

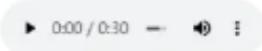
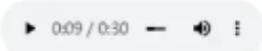
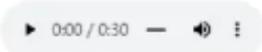
谷歌“狂飙”生成式AI赛道：将应用场景扩展到了音乐圈

1月28日消息，在生成式AI模型的赛道上，谷歌正一路“狂飙”。继文字生成AI模型Wordcraft、视频生成工具Imagen Video之后，谷歌将生成式AI的应用场景扩展到了音乐圈。

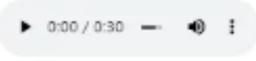
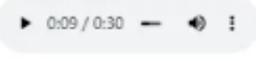
当地时间1月27日，谷歌发布了新的AI模型——MusicLM，该模型可以从文本甚至图像中生成高保真音乐，也就是说可以把一段文字、一幅画转化为歌曲，且曲风多样。

谷歌在相关论文中展示了大量案例，如输入字幕“雷鬼和电子舞曲的融合，带有空旷的、超凡脱俗的声音，引发迷失在太空中的体验，音乐的设计旨在唤起一种惊奇和敬畏的感觉，同时又适合跳舞”，MusicLM便生成了30秒的电子音乐。

从丰富的字幕生成音频

标题	生成的音频
街机游戏的主要配乐。它节奏快且乐观，带有朗朗上口的电吉他即兴重复段。音乐是重复的，容易记住，但有意想不到的声音，如铙钹撞击声或铃声。	
雷鬼和电子舞曲的融合，带有空旷的、超凡脱俗的声音。引发迷失在太空中的体验，音乐的设计旨在唤起一种惊奇和敬畏的感觉，同时又适合跳舞。	
上升合成器正在演奏带有大量混响的琶音。它由打击垫、次低音线和软鼓支持。这首歌充满了合成器的声音，营造出一种舒缓和冒险的氛围。它可能会在音乐节上播放两首歌曲以进行积累。	
慢节奏、贝司和鼓主导的雷鬼歌曲。持续的电吉他。带有铃声的高音手鼓。人声轻松有悠闲的感觉，很有表现力。	

又如以世界名画《跨越阿尔卑斯山圣伯纳隘口的拿破仑》为“题”，MusicLM生成的音乐庄重典雅，将冬日的凌厉肃杀和英雄主义色彩体现地淋漓尽致。写实油画之外，《舞蹈》《呐喊》《格尔尼卡》《星空》等抽象派画作均可为题。

画名及作者	绘画图像 (来自维基百科)	绘画说明	生成的音频
记忆的永恒——萨尔瓦多·达利		“他融化的时钟图像嘲笑时间的僵硬。手表本身看起来像软奶酪——事实上，根据 Dalí 自己的说法，他们的灵感来自于吃了卡门·福尔的甜食的幻觉。在图片的中央，在其中一只手表的下面，是一张扭曲的人脸。盘子里的蚂蚁代表腐烂。” 格罗姆利, 杰西卡, “记忆的持久性”, 大英百科全书, 2022年4月14日.	
拿破仑穿越阿尔卑斯山 - 雅克-路易·大卫		“这幅作品展示了拿破仑和他的军队于 1800 年 5 月通过大圣伯纳山口穿越阿尔卑斯山的真实穿越的强烈理想化景象。” 通过<a>维基百科	
格尔尼卡 - 巴勃罗·毕加索		“这幅灰色、黑色和白色的画作画在一块高 3.49 米、宽 7.76 米的画布上，描绘了暴力和混乱造成的苦难。画面中突出的是一匹被刺伤的马，一头公牛，尖叫的妇女，一个死去的婴儿，一个被肢解的士兵，还有火焰。” 通过<a>维基百科	
星夜 - 文森特·梵高		“星夜之夜 (荷兰语: De sterrennacht) 是荷兰后印象派画家文森特·梵高的一幅布面油画。作于 1889 年 6 月，描绘了他在圣彼得堡精神病院的收容所东窗外的景色 - Rémy-de-Provence，就在日出之前，加上一个想象中的村庄。” 通过<a>维基百科	

MusicLM甚至来个音乐串烧，在故事模式下将不同风格的曲子混杂在一

起。即便要求生成5分钟时长的音乐，MusicLM也不在话下。

故事模式

音频是通过提供一系列文本提示生成的。这些影响模型如何继续从先前的标题中派生的语义标记。

The screenshot displays the 'Story Mode' interface for MusicLM. It features a list of text prompts on the left and corresponding audio playback controls on the right. The prompts are organized into three sections, each with a play button and a progress bar. The first section includes prompts like '冥想时间(0:00-0:15)', '醒来时间(0:15-0:30)', '跑步时间(0:30-0:45)', and '100% 付出时间(0:45-0:60)'. The second section includes '电子游戏中播放的电子歌曲(0:00-0:15)', '河边播放的冥想歌曲(0:15-0:30)', '火(0:30-0:45)', and '烟花(0:45-0:60)'. The third section includes '爵士歌曲(0:00-0:15)', '流行歌曲(0:15-0:30)', '摇滚歌曲(0:30-0:45)', '死亡金属歌曲(0:45-1:00)', '说唱歌曲(1:00-1:15)', '弦乐四重奏与小提琴(1:15-1:30)', '史诗电影配乐与鼓(1:30-1:45)', and '苏格兰民歌与传统乐器(1:45-2:00)'. The audio controls show a play button, a progress bar (e.g., 0:00 / 1:00), a volume icon, and a settings icon.

另外，MusicLM具备强大的辅助功能，可以规定具体的乐器、地点、流派、年代、音乐家演奏水平等，对生成的音乐质量进行调整，从而让一段曲子幻化出多个版本。

MusicLM并非第一个生成歌曲的AI模型，同类型产品包括Riffusion、Dance Diffusion等，谷歌自己也发布过AudioML，时下最热门的聊天机器人“ChatGPT”的研发者OpenAI则推出过Jukebox。

MusicLM有何独到之处？

它其实是一个分层的序列到序列（Sequence-to-Sequence）模型。根据人工智能科学家Keunwoo Choi的说法，MusicLM结合了MuLan+AudioLM和MuLan+w2b-Bert+Soundstream等多个模型，可谓集大成者。

其中，AudioLM模型可视作MusicLM的前身，MusicLM就是利用了AudioLM的多阶段自回归建模作为生成条件，可以通过文本描述，以24kHz的频率生成音乐，并在几分钟内保持这个频率。

相较而言，MusicLM的训练数据更多。研究团队引入了首个专门为文本-音乐生成任务评估数据MusicCaps来解决任务缺乏评估数据的问题。MusicCaps由专业人士共建，涵盖5500个音乐-文本对。

基于此，谷歌用280000小时的音乐数据集训练出了MusicLM。

谷歌的实验表明，MusicLM在音频质量和对文本描述的遵守方面都优于以前的模型。

不过，MusicLM也有着所有生成式AI共同的风险——技术不完善、素材侵权、道德争议等。

对于技术问题，比方说当要求MusicLM生成人声时，技术上可行，但效果不佳，歌词乱七八糟、意义不明的情况时有发生。MusicLM也会“偷懒”——起生成的音乐中，约有1%直接从训练集的歌曲中复制。

另外，由AI系统生成的音乐到底算不算原创作品？可以受到版权保护吗？能不能和“人造音乐”同台竞技？相关争议始终未有一致见解。

这些都是谷歌没有对外发布MusicLM的原因。“我们承认该模型有盗用创意内容的潜在风险，我们强调，需要在未来开展更多工作来应对这些与音乐生成相关的风险。”谷歌发布的论文写道。

本文链接：<https://dqcm.net/zixun/16750011596760.html>