

# Projeto de IA Geradora de Áudio

O objetivo deste projeto é desenvolver uma Inteligência Artificial (IA) capaz de gerar áudios de voz humana de alta qualidade com base em textos ou comandos específicos (como outro áudio). Esta IA será projetada para síntese da fala. O projeto explorará técnicas avançadas de processamento de linguagem natural (NLP) e aprendizado profundo para criar um sistema robusto e eficiente.

## Objetivos

- Desenvolver um modelo de IA que possa gerar voz humana a partir de textos e outros comandos.
- Avaliar a qualidade e a naturalidade do áudio gerado.
- Testar e validar o modelo em diferentes cenários de aplicação e diferentes entradas, podendo ser possível especificar até intonação da voz.

## Conjunto de Dados

- Áudios gravados por e transcritos por nós para testes iniciais.
- Áudios retirados de vídeos de várias fontes e transcritos por nós.
- Uso do banco de dados Common Voice do mozilla, que possui muitos áudios já transcritos e validados.

## Preparação dos Dados

- Pré-processar os áudios ajustando a normalização do volume, remoção de ruídos e segmentação dos arquivos de áudio.
- Garantir que todos os dados de áudio estejam anotados corretamente com transcrições textuais e etiquetas de tempo.
- Os dados serão transcritos com o uso de ferramentas de I.A.
- Separar os dados em conjuntos de treinamento, validação e teste.

## Modelos e Algoritmos

- Estamos buscando formas diferentes de gerar e pré-processar os áudios. Atualmente estamos utilizando:
  - Tocotron 2 para gerar os espectrogramas mel.
  - Arquivos de áudio do tipo WAV.
  - Áudios gravados e transcritos por nós.

## Treinamento

- O treinamento ainda está em fase inicial, estamos testando algumas alternativas de código a fim de encontrar a que obtenha o resultado mais próximo do esperado.

## **Métricas de Avaliação**

- Teste de qualidade do output gerado, sendo o resultado ótimo, o que se aproxima mais da voz humana.
- Analisar se o áudio gerado é fiel ao texto da entrada.
- Avaliar a qualidade do espectrograma gerado antes do resultado final (o áudio final é gerado a partir de um espctrograma).

## **Resultados Esperados**

Esperamos alcançar uma qualidade de áudio que seja semelhante da voz humano em termos de naturalidade e clareza, buscando uma voz viva e menos robotizada.