让 AI 的声音哭出来

```
{
  "request_id": "tts-emotion-experiment-001",
  "voice_id": "doubao_v3_pro",
  "text_content": {
    "text": "I never thought it would end like this... with nothing but the silence and the memories. Please, don't leave me here.",
    "emotion_params": {
      "emotion": "intense_sorrow",
      "intensity": 0.95,
      "prosody": {
        "pitch": -0.4,
        "speed": 0.7,
        "pause_duration": 1.2
      }
    }
  },
  "streaming": true,
  "format": "wav",
  "sample_rate": 24000
}
```

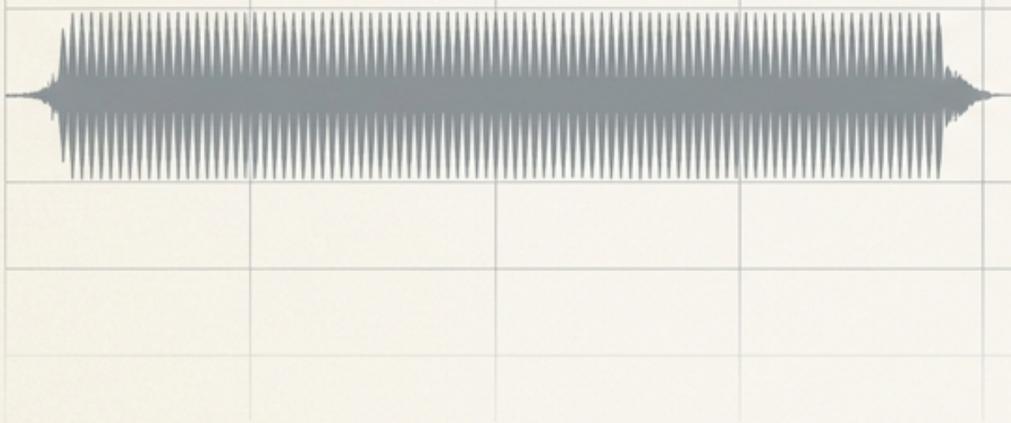
豆包 TTS 情感表现力深度实验

实验日志
/ 30 个样本
破局点

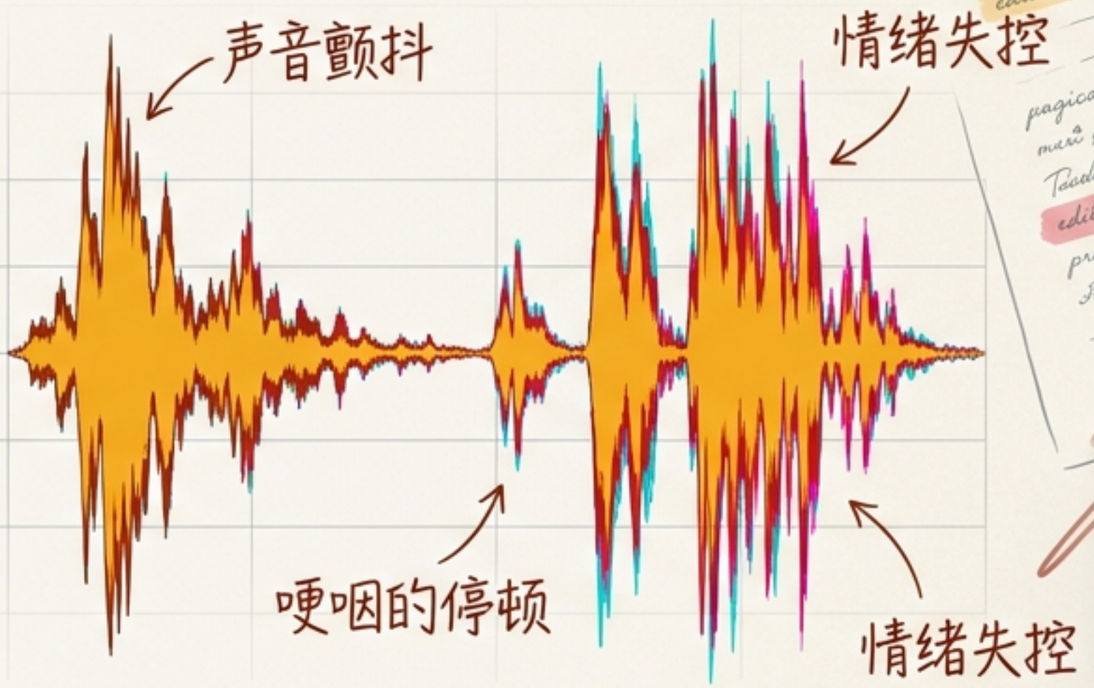
NotebookLM

我们需要真正的碎心，而不是假哭。

speed=0.8 pause=2s pitch=-1



机器的伪装：减速与停顿，听感冰冷。



人类的共鸣：让你刷到一半停下来，确认对方是否安好。

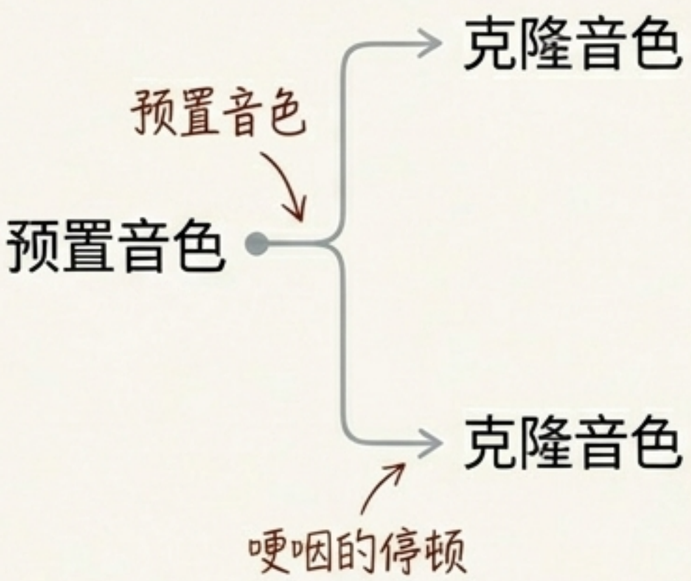
NotebookLM

实验蓝图：探寻情感的绝对边界

30

个音频样本

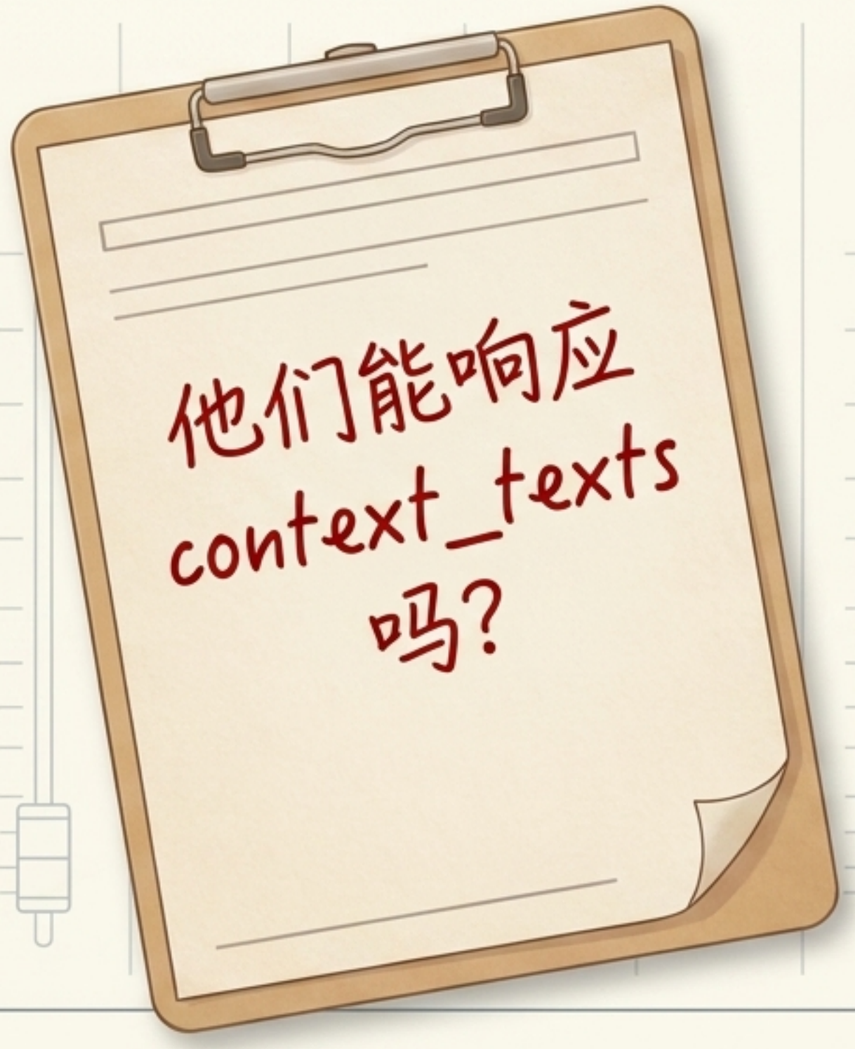
3 轮系统性压力测试



目标：抛开官方文档的表面说辞，测出到底什么管用。

NotebookLM

实验一：挑选演员



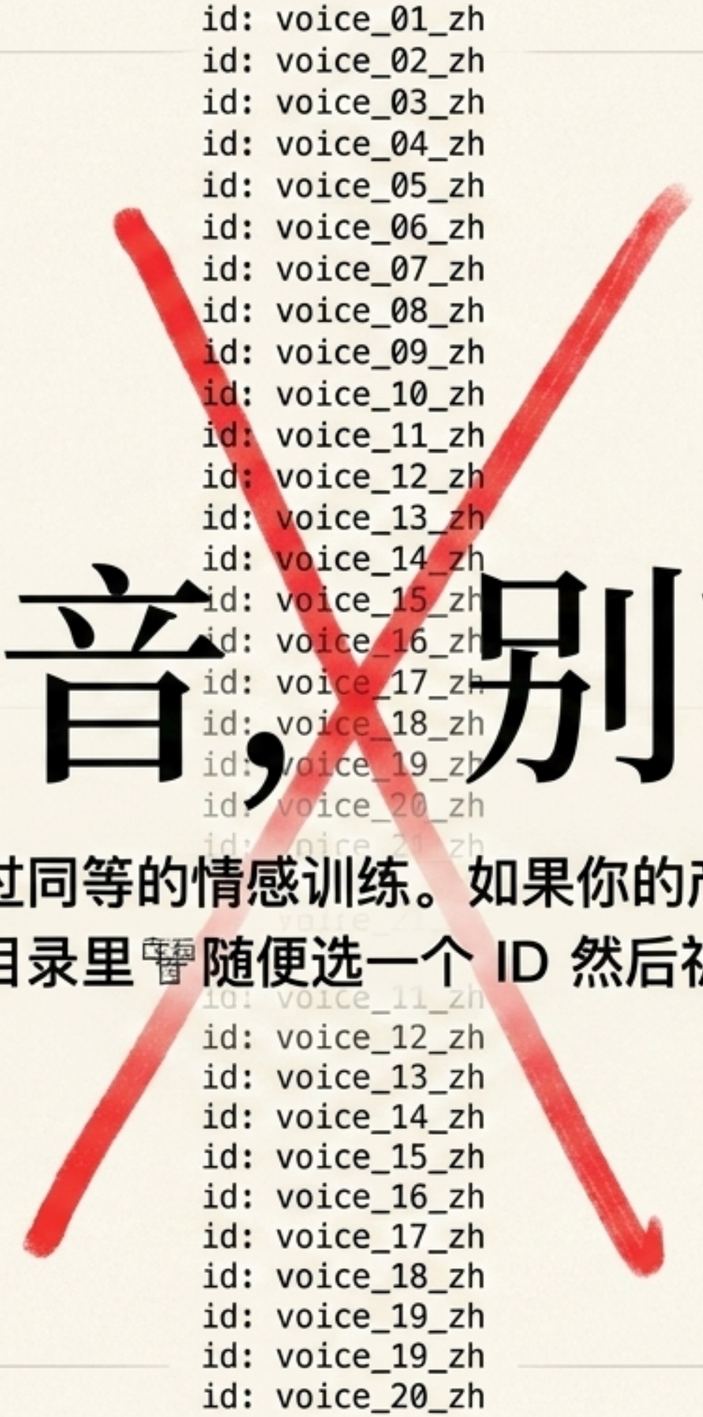
豆包 2.0 有几十个预置音色。文档说支持情感描述，但“支持”和“有反应”是两回事。有些极具表现力，有些只是木头。

NotebookLM

试镜结果：情感响应度对比

Vivi (zh_female_vv_…)	基准线女声	70%	稳定接住情感提示
刘飞 (zh_male_liufei_…)	成熟男声	60%	含蓄，伤心时有重力感
云舟 / M191	最终入选	100%	极致表现力，脆弱且心碎
桃城小天	年轻男声	30%	太轻，撑不起心碎场景

NotebookLM



“先试音，别盲选”

并非所有音色都经过同等的情感训练。如果你的产品需要情感深度，
绝对不能从目录里随便选一个 ID 然后祈祷它管用。

PLAY

NotebookLM

实验二：克隆的声音，能哭吗？

官方文档

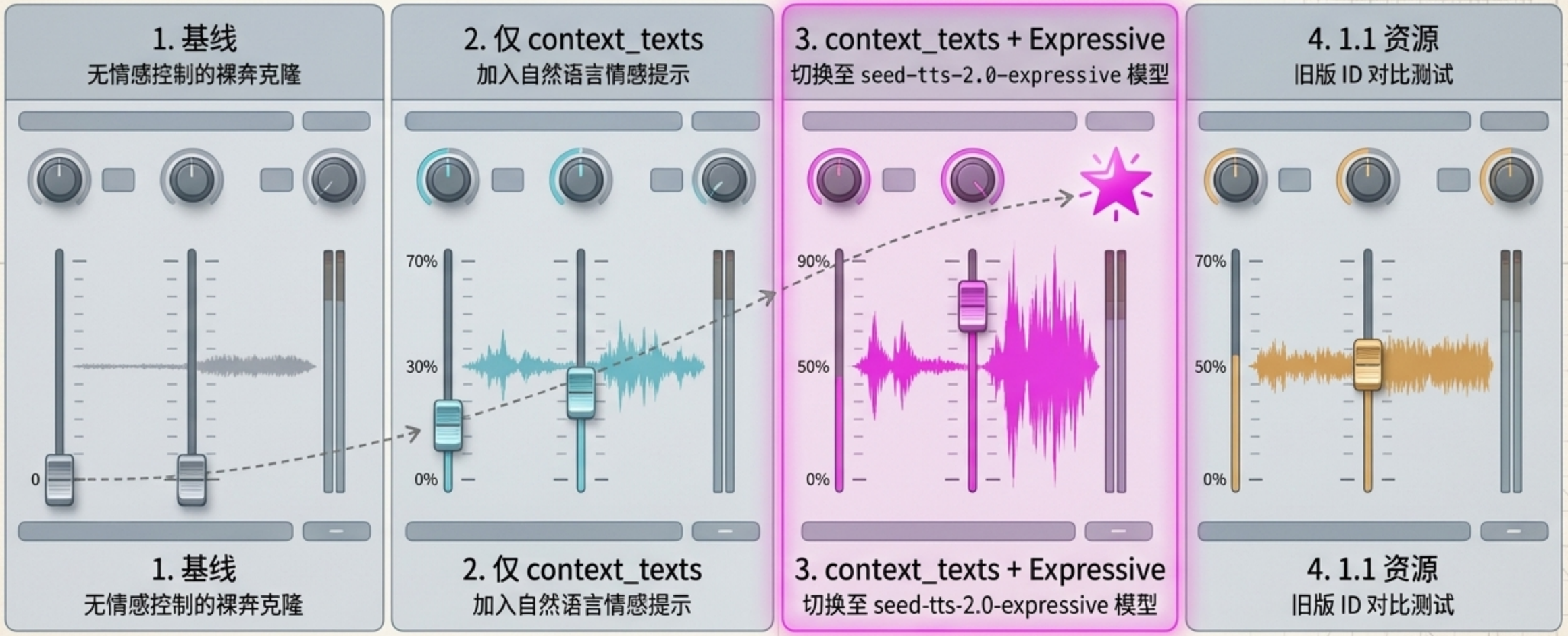
2.0 预置音色支持 context_texts。

```
{
    ...
    speaker_id: "cloned_voice_X7Y8Z9"
    ...
}
```

文档没写，但我们测了再说。

NotebookLM

克隆音色：四大控制变量实验



NotebookLM

文档没写，但确实有效！

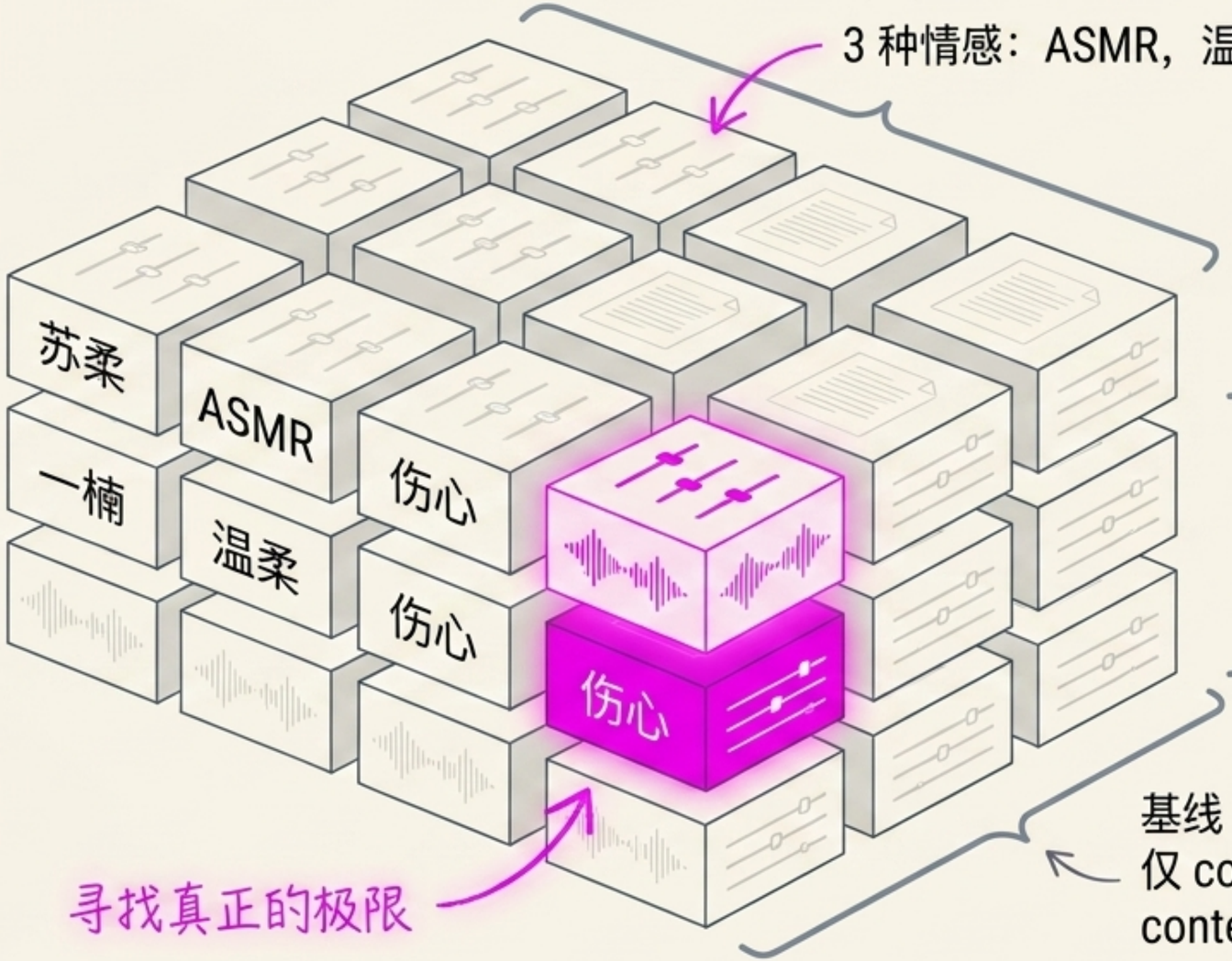
核心发现：你在使用克隆音色时，并不会被困在一个平板的声音里。克隆音色能够**完美响应自然语言的情感提示。**

Key takeaway: emotion transfer is real.

NotebookLM

Studio Daylight

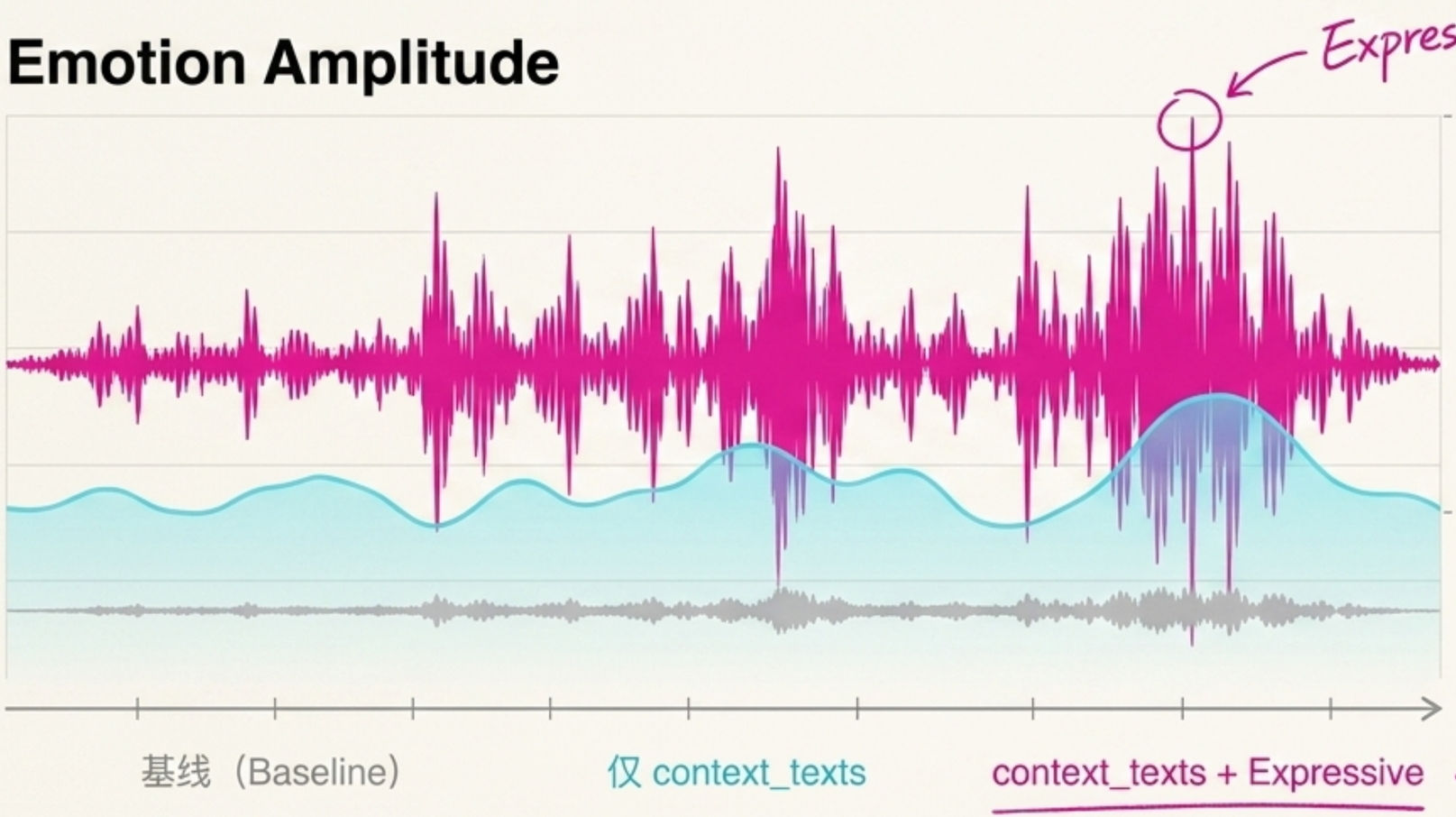
全面压力测试：18 个样本的对决



NotebookLM

这不是安慰剂：情感波幅数据分析

Emotion Amplitude



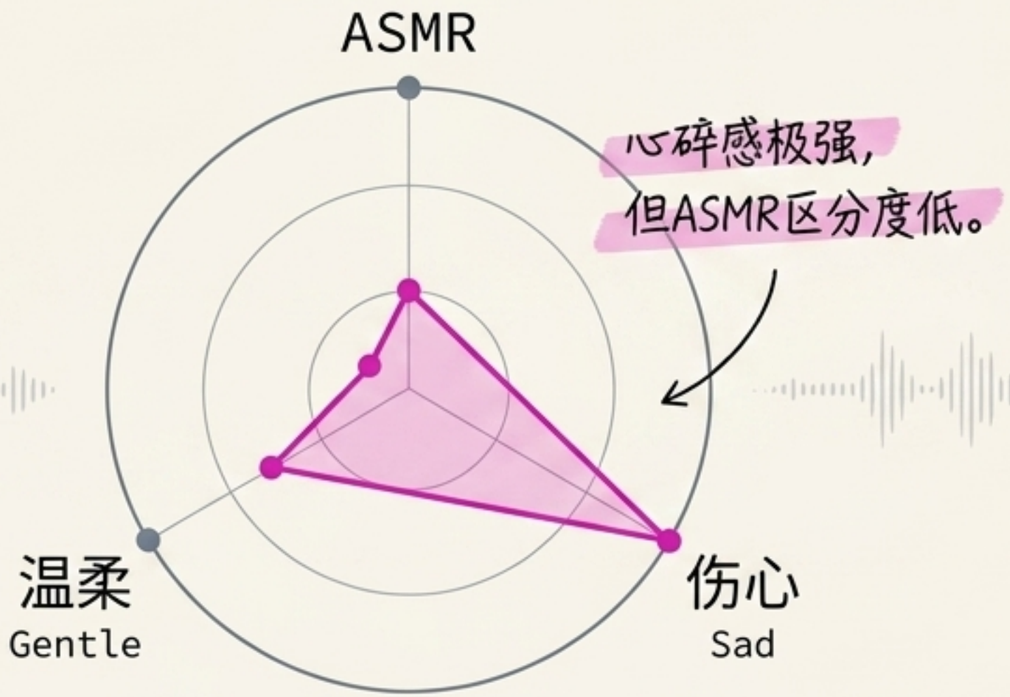
~26%

Expressive 模型在纯文本提示的基础上，增加了一层可量化的情感跳跃。

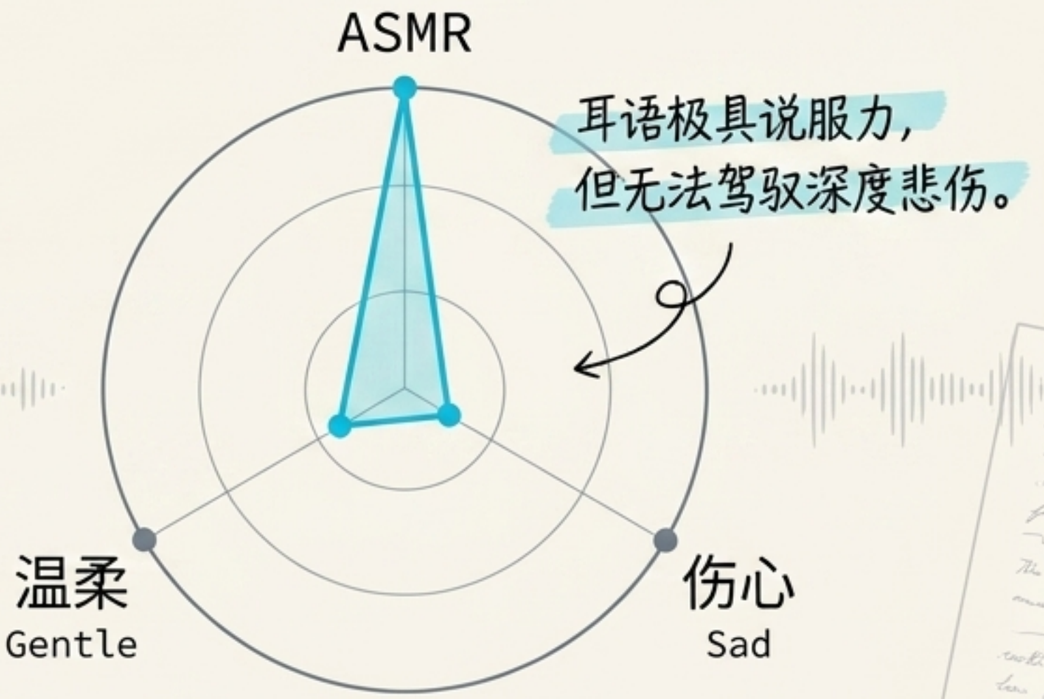
NotebookLM

COMPARATIVE WARNING: VOICE-EMOTION ASYMMETRY

苏柔（女声）



一楠（男声）



测你自己的场景。每个音色都有自己的绝对强项，没有全能王。

NotebookLM

真正的突破：别打标签，去导戏

THE CODER'S FAILURE

<emotion="sad">

No emotional nuance.

模型不响应分类标签。

THE DIRECTOR'S SUCCESS

声音在发抖，强忍着泪水，
不想让人看出来...

生动的！
(Vivid!)

有画面感
(Visuals)

模型响应的是**生动的场景描述**。

你不是在触发参数，
你是在导演一场表演。

Shift your mindset.

NotebookLM

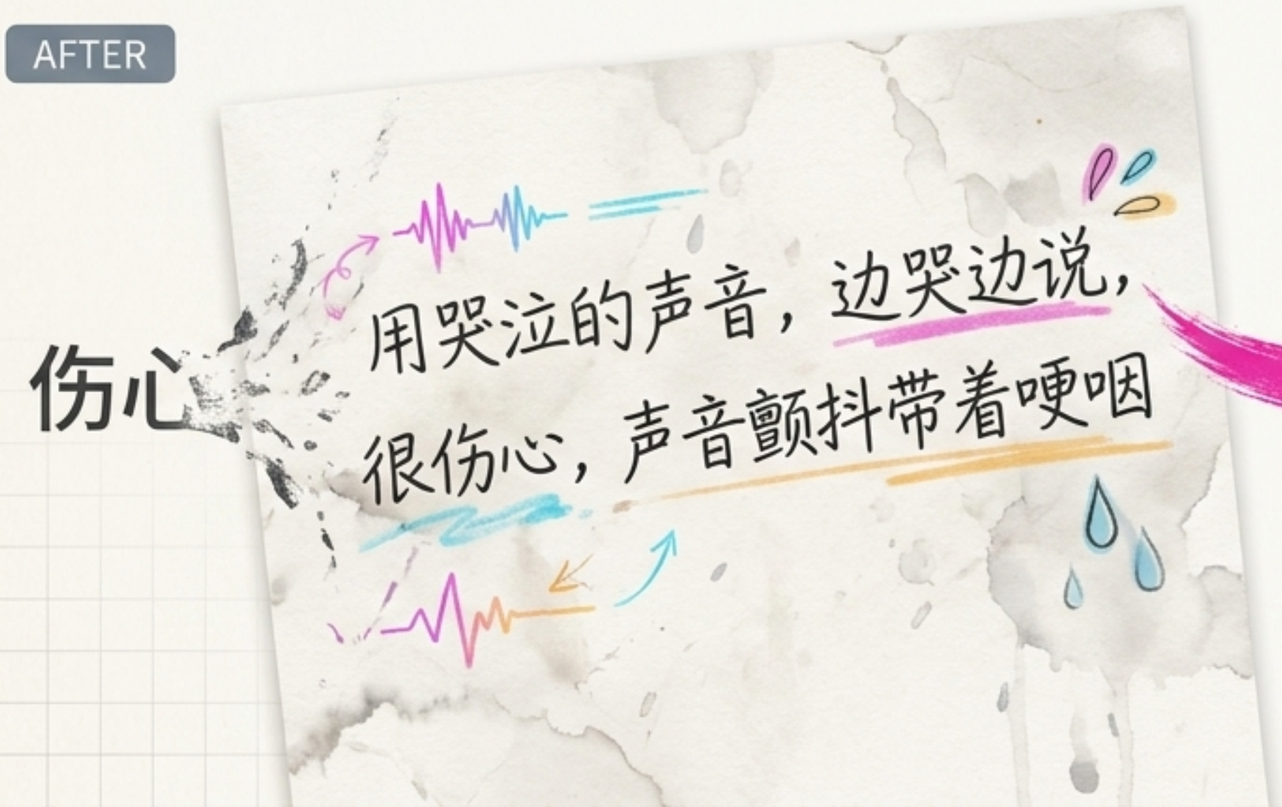
提示词的进化：从词汇到画面

BEFORE

伤心

结果：略微伤心（毫无波澜）

AFTER



结果：天和地
（震撼的真情实感）

NotebookLM

完美的导演提示词公式

Editorial editorial.

[动词 / Action]

用哭泣的声音

[身体反应 / Physicality]

声音颤抖带着哽咽

Synvenaniah

[场景描述 / Context]

像在跟男朋友撒娇

Mixing desk

用哭泣的声音，声音颤抖带着哽咽，像在跟男朋友撒娇

规律：动作 + 身体 + 场景。给模型一个要演出来的剧本，

而不是一个要选择的分类。

NotebookLM

导演的调色板（The Director's Palette）

压抑隐忍

用压抑悲伤的声音，故意克制但声音微微颤抖，不让人看出来来。

愤怒嫌弃

用愤怒嫌弃的语气，非常不满在骂人，声音拔高。

慵懒犯困

用疲惫慵懒的声音，边打哈欠边撒娇，声音软绵绵的。

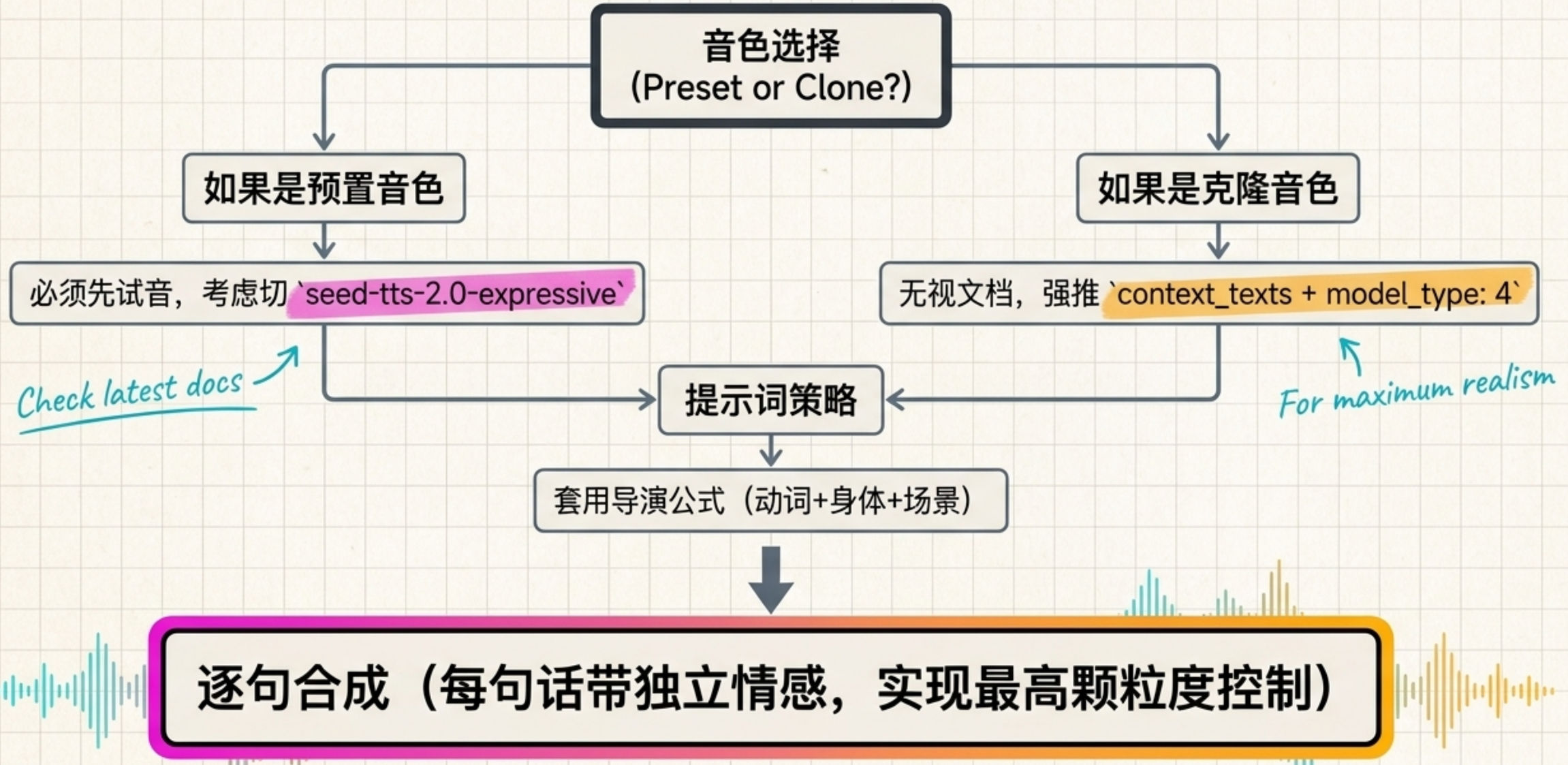
细节在于身体反应

细节在于身体反应

边打哈欠边撒娇

NotebookLM

产品构建者的最终执行路径



NotebookLM

导演不会对演员说‘伤心’。
导演会说‘你想忍住不哭但声音控制不住在发抖’。TTS 模型也一样。

AI 也有演技，前提是你给它它足够生动的剧本。

NotebookLM