






# CODA·BR



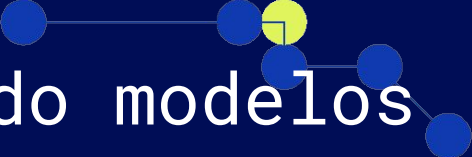

↑ ESCOLA DE DADOS  
→ x



**Texto, áudio e imagens:** usando modelos  
de IA gratuitos e de código-aberto com  
dados não-estruturados

**ADRIANO BELISARIO** • [adriano@belisario.website](mailto:adriano@belisario.website) • [belisario.website](http://belisario.website)





**Texto, áudio e imagens:** usando modelos  
de IA gratuitos e de código-aberto com  
dados não-estruturados

[bit.ly/ia-coda-24](https://bit.ly/ia-coda-24)



## NESTE WORKSHOP, VAMOS VER...

- Introdução a modelos de IA de código-aberto;
- Trabalhando com textos (para além do ChatGPT);
- Classificação e outras tarefas com imagens;
- Soluções para fazer transcrições de áudio;

# IA de código-aberto

---

Dicas para começar a implementar soluções

grandes  
proprietários  
genéricos  
na "nuvem"  
tamanho "único"

---

~~grandes~~  
~~proprietários~~  
~~genéricos~~  
~~na "nuvem"~~  
~~tamanho único~~

---

**pequenos**  
**abertos**  
**específicos**  
**auto-hospedados**  
**customizados**

# IA DE CÓDIGO ABERTO

## De programas a modelos: re-enquadramento da definição de [código aberto](#)

### What is Open Source AI

When we refer to a “system,” we are speaking both broadly about a fully functional structure and its discrete structural elements. To be considered Open Source, the requirements are the same, whether applied to a **system**, a **model**, **weights and parameters**, or other structural elements.

An *Open Source AI* is an AI system made available under terms and in a way that grant the freedoms<sup>1</sup> to:

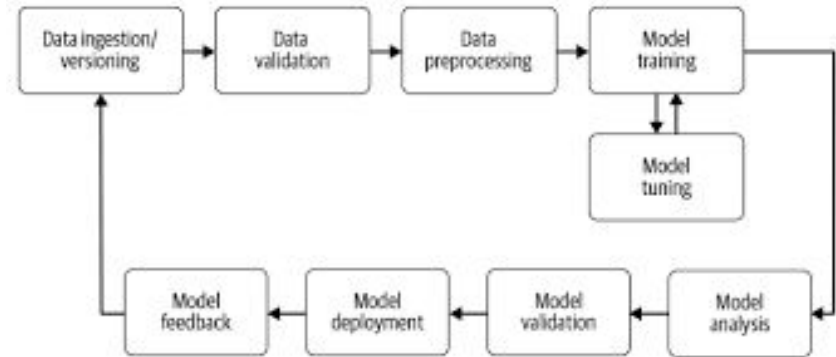
- **Use** the system for any purpose and without having to ask for permission.
- **Study** how the system works and inspect its components.
- **Modify** the system for any purpose, including to change its output.
- **Share** the system for others to use with or without modifications, for any purpose.



# IA DE CÓDIGO ABERTO

A barreira de entrada diminuiu, mas implementar modelos de IA abertos ainda demanda conhecimento de programação;

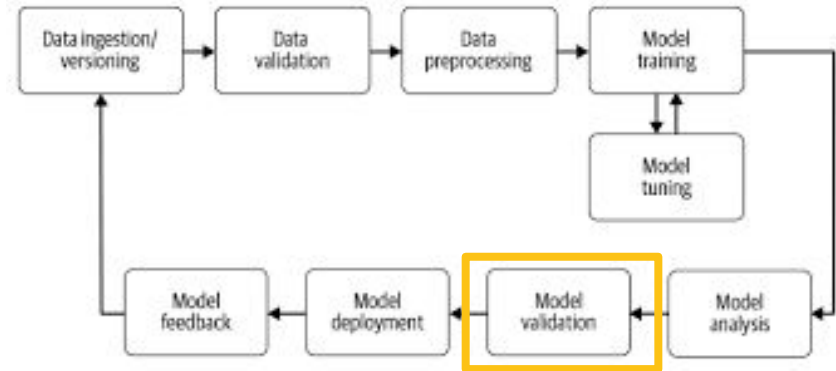
GPUs: notebooks com Python no Google Colab permitem prototipar soluções rápida e gratuitamente;



# AValiação DE MODELOS

Conheça o corpus de treinamento;

Defina uma linha base de avaliação (trabalhos prévios, modelos dummy, etc);



# AVALIAÇÃO DE MODELOS

Use estatísticas de concordâncias, como coeficiente kappa, para avaliar o grau de concordância de anotações coletivas;

Garanta a qualidade dos dados usados para avaliação do modelo;

# HUGGINGFACE

Porta de entrada para desenvolver soluções customizadas.

The image shows a screenshot of the Hugging Face model hub interface. On the left, there are two main categories: 'Multimodal' and 'Computer Vision'. Under 'Multimodal', there are buttons for 'Audio-Text-to-Text', 'Image-Text-to-Text', 'Visual Question Answering', 'Document Question Answering', 'Video-Text-to-Text', and 'Any-to-Any'. Under 'Computer Vision', there are buttons for 'Depth Estimation', 'Image Classification', 'Object Detection', 'Image Segmentation', 'Text-to-Image', 'Image-to-Text', 'Image-to-Image', 'Image-to-Video', 'Unconditional Image Generation', 'Video Classification', 'Text-to-Video', 'Zero-Shot Image Classification', 'Mask Generation', 'Zero-Shot Object Detection', 'Text-to-3D', and 'Image-to-3D'. On the right, there is a list of model cards. The first card is 'Djrrango/Qwen2v1-Flux' (Text-to-Image, Updated 5 days ago, 313 likes). The second is 'AIDC-AI/Marco-o1' (Text Generation, Updated 9 days ago, 8.33k downloads, 586 likes). The third is 'Lightricks/LTX-Video' (Image-to-Video, Updated 9 days ago, 30.5k downloads, 536 likes). The fourth is 'OuteAI/OuteTTS-0.2-500M' (Text-to-Speech, Updated about 1 hour ago, 9.58k downloads, 210 likes). The fifth is 'HuggingFaceTB/SmolVLM-Instruct' (Image-Text-to-Text, Updated about 16 hours ago, 18.9k downloads, 193 likes). The sixth is 'black-forest-labs/FLUX.1-dev' (Text-to-Image, Updated Aug 16, 1.37M downloads, 6.86k likes). The seventh is 'ginipnick/flux-lora-eric-cat'.

Multimodal

- Audio-Text-to-Text
- Image-Text-to-Text
- Visual Question Answering
- Document Question Answering
- Video-Text-to-Text
- Any-to-Any

Computer Vision

- Depth Estimation
- Image Classification
- Object Detection
- Image Segmentation
- Text-to-Image
- Image-to-Text
- Image-to-Image
- Image-to-Video
- Unconditional Image Generation
- Video Classification
- Text-to-Video
- Zero-Shot Image Classification
- Mask Generation
- Zero-Shot Object Detection
- Text-to-3D
- Image-to-3D
- Image Feature Extraction
- Keypoint Detection

**Djrrango/Qwen2v1-Flux**  
Text-to-Image • Updated 5 days ago • ❤️ 313

**AIDC-AI/Marco-o1**  
Text Generation • Updated 9 days ago • 📄 8.33k • ❤️ 586

**Lightricks/LTX-Video**  
Image-to-Video • Updated 9 days ago • 📄 30.5k • ❤️ 536

**OuteAI/OuteTTS-0.2-500M**  
Text-to-Speech • Updated about 1 hour ago • 📄 9.58k • ❤️ 210

**HuggingFaceTB/SmolVLM-Instruct**  
Image-Text-to-Text • Updated about 16 hours ago • 📄 18.9k • ❤️ 193

**black-forest-labs/FLUX.1-dev**  
Text-to-Image • Updated Aug 16 • 📄 1.37M • ⚡ • ❤️ 6.86k

**ginipnick/flux-lora-eric-cat**



## MODOS DE USAR

**Inferência zero-shot:** uso de um modelo generativo pré-treinado sem o provimento de exemplos;

**Few-shot:** uso de um modelo generativo pré-treinado provendo alguns exemplos "anotados";

## MODOS DE USAR

**Ajuste fino (fine-tuning):** altera os pesos do modelo com base em um conjunto razoável de dados anotados;

**RAGs:** usa uma coleção de documentos para "embasar" as respostas do modelo;

# TRABALHANDO COM TEXTO

---

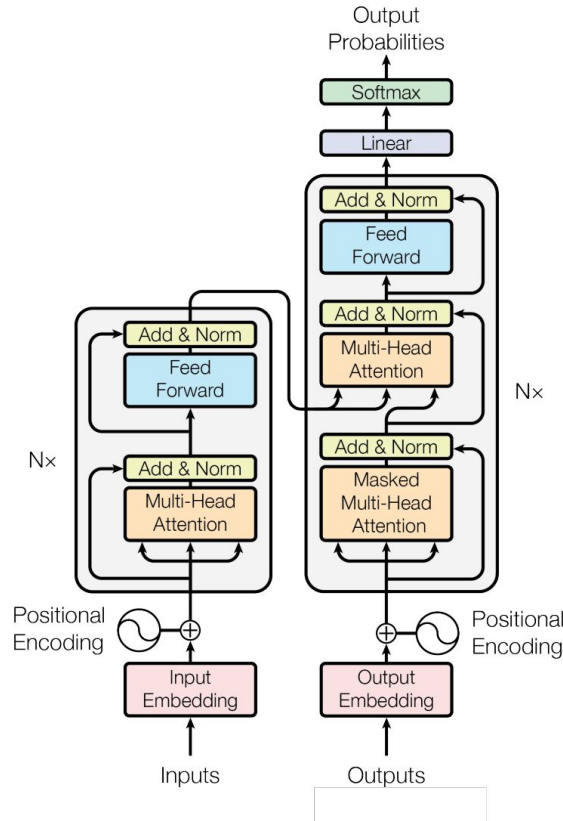
Muito além do ChatGPT



# MODELOS TRANSFORMER

## BERT

Encoder



## GPT

Decoder

# BERT

Não é um modelo generativo;

Faz tarefas como classificação de texto e extração de entidades nomeadas;

São leves e exigem poucos recursos;

Ideal para tarefas que exigem entendimento da linguagem;

# BERT

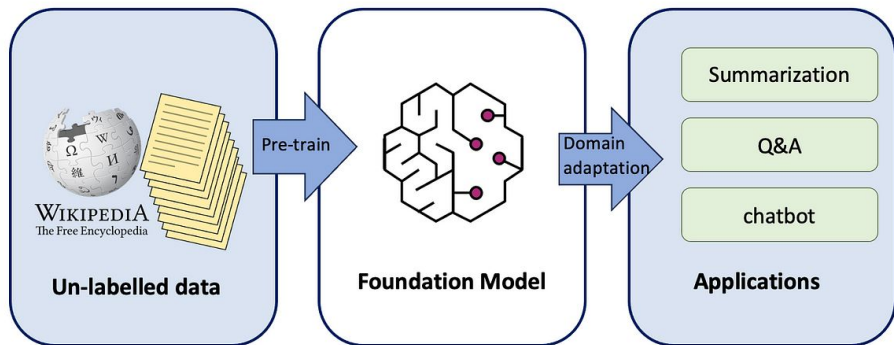


# GPT

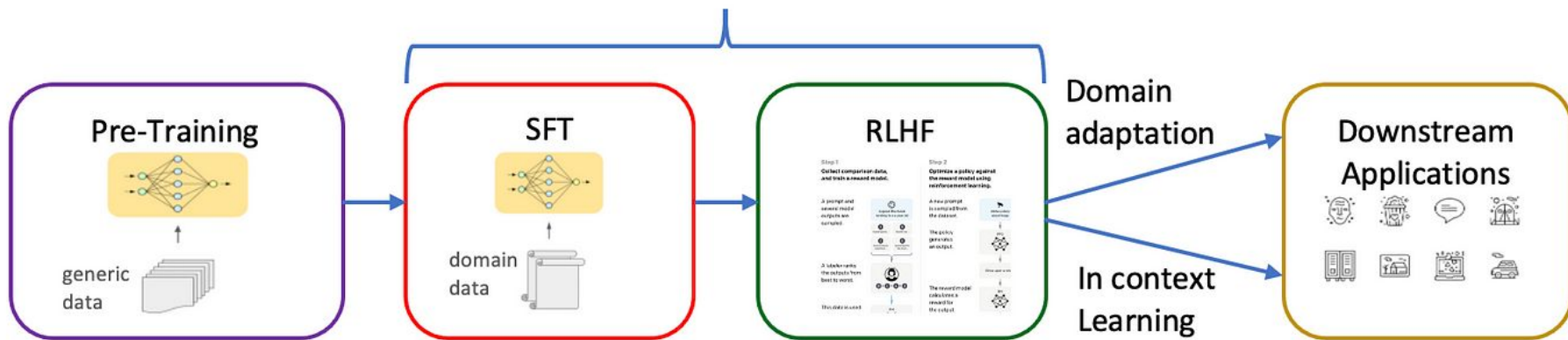
Modelo para geração de texto com "capacidades emergentes" quando ganha escala;

Generalista e multi-propósito, mas pesado e custoso;

# GPT



## Alignment



## EM RESUMO

BERT	GPT
Entendimento da linguagem	Geração de textos
Pequenos ou médios	Grandes
Bidirecional	Autoregressivo

# **CLASSIFICAÇÃO DE PROJETOS DE LEI**

# PRÉ-TREINO CONTINUADO

The screenshot shows the Hugging Face interface for the model 'congretribau' by user 'belisards'. The top navigation bar includes the Hugging Face logo, a search bar, and links for Models, Datasets, Spaces, Posts, Docs, Enterprise, and Pricing. The model card header features the user name 'belisards', the model name 'congretribau', and a 'like' button with a count of 0. Below the header are tags for 'Fill-Mask', 'Transformers', 'TensorBoard', 'Safetensors', 'belisards/ementas\_senado\_1946\_2024', 'belisards/ementas\_camarabr\_1934\_2024', 'Portuguese', and 'bert'. A 'Generated from Trainer' badge and 'Inference Endpoints' link are also present. The main navigation tabs include 'Model card', 'Files and versions', 'Training metrics', 'Community', and 'Settings'. The 'Model card' tab is active, showing a description of the model as a fine-tuned version of BERTimbau on Brazilian law proposals. It lists evaluation metrics such as eval\_loss, eval\_runtime, eval\_samples\_per\_second, eval\_steps\_per\_second, epoch, and step. A 'Downloads last month' chart shows 43 downloads. The 'Safetensors' section indicates a model size of 334M params and a tensor type of F32. The 'Inference Examples' section lists 'Fill-Mask' and notes that the model does not have enough activity for deployment. The 'Model tree' section shows the model's lineage from 'neuralmind/bert-base-portuguese-cased' through 99 finetunes to 'this model'.

**congretribau**

This model is a fine-tuned version of [BERTimbau](#) on a dataset with bills of Brazilian law proposals. It achieves the following results on the evaluation set:

- eval\_loss: 0.4885
- eval\_runtime: 798.5704
- eval\_samples\_per\_second: 169.279
- eval\_steps\_per\_second: 1.324
- epoch: 2.3669
- step: 10000

**Training and evaluation data**

Data from the Chamber of Deputies and the Federal Senate.

Downloads last month: **43**

**Safetensors** Model size: **334M params** Tensor type: **F32**

**Inference Examples**

Fill-Mask

This model does not have enough activity to be deployed to Inference API (serverless) yet. Increase its social visibility and check back later, or deploy to Inference Endpoints (dedicated) instead.

**Model tree for belisards/congretribau**

Base model: neuralmind/bert-base-portuguese-cased

Finetuned (99): **this model**

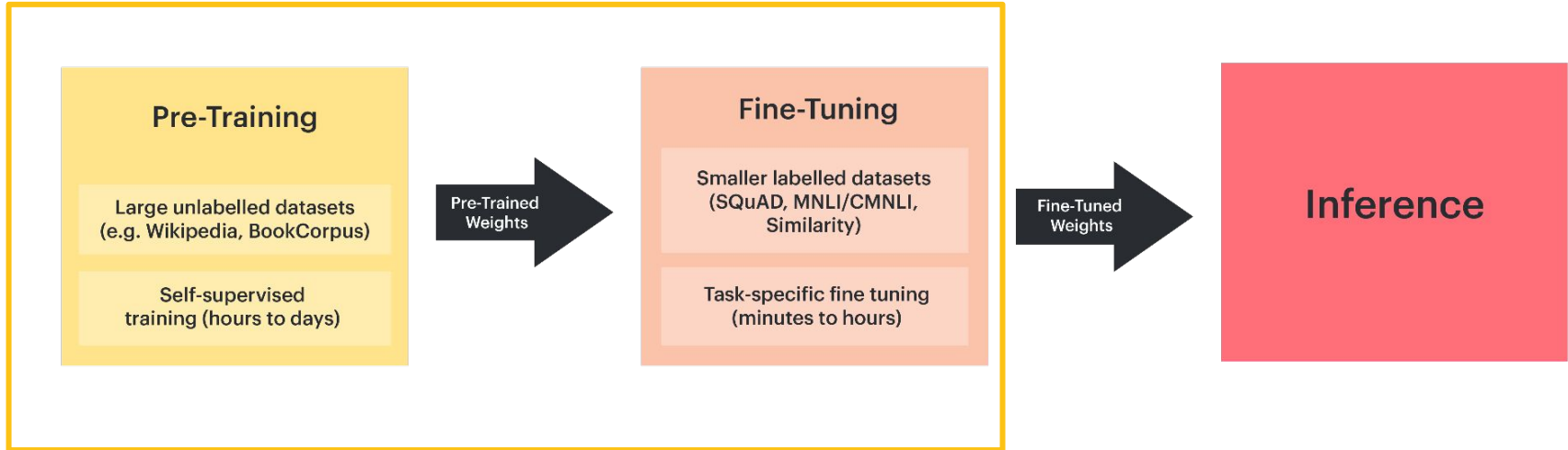
Finetunes: 2 models

Congretribau

<https://huggingface.co/belisards/congretribau>



# PRÉ-TREINO CONTINUADO



Congretimbau

<https://huggingface.co/belisards/congretimbau>

# PRÉ-TREINO CONTINUADO

The screenshot shows the Hugging Face interface for the model 'ia-feminista-bert-tema'. At the top, it displays the user 'azmina', a like button (1), and a 'Following' button for 'Instituto AzMina' (3). Below this are tags for 'Text Classification', 'Transformers', 'Safetensors', 'azmina/ementas\_congresso', 'Portuguese', 'bert', 'congresso', 'direito', 'politica', and 'Inference Endpoints'. The license is listed as 'mit'. Navigation options include 'Model card', 'Files and versions', 'Community', and 'Settings'. Action buttons for 'Train', 'Deploy', and 'Use this model' are visible. The main content area is titled 'IA Feminista' and contains a description of the model's purpose: thematic classification of law projects by gender. It details the development process, mentioning fine-tuning of the 'BERTimbau' model. Performance metrics are listed: Loss: 0.5892, Accuracy: 0.8155, F1: 0.7939, and Recall: 0.7935. On the right sidebar, there is a 'Downloads last month' section showing 85 downloads with a line graph and a 'View full history' link. Below that, 'Safetensors' information is shown: Model size 334M params, Tensor type F32. The 'Inference Examples' section notes that the model is not yet active for deployment. The 'Model tree' section shows the model's lineage: 'neuralmind/bert-base-portuguese-cased' as the base model, 'belisards/congretimbau' as a finetuned model, and 'this model' as a second finetuned model.

azmina **ia-feminista-bert-tema** like 1 Following Instituto AzMina 3

Text Classification Transformers Safetensors azmina/ementas\_congresso Portuguese bert congresso direito politica Inference Endpoints

License: mit

Model card Files and versions Community Settings

Train Deploy Use this model

Edit model card

## IA Feminista

Este modelo faz classificação temática de projetos de lei relacionados a gênero a partir da ementa disponibilizada pela Câmara e o Senado brasileiro.

Ele foi desenvolvido a partir do ajuste fino (fine-tuning para classificação) do Congretimbau, que por sua vez é uma versão do BERTimbau adaptada (com técnicas de pré-treino continuado via modelagem de linguagem mascarada) para ementas de projetos de lei. O modelo deve ser usado em conjunto com o tokenizador do BERTimbau, conforme o exemplo abaixo.

O modelo atinge as seguintes métricas no conjunto de dados de teste:

- Perda(loss): 0.5892
- Acurácia: 0.8155
- F1: 0.7939
- Recall: 0.7935

Downloads last month  
**85**  
NEW View full history

Safetensors Model size 334M params Tensor type F32

### Inference Examples

Text Classification

This model does not have enough activity to be deployed to Inference API (serverless) yet. Increase its social visibility and check back later, or deploy to Inference Endpoints (dedicated) instead.

### Model tree for azmina/ia-feminista-bert-tema

Base model neuralmind/bert-base-portuguese-cased

Finetuned belisards/congretimbau

Finetuned (2) this model

<https://huggingface.co/azmina/ia-feminista-bert-tema>

# TAMANHO NÃO É TUDO

Modelo	Precisão	Recall	F1
Naive-Bayes	0.54	0.16	0.16
<a href="#">mDeBERTa-v3-base-mnli-xnli</a> (zero-shot)	0.32	0.28	0.27
<a href="#">facebook/bart-large-mnli</a> (zero-shot)	0.34	0.25	0.25
<a href="#">legal-bert-base-cased-ptbr</a>	0.75	0.63	0.66
<a href="#">DeBERTina</a>	0.82	0.74	0.75
<a href="#">BERTimbau large</a>	0.82	0.74	0.75
<a href="#">Gemma-9b</a>	0.70	0.70	0.69
<a href="#">LLama3-8b</a>	0.66	0.61	0.61
<a href="#">Congretimbau</a>	<b>0.80</b>	<b>0.79</b>	<b>0.79</b>

# TRABALHANDO COM IMAGENS

---

Usando CLIP e outros modelos

# IMAGENS COMO DADOS

Abordagens: vetorizar imagens diretamente ou transformar em texto;

Algumas tarefas: responder perguntas baseado em imagens, classificar imagens, **segmentar imagens**, detectar objetos ou pessoas, descrever imagens, etc;



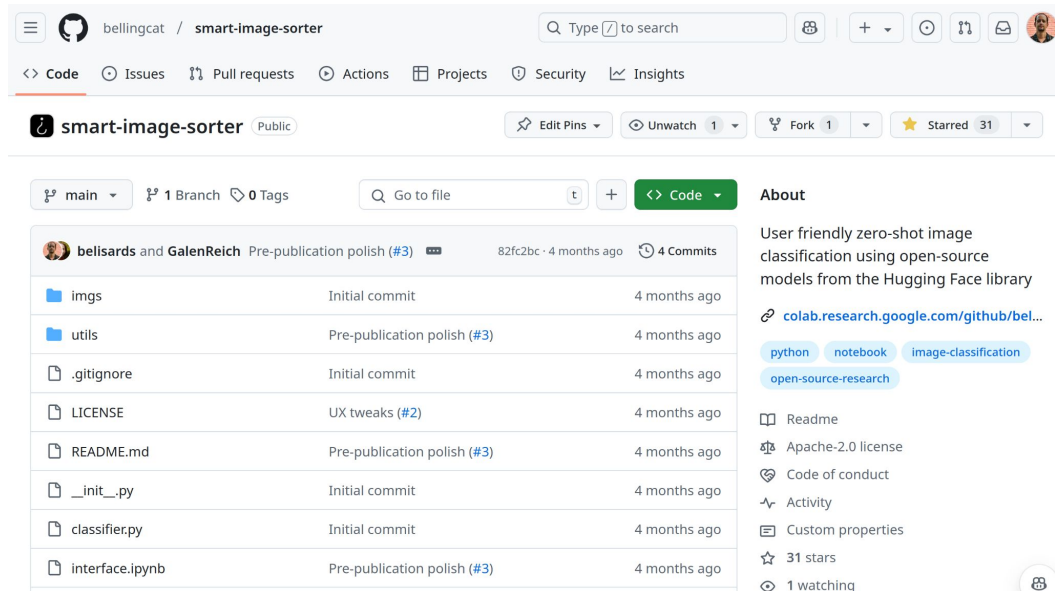
<https://segment-anything.com/>  
<https://samgeo.gishub.org/>

# SMART IMAGE SORTER

Notebooks com interface gráfica ou script Python;

Classificação zero-shot de coleções de imagens;

Guia [traduzido para português](#);



The screenshot shows the GitHub repository page for 'smart-image-sorter' by user 'bellingcat'. The repository is public and has 31 stars and 1 fork. The main branch is selected, and there are 4 commits. The repository description is 'User friendly zero-shot image classification using open-source models from the Hugging Face library'. A link to a Colab notebook is provided: [colab.research.google.com/github/bel...](https://colab.research.google.com/github/bellingcat/smart-image-sorter/blob/main/notebook.ipynb). The repository contains several files and folders:

File/Folder	Commit Message	Time Ago
imgs	Initial commit	4 months ago
utils	Pre-publication polish (#3)	4 months ago
.gitignore	Initial commit	4 months ago
LICENSE	UX tweaks (#2)	4 months ago
README.md	Pre-publication polish (#3)	4 months ago
__init__.py	Initial commit	4 months ago
classifier.py	Initial commit	4 months ago
interface.ipynb	Pre-publication polish (#3)	4 months ago

Additional repository details include: Readme, Apache-2.0 license, Code of conduct, Activity, Custom properties, 31 stars, and 1 watching.

<https://github.com/bellingcat/smart-image-sorter>

# SMART IMAGE SORTER

bellingcat

Investigations Guides Ukraine Justice & Accountability Workshops Q EN ▾

Donate



## Adriano Belisario

Adriano Belisario is a 2024 Bellingcat Tech Fellow. He is passionate about using data, journalism and open source tools to address pressing social issues.

## Easy AI: A Simplified Approach to Classifying Images with Off-the-Shelf AI Models

August 15, 2024 AI Imagery Analysis

Suppose you've scraped thousands of images from a Telegram group or social media site, some of which may be crucial to an investigation of an ongoing conflict. You're looking specifically for photos and videos of weapons, but these are mixed in with memes, screenshots and other unrelated material, and manually reviewing and categorising the images would take more time than you have. What do you do?

In this guide, we show you how you can use artificial intelligence (AI) models to speed up such tasks – even if you don't know how to code – with the help of the [Smart Image Sorter](#), an open-source tool we created.

AI image classification has proven useful in previous investigations, such as those involving war crimes in Yemen or illegal mining in the Amazon rainforest

<https://www.bellingcat.com/resources/how-tos/2024/08/15/easy-ai-zero-shot-ai-image-classification-smart-image-sorter/>

# TRABALHANDO COM ÁUDIOS

---

Usando modelos Whisper

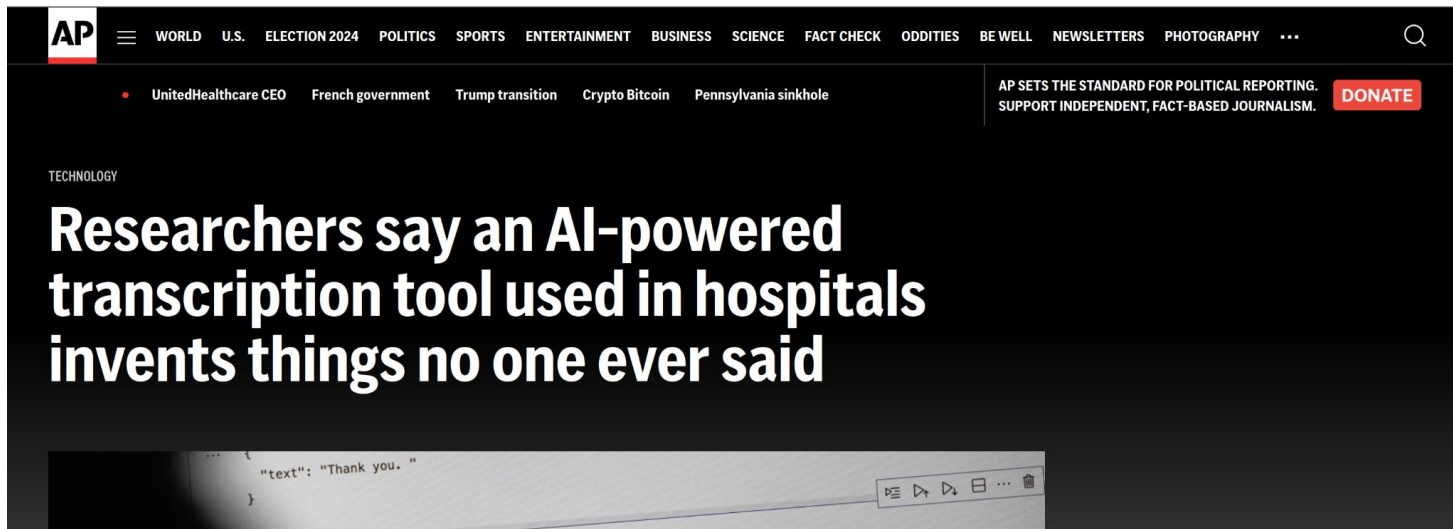


# WHISPER

Notebook para [transcrição de conteúdos do YouTube e vídeos](#) ou [arquivos de áudio](#);

Há também diferentes implementações do modelo, como o ['insanely-fast-whisper'](#)

# RISCO E LIMITAÇÕES



The screenshot shows the top navigation bar of the AP News website. The AP logo is on the left, followed by a menu icon and a list of categories: WORLD, U.S., ELECTION 2024, POLITICS, SPORTS, ENTERTAINMENT, BUSINESS, SCIENCE, FACT CHECK, ODDITIES, BE WELL, NEWSLETTERS, PHOTOGRAPHY, and a search icon. Below the navigation bar, there are several topic links: UnitedHealthcare CEO, French government, Trump transition, Crypto Bitcoin, and Pennsylvania sinkhole. On the right side, there is a promotional banner for AP's political reporting with a 'DONATE' button. The main article section features the word 'TECHNOLOGY' and a large headline: 'Researchers say an AI-powered transcription tool used in hospitals invents things no one ever said'.

AP

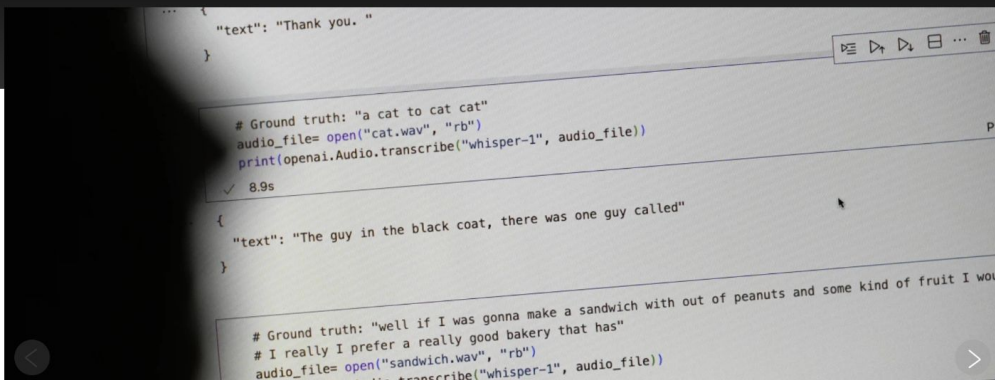
WORLD U.S. ELECTION 2024 POLITICS SPORTS ENTERTAINMENT BUSINESS SCIENCE FACT CHECK ODDITIES BE WELL NEWSLETTERS PHOTOGRAPHY ...

UnitedHealthcare CEO French government Trump transition Crypto Bitcoin Pennsylvania sinkhole

AP SETS THE STANDARD FOR POLITICAL REPORTING. SUPPORT INDEPENDENT, FACT-BASED JOURNALISM. [DONATE](#)

TECHNOLOGY

## Researchers say an AI-powered transcription tool used in hospitals invents things no one ever said



The screenshot shows a code editor with Python code for audio transcription. The code includes comments for ground truth and the use of the OpenAI Whisper API. The output shows a transcription of an audio file, which is a classic example of a transcription error where the AI invents content not present in the original audio.

```
... {  
    "text": "Thank you. "  
}  
}  
  
# Ground truth: "a cat to cat cat"  
audio_file= open("cat.wav", "rb")  
print(openai.Audio.transcribe("whisper-1", audio_file))  
✓ 8.9s  
  
{  
    "text": "The guy in the black coat, there was one guy called"  
}  
}  
  
# Ground truth: "well if I was gonna make a sandwich with out of peanuts and some kind of fruit I wou"  
# I really I prefer a really good bakery that has"  
audio_file= open("sandwich.wav", "rb")  
... audio.transcribe("whisper-1", audio_file))
```

## OUTRAS REFERÊNCIAS

- [3Blue1Brown](#): playlists sobre fundamentos de redes neurais e modelos Transformer;
- [Practical AI for Investigative Journalism](#): playlist de vídeo tutoriais sobre LLMs e outras ferramentas de IA;
- [Argilla](#): ferramenta para anotação colaborativa de documentos;
- Link da apresentação: [bit.ly/ia-coda-24](https://bit.ly/ia-coda-24)