

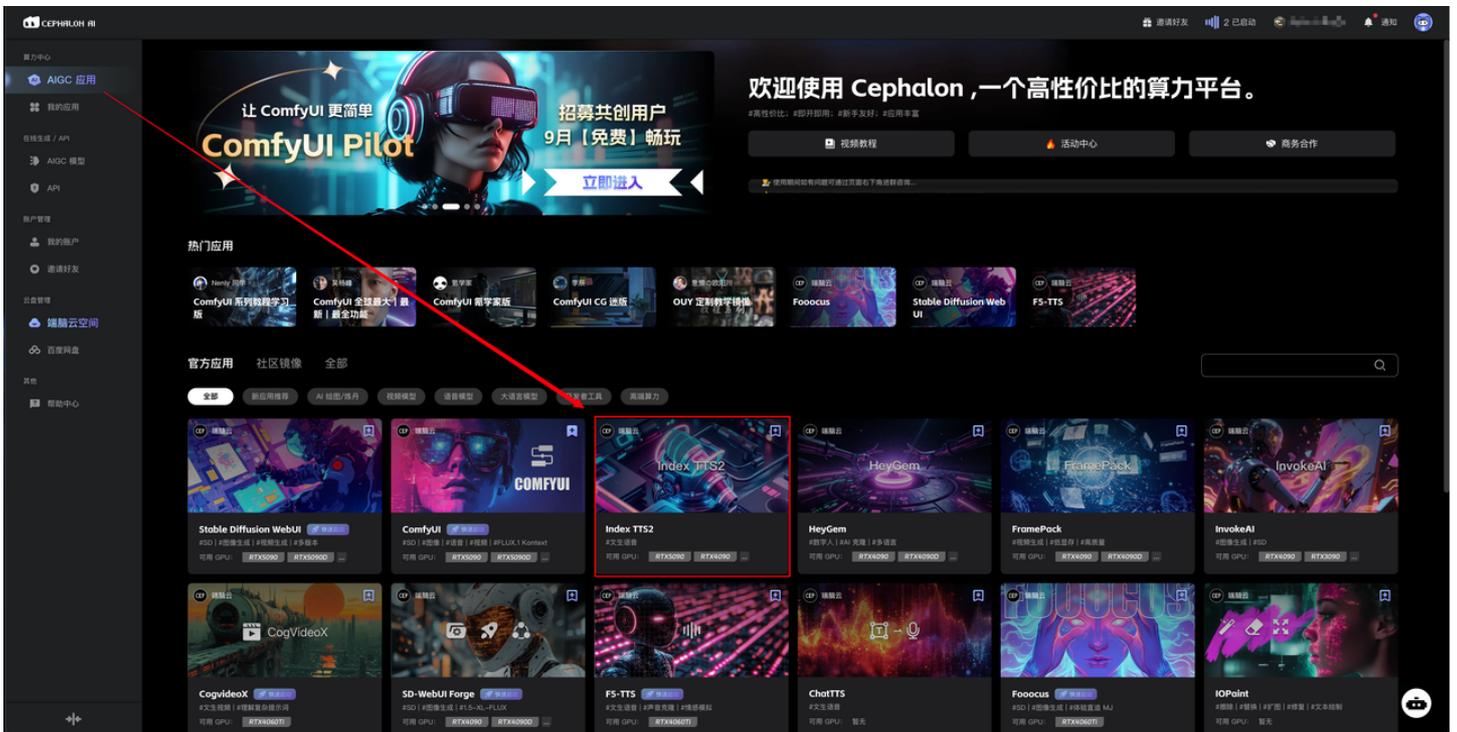
IndexTTS2 使用说明

一、简介

IndexTTS2 是哔哩哔哩开源的新一代零样本文本转语音模型，能够通过简短的参考音频瞬间克隆目标音色，并支持对生成语音的时长和情感进行精准的独立控制。

二、应用地图

端脑云首页，选择“IndexTTS2”创建应用



三、核心亮点

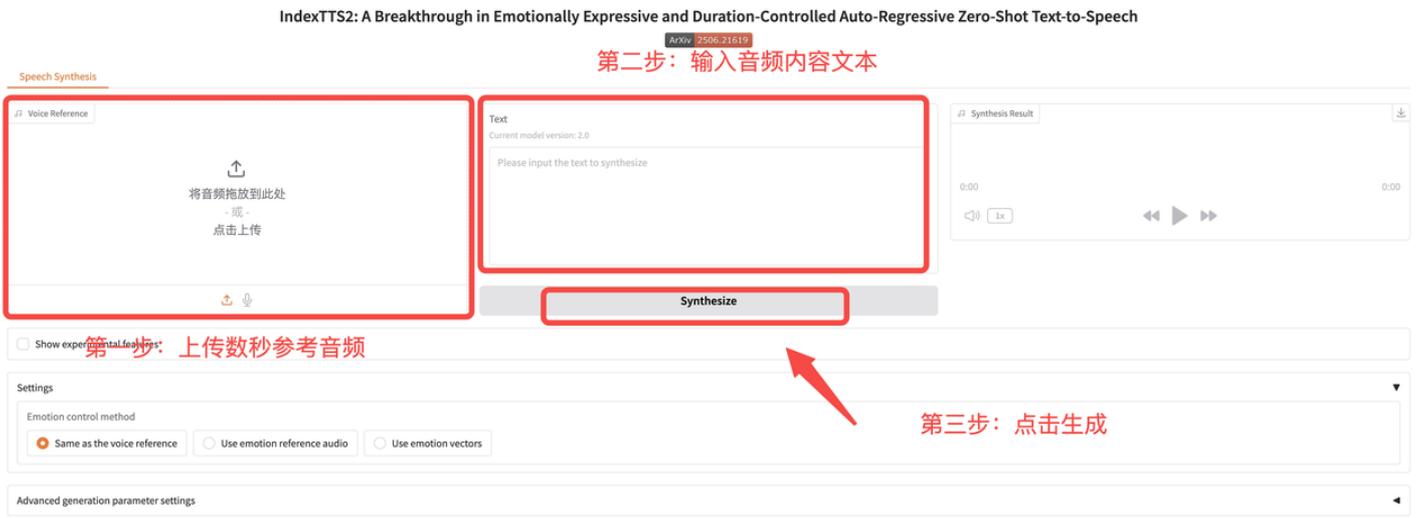
- 高效的零样本语音克隆能力：仅需**数秒参考音频**，无需任何模型微调，即可**实时生成与目标音色高度相似的语音**，实现近乎即时的声音复刻。
- 精确控制语音时长：不同于其他传统TTS模型，IndexTTS2 很创新的一点是**可以精确控制语音时长**。用户可以随意指定语音时长，IndexTTS2能生成**时长分毫不差的语音**，完美匹配视频里人物的**口型，实现音画同步**。
- 音色与情感的独立调节功能：用户可**分别指定目标音色和情感风格**，例如，使用A声音的音色，同时注入B声音的情感特征，或**直接通过文本指令**（如“欢快的”、“沉稳的”）来精确控制情感表达。

四、仓库地址

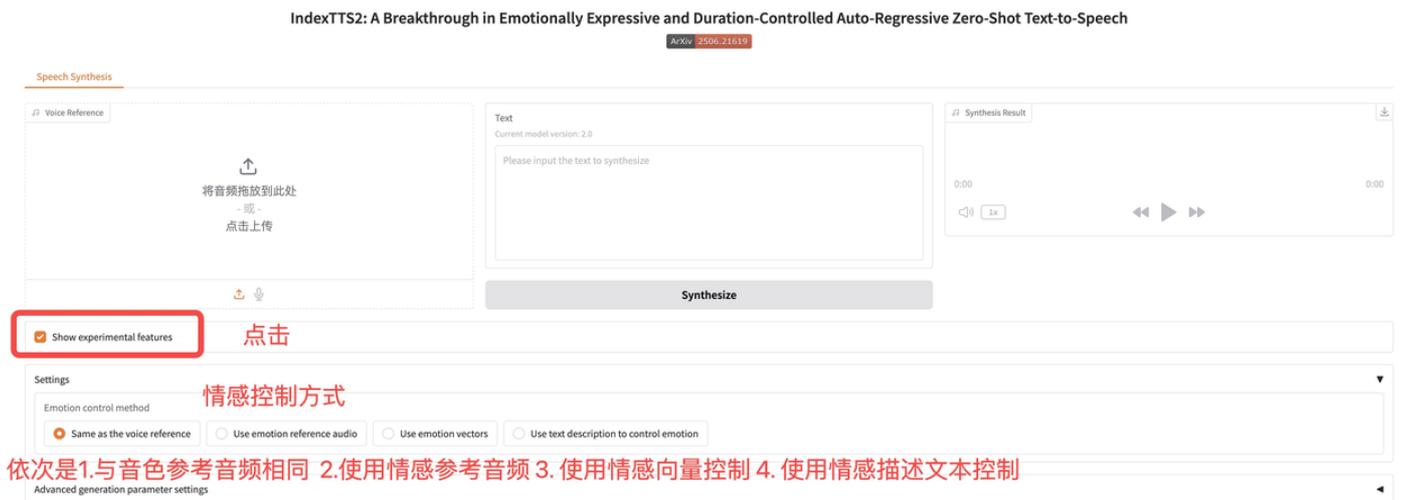
IndexTTS2的 GitHub 仓库地址为：<https://github.com/index-tts/index-tts>

五、操作指引

- 语音克隆音频生成：第一步，上传**数秒参考音频**，为获得最佳效果，最好使用清晰、高品质、无背景音等杂音的所需声音干声音频。第二步，输入想要生成的音频内容文本。第三步，点击生成按钮。



- 音频情感设置：点击 `Show experimental features`，`Settings` 处出现四个情感设置按钮，用户可以根据喜好方式设置生成音频的情感。



`Same as the voice reference`：与音色参考音频相同，默认设置。

`Use emotion reference audio`：使用情感参考音频，用户需额外上传一段情感参考音频。

`Use emotion vectors`：使用情感向量控制，用户可以通过调节不同的情绪按钮控制情感。

Randomize emotion sampling 随机情绪采样

Happy 快乐的	<input type="range" value="0"/>	Disgusted 厌恶的	<input type="range" value="0"/>
Angry 生气的	<input type="range" value="0"/>	Melancholic 忧郁的	<input type="range" value="0"/>
Sad 悲伤的	<input type="range" value="0"/>	Surprised 惊讶的	<input type="range" value="0"/>
Afraid 害怕的	<input type="range" value="0"/>	Calm 平静的	<input type="range" value="0"/>

Emotion control weight 情绪控制权重

Use text description to control emotion：使用情感描述文本控制，用户可以输入像‘很悲伤，很难过’等情绪形容词来控制情感。

- 高级参数设置：点击 `Advanced generation parameter settings`，出现高级参数设置板块，具体设置感兴趣可以参考‘视频教程推荐’。

六、视频教程推荐

https://www.bilibili.com/video/BV12gYjzxEix/?spm_id_from=333.337.search-card.all.click&vd_source=d4f6447b97dbdc894ae9db27dee092b1

说明：视频来自「[Jack-Cui](#)」 ← 点击可访问首页，视频内容不代表平台观点，如有侵权请联络平台删除