

CLONAGEM DE VOZ LOCAL

Qwen3-TTS 1.7B + transcricao com Qwen3-ASR 0.6B

Um guia simples para Windows, pensado para quem nunca instalou um modelo de inteligencia artificial.

CRIADO PARA O CANAL

NEIDSON LIMA

Versao do tutorial: agosto de 2026

Antes de comecar

Este sistema funciona no proprio computador. O painel visual e o Native Studio, incluido no projeto open source audio.cpp. O Qwen3-TTS gera uma nova fala com a voz de referencia, enquanto o Qwen3-ASR transforma audio em texto.

O que sera instalado

- audio.cpp Native Studio: painel local acessado pelo navegador.
- Qwen3-TTS 12Hz 1.7B Base Q8: clonagem e geracao de voz.
- Qwen3-ASR 0.6B Q8: transcricao de audio em portugues.
- Python 3.10 ou mais recente: usado pelo gerenciador de downloads dos modelos.

Item	Recomendacao simples
Sistema	Windows 10 ou Windows 11, 64 bits
Memoria RAM	16 GB no minimo; 24 GB ou mais e melhor
Espaco livre	Pelo menos 12 GB
Processador	Intel ou AMD moderno
GPU	Opcional neste tutorial; o modo CPU e mais compativel

Importante: o modelo 1.7B e pesado. Em uma maquina com 16 GB de RAM, feche jogos, editores de video e outros programas antes de gerar audio.

Instalacao passo a passo

PASSO 1

Instale o Python

Acesse python.org/downloads/windows. Baixe o instalador de 64 bits. Na primeira tela, marque Add Python to PATH e conclua a instalacao.

PASSO 2

Baixe o audio.cpp para Windows

Acesse github.com/0xShug0/audio.cpp/releases. Abra a versao mais recente e baixe o arquivo com nome semelhante a audio-vX.X.X-bin-windows-x64-cpu.zip.

PASSO 3

Crie uma pasta simples

Crie a pasta C:\voz-ia. Extraia todo o ZIP do audio.cpp dentro dela. Evite nomes com acentos ou caminhos sincronizados pelo OneDrive, pois alguns executaveis podem ter dificuldade para ler esses caminhos.

Ao final, confira se existe:

```
C:\voz-ia\audiocpp_server.exe
```

PASSO 4

Abra o PowerShell na pasta

Abra a pasta C:\voz-ia no Explorador. Clique na barra de endereço, digite powershell e pressione Enter.

PASSO 5

Baixe o modelo de clonagem

Copie o comando abaixo, cole no PowerShell e pressione Enter. Aguarde o download terminar.

```
python .\tools\model_manager_v2.py install --models-root C:\voice-models  
qwen3_tts_1_7b_base_q8_0
```

PASSO 6

Baixe o modelo de transcrição

Depois que o primeiro download terminar, execute:

```
python .\tools\model_manager_v2.py install --models-root C:\voice-models qwen3_asr_0_6b_q8_0
```

Os arquivos ficam em C:\voice-models. Não feche a janela enquanto houver download em andamento.

Configurando o painel

PASSO 7

Crie o arquivo painel.json

Abra o Bloco de Notas, cole o conteúdo abaixo e salve como C:\voz-ia\painel.json. Em Tipo, escolha Todos os arquivos para evitar painel.json.txt.

```
{
  "host": "127.0.0.1",
  "port": 8081,
  "backend": "cpu",
  "threads": 8,
  "lazy_load": true,
  "max_loaded_models": 1,
  "models": [
    { "id": "qwen3-tts-1.7b", "family": "qwen3_tts",
      "path":
"C:/voice-models/Qwen3-TTS-12Hz-1.7B-Base-GGUF/qwen3-tts-12hz-1.7b-base-q8_0_v2.gguf",
      "task": "tts", "mode": "offline" },
    { "id": "qwen3-asr-0.6b", "family": "qwen3_asr",
      "path": "C:/voice-models/Qwen3-ASR-0.6B-GGUF/qwen3-asr-0.6b-q8_0.gguf",
      "task": "asr", "mode": "offline" }
  ]
}
```

PASSO 8

Inicie o painel

No PowerShell aberto em C:\voz-ia, execute:

```
.\audiocpp_server.exe --config .\painel.json --ui --backend cpu --max-loaded-models 1
```

Quando aparecer a mensagem de que o servidor esta ouvindo, nao feche o PowerShell. Abra o navegador e acesse:

```
http://127.0.0.1:8081/
```

Esse endereco funciona somente no proprio computador. Ele nao envia o audio para um site externo.

Como transcrever

- Clique em ASR / Transcription.
- Escolha Qwen3-ASR 0.6B.
- Envie o arquivo de audio e informe Portuguese ou pt no idioma.
- Clique em Run e aguarde.
- Copie a transcricao produzida.

Como clonar uma voz

1	Clique em Text to speech e selecione Qwen3-TTS 1.7B Base.
2	No campo Text, escreva a nova frase que a voz devera falar.
3	Em Reference voice, envie uma gravacao limpa de 8 a 15 segundos.
4	Em Reference transcript, cole exatamente o que foi falado na referencia.
5	Use Language = Portuguese e Maximum tokens = 4096.
6	Para texto curto, desative Split and merge long text.
7	Clique em Run e aguarde o arquivo aparecer em Output.

Parametros recomendados

Campo	Valor inicial
temperature	0.8
top_k	40
top_p	0.9
repetition_penalty	1.05
do_sample	Enabled
instruct / speaker	Deixar vazio no modelo Base
Additional options	{}

Problemas comuns

O audio saiu cortado

Use referencia de 8 a 15 segundos, transcricao exata, Maximum tokens 4096 e texto curto sem divisao.

O computador travou

O modo CPU usa muita RAM. Feche outros programas. Para interromper, feche o servidor no PowerShell com Ctrl+C.

Model path does not exist

Use caminhos simples, sem acentos. Confirme que os modelos estao em C:\voice-models.

A voz ficou diferente

Grave em ambiente silencioso, com o mesmo ritmo e emocao desejados no resultado.

Nao aparece transcricao

Abra ASR / Transcription. O botao nao fica dentro da tela Text to speech.

Uso responsavel

Clone somente sua propria voz ou uma voz para a qual exista autorizacao clara. Informe ao publico quando um audio for sintetico. Nao use clonagem para se passar por outra pessoa, enganar, aplicar golpes ou criar declaracoes falsas.

Creditos e licencas

- Painel e mecanismo: audio.cpp - github.com/0xShug0/audio.cpp
- Modelos: Qwen3-TTS e Qwen3-ASR, publicados pela equipe Qwen sob Apache 2.0.
- Tutorial em portugues: Neidson Lima.

Neidson Lima

Conteudo simples, local e acessivel para todos.