 1 | 75 |
| 40 | 18 | 650 | 0 | 275 |
| 130 | 65 | 380 | 0 | 60 |
| 3100 | 2100 | 60 | 1 | 14 |
| 300 | 150 | 420 | 0 | 85 |

VALIDAÇÃO
(Não usados no treino)

M You

ERREI,, mas não fez diferença



Usando as 25 primeiras colunas como treino e as 5 últimas como validação, crie um modelo random forest para fazer a previsão das vendas

ChatGPT

Memory updated

Para criar um modelo de random forest que preveja as vendas com base nos dados fornecidos, seguiremos estes passos:

1. **Preparação dos dados:** Dividiremos os dados em um conjunto de treinamento (as 25 primeiras linhas) e um conjunto de validação (as 5 últimas linhas).
2. **Treinamento do modelo:** Usaremos o conjunto de treinamento para treinar o modelo de random forest.
3. **Validação do modelo:** Testaremos o modelo no conjunto de validação para avaliar seu desempenho.

Vamos começar com essas etapas:

! Error analyzing ▾

????

AGENTIC AI

! Error analyzing ▾

! Error analyzing ▾

! Error analyzing ▾

 Error analyzing

Ocorreu um erro porque o `LabelEncoder` não foi definido nesta sessão. Preciso importar e inicializar o `LabelEncoder` antes de usar. Vou corrigir isso e executar o código novamente. [-]
O modelo de random forest foi treinado e testado, resultando em um erro quadrático médio (MSE) de aproximadamente 296.40 nas vendas mensais. Isso indica o desempenho do modelo ao prever as vendas com base nos dados fornecidos. [-]

 You

Por favor, mostre as predições no dataset de validação

 ChatGPT

Aqui estão as predições do modelo no conjunto de validação, comparadas com os valores reais de vendas mensais:

| | Vendas_Reais | Vendas_Preditas |
|----|--------------|-----------------|
| 25 | 75 | 78.10 |
| 26 | 275 | 242.90 |
| 27 | 60 | 55.20 |
| 28 | 14 | 16.19 |
| 29 | 85 | 64.65 |

As predições estão relativamente próximas aos valores reais, com algumas variações, o que é esperado em um modelo de previsão. [-]



You

Agora crie um modelo de redes neurais com 2 camadas internas para o mesmo problema e me diga quais foram os resultados



ChatGPT

O modelo de rede neural com 2 camadas internas de 100 neurônios cada resultou em um erro quadrático médio (MSE) de aproximadamente 517.17 nas vendas mensais. Aqui estão as previsões feitas pelo modelo no conjunto de validação:

| | Vendas_Preditas |
|----|-----------------|
| 25 | 101.49 |
| 26 | 240.50 |
| 27 | 47.90 |
| 28 | 16.12 |
| 29 | 61.69 |

O MSE é maior do que o obtido pelo modelo de random forest, o que sugere que o modelo de random forest pode ter desempenhado melhor neste caso específico. Além disso, o modelo de rede neural não convergiu com o número máximo de iterações definido, indicando que pode ser necessário um ajuste adicional nos parâmetros do modelo ou mais iterações para melhorar a convergência e o desempenho. [-]



You

Digo, faça um modelo de decision tree e me passe o código equivalente



You

E qual seria o resultado da validação desse modelo?



ChatGPT

O modelo de árvore de decisão resultou em um erro quadrático médio (MSE) de aproximadamente 1210.2 na validação, indicando o desempenho do modelo ao prever as vendas mensais com base no conjunto de dados fornecido. [–]

Por fim, pedi que ele empacotasse tudo o que falamos aqui em um código Python para que eu pudesse comparar melhor as predições



ChatGPT

Aqui está o código Python que cobre os passos solicitados:

1. A tabela de dados.
 2. O treinamento e o teste do modelo de Random Forest com 100 árvores.
 3. O treinamento e o teste do modelo de rede neural com 2 camadas intermediárias.
 4. O treinamento e o teste do modelo de árvore de decisão.

python

 Copy code

```
import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.ensemble import RandomForestRegressor
from sklearn.neural_network import MLPRegressor
from sklearn.tree import DecisionTreeRegressor
from sklearn.metrics import mean_squared_error
from sklearn.preprocessing import LabelEncoder

# Dados
data = {
    'Preço': [1200, 200, 50, 100, 3000, 150, 30, 80, 2200, 350, 60, 120, 2500, 100, 1500],
    'Custo': [800, 100, 20, 50, 2000, 75, 10, 40, 1500, 200, 30, 60, 1700, 50, 25, 550],
    'Estoque': [100, 150, 500, 250, 50, 300, 600, 200, 90, 400, 450, 350, 75, 500, 550],
    'Promoção': ['Não', 'Sim', 'Não', 'Não', 'Sim', 'Não', 'Sim', 'Não', 'Sim', 'Não'],
    'Vendas_Mensais': [30, 50, 200, 40, 15, 60, 250, 30, 20, 70, 300, 50, 18, 80, 220]
}

df = pd.DataFrame(data)
label_encoder = LabelEncoder()
```

Esse é o começo

+/- 30
linhas de
código
PYTHON

mfilipak / AUTOCOM2024

Code Issues Pull requests Actions Projects Wiki Security Insights Settings

main AUTOCOM2024 / ChatGP_Randomforest.ipynb Go to file ...

mfilipak Created using Colaboratory 8428a80 · 1 hour ago History

795 lines (795 loc) · 341 KB

Preview Code Blame Raw

[Open in Colab](#)

In [32]:

```
import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.ensemble import RandomForestRegressor
from sklearn.neural_network import MLPRegressor
from sklearn.tree import DecisionTreeRegressor
from sklearn.metrics import mean_squared_error
from sklearn.preprocessing import LabelEncoder

# Dados
data = {
    'Preço': [1200, 200, 50, 100, 3000, 150, 30, 80, 2200, 350, 60, 120, 2500, 100, 45, 110, 2000, 250, 3,
    'Custo': [800, 100, 20, 50, 2000, 75, 10, 40, 1500, 200, 30, 60, 1700, 50, 25, 55, 1300, 125, 15, 45,
    'Estoque': [100, 150, 500, 250, 50, 300, 600, 200, 90, 400, 450, 350, 75, 500, 550, 300, 85, 450, 700
    'Promoção': ['Não', 'Sim', 'Não', 'Sim', 'Não', 'Sim', 'Sim', 'Não', 'Sim', 'Não', 'Sim', 'Não', 'Sim', 'Não',
    'Vendas_Mensais': [30, 50, 200, 40, 15, 60, 250, 30, 20, 70, 300, 50, 18, 80, 220, 35, 22, 65, 320, 4
}

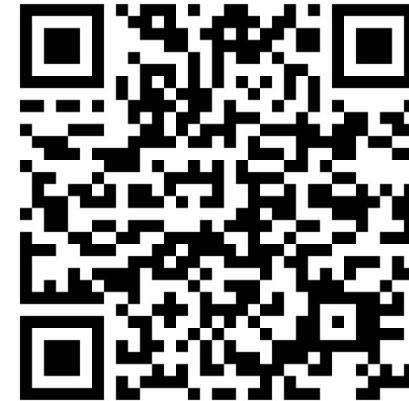
df = pd.DataFrame(data)
label_encoder = LabelEncoder()
df['Promoção'] = label_encoder.fit_transform(df['Promoção'])

# Divisão dos dados em treino e teste
X = df.drop('Vendas_Mensais', axis=1)
y = df['Vendas_Mensais']
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.17, random_state=42)

X_train, X_test, y_train, y_test = X[:-5], X[-5:], y[:-5], y[-5:]
```

O código gerado no último slide, com alguns pequenos ajustes que fiz pode ser rodado até mesmo do seu smartphone nesse link:

<https://bit.ly/3VMhK5o>



PREDIÇÕES

| Venda Mensal | Modelo Random Forest | Modelo Rede Neural | Árvore de Decisão |
|--------------|----------------------|--------------------|-------------------|
| 75 | 78.10 | 101.485248 | 65.0 |
| 275 | 242.90 | 240.500573 | 200.0 |
| 60 | 55.20 | 47.897833 | 70.0 |
| 14 | 16.19 | 16.117632 | 15.0 |
| 85 | 64.65 | 61.693126 | 70.0 |

MENORES DIFERENÇAS

| Venda Mensal | Modelo Random Forest | Modelo Rede Neural | Árvore de Decisão |
|--------------|----------------------|--------------------|-------------------|
| 75 | 3.100000 | 26.485248 | 10.000000 |
| 275 | 32.100000 | 34.499427 | 75.000000 |
| 60 | 4.800000 | 12.102167 | 10.000000 |
| 14 | 2.190000 | 2.117632 | 1.000000 |
| 85 | 20.350000 | 23.306874 | 15.000000 |

E como eu uso esse modelo que foi melhor que uma rede neural?

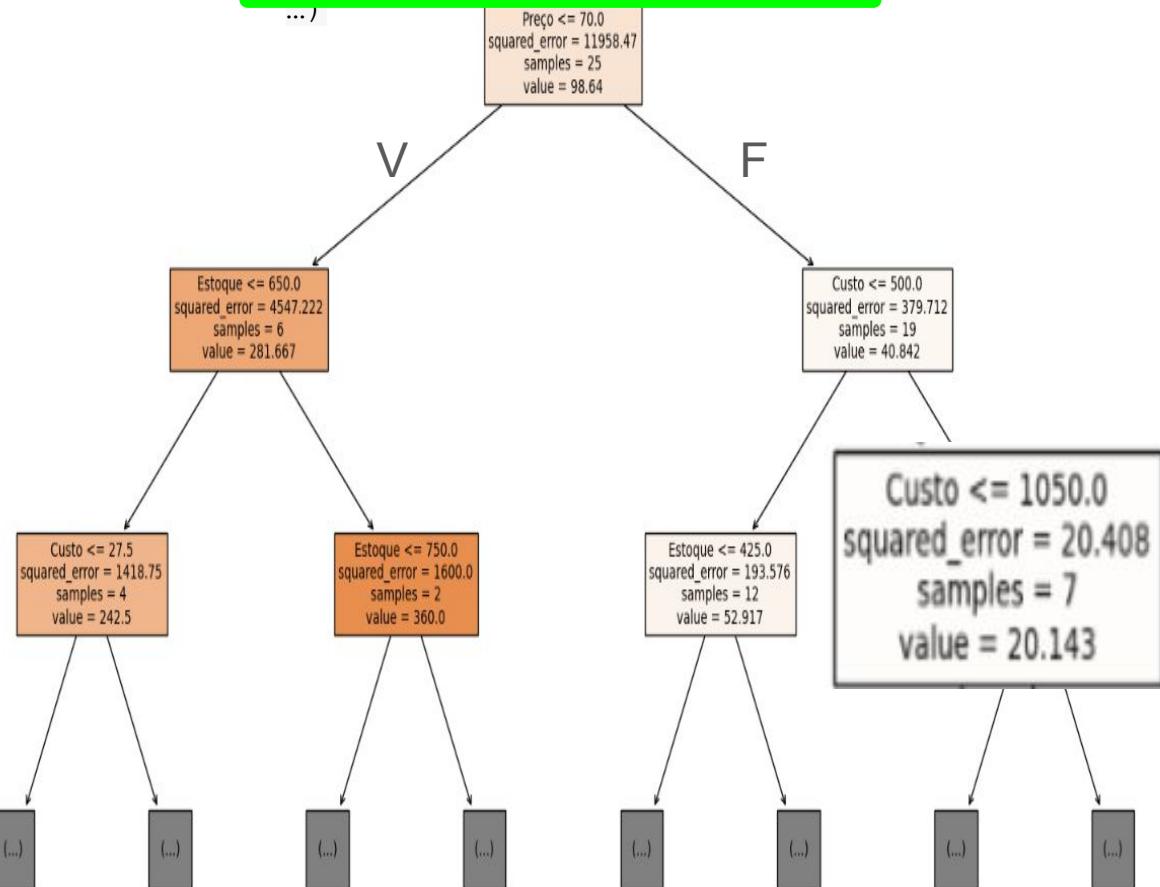
No código gerado aqui, o modelo de árvore de decisão se chama `dt_model` o random forest `rf_model` e por aí vai...

A chamada pra fazer uma predição seria algo como:

```
previsão_de_vendas = dt_model.predict([preço, custo, estoque, promoção]),
```

```
tree.plot_tree(dt_model,
```

```
...)
```



E se eu tiver que explicar pra alguém como esse modelo funciona?

Aquela funçãozinha `tree.plot...` gera a árvore.

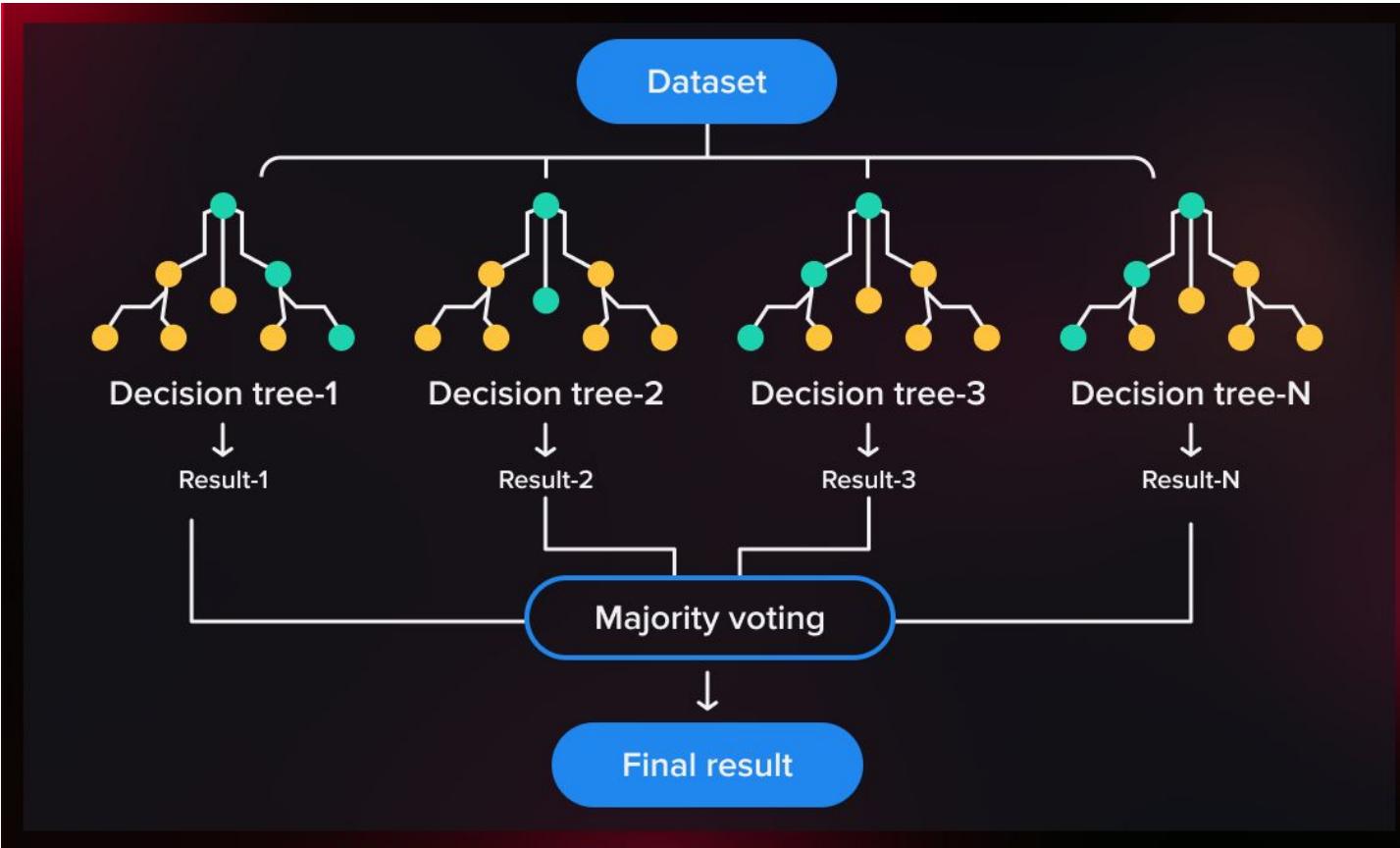
Se preço > 70 então
se custo > 500 então
venda_mensal = 20.14

...

Se parasse aqui, essa predição teria um erro provável (MSE de 20,4), tirando a raiz quadrada seria 4.5.

O resultado mais confiável seria:
20,14 +/- 4.5

Mas e esse Random Forest, que foi disparado o melhor?



É composto de várias árvores simples treinadas em subconjuntos menores do dataset original.

Ao fim é aplicada uma operação, por exemplo, a média ou a maioria vence, e gera um resultado por isso mais robusto.

Inteligência Artificial não é magia negra.

CHECKPOINT

- De forma geral, dados tabulares esparsos não funcionam muito bem com redes neurais e redes complexas.
- O exemplo aqui foi lúdico, mas até mesmo com tabelas da ordem de dezenas de milhares, random forest costuma ser melhor que redes neurais.
- Mas quando temos big data (ou modelos pré-treinados em big data que possam ser usados nos nossos dados) aí o cenário tende a ser favorável às redes neurais.
- E às vezes o melhor é fazer uma composição com mais de um tipo de modelo (random forest já é meio que isso).

I.A NA PRÁTICA

com alguns exemplos na automação comercial.



- 1 ~~Use agentes (LLM) pra quase tudo.~~
- 2 ~~Inteligência Artificial não é magia negra.~~
- 3 - Entenda seu problema - O que já fiz de “errado”.**
- 4 - O inesperado acontece. O que são LLMs e o que é I.A generativa?
- 5 - Agentic A.I é o futuro?
- 6 - Considerações finais. Alguns exemplos na automação comercial.

Bônus 1 - Sugestões para aprendizado de I.A.

Bônus 2 - Uma breve história da I.A. Marcos importantes. Quem é Quem?

O que já fiz de “errado”

Projeto: Usar câmeras instaladas numa loja para conhecer melhor os clientes.

Arquitetura: Imagens capturadas enviadas para nuvem, o backend faz o tratamento do reconhecimento facial e devolve para a aplicação local.

Como protótipo funcionou bem.

O que já fiz de “errado”

Problemas:

Custo, especialmente porque era um produto que deveria ser barato.

Dava umas engasgadas especialmente em WIFI.

Dependência de WIFI sempre.

Crença: (que inclusive PhDs em deep learning de institutos de pesquisa afirmavam)

Não dá pra rodar localmente

Realidade:

Rodaria muito melhor localmente (mesmo numa RASPBERRY sem GPU)

Isso aconteceu por volta de 2016/2017, infelizmente o projeto foi abortado por falta de viabilidade.

Exemplo atual

Uma chamada à API Nota Certa (na nuvem) leva tipicamente 0.1s (e oscila)



result

```
{'data': [{label': '22021000',
  'confidences': [{label': '22021000', 'confidence': 0.5184096693992615},
    {'label': '22011000', 'confidence': 0.03791089728474617},
    {'label': '22030000', 'confidence': 0.03218604251742363},
    {'label': '21069090', '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}
```

Uma predição local é 1000 vezes mais rápida no pior PC que você encontrar, podendo chegar a 100000 vezes se você tiver bastante memória e GPU local.

Recomendações

- 1 - Analise criticamente o que pode rodar localmente e o que precisa rodar na nuvem.
- 2 - Se achar que não dá pra rodar localmente, analise de novo pois modelos de I.A são mais fáceis de portar que código (São basicamente só matrizes numéricas e regras de operações matemáticas).
- 3 - Rodar na nuvem pode ser mais simples, portanto prototipar nela é muito interessante.
- 4 - Pondere, pois às vezes mesmo sendo mais caro e ineficiente que rodar localmente, pode ser muito mais simples e gerenciável rodar na nuvem.
- 5 - Modelos muito grandes podem exigir rodar na nuvem para não trazerem dependências muito específicas de hardware local.

Via de regra: Treine na nuvem e rode os modelos localmente.

I.A NA PRÁTICA

com alguns exemplos na automação comercial.



- 1 ~~Use agentes (LLM) pra quase tudo.~~
- 2 ~~Inteligência Artificial não é magia negra.~~
- 3 ~~Entenda seu problema - O que já fiz de “errado”.~~
- 4 - O inesperado acontece. O que são LLMs e o que é I.A generativa?**
- 5 - Agentic A.I é o futuro?
- 6 - Considerações finais. Alguns exemplos na automação comercial.

Bônus 1 - Sugestões para aprendizado de I.A.

Bônus 2 - Uma breve história da I.A. Marcos importantes. Quem é Quem?

Vídeo Obrigatório



A screenshot of a YouTube video player. On the left is a vibrant, colorful illustration of Earth surrounded by numerous icons and symbols representing various fields of knowledge and technology. To the right of the illustration, the title "The busy person's intro to LLMs" is displayed in a white sans-serif font. Below the title, a small video frame shows Andrej Karpathy from the chest up, wearing a dark t-shirt and looking directly at the camera. The video player interface includes standard controls like play/pause, volume, and a progress bar indicating the video is at 0:00 / 59:47. The overall theme is an introduction to Large Language Models (LLMs) for professionals.

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

[1hr Talk] Intro to Large Language Models



Andrej Karpathy
427 mil inscritos

Inscrito ▾

59 mil



Compartilhar

...



Andrej Karpathy

I like to train deep neural nets on large datasets 🧠🤖💥



2024



coming soon 🎉

2023 - 2024

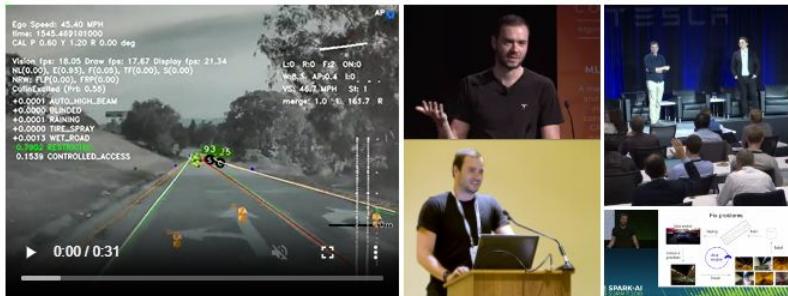


Back to [OpenAI](#). Built a small team, launched a model to ChatGPT, great pleasure to build with the top notch talent within.

2017 - 2022



I was the Sr. Director of AI at Tesla, where I led the computer vision team of [Tesla Autopilot](#). This includes in-house data labeling, neural network training, the science of making it work, and deployment in production running on our custom inference chip. Today, the Autopilot increases the safety and convenience of driving, but the team's goal is to develop and deploy [Full Self-Driving](#) to our rapidly growing fleet of millions of cars. Our Aug 2021 [Tesla AI Day](#) provides the most detailed and up-to-date overview of this effort.



2015 - 2017



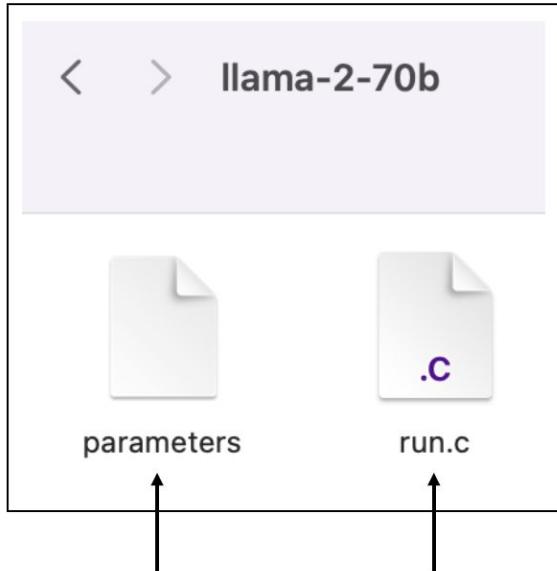
I was a research scientist and a founding member at [OpenAI](#).



https://www.youtube.com/watch?v=zjkBMFhNj_g&t

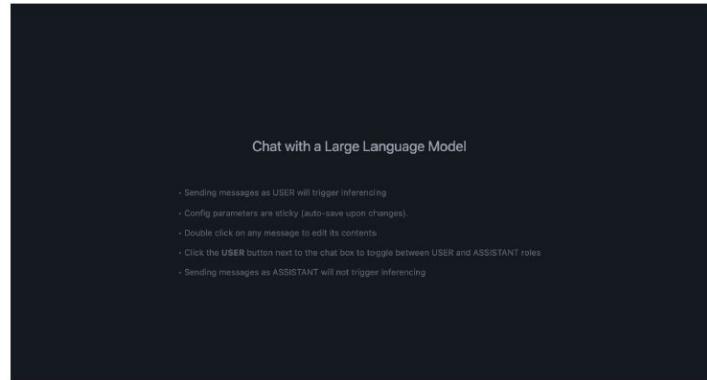
Large Language Model (LLM)

Llama é o LLM da META (Facebook)



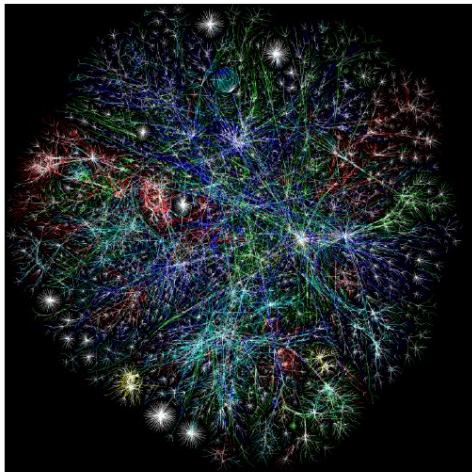
~500 lines
of C code

MacBook



TREINÁ-LOS É MAIS COMPLEXO...

Pense neles como uma “compressão” da Internet



Chunk of the internet,
~10TB of text



6,000 GPUs for 12 days, ~\$2M
~1e24 FLOPS



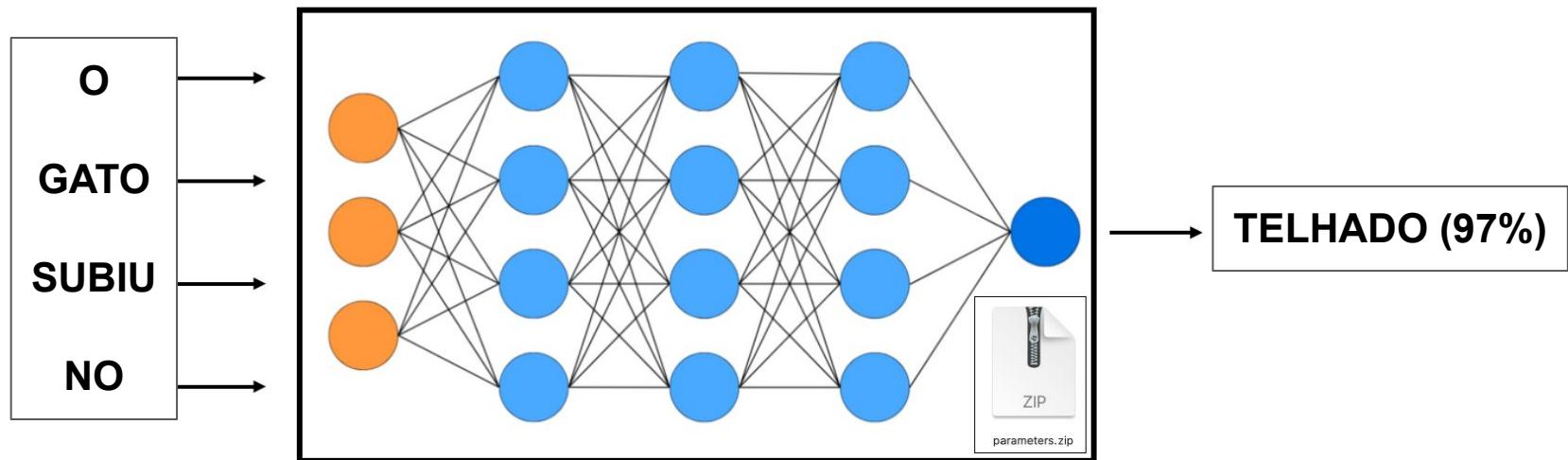
ZIP
parameters.zip

~140GB file

*numbers for Llama 2 70B

Neural Network

Predicts the next word in the sequence.



e.g. context of 4 words

predict next word

Da Wikipédia, a enciclopédia livre

Aprender a próxima palavra força a rede neural a aprender muito sobre o mundo.

Geoffrey Everest Hinton CC FRS FRSC [12] (nascido em 6 de dezembro de 1947) é um cientista da computação e psicólogo cognitivo britânico-canadense, mais conhecido por seu trabalho em redes neurais artificiais. De 2013 a 2023, ele dividiu seu tempo trabalhando para o Google (Google Brain) e para a Universidade de Toronto, antes de anunciar publicamente sua saída do Google em maio de 2023, citando preocupações sobre os riscos da tecnologia de inteligência artificial (IA). [13] Em 2017, ele cofundou e se tornou o principal consultor científico do Vector Institute em Toronto. [14] [15]

Com David Rumelhart e Ronald J. Williams, Hinton foi coautor de um artigo altamente citado publicado em 1986 que popularizou o algoritmo de retropropagação para treinamento de redes neurais multicamadas, [16] embora não tenham sido os primeiros a propor a abordagem. [17] Hinton é visto como uma figura importante na comunidade de aprendizagem profunda. [18] [19] [20] [21] [22] O marco dramático de reconhecimento de imagem do AlexNet

Geoffrey Hinton
CC FRS FRSC



Hinton em 2023

Nascer

Geoffrey Everest Hinton
6 de dezembro de 1947 (76 anos) [11]

M

You

User

Continue de onde esse texto parar:

Geoffrey Everest Hinton CC FRS FRSC [12] (nascido em 6 de dezembro de 1947) é um cientista da computação e psicólogo cognitivo britânico-canadense ,



ChatGPT

Geoffrey Everest Hinton, conhecido principalmente por seu trabalho em redes neurais artificiais, é um dos pioneiros da área de aprendizado profundo. Após sua graduação na Universidade de Cambridge, ele continuou seus estudos e pesquisa, contribuindo significativamente para o desenvolvimento de algoritmos que permitiram avanços em machine learning e inteligência artificial.

Hinton, juntamente com colegas como Yann LeCun e Yoshua Bengio, é frequentemente referido como um dos "padrinhos da inteligência artificial", devido ao seu papel fundamental no desenvolvimento de redes neurais profundas. Suas pesquisas tiveram um impacto considerável em diversos campos, incluindo reconhecimento de fala, visão computacional e processamento de linguagem natural.

MARCO EM MODELOS DE LINGUAGEM

2017 - Paper escrito por equipe do Google Brain

Tem somente 9 páginas

Attention Is All You Need

Ashish Vaswani*
Google Brain
avaswani@google.com

Noam Shazeer*
Google Brain
noam@google.com

Niki Parmar*
Google Research
nikip@google.com

Jakob Uszkoreit*
Google Research
usz@google.com

Llion Jones*
Google Research
llion@google.com

Aidan N. Gomez* †
University of Toronto
aidan@cs.toronto.edu

Łukasz Kaiser*
Google Brain
lukaszkaiser@google.com

Illia Polosukhin* ‡
illia.polosukhin@gmail.com

Objetivos do paper

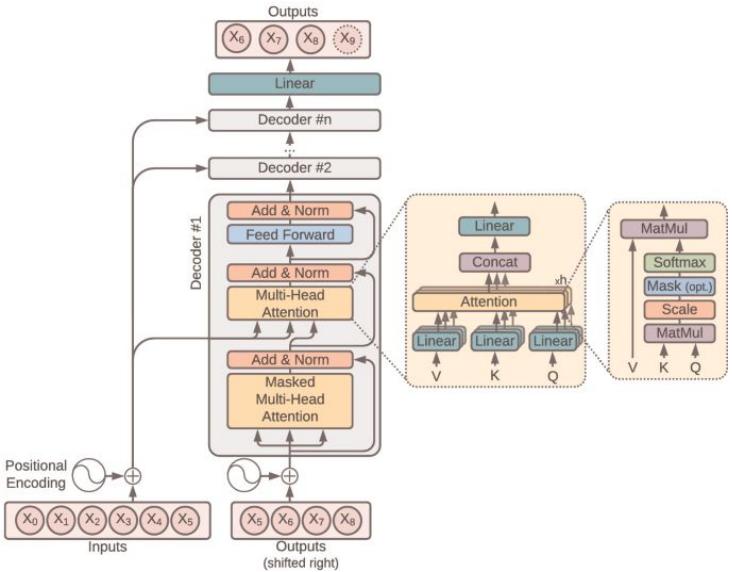
Os modelos dominantes de **tradução de sequências** são baseados em redes neurais recorrentes ou convolucionais complexas que incluem um codificador e um decodificador. Os modelos de melhor desempenho também conectam o codificador e o decodificador por meio de um mecanismo de atenção. **Propomos uma nova arquitetura de rede simples, o Transformer, baseada exclusivamente em mecanismos de atenção**, dispensando totalmente a recorrência e as convoluções. Experimentos em duas tarefas de tradução automática mostram que esses modelos são superiores em qualidade, sendo mais paralelizáveis e exigindo **significativamente menos tempo para treinamento**. Nosso modelo alcança 28,4 BLEU na tarefa de **tradução do inglês para o alemão** do WMT 2014, melhorando os melhores resultados existentes, incluindo conjuntos, em mais de 2 BLEU. Na tarefa de tradução do inglês para o francês do WMT 2014, nosso modelo estabelece um novo recorde de pontuação BLEU para um único modelo de 41,0, **após treinar por 3,5 dias em oito GPUs**, uma pequena fração dos custos de treinamento dos melhores modelos da literatura.

Eis que o inesperado acontece...

- Um modelo criado pelo Google focado em tradução.
- Muito mais simples, rápido (e barato) de treinar.
- Viabilizou o treino para prever a próxima palavra no big data da Internet
- Descobriu-se que podia criar “raciocínios” aparentemente complexos
- O CHAT GPT (da Open AI) popularizou-se rapidamente.
- Criou uma “corrida” pra construir o melhor modelo.

**Nada disso foi planejado.
Nada disso era esperado.**

Como os LLMs funcionam?



- Cada detalhe da arquitetura é conhecido.
- Cada detalhe de como treinar e refinar o conhecimento é conhecido.
- Eles constroem e retém um tipo de base de dados de conhecimento. Mas é um pouco estranho e imperfeito.



M

You

Ana Carolina é a mãe de Paulo Afonso.



ChatGPT

Memory updated

Entendi, Ana Carolina é a mãe de Paulo Afonso. Há algo mais que você gostaria de compartilhar ou perguntar sobre isso?



M

You

João Francisco é filho de Ana Carolina



ChatGPT

Memory updated

Entendi, Ana Carolina tem um filho chamado João Francisco. Posso te ajudar com mais alguma coisa?



M

You

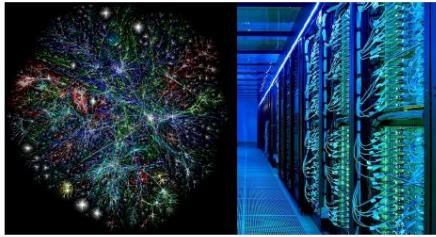
Quem é o irmão de João Francisco?



ChatGPT

Com base no que sabemos até agora, não temos informações específicas sobre quem é o irmão de João Francisco. Seria interessante descobrir mais detalhes sobre a família para responder com precisão.

Como treinar seu próprio CHAT GPT



1x por
ano

Estágio 1: Pré-treino

1. Baixe ~10TB de texto.
2. Obtenha um cluster com ~6000 GPUs
3. “Comprima” o texto em uma rede neural ~R\$10.000.000 e espere 12 dias.
4. Você obteve o modelo base.



<USER>

Can you write a short introduction about the relevance of the term "monopsony" in economics? Please use examples related to potential monopsonies in the labour market and cite relevant research.

<ASSISTANT>

"Monopsony" refers to a market structure where there is only one buyer for a particular good or service. In economics, this term is particularly relevant in the labor market, where a monopsony employer has significant power over the wages and working conditions of their employees. The presence of a monopsony can result in lower wages and reduced employment opportunities for workers, as the employer has little incentive to increase wages or provide better working conditions...

RLHF

1x por
semana

Estágio 2: Sintonia fina (finetuning)

1. Escreva instruções para anotação
2. Contrate pessoas (ou use serviços de anotação como scale.ai), cole umas 100.000 P&R e/ou comparações.
3. “Sintonize” seu modelo base, +/-1 dia de treino.
4. Você obteve o “modelo assistente”.
5. Teste
6. Disponibilize publicamente
7. Monitore, cole desvios, vá para o passo 1

Onde encontro modelos base para customizar?

Finetuning = “sintonia fina”

- Comece com o Llama2 da META. (Tem muitos tutoriais no youtube)
- Consulte a chatbot arena (próximo slide).
<https://huggingface.co/spaces/lmsys/chatbot-arena-leaderboard>
- Huggingface é um local bastante indicado para encontrar modelos de IA em geral.
- Também é interessante para disponibilizar modelos e web services como foi mostrado no caso da NOTA CERTA.
- Busque por LLM finetuning, hoje tem até produtos voltados a isso, momento de ebulação.

LMSYS Chatbot Arena Leaderboard

| Rank | Model | Arena Elo | 95% CI | Votes | Organization | License | Knowledge Cutoff |
|------|------------------------------------|-----------|--------|-------|--------------|-----------------|------------------|
| 1 | Claude 3.0 Opus | 1255 | +3/-4 | 37663 | Anthropic | Proprietary | 2023/8 |
| 1 | GPT-4-1106-preview | 1252 | +3/-3 | 56936 | OpenAI | Proprietary | 2023/4 |
| 1 | GPT-4-0125-preview | 1249 | +3/-4 | 38105 | OpenAI | Proprietary | 2023/12 |
| 4 | Bard (Gemini Pro) | 1204 | +5/-5 | 12468 | Google | Proprietary | Online |
| 4 | Claude 3. Sonnet | 1200 | +3/-4 | 40389 | Anthropic | Proprietary | 2023/8 |
| 6 | GPT-4-0314 | 1185 | +4/-4 | 35803 | OpenAI | Proprietary | 2021/9 |
| 7 | Claude 3. Haiku | 1177 | +3/-4 | 26773 | Anthropic | Proprietary | 2023/8 |
| 8 | GPT-4-0613 | 1160 | +3/-5 | 54509 | OpenAI | Proprietary | 2021/9 |
| 8 | Mistral-Large-2402 | 1157 | +5/-4 | 28356 | Mistral | Proprietary | Unknown |
| 9 | Qwen1.5-72B-Chat | 1149 | +4/-5 | 21981 | Alibaba | Qianwen LICENSE | 2024/2 |

Qwen1.5-72B-Chat

Ana Carolina é a mãe de Paulo Afonso.

Quem é o irmão de João Francisco?

Portanto, Paulo Afonso é filho de Ana Carolina.

João Francisco é filho de Ana Carolina

Então, Ana Carolina é a mãe de João Francisco.

Não tenho informações suficientes para responder quem é o irmão de João Francisco com certeza. Aparentemente, Paulo Afonso e João Francisco têm a mesma mãe, Ana Carolina, mas não se sabe se eles são irmãos, a menos que você forneça mais informações ou confirmar que ambos têm o mesmo pai.

Qwen1.5-72B-Chat

Qwen1.5-72B-Chat is the 72-billion parameter chat model of the Qwen series.

System

You are a helpful assistant.

 Set system prompt and clear history.

Qwen1.5-72B-Chat

Fale sobre a entidade Brasileira AFRAC.

Fale sobre a entidade Brasileira AFRAC.

AFRAC, sigla em português para Associação dos Fabricantes de Roupas e Acessórios do Cachorro, é uma entidade representativa do setor de roupas e acessórios para animais de

AFRAC, sigla em português para Associação dos Fabricantes de Roupas e Acessórios do Cachorro, é uma entidade representativa do setor de roupas e acessórios para animais

- 1. Advocacia:** Defende os interesses dos membros diante das autoridades governamentais, promovendo políticas públicas favoráveis à indústria de roupas e acessórios para cães.
- 2. Eventos:** Organiza ou participa de feiras, seminários e exposições relacionadas ao



Esse sujeito está
tendo
ALUCINAÇÕES
!!!

Cuide bem do seu modelo
para que ele não tenha
"surtos psicóticos".

- Realize sintonias finas regularmente.
- Avalie com rigor, tenha bons métodos de Q.A.
- Colete feedbacks dos usuários (likes, dislikes, ...) para a próxima sintonia.

I.A NA PRÁTICA

com alguns exemplos na automação comercial.



- 1 ~~Use agentes (LLM) pra quase tudo.~~
- 2 ~~Inteligência Artificial não é magia negra.~~
- 3 ~~Entenda seu problema - O que já fiz de “errado”.~~
- 4 ~~O inesperado acontece. O que são LLMs e o que é I.A generativa?~~
- 5 - Agentic A.I é o futuro?**
- 6 - Considerações finais. Alguns exemplos na automação comercial.

Bônus 1 - Sugestões para aprendizado de I.A.

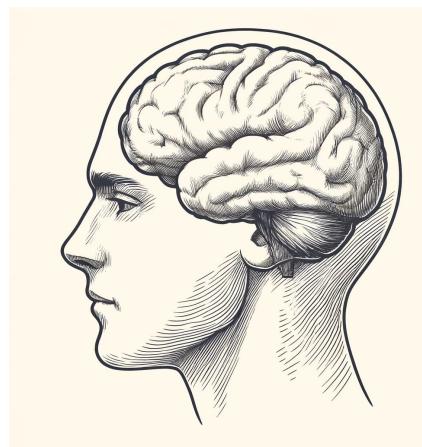
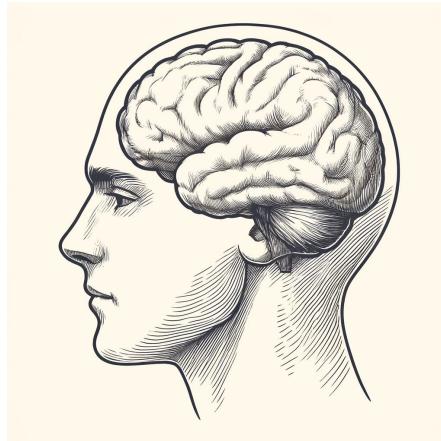
Bônus 2 - Uma breve história da I.A. Marcos importantes. Quem é Quem?

"Uma obra-prima."
Financial Times

RÁPIDO
E DEVAGAR
DUAS FORMAS DE PENSAR



DANIEL
KAHNEMAN
PRÊMIO NOBEL DE ECONOMIA



SISTEMA 1

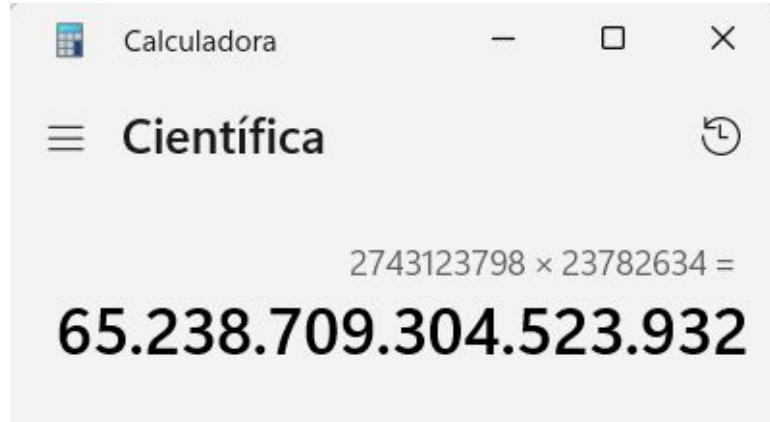
$$2 + 2 = ?$$

Rápido, Emocional,
Instintivo, Automático, Sem
esforço, Inconsciente

SISTEMA 2

$$24 \times 17 = ?$$

Lento, Racional, Decisões
complexas, Muito esforço,
Consciente



— ChatGPT 3.5 ▾

M You
Quanto é $2743123798 * 23782634$

ChatGPT
O resultado da multiplicação entre 2743123798 e 23782634 é
65229158772883252.

Qwen1.5-72B-Chat

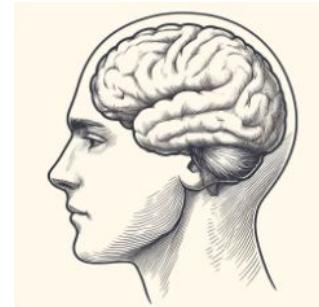
irmãos, a menos que você forneça mais informações ou confirmar que ambos têm o mesmo pai.

Quanto é $2743123798 * 23782634$

Para calcular esse produto, podemos usar uma calculadora ou fazer os cálculos manualmente com several etapas. Aqui está o resultado:

$2743123798 * 23782634 = 6,533,086,313,039,112$

Portanto, o produto é 6,533,086,313,039,112.



SISTEMA 1
 $2 + 2 = ?$

Rápido, Emocional,
Instintivo, Automático, Sem
esforço, Inconsciente



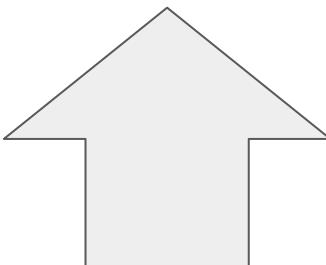
You

Quanto é $2743123798 * 23782634$



ChatGPT

O resultado de $2,743,123,798 * 237,826,34$ é
 $65,238,709,304,523,932$. [–]



You

Quanto é $2 + 2$



ChatGPT

$2 + 2$ é igual a 4.

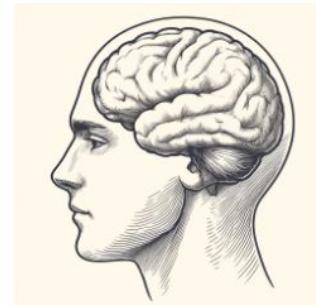


Analysis

python

 $2743123798 * 23782634$

Result

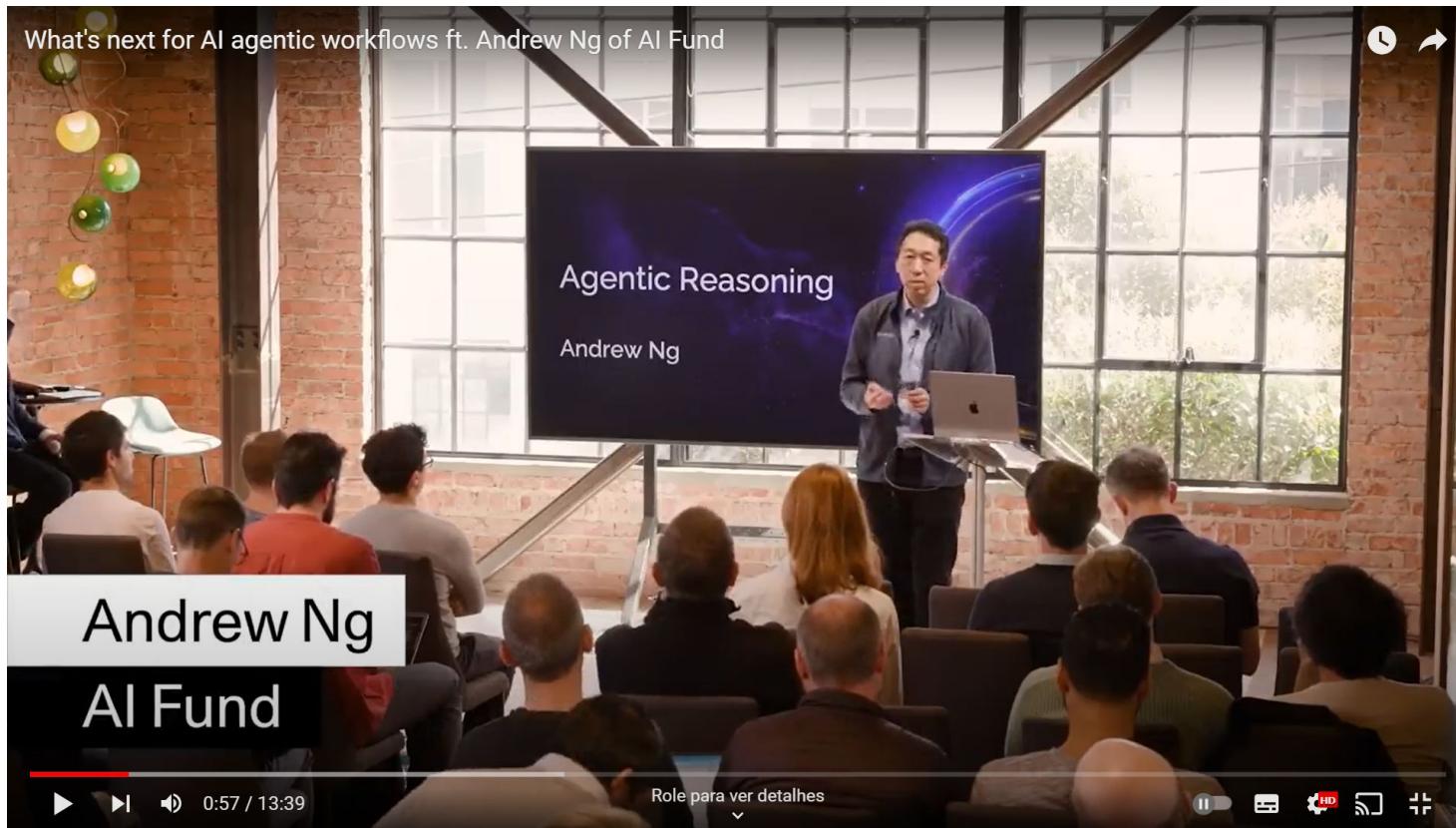
 65238709304523932 

SISTEMA 2
 $24 \times 17 = ?$

Lento, Racional, Decisões complexas, Muito esforço, Consciente

Vídeo Obrigatório

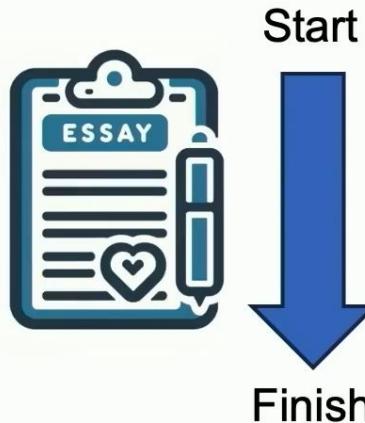
<https://youtu.be/sal78ACtGTc>



LLM-based agents

Non-agentic workflow (zero-shot):

Please type out an essay on topic X from start to finish in one go, without using backspace.



Agentic workflow:

Write an essay outline on topic X

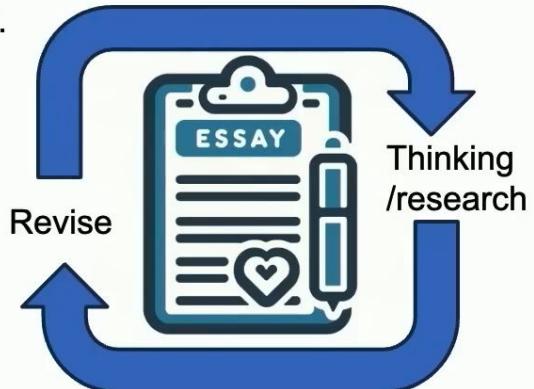
Do you need any web research?

Write a first draft.

Consider what parts need revision or more research.

Revise your draft.

....



! Error analy

Quando pedimos, no início da apresentação, ao chat gpt 4 para elaborar um modelo mais complexo ele claramente usou um workflow mais complexo do que somente uma predição da LLM.

! Error analyzing ▾

????

AGENTIC AI

! Error analyzing ▾

! Error analyzing ▾

! Error analyzing ▾

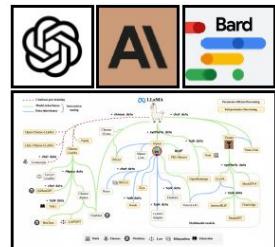
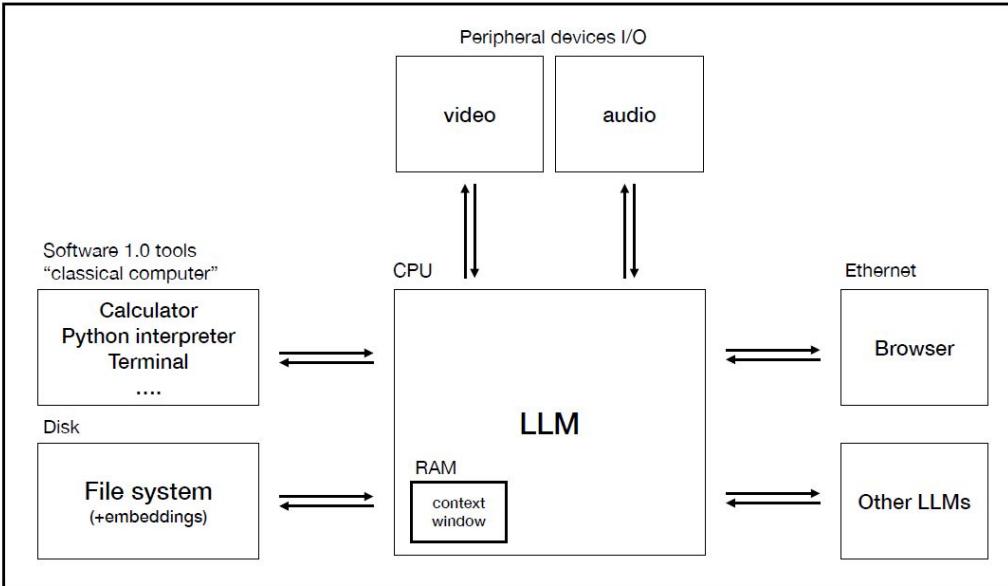
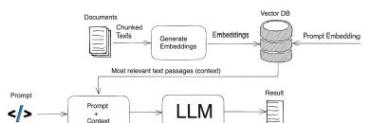
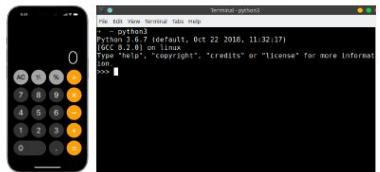
Agentic A.I é o futuro?

- Certamente vai ser muito relevante, pois pode superar as limitações das LLMs e delas emergirem novas surpresas.
- Estudos (ver vídeo do Andre Ng) demonstram que um workflow agentic para geração de código com CHAT GPT 3.5 supera o CHAT GPT 4 sem workflow agentic.
- Procurem vídeos sobre o tema. Existem implementações que usam vários LLMs colaborando entre si para resolver o problema.
- Abstrações como criar uma “Empresa de criação” composta de vários atores (LLMs isolados representando o CEO, o Diretor de Criação, o cliente, o revisor, ...) funciona perfeitamente e gera resultados muitas vezes impressionantes.
- Tem mais aderência ao estado de conhecimento da ciência cognitiva humana. Onde temos partes rápidas (inconscientes e paralelizáveis) e partes lentas (conscientes e seriais) orquestrando tudo.

Thank you!

LLM OS

Thank you!



I.A NA PRÁTICA

com alguns exemplos na automação comercial.

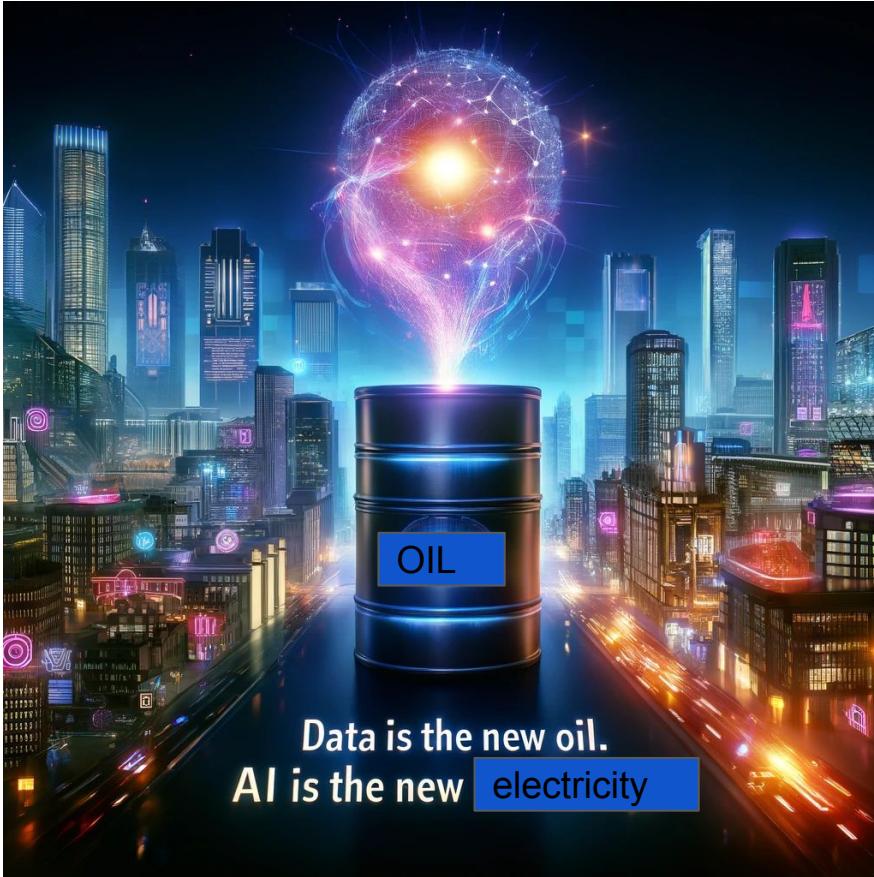


- 1 Use agentes (LLM) pra quase tudo.
- 2 Inteligência Artificial não é magia negra.
- 3 Entenda seu problema - O que já fiz de “errado”.
- 4 O inesperado acontece. O que são LLMs e o que é I.A generativa?
- 5 Agentic A.I é o futuro?
- 6 - Considerações finais. Alguns exemplos na automação comercial.**

Bônus 1 - Sugestões para aprendizado de I.A.

Bônus 2 - Uma breve história da I.A. Marcos importantes. Quem é Quem?

Considerações finais. Alguns exemplos na automação comercial.



- I.A tem uma abrangência cada vez maior, como a eletricidade 100 anos atrás ela pode ser usada e transformará quase tudo.
- Parece magia negra. No começo a eletricidade também parecia.
- Nem sempre o modelo mais complexo é o mais adequado.
- A computação especialmente nos smartphones tem viabilizado cada vez mais rodar as coisas localmente.
- Em 2012 o senso comum era de que redes neurais eram uma tecnologia limitada e de nicho. É nova e é relativamente simples. Domine-a

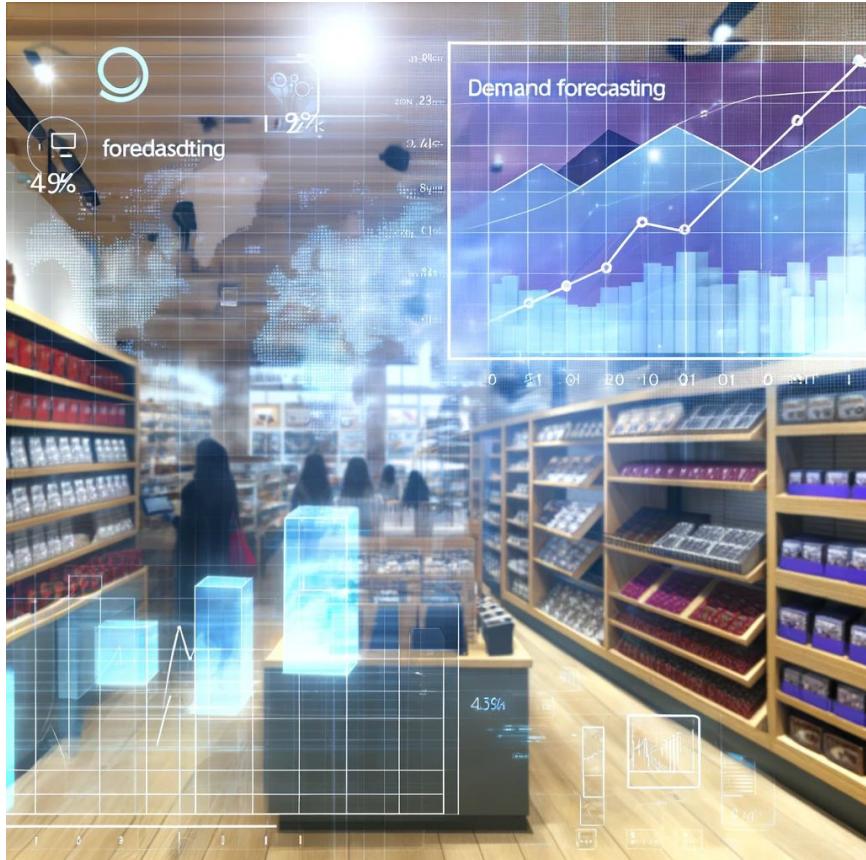
E os exemplos de I.A na automação comercial?

Pense no seu PDV, no seu software, na sua balança sem eletricidade ou sem bateria.

Em breve, essa será a automação comercial sem I.A.



Alguns exemplos



Fizemos um exemplo lúdico durante a apresentação.

Algoritmos similares ao random forest como catboost e xgboost costumam resolver bem esses cenários.

Criar um data lake e daí gerar uma boa tabulação com dados enriquecidos (ex: tempo, cotação do dólar) costuma fazer a diferença.

E que tal usar uma LLM para ler as notícias e alimentar suas tabelas?

Alguns exemplos

The screenshot shows the AFRAC Nota Certa app interface. At the top, there is a header with the AFRAC logo and a search bar containing the text "mandioca kg". Below the search bar, a blue callout box displays the NCM code "07141000" and a probability of "probabilidade 11%". The detailed description in the box reads: "Produtos hortícolas, plantas, raízes e tubérculos, comestíveis. Raízes de mandioca, de araruta e de salepo, tupinambás, batatas-doces e raízes ou tubérculos semelhantes, com elevado teor de fécula ou de inulina, frescos, refrigerados, congelados ou secos, mesmo cortados em pedaços ou em pellets; medula de sagueiro. - Raízes de mandioca". At the bottom of the screen, there is a navigation bar with icons for Início, NotaCerta (highlighted in red), AFRAC, Notificações, and Mais.

Manutenção de cadastro de produtos.
Verificação de compliance fiscal.
Agentes (I.A) especialista na legislação tributária.
Não seria o caso do governo criar esses agentes?

...

Alguns exemplos

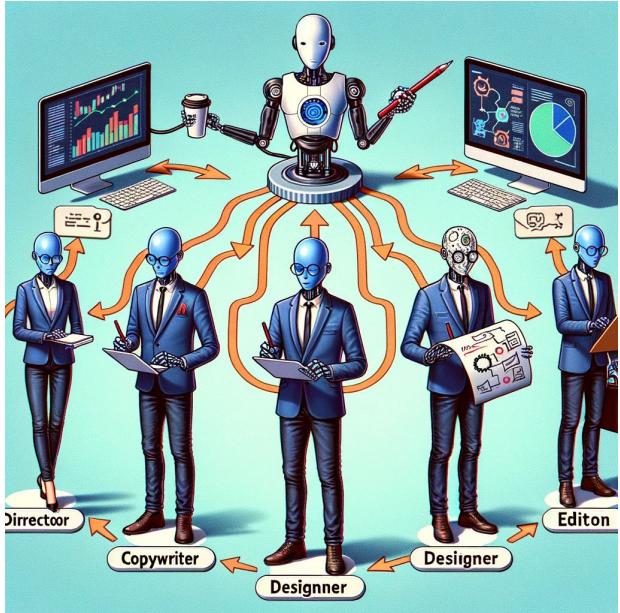


Visão computacional tem vários modelos maduros.

Contagem de objetos com Yolo V8/V9.
Reconhecimento facial e biométrico.
Segurança contra furtos.
Lojas autônomas (+/-)
Detecção de emoções
Visão de robôs e veículos em operações logísticas

...

Alguns exemplos



Comece a explorar e observar a evolução da “Agentic A.I”

De forma ilustrativa, é concebível criar um workflow com agentes especializados em marketing, design, revisão de conteúdo e criar um fluxo interativo entre esses agentes para criar por exemplo um plano de promoção de produtos.

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.



A cross-sectional diagram of Lake Michigan. The lake is depicted as a wide, shallow basin on the left and a deep, narrow channel on the right. The water is light blue, and the surrounding land is brown. On the far left, there's a small silhouette of a city skyline. On the far right, there's a small patch of green grass at the water's edge.

1940

1

calcs/second

Mother Jones

[Home](#) > Inteligência Artificial

Você não será substituído por IA, mas por quem usa bem IA

Inteligência artificial não é sobre o ostracismo humano, mas sobre ganhos de produtividade para aqueles que dominam a tecnologia

É verdade, mas quero encerrar de forma mais positiva.

- A Inteligência Artificial vai nos libertar ainda mais das tarefas mais simples.
- Como toda tecnologia poderosa precisa de atenção e cuidado.
- Está trazendo acesso à educação, saúde, segurança e até mesmo à justiça a quem hoje não tem nenhuma e trará mudanças muito positivas à sociedade.



I.A NA PRÁTICA

com alguns exemplos na automação comercial.



OBRIGADO!!!

Marcelo Filipak

mfilipak@hotmail.com

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

<https://github.com/mfilipak/AUTOCOM2024>

I.A NA PRÁTICA

com alguns exemplos na automação comercial.



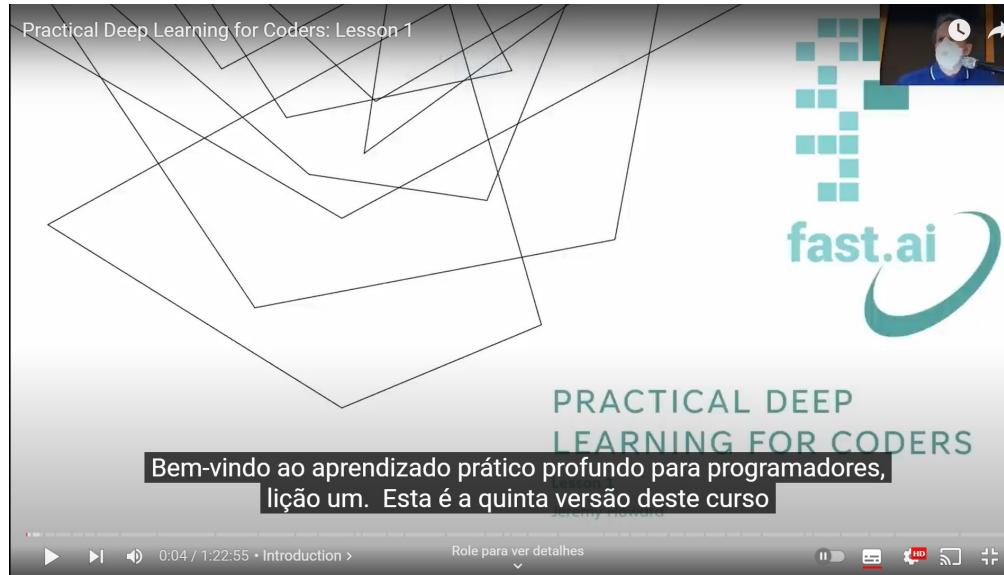
- 1 Use agentes (LLM) pra quase tudo.
- 2 Inteligência Artificial não é magia negra.
- 3 Entenda seu problema - O que já fiz de “errado”.
- 4 O inesperado acontece. O que são LLMs e o que é I.A generativa?
- 5 Agentic A.I é o futuro?
- 6 Considerações finais. Alguns exemplos na automação comercial.

Bônus 1 - Sugestões para aprendizado de I.A.

Bônus 2 - Uma breve história da I.A. Marcos importantes. Quem é Quem?

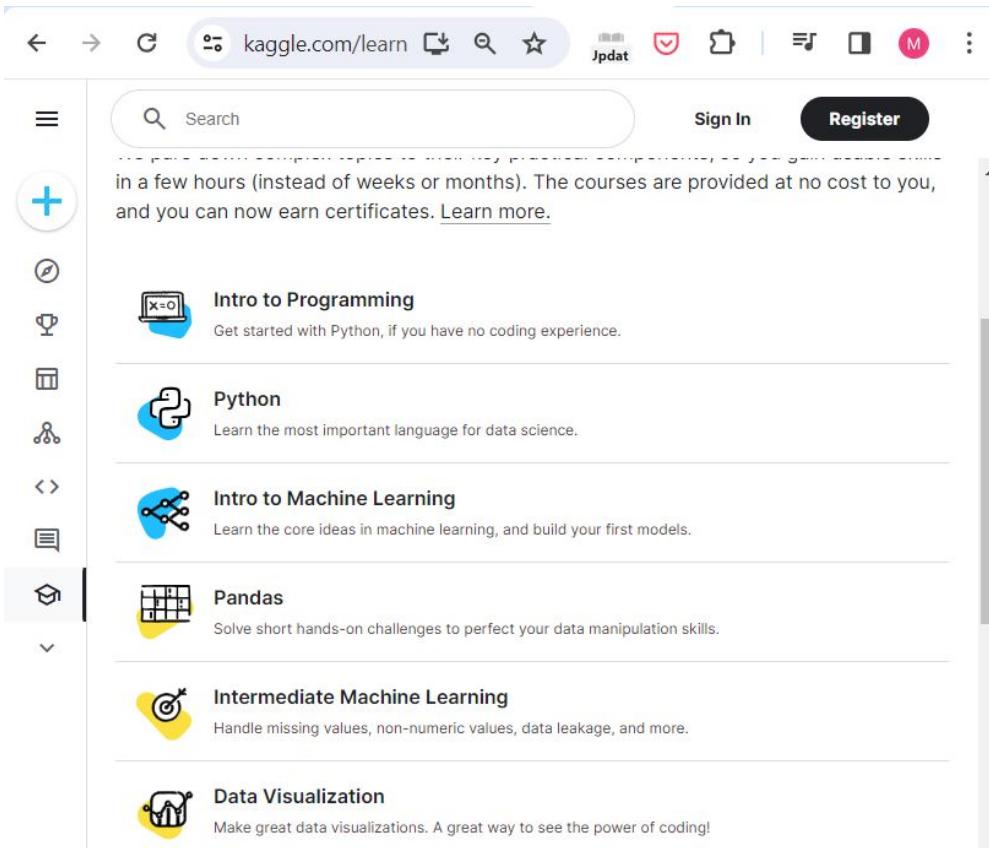
Bônus 1 - Sugestões para aprendizado de I.A.

https://www.youtube.com/watch?v=8SF_h3xF3cE



Fast.ai é um ótimo ponto de partida, especialmente para um primeiro contato.

Bônus 1 - Sugestões para aprendizado de I.A.



A screenshot of the Kaggle Learn homepage. The URL in the address bar is [kaggle.com/learn](https://www.kaggle.com/learn). The page features a sidebar with various icons for navigation. The main content area displays several free courses:

- Intro to Programming**: Get started with Python, if you have no coding experience.
- Python**: Learn the most important language for data science.
- Intro to Machine Learning**: Learn the core ideas in machine learning, and build your first models.
- Pandas**: Solve short hands-on challenges to perfect your data manipulation skills.
- Intermediate Machine Learning**: Handle missing values, non-numeric values, data leakage, and more.
- Data Visualization**: Make great data visualizations. A great way to see the power of coding!

<https://www.kaggle.com/learn>

Bons cursos gratuitos.

Bônus 1 - Sugestões para aprendizado de I.A.

PARTICIPE OU ACOMPANHE ALGUM PROJETO COLABORATIVO
GRUPO DE TRABALHO DE IOT & AI DA AFRAC

https://github.com/mfilipak/AFRAC_IA

The screenshot displays the GitHub repository page for `mfilipak/AFRAC_IA`. The repository is public and has one branch, `main`, with no tags. The commit history shows several contributions from `mfilipak`, including a merge from the `main` branch and various commits related to datasets, models, and EDA notebooks. The files listed include `DATASET`, `MODELS`, `templates`, `.gitignore`, and several Jupyter notebooks like `001_EDA_(Dados_do_Portal_da_cidadânia.ipynb)`, `002_EMBEDDINGS1_Portal_da_Transparéncia.ipynb`, `003_Visualizações_Snippets_usando_plotly.ipynb`, and `004_Visualizações_usando_modelo_pré_treinado.ipynb`.

| File | Description | Last Commit |
|---|--|-------------|
| <code>DATASET</code> | Bugfix - Tabela descrição ncm | 2 years ago |
| <code>MODELS</code> | Modelo treinado com o notebook 005B. Predição do capítulo... | 2 years ago |
| <code>templates</code> | Create index.html | 3 years ago |
| <code>.gitignore</code> | Geração de tabela para os códigos NCM | 2 years ago |
| <code>001_EDA_(Dados_do_Portal_da_cidadânia.ipynb)</code> | EDA - Transparéncia | 2 years ago |
| <code>002_EMBEDDINGS1_Portal_da_Transparéncia.ipynb</code> | Quase lá | 2 years ago |
| <code>003_Visualizações_Snippets_usando_plotly.ipynb</code> | Created using Colaboratory | 2 years ago |
| <code>004_Visualizações_usando_modelo_pré_treinado.ipynb</code> | Visualizações com redução de dimensão | 2 years ago |

 Introductory

**Machine Learning
Specialization**

 3 Courses

 Introductory

 Andrew Ng, Eddy Shyu, Aarti Bagul, Geoff Ladwig

 Stanford Online



 Intermediate

**Deep Learning
Specialization**

 5 Courses

 Intermediate

 Andrew Ng, Kian Katanforoosh, Younes Bensouda Mourri



**Generative AI with
LLMs**

 1 Course

 Intermediate

 Antje Barth, Chris Fregly, Shelbee Eigenbrode, Mike Chambers



<https://www.youtube.com/@AndrejKarpathy>

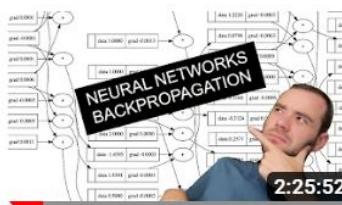


[1hr Talk] Intro to Large Language Models

Andrej Karpathy • 1,7 mi de visualizações • há 4 meses

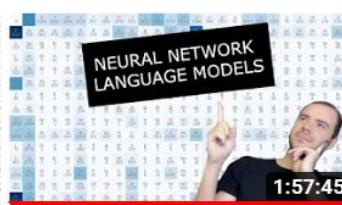
This is a 1 hour general-audience introduction to Large Language Models: the core technical component behind systems like ChatGPT, Claude, and Bard. What they are, where they are headed, comparison...

Neural Networks: Zero to Hero ► Reproduzir tudo



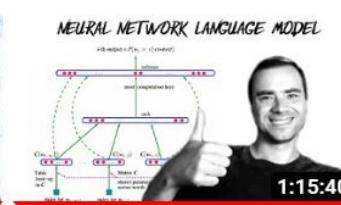
The spelled-out intro to
neural networks and...

Andrej Karpathy
1,4 mi de visualizações • há 1 ano



The spelled-out intro to
language modeling: buildin...

Andrej Karpathy
556 mil visualizações • há 1 ano



Building makemore Part 2:
MLP

Andrej Karpathy
260 mil visualizações • há 1 ano



Building makemore Part 3:
Activations & Gradients,...

Andrej Karpathy
230 mil visualizações • há 1 ano



Building makemore Part 4:
Becoming a Backprop Ninja

Andrej Karpathy
160 mil visualizações • há 1 ano

Meu conteúdo

<https://github.com/mfilipak/FILIPAK-DATA-SCIENCE>

 **FILIPAK-DATA-SCIENCE** Public

 main  1 Branch  0 Tags  Go to file  Add file  Code

|  mfilipak | Add files via upload | eeace20 · last week |  65 Commits |
|--|----------------------|---------------------------------|--|
|  | DATA | Delete DATA/nothing.txt | last week |
|  | MODELS | Add files via upload | 3 weeks ago |
|  | DIA1.ipynb | Created using Colaboratory | 2 months ago |
|  | DIA2.ipynb | Created using Colaboratory | 2 months ago |
|  | DIA3.ipynb | Created using Colaboratory | 2 months ago |
|  | DIA4.ipynb | Rename DAY4.ipynb to DIA4.ipynb | last month |

Meu conteúdo

<https://github.com/mfilipak/ACBR-2023>



The AFRAC logo is displayed on a white rectangular background. The logo features a blue stylized 'A' icon followed by the acronym 'AFRAC' in bold blue capital letters. Below the acronym, the full name 'Associação Brasileira de Tecnologia para o Comércio e Serviços' is written in a smaller, sans-serif font.



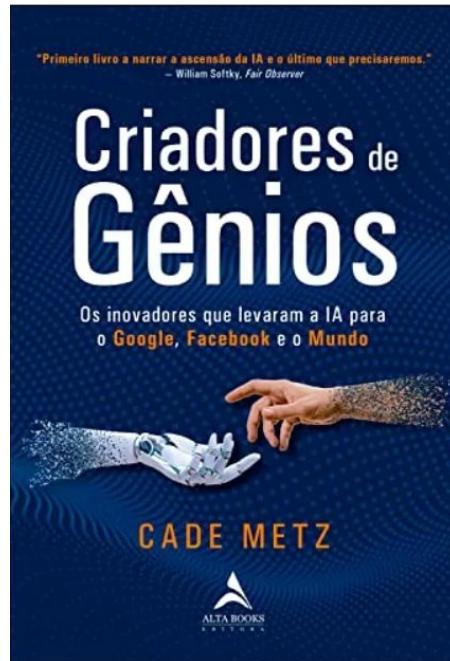
A photograph of a man's face is overlaid with a green wireframe grid. The grid highlights specific facial features, such as the eyes, nose, and mouth, with colored dots (green for the eyes, red for the nose, and blue for the mouth) indicating detection points for a facial recognition system.

IA NA PRÁTICA

COMO CRIAR OU ENCONTRAR MODELOS E INTEGRÁ-LOS
À SUA APLICAÇÃO.
CASE AFRAC-SEFA-PR - API NOTA CERTA

Dia do ACBr
2 0 2 3

Bônus 2 - Uma breve história da I.A. Marcos importantes. Quem é Quem?



Bônus 2 - Uma breve história da I.A. Marcos importantes. Quem é Quem?

1. **1957 - Perceptron**: Modelo de rede neural artificial desenvolvido por Frank Rosenblatt.
2. **1960s - Gradiente Descendente**: Método de otimização usado para o treinamento de redes neurais.
3. **1970s - Backpropagation**: Algoritmo para o treinamento de redes neurais multicamadas.
4. **Anos 1970 e 1990 - Inverno da IA**: Períodos de redução no interesse e financiamento para IA.
5. **1980 - Redes Convolucionais**: Especializadas para processar dados como imagens.
6. **2009 - ImageNet**: Dataset de imagens para treinar sistemas de visão computacional.
7. **2012 - AlexNet**: Rede convolucional que foi significativa no reconhecimento de imagens.
8. **2017 - Transformers**: Modelo baseado em atenção para processamento de linguagem natural.
9. **Anos 2010s - IA Generativa**: Técnicas como GANs e VAEs para gerar dados sintéticos.
10. **2020s - Stable Diffusion**: Modelo generativo que cria imagens a partir de descrições textuais.
11. **2020s - Modelos de Linguagem de Grande Escala (LLMs)**: Como o GPT-3, avançados na compreensão e geração de texto.

EM CONSTRUÇÃO

