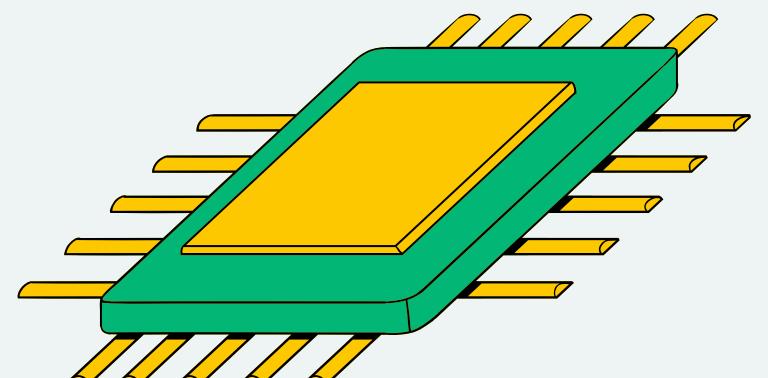


DESBRAVANDO IA LIVRE: HUGGING FACE E MODELOS DE LINGUAGEM TRANSFORMERS

JOÃO DAVI COSTA LIMA



ROTEIRO DA PALESTRA

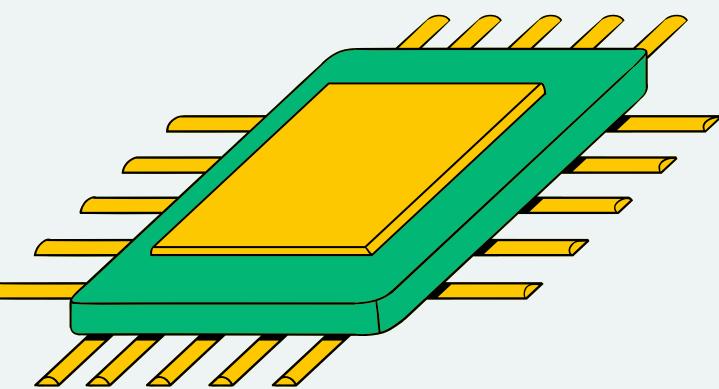


1. Apresentação do palestrante
2. Introdução
3. Histórico da IA e Machine Learning
4. O que é a Hugging Face?
 - a. Modelos da Hugging Face
 - b. Quais as vantagens de utilizar?
5. Modelos de Linguagem Transformers
 - a. Codificação Posicional
 - b. Mecanismo de Autoatenção
 - c. Principais modelos Transformers
6. Estudos de Caso
7. Considerações Finais
8. Perguntas e Agradecimentos



APRESENTAÇÃO DO PALESTRANTE

- Formando em Ciência da Computação pela UFPA;
- Pesquisador PIBIC em PLN com aplicações em Análise de Sentimentos e Reconhecimento de Entidades Nomeadas;
- Organizador da comunidade Açaí Com Dados;



INTRODUÇÃO

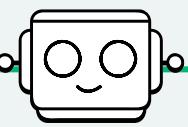
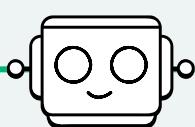
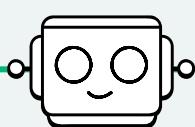
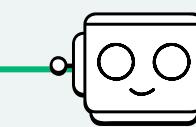
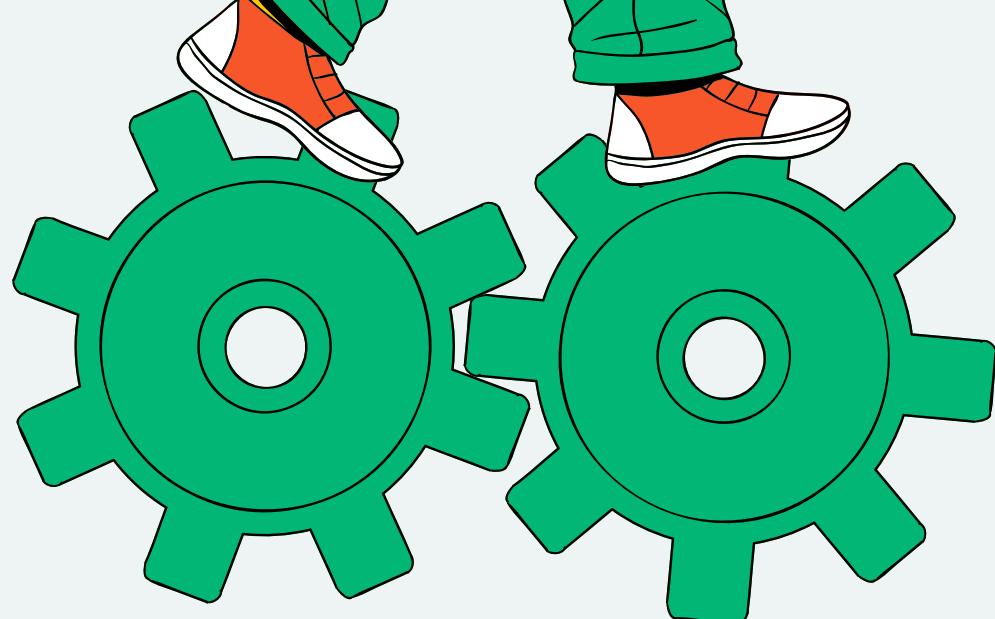
Imagine um mundo onde os avanços em Inteligência Artificial estão confinados aos domínios de instituições de pesquisa de elite ou gigantes da tecnologia exclusivos.



Nesta apresentação, exploraremos de que forma a IA Livre está se desenvolvendo, guiada pelos esforços pioneiros da plataforma Hugging Face. Ao defender o acesso aberto à Inteligência Artificial, a Hugging Face está revolucionando o desenvolvimento, distribuição e utilização de modelos de linguagem Transformer de última geração por meio de software livre e de código aberto.



HISTÓRICO DA IA E ML



1943

Walter Pitts e Warren McCulloch introduziram o primeiro modelo matemático de redes neurais.

1949

O neuropsicólogo Donald Hebb publicou “The Organization of Behavior”, que delineou o conceito de “interação de células cerebrais”.

1950

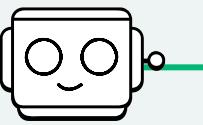
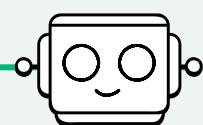
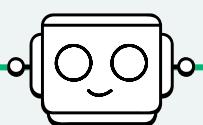
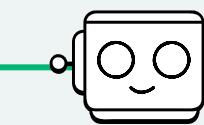
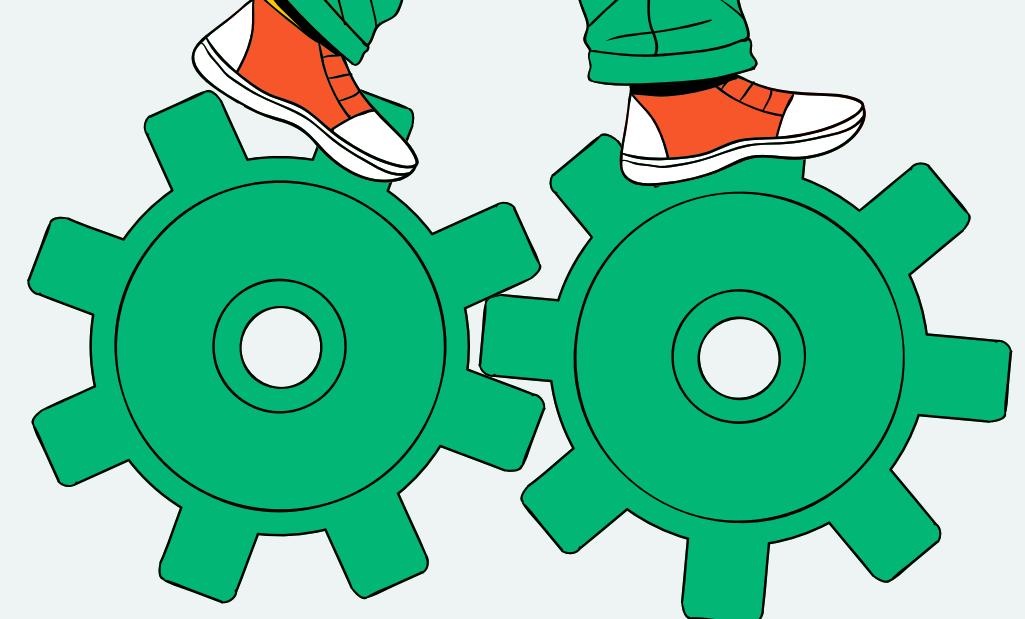
O termo “Machine Learning” foi cunhado por Arthur Samuel, em meados da década de 1950, o qual a definiu como: “a habilidade das máquinas em aprender sem serem explicitamente programadas.”

1960

Os métodos bayesianos foram introduzidos para inferência probabilística na ML, fornecendo uma maneira de atualizar probabilidades com base na observação dos dados



HISTÓRICO DA IA E ML



1970

'Inverno da IA': período de desilusão em relação à eficácia da ML. A pesquisa continuou e a década viu o desenvolvimento de novas técnicas e algoritmos, incluindo a invenção das Redes Neurais Artificiais (RNNs).

1980

Marcada pelo uso crescente da retropropagação, método de treinamento de redes neurais que envolve o ajuste dos pesos das conexões neuronais com base no erro produzido pela rede

1990

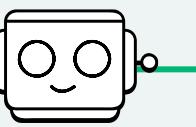
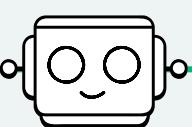
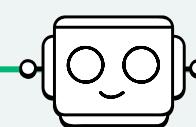
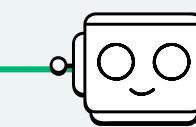
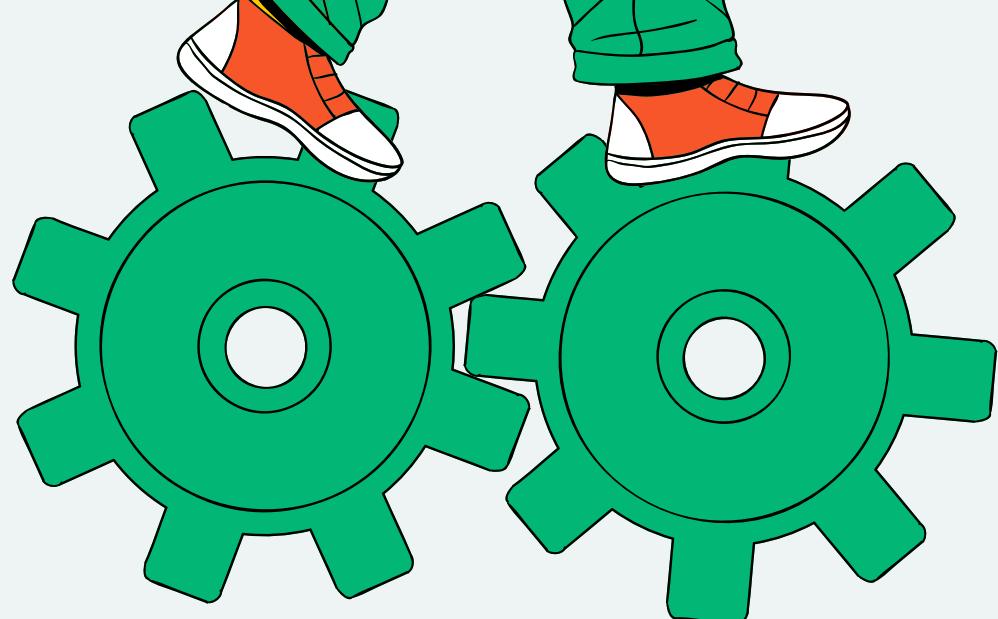
O trabalho em aprendizagem de máquina mudou de uma abordagem baseada em conhecimento para uma abordagem baseada em dados.

2000

Durante este período, o deep learning, um subcampo da aprendizagem de máquina focado em grandes redes neurais com múltiplas camadas, começou a ganhar popularidade



HISTÓRICO DA IA E ML



2010

A aprendizagem de máquina expandiu seus horizontes com a prática de aprendizagem não supervisionada e aprendizagem por reforço.

2016

A Hugging Face foi fundada em Nova York como uma empresa que desenvolveu um aplicativo de chatbot, e posteriormente, pra uma plataforma focada em aprendizado de máquina open source.

2017

Vaswani et al. publicaram o artigo “Attention is All You Need”, que introduziu o modelo Transformer, o qual revolucionou o campo do processamento de linguagem natural.

2020

A IA generativa levou a modelos revolucionários, criando uma proliferação de modelos de base tanto proprietários quanto de código aberto



O QUE É A HUGGING FACE?

A Hugging Face é uma plataforma de IA que democratiza o acesso a modelos de linguagem Transformer avançados.

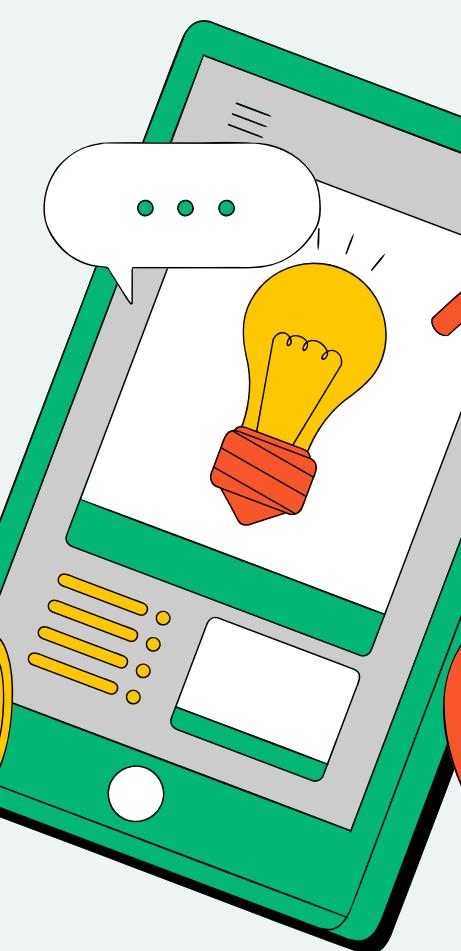
Oferecem bases de dados, modelos pré-treinados, ferramentas e recursos gratuitos em um ecossistema de código aberto, promovendo a colaboração e facilitando a integração em diversas aplicações de análise de sentimentos, tradução automática e mais.



HUGGING FACE



MODELOS NA HUGGING FACE



● Tasks

Hugging Face is the home for all Machine Learning tasks. Here you can find what you need to get started with a task: demos, use cases, models, datasets, and more!

Computer Vision

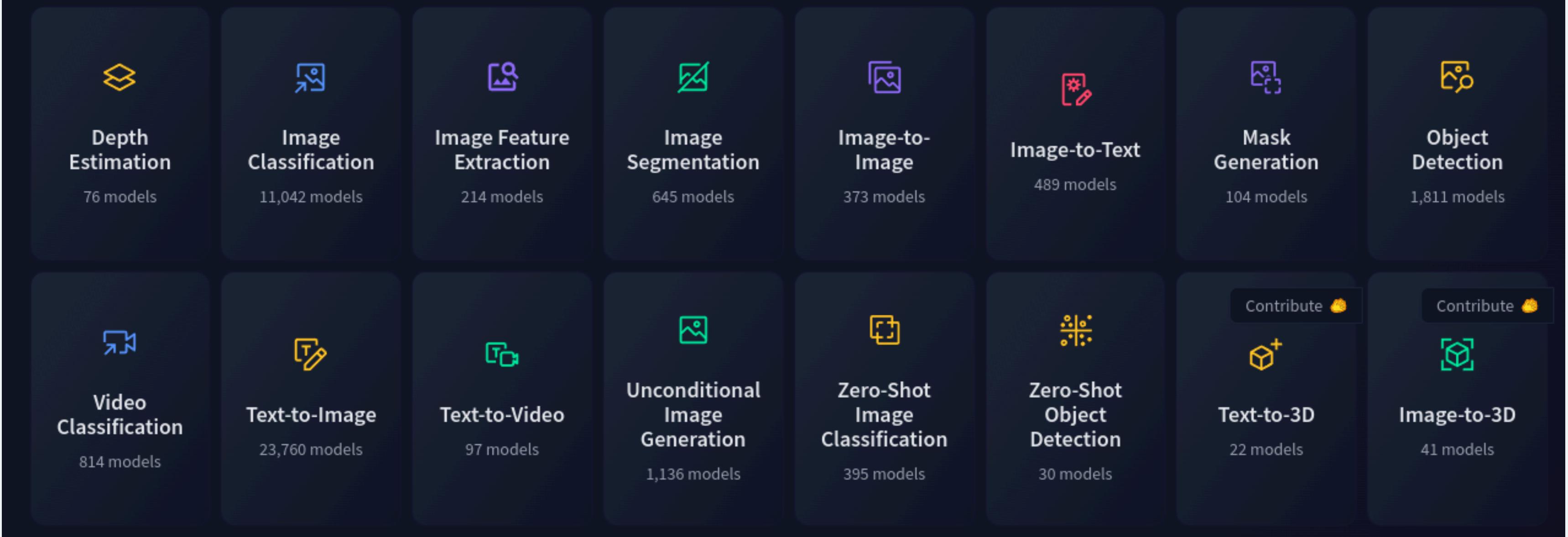


Figura 1: Tarefas de Visão Computacional na Hugging Face

Natural Language Processing

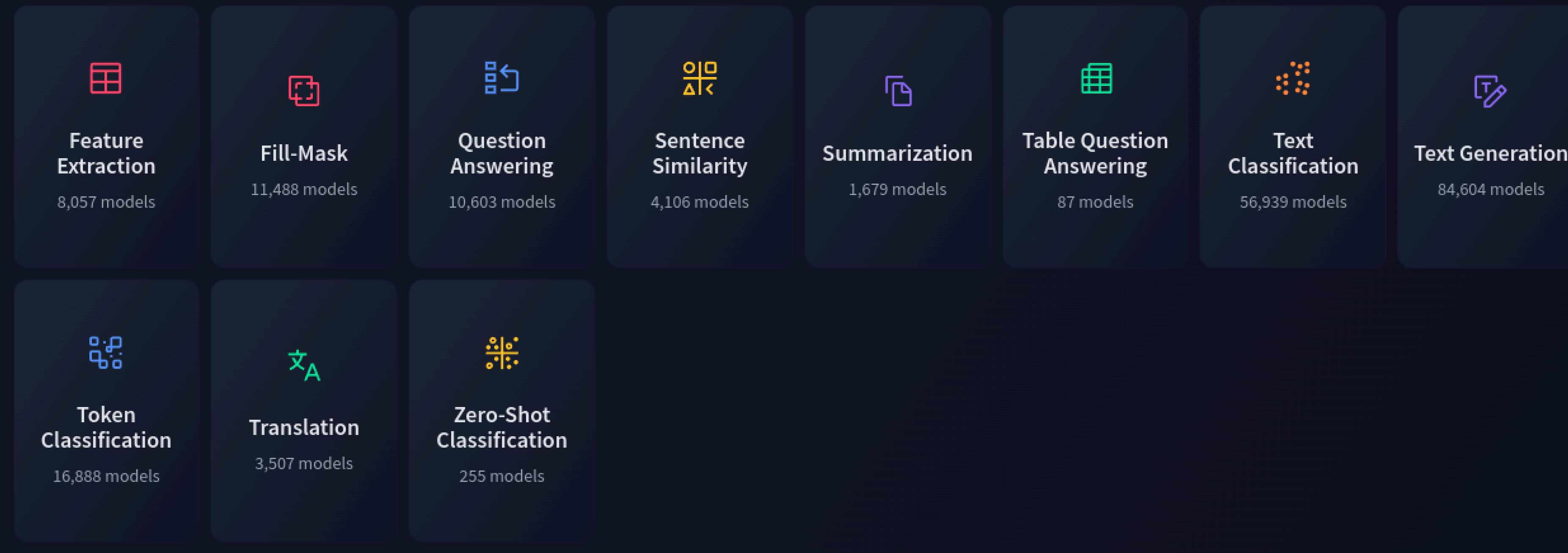


Figura 2: Tarefas de Processamento de Linguagem Natural na Hugging Face

openai/whisper-large-v3 x +

huggingface.co/openai/whisper-large-v3?

Hugging Face Search models, datasets, users...

Models Datasets Spaces Posts Docs Pricing

openai/whisper-large-v3 like 2.36k

Automatic Speech Recognition Transformers PyTorch JAX Safetensors 99 languages whisper audio hf-asr-leaderboard Inference Endpoints

arxiv:2212.04356 arxiv:2311.00430 License: apache-2.0

Model card Files and versions Community 105

Edit model card

Downloads last month 1,966,369

Safetensors Model size 1.54B params Tensor type FP16

Inference API

Automatic Speech Recognition Examples

Browse for file or Record from browser or

Realtime speech recognition

LibriSpeech sample 1

Whisper

Whisper is a pre-trained model for automatic speech recognition (ASR) and speech translation. Trained on 680k hours of labelled data, Whisper models demonstrate a strong ability to generalise to many datasets and domains **without** the need for fine-tuning.

Whisper was proposed in the paper [Robust Speech Recognition via Large-Scale Weak Supervision](#) by Alec Radford et al. from OpenAI. The original code repository can be found [here](#).

Whisper large-v3 has the same architecture as the previous large models except the following minor differences:

Vídeo 1: Exemplo de utilização de modelo de ASR da Hugging Face – WhisperLargeV3

Atividades Google Chrome ▾ 26 de abr 11:26 17 %

recogna-nlp/ptt5-base-s x +
huggingface.co/recogna-nlp/ptt5-base-summ-cstnews

Hugging Face Search models, datasets, users... Models Datasets Spaces Posts Docs Pricing

recogna-nlp/ptt5-base-summ-cstnews like 8
Summarization Transformers PyTorch Safetensors Portuguese t5 text2text-generation pt-br abstractive summarization Inference Endpoints
text-generation-inference License: mit

Model card Files and versions Community 1 Edit model card

Train Deploy Use in Transformers

Portuguese T5 for Abstractive Summarization (PTT5 Summ)

Downloads last month 69

Safetensors Model size 223M params Tensor type F32

Inference API Examples

Summarization

A Hugging Face é uma plataforma de Inteligência Artificial (IA) que se destaca por democratizar o acesso aberto a modelos de linguagem Transformer de última geração. Fundada com a missão de tornar a IA acessível a todos, a Hugging Face oferece uma ampla

Vídeo 2: Exemplo de utilização de modelo de Summarization da Hugging Face - PTT5 Summ

QUAIS AS VANTAGENS DE UTILIZAR A HUGGING FACE?

01

ACESSO DEMOCRÁTICO

Democratiza e desmistifica a Inteligência Artificial, oferecendo acesso aberto e gratuito a uma vasta gama de modelos de linguagem Transformer de última geração.

02

ECOSISTEMA ABERTO

Operando em um ecossistema de software livre e código aberto, permite que estudantes, pesquisadores, desenvolvedores e empresas explorem e utilizem esses modelos sem barreiras financeiras, promovendo a inovação e o avanço da área.

03

VARIEDADE DE APLICAÇÕES

Os modelos de linguagem Transformer disponíveis na Hugging Face podem ser utilizados em uma variedade de aplicações, desde análise de sentimentos até tradução automática e geração de texto.



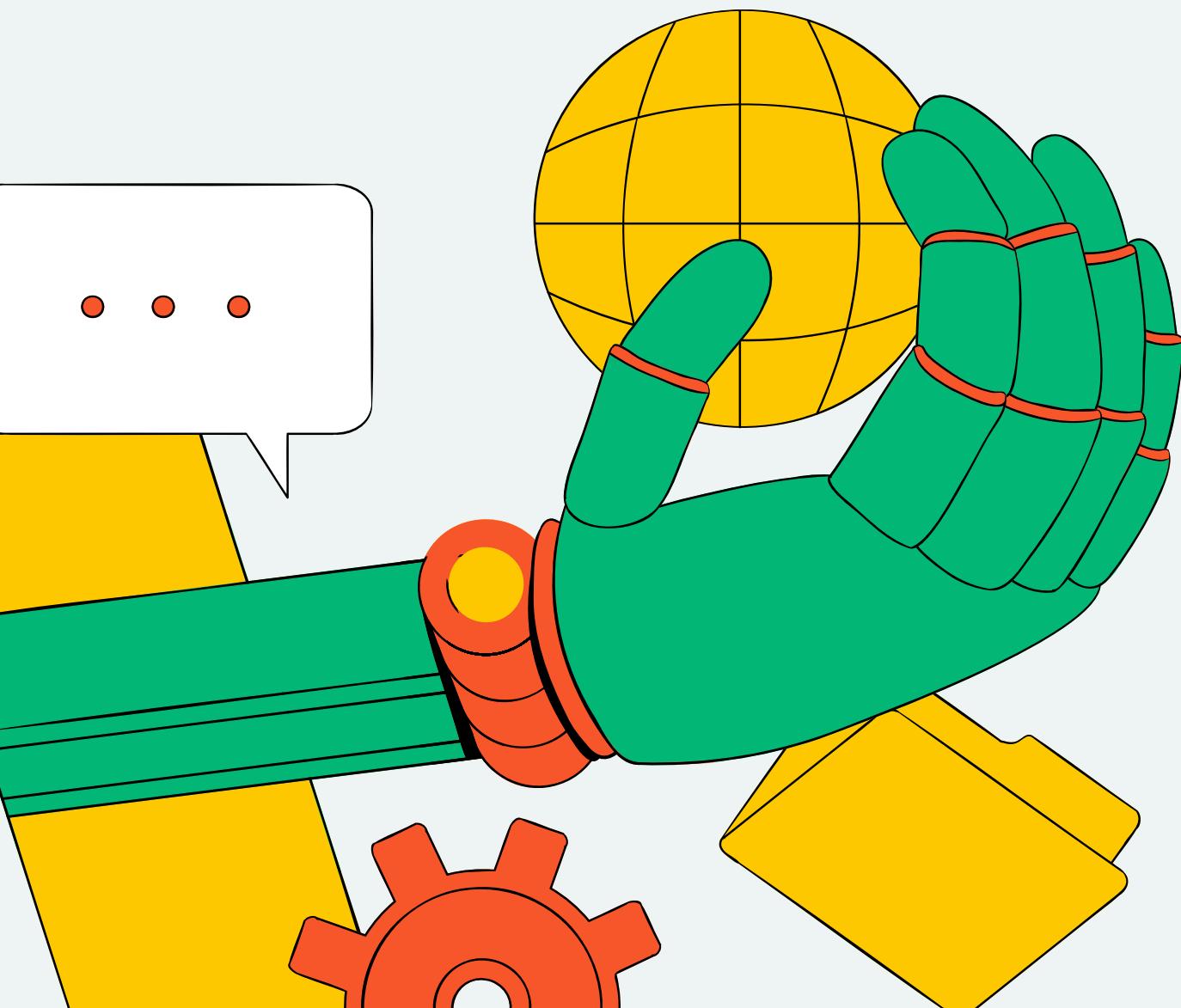
MODELOS DE LINGUAGEM TRANSFORMERS

Antes de sua ascensão, as abordagens tradicionais muitas vezes enfrentavam desafios significativos em capturar efetivamente o contexto de longo alcance em textos.

Por meio do artigo de Vaswani et al., publicado em 2017, a arquitetura Transformers foi apresentada, e analisaremos suas duas principais inovações: **Autoatenção e Codificação Posicional**.



MECANISMO DE AUTOATEAÇÃO

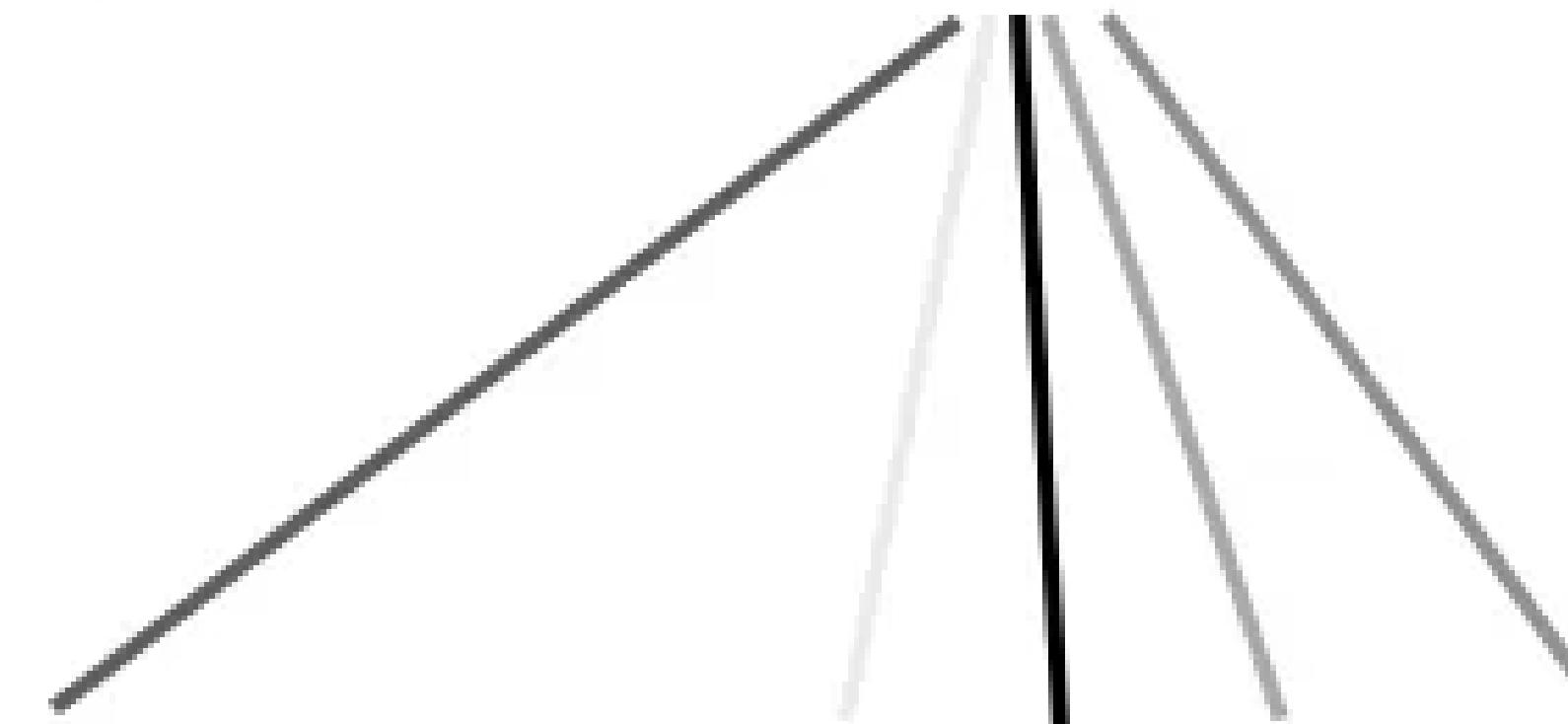


Atenção é um mecanismo que calcula pesos para cada palavra em uma frase conforme eles se relacionam com todas as outras palavras da frase, para que o modelo possa prever palavras que provavelmente serão usadas em sequência.

Esse entendimento do contexto é aprendido ao longo do tempo, à medida que um modelo é treinado em muitos dados.



The young boy always carries his teddy bear with him.



The young boy always carries his teddy bear with him.

Figura 3: Exemplo do mecanismo de auto-atenção da palavra **his** (seu). As linhas indicam a atenção que a palavra “dedica” pras outras palavras da frase.

CODIFICAÇÃO POSICIONAL

A codificação posicional é uma técnica que permite ao modelo entender a ordem das palavras em uma frase.

É implementada adicionando-se um vetor único a cada palavra, que é gerado usando funções seno e cosseno de diferentes frequências.

Por exemplo, na frase “O gato comeu o rato”, a codificação posicional permitiria ao modelo entender que “gato” vem antes de “comeu”, e “comeu” vem antes de “rato”, o que é essencial para compreender o significado da frase.



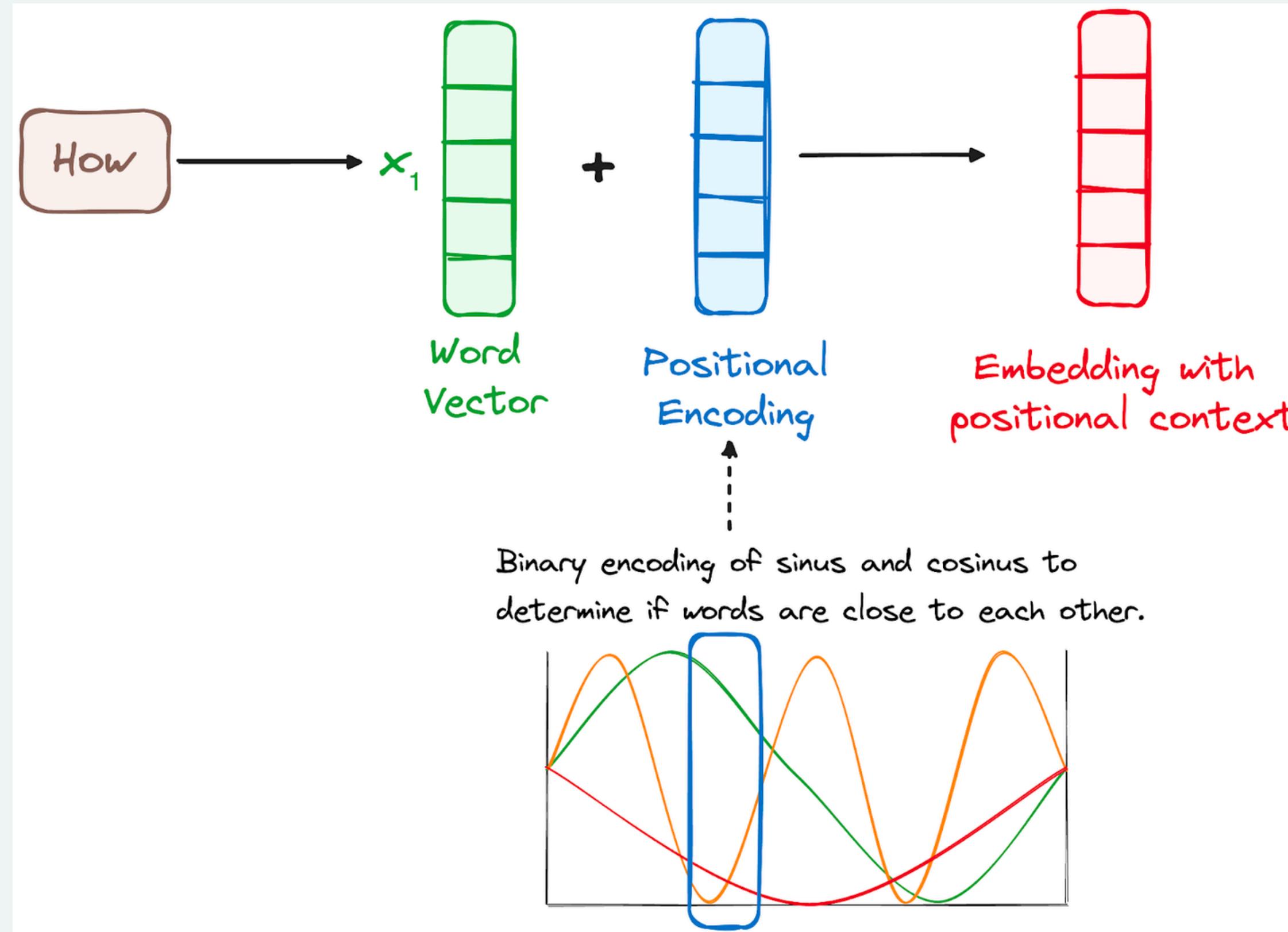


Figura 4: Ilustração do processo de codificação posicional, desde a vetorização da palavra, até a obtenção do contexto posicional.

PRINCIPAIS MODELOS TRANSFORMERS

GPT

- Desenvolvido pela OpenAI;
- Utiliza uma arquitetura de Transformer autoregressiva;
- Utilizado para diversas tarefas de geração de texto;

BERT

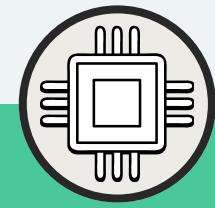
- Desenvolvido pelo Google;
- Utiliza uma arquitetura de Transformer bidirecional.
- Altamente eficaz em várias tarefas de PLN, como classificação de texto e question answering.

LLAMA

- Desenvolvido pela Microsoft e META;
- Incorpora tarefas de nivelamento linguístico para abordar o viés e a desigualdade linguística em modelos de linguagem.

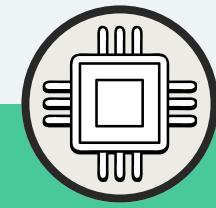


ESTUDOS DE CASO



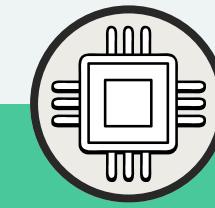
HUGGING CHAT

Chatbots Open
Source e Assistência
virtual



HUGGING FACE AWS

Aplicação dos
modelos Hugging
Face na AWS



AUTO TRAIN

Automatização do
treinamento de
modelos



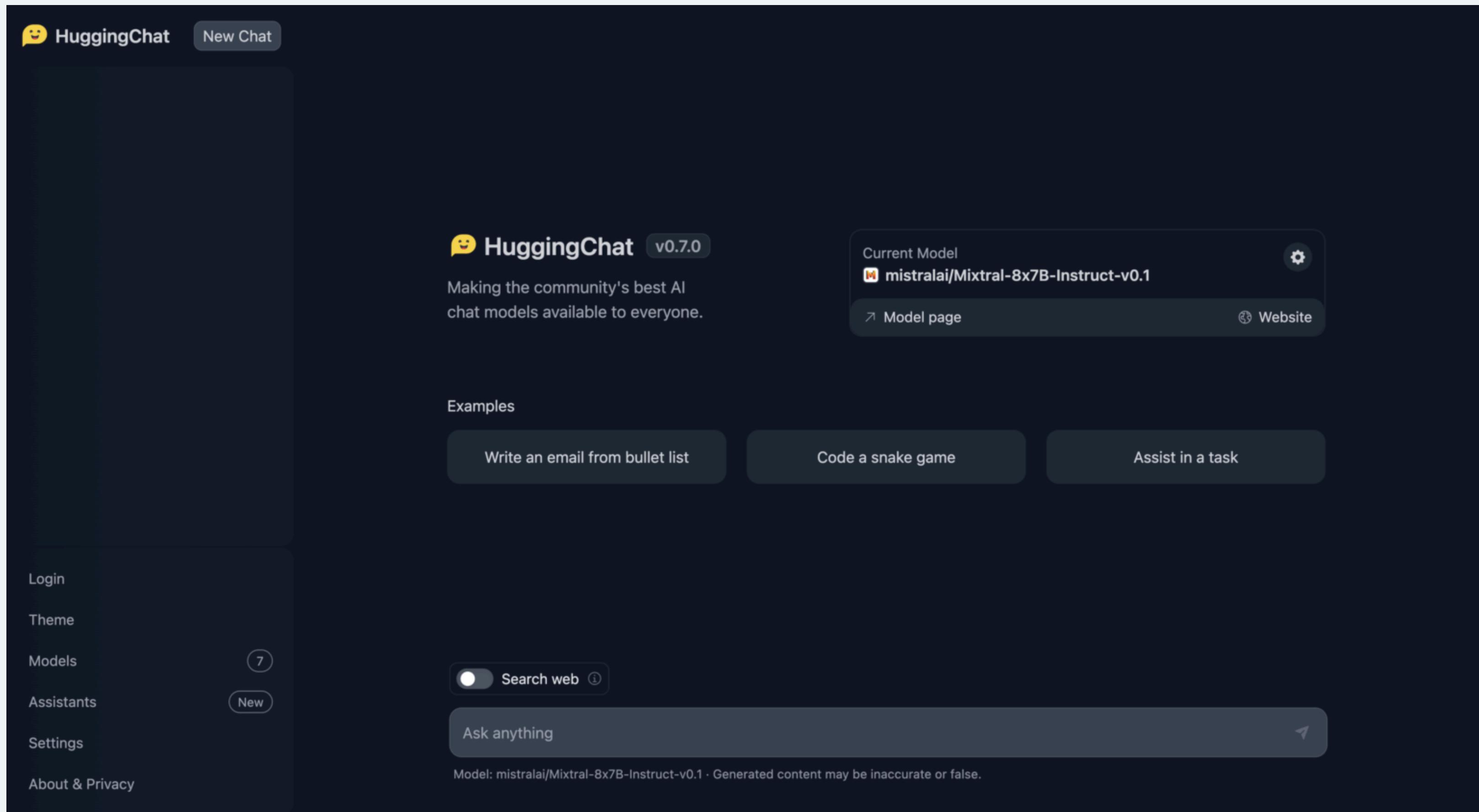


Figura 5: IA Conversacional open source e de fácil utilização

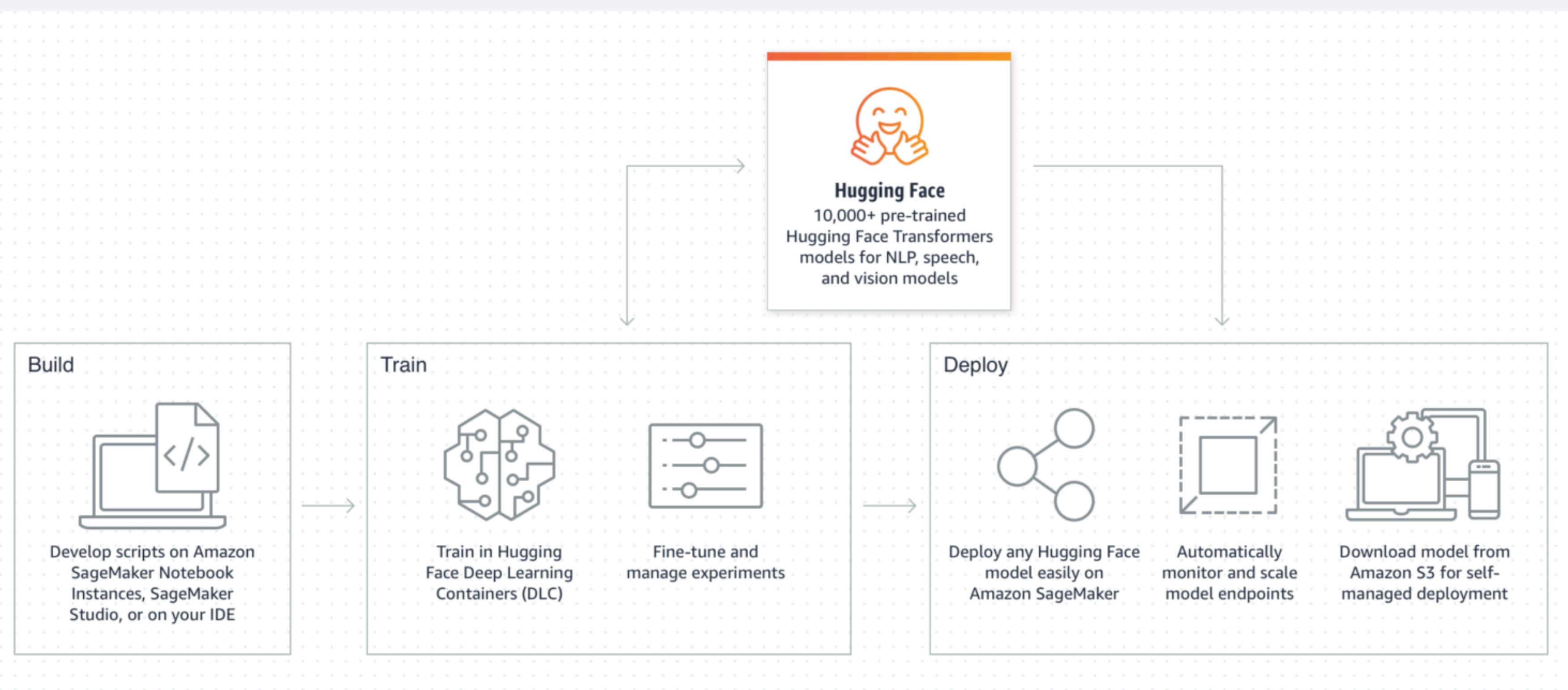


Figura 6: Em colaboração com a Amazon AWS, Hugging Face disponibiliza containers de Deep Learning contendo mais de 10.000 modelos pré-treinados

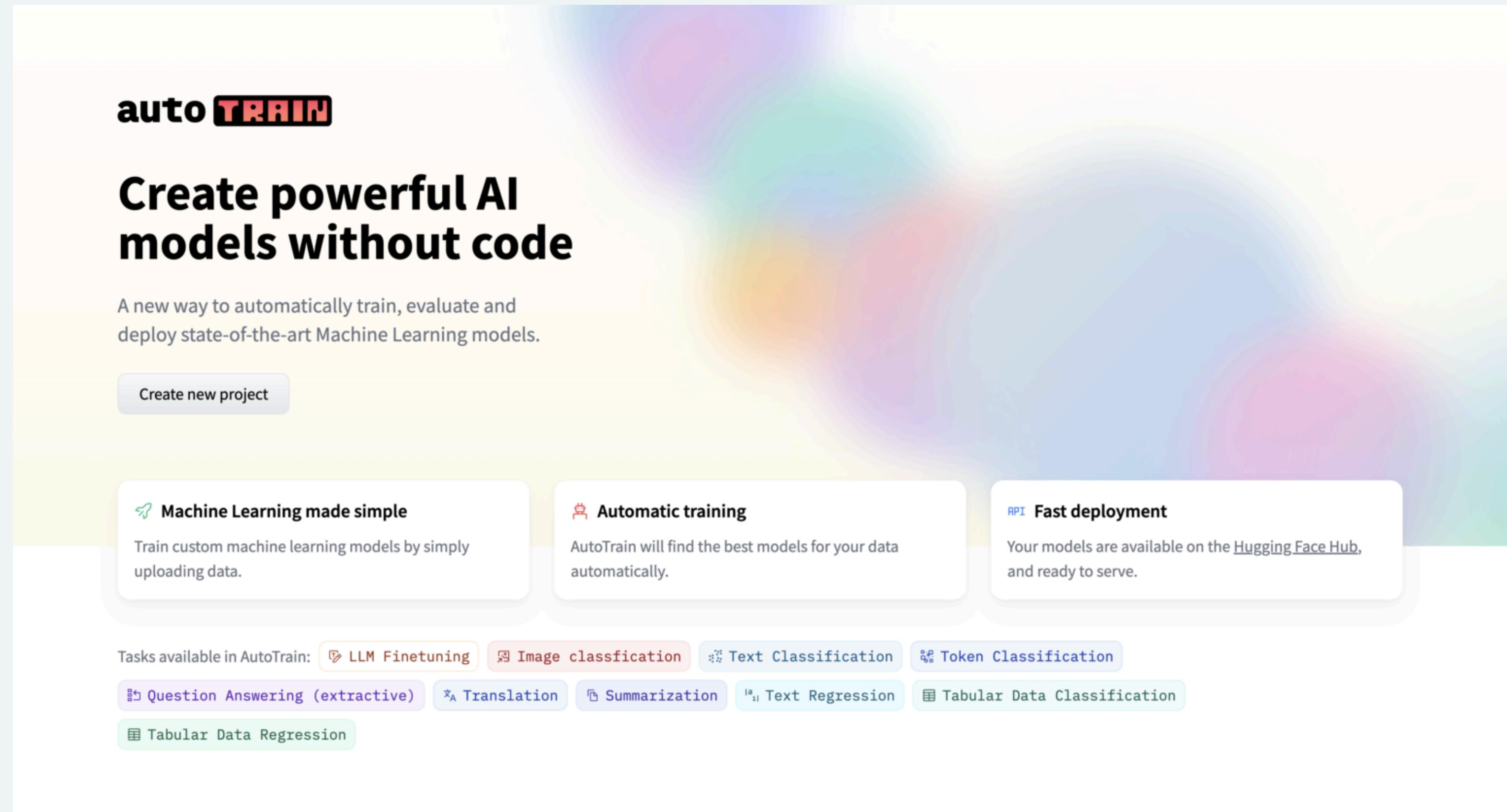


Figura 7: Baseado nos dados de envio do usuário, o sistema realiza automaticamente o treinamento de um modelo, otimizando-o para a tarefa específica em questão.

CONSIDERAÇÕES FINAIS



Ao explorarmos os estudos de caso, podemos ver claramente a diversidade de aplicações e de possibilidades oferecidas pelos modelos de linguagem Transformers da Hugging Face.

Desde a criação de chatbots inteligentes até a implantação escalável na nuvem e a automação do treinamento de modelos personalizados, esses exemplos ilustram como a Hugging Face está liderando a inovação e democratizando o acesso à inteligência artificial livre e open-source.



FAÇA PARTE DA COMUNIDADE AÇAI COM DADOS!



Figura 8: QR Code do grupo da Comunidade Açaí com Dados no Whatsapp

OBRIGADO PELA ATEÇÃO! DÚVIDAS?

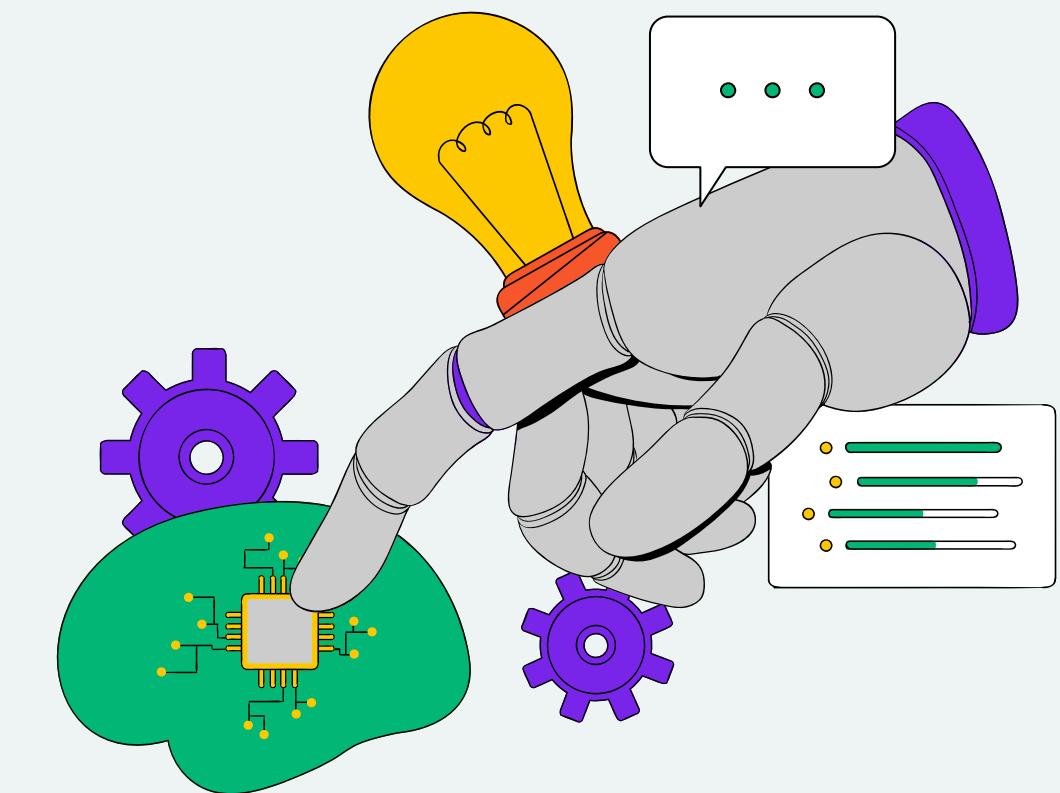
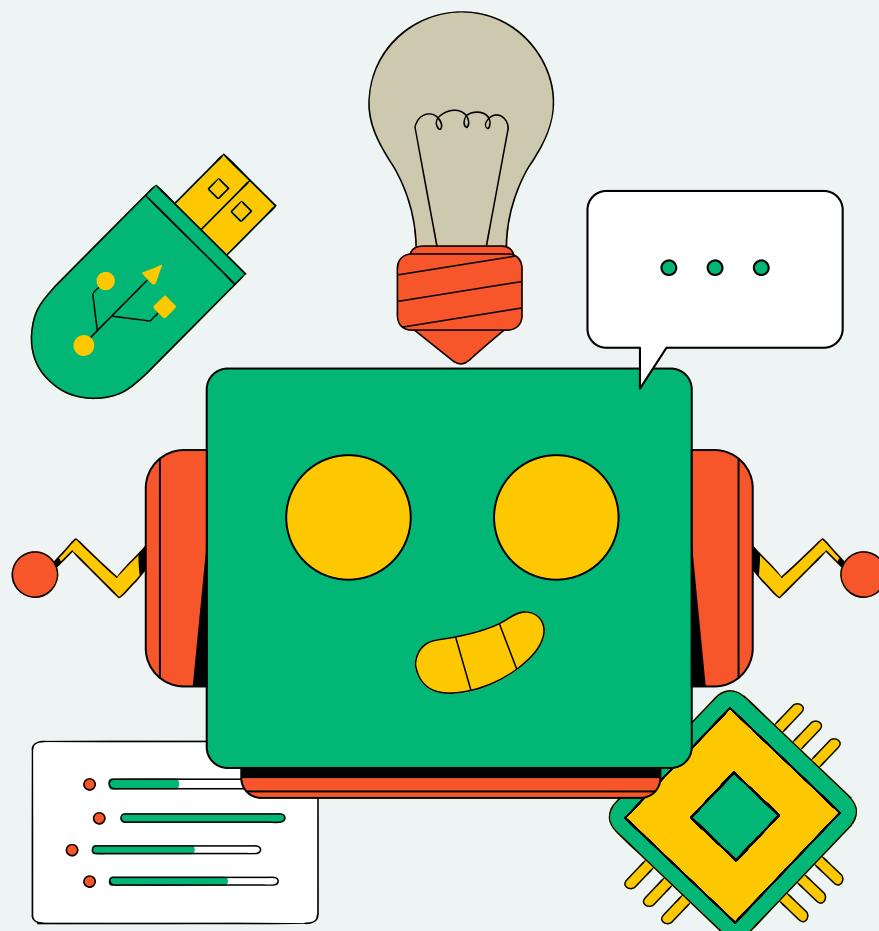


Figura 9: Página do Linktree – Redes Sociais e Meios para contato

REFERÊNCIAS BIBLIOGRÁFICAS

- What is a transformer model | IBM. Disponível em: <<https://www.ibm.com/topics/transformer-model>>.
- The Evolution of Machine Learning: A Brief History and Timeline. Disponível em: <<https://machinelearningmodels.org/the-evolution-of-machine-learning-a-brief-history-and-timeline/>>.
- FERRER, J. How Transformers Work: A Detailed Exploration of Transformer Architecture. Disponível em: <<https://www.datacamp.com/tutorial/how-transformers-work>>. Acesso em: 22 abr. 2024.
- KAPOOR, M. Every Hugging Face Statistics You Need to Know (2024). Disponível em: <[https://www.greatai\(prompts\).com/guide/huggingface-statistics/](https://www.greatai(prompts).com/guide/huggingface-statistics/)>. Acesso em: 23 abr. 2024.
- PEREIRA, T. Guia de Introdução aos Modelos GPT. Disponível em: <<https://blog.dsacademy.com.br/guia-de-introducao-aos-modelos-gpt/#:~:text=A%20gera%C3%A7%C3%A3o%20de%20texto%20com>>. Acesso em: 23 abr. 2024.
- Brand assets – Hugging Face. Disponível em: <<https://huggingface.co/brand>>. Acesso em: 24 abr. 2024.

REFERÊNCIAS BIBLIOGRÁFICAS

- PAIOLA, Pedro H.; DE ROSA, Gustavo H.; PAPA, João P. Deep Learning-Based Abstractive Summarization for Brazilian Portuguese Texts. In: XAVIER-JUNIOR, João Carlos; RIOS, Ricardo Araújo (Ed.). BRACIS 2022: Intelligent Systems. Cham: Springer International Publishing, 2022. p. 479–493. ISBN 978-3-031-21689-3.
- RADFORD, Alec; KIM, Jong Wook; XU, Tao; BROCKMAN, Greg; MCLEAVEY, Christine; SUTSKEVER, Ilya. Robust Speech Recognition via Large-Scale Weak Supervision. arXiv, 2022. Disponível em: <<https://arxiv.org/abs/2212.04356>>. DOI: 10.48550/ARXIV.2212.04356.
- HELENE. Attention and Transformer Models. Disponível em: <<https://towardsdatascience.com/attention-and-transformer-models-fe667f958378>>. Acesso em: 25 abr. 2024.
- EXPERT ACADEMY, I. Curso Processamento de Linguagem Natural com BERT e Python. , 7 maio 2020. Disponível em: <<https://www.youtube.com/watch?v=pHkwwIRj-e0>>. Acesso em: 25 abr. 2024
- VASWANI, Ashish; SHAZEEER, Noam; PARMAR, Niki; USZKOREIT, Jakob; JONES, Llion; GOMEZ, Aidan N.; KAISER, Lukasz; POLOSUKHIN, Illia. Attention Is All You Need. arXiv, 2017. Disponível em: <<https://arxiv.org/abs/1706.03762>>. DOI: 10.48550/arXiv.1706.03762.

REFERÊNCIAS BIBLIOGRÁFICAS

- FIRICAN, G. The history of Machine Learning. Disponível em: <<https://www.lightsondata.com/the-history-of-machine-learning/>>. Acesso em: 25 abr. 2024.
- SALEEM, A. Transformer models: the future of natural language processing. Disponível em: <<https://datasciencedojo.com/blog/transformer-models/>>. Acesso em: 25 abr. 2024.
- HUGGING FACE. Disponível em: <<https://huggingface.co>>. Acesso em: 26 abr. 2024.
- Devlin, J.; Chang, M. W.; Lee, K.; Toutanova, K. (2018). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. arXiv preprint arXiv:1810.04805, 2018.

Obs: Caso queira visualizar essa apresentação no canva com as animações e vídeos, basta acessar esse link:

https://www.canva.com/design/DAGDdcXp-SA/VWbG6IVHe69Ew4TSZaSFqA/view?utm_content=DAGDdcXp-SA&utm_campaign=designshare&utm_medium=link&utm_source=editor