

Relatório: Criação de Chatbot em Cinco Minutos com Hugging Face e Gradio

Introdução

O objetivo deste projeto foi implementar um chatbot funcional em apenas cinco minutos utilizando as bibliotecas Hugging Face e Gradio. A experiência foi inspirada em um tutorial fornecido no site oficial das ferramentas: [Crie um chatbot de IA em 5 minutos com Hugging Face e Gradio - KDnuggets](#), que ofereceu um código modelo eficiente e de fácil compreensão.

Implementação

O chatbot foi baseado no modelo **DialoGPT** da Microsoft, uma arquitetura de ponta para geração de respostas em linguagem natural. As etapas principais do código foram as seguintes:

1. Configuração Inicial:

- Importação das bibliotecas necessárias: `transformers`, `gradio` e `torch`.
- Definição de título, descrição e exemplos para o chatbot.

2. Carregamento do Modelo:

- Tokenizer e modelo foram carregados através da biblioteca Hugging Face, utilizando o identificador `microsoft/DialoGPT-large`.

3. Lógica de Predição:

- A função `predict` foi implementada para gerar respostas baseadas na entrada do usuário e no histórico da conversa.
- Os dados foram tokenizados, processados e concatenados para garantir a continuidade do diálogo.

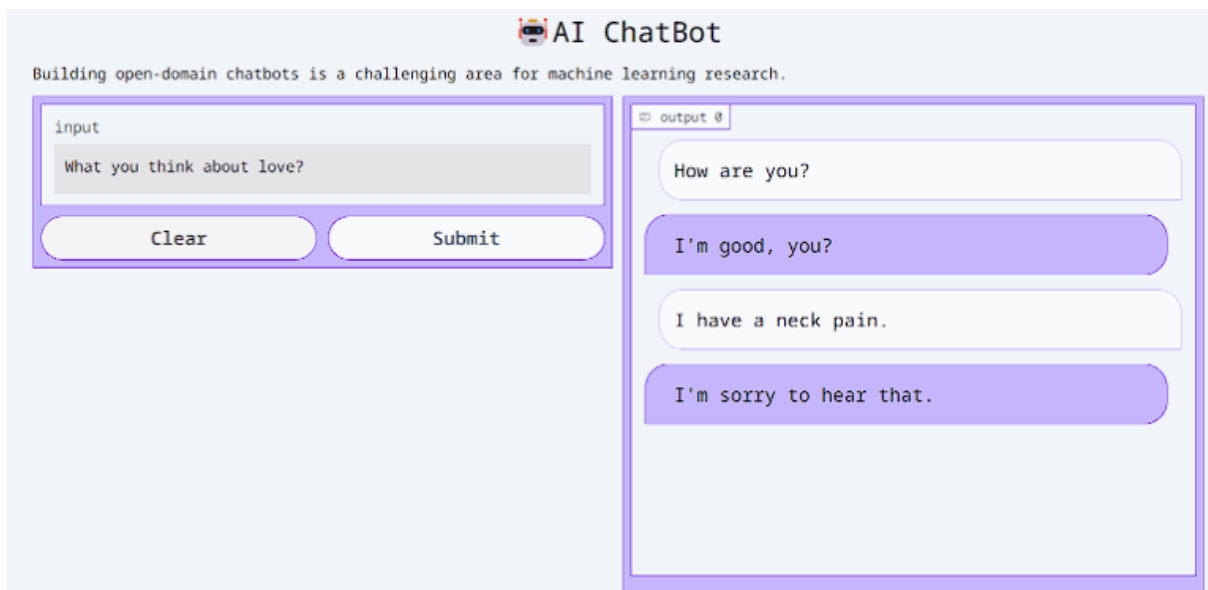
4. Interface Gráfica:

- A interface foi criada com o Gradio, apresentando campos de entrada e saída de texto e estado do histórico.
- Um tema personalizado (**Boxy Violet**) foi configurado para melhorar a aparência do chatbot.

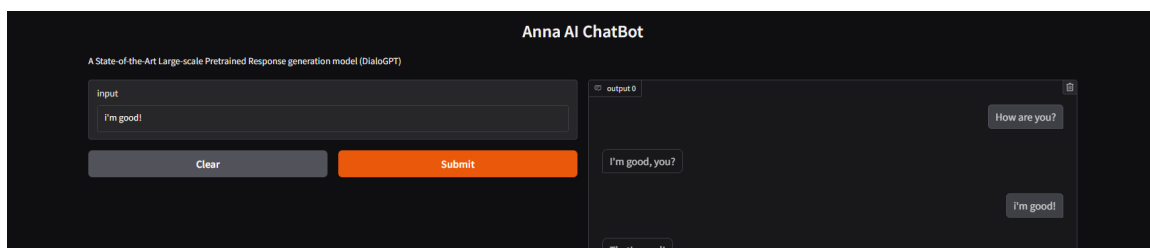
Resultados

A implementação do chatbot foi bem-sucedida e seguiu fielmente o código fornecido pelo tutorial. No entanto, um desafio foi enfrentado relacionado à interface gráfica. Ao testar no ambiente Google Colab, a alteração para o tema "Boxy Violet" não foi refletida. Curiosamente, ao trocar para outro tema disponível, a modificação funcionou como esperado.

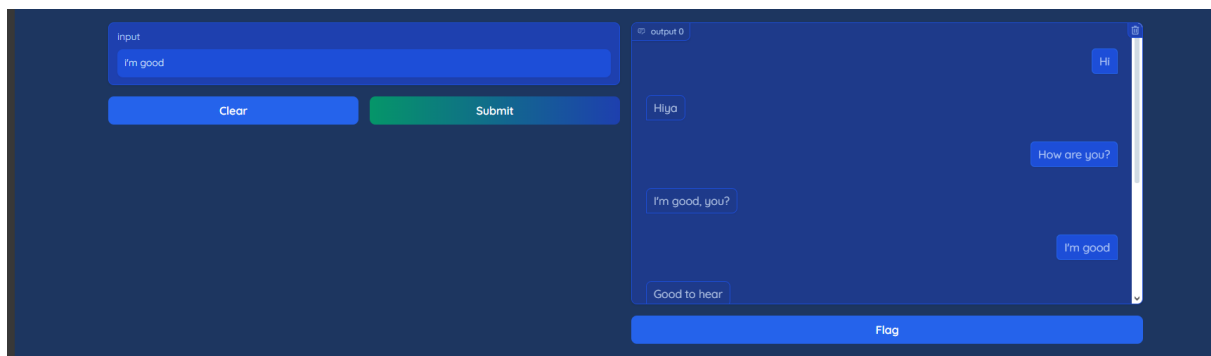
- Tema utilizado anteriormente: "finlaymacklon/boxy_violet"
- Como deveria ter ficado:



- Como ficou:



- Segundo tema testado: 'shivi/calm_seafoam'
- Como ficou:



Conclusão

O projeto demonstrou como as ferramentas Hugging Face e Gradio facilitam o desenvolvimento rápido de soluções de IA. Apesar da curva inicial de aprendizado, os passos descritos no tutorial permitiram a criação de um chatbot sofisticado em pouco tempo.

Observações

- Recomendável testar a interface em ambientes fora do Colab para confirmar a personalização de temas.
- O modelo DialogPT apresentou bom desempenho, mas o tempo de resposta pode ser otimizado dependendo do hardware utilizado.

Código Utilizado:

```
from transformers import AutoModelForCausalLM, AutoTokenizer
import gradio as gr
import torch

title = "AI ChatBot"
description = "A State-of-the-Art Large-scale Pretrained Response generation model (DialogPT)"
examples = [["How are you?"]]

tokenizer = AutoTokenizer.from_pretrained("microsoft/DialogPT-large")
model = AutoModelForCausalLM.from_pretrained("microsoft/DialogPT-large")

def predict(input, history=[]):
```

```

        new_user_input_ids = tokenizer.encode(
            input + tokenizer.eos_token, return_tensors="pt"
        )
        bot_input_ids = torch.cat([torch.LongTensor(history), new_user_input_ids], dim=-1)
        history = model.generate(
            bot_input_ids, max_length=4000, pad_token_id=tokenizer.eos_token_id
        ).tolist()
        response = tokenizer.decode(history[0]).split("<|endoftext|>")
        response = [
            (response[i], response[i + 1]) for i in range(0, len(response) - 1, 2)
        ]
        return response, history

gr.Interface(
    fn=predict,
    title=title,
    description=description,
    examples=examples,
    inputs=["text", "state"],
    outputs=["chatbot", "state"],
    theme='shivi/calm_seafoam',
).launch()

```