



MAICON QUERINO JESUS DE SOUZA - 202210115

RELATÓRIO FINAL TRANSCRIÇÃO DE ÁUDIO COM IA

LAVRAS

2025

RELATÓRIO FINAL

PROJETO TRANSCRIÇÃO DE ÁUDIO COM IA
Disciplina: Sistemas Distribuídos
Professor: André Lima Salgado.

LAVRAS

2025

Introdução

A transcrição de áudio para texto tem se tornado uma prática essencial em diversos setores, como educação, saúde, mídia e negócios. Essa técnica facilita o acesso e a análise de informações contidas em formatos de áudio, permitindo uma busca mais eficiente por conteúdos específicos e promovendo a acessibilidade para pessoas com deficiência auditiva. Além disso, a transcrição possibilita a reutilização do conteúdo em diferentes plataformas, como blogs, jornais e infográficos, ampliando seu alcance e utilidade.

Com os avanços na área de Inteligência Artificial (IA), especialmente em modelos de reconhecimento de fala, a transcrição automática de áudio tornou-se mais precisa e eficiente. Ferramentas baseadas em IA são capazes de lidar com diferentes idiomas e sotaques, reduzindo o tempo e os custos associados à transcrição manual. Além disso, essas ferramentas podem filtrar ruídos de fundo, melhorando a qualidade da transcrição mesmo em ambientes com áudio de baixa qualidade.

Este projeto visa desenvolver uma aplicação que permite ao usuário gravar um áudio e, utilizando técnicas de IA, transcrevê-lo automaticamente para texto. A implementação de uma solução como esta pode beneficiar profissionais que lidam com grandes volumes de informações em áudio, como jornalistas, pesquisadores e médicos, ao proporcionar uma maneira rápida e precisa de converter fala em texto. Além disso, ao facilitar a transcrição, promove-se a inclusão e a acessibilidade, garantindo que mais pessoas possam acessar e compreender o conteúdo originalmente disponível apenas em formato de áudio.

Desenvolvimento do Projeto

O projeto foi desenvolvido utilizando o microframework Flask, que facilita a criação de APIs RESTful em Python. Para a transcrição de áudio, integrou-se o modelo Whisper da OpenAI, conhecido por sua precisão e capacidade de lidar com diversos idiomas e sotaques.

Tecnologias Utilizadas:

- **Flask:** Framework leve e flexível para desenvolvimento de aplicações web em Python, ideal para a criação de APIs RESTful.
- **Whisper:** Modelo de transcrição de áudio da OpenAI, capaz de converter fala em texto com alta precisão.
- **Sounddevice e SciPy:** Bibliotecas para captura de áudio do microfone e manipulação de arquivos WAV.

Repositório no GitHub:

O desenvolvimento foi feito usando o sistema de versionamento do git. O código-fonte completo do projeto está disponível no GitHub, permitindo acesso, colaboração e possíveis contribuições de outros desenvolvedores. Link do repositório:

<https://github.com/Maicon-Souza/Projeto-Audios-com-IA>

Relatório de Impacto à Proteção de Dados Pessoais:

Hoje a LGPD atua como a lei que defende os direitos de exposições de dados na internet e suas dependências, dessa maneira foi realizado um relatório de impacto à proteção de dados pessoais que tem como finalidade avaliar os riscos associados ao tratamento de dados pessoais no projeto de conversão de texto em fala. Segue o link do relatório:

<https://drive.google.com/file/d/18nukbhtU2HsdHMYJTNFbYMjy23PR-AZ/view?usp=sharing>

Dificuldades Encontradas:

Durante o desenvolvimento, enfrentaram-se desafios relacionados à integração do modelo Whisper com o Flask, especialmente na gestão de arquivos temporários e no tratamento de erros. Além disso, as versões do python da minha máquina eram diferentes da requerida para usar as funcionalidades do modelo Whisper, dessa maneira, ao fazer o downgrade de versão acabou me gerando muitos erros. No final tentei realizar a containerização do projeto, tudo foi feito e consegui colocar o servidor no ar, no entanto ao acessar os endpoints não era possível realizar as ações desejadas, o primeiro endpoint era de gravar áudio, no entanto não foi possível fazer a integração da entrada de áudio da minha máquina com o container docker, mas funcionava ao usar um áudio já pronto, o outro endpoint de transcrição também deu erro pois o download do modelo whisper estava corrompido dentro do container, devido ao tempo e para manter as funcionalidades principais funcionando decidi terminar a aplicação sem usar a containerização.

Conclusão

O desenvolvimento deste projeto proporcionou uma compreensão aprofundada dos desafios e das oportunidades associadas à transcrição de áudio utilizando Inteligência Artificial (IA). A integração do modelo Whisper da OpenAI demonstrou ser eficaz na conversão de fala em texto, evidenciando o potencial da IA em automatizar processos que tradicionalmente exigiriam intervenção manual.

No entanto, o projeto também evidenciou desafios significativos, especialmente relacionados à precisão da transcrição em contextos com ruído de fundo ou sotaques diversos e a parte de docker. A necessidade de ajustes finos nos parâmetros do modelo e a implementação de técnicas de pré-processamento de áudio para melhorar a qualidade da entrada foram aspectos cruciais para alcançar resultados satisfatórios.

Em conclusão, este projeto não apenas demonstrou a viabilidade da transcrição de áudio automatizada utilizando IA, mas também destacou a importância de abordagens integradas que considerem as especificidades do idioma e do contexto. A contínua evolução das tecnologias de IA promete aprimorar ainda mais a precisão e a aplicabilidade dessas soluções, ampliando seu impacto em diversas áreas.