

# AI音乐应用产业报告

# 目 录

01 技术篇

02 产业篇

03 展望篇

04 产品案例

## 核心结论

- ◆ 音乐符号模型打开AI音乐生成的大门，音频模型找到AI落地音乐应用的方式
- ◆ 音频路线成为研究热点，为应用爆发奠定技术基础
- ◆ 生成式AI改变音乐产业，简化音乐制作流程，提供音乐创作新思路
- ◆ 流媒体平台或成为AI音乐商业化中最确定的一方，传统的音乐工程在此轮变革中是受冲击最大的一方
- ◆ 数据是技术迭代和商业化过程中最关键的一环，生成音乐中对情感表达的把控是产品迭代的关键

01

技术篇

定义：AI音乐生成是指通过对大量音乐数据进行学习和分析，掌握音乐的基本规律和风格特征，从而创作出音乐片段或完整的音乐。

- 谷歌Megenta项目  
音乐生成进入神经网络时代
- OpenAI发布JukeBox  
奠定了Transformer文生音乐的基本架构；  
采用压缩技术让Jukebox能够处理更复杂的音乐，更好地理解音乐文本
- 谷歌发布MusicLM  
2023年1月，谷歌发布MusicLM，搭建在之前发布的语音模型AudioLM之上
- Meta入局  
2023年6月，Meta推出AudioCraft系列产品，包括负责生成音效的AudioGen，生成音乐的MusicGen以及一个全新的压缩编码方式Codec
- Suno  
音乐生成领域的「集大成者」，实现AI生成完整歌曲；  
2024年3月发布V3模型
- Udio  
4月11日，AI音乐生成工具Udio完成测试，正式上线
- 天工SkyMusic  
4月，昆仑万维发布基于「天工3.0」打造的「天工SkyMusic」产品

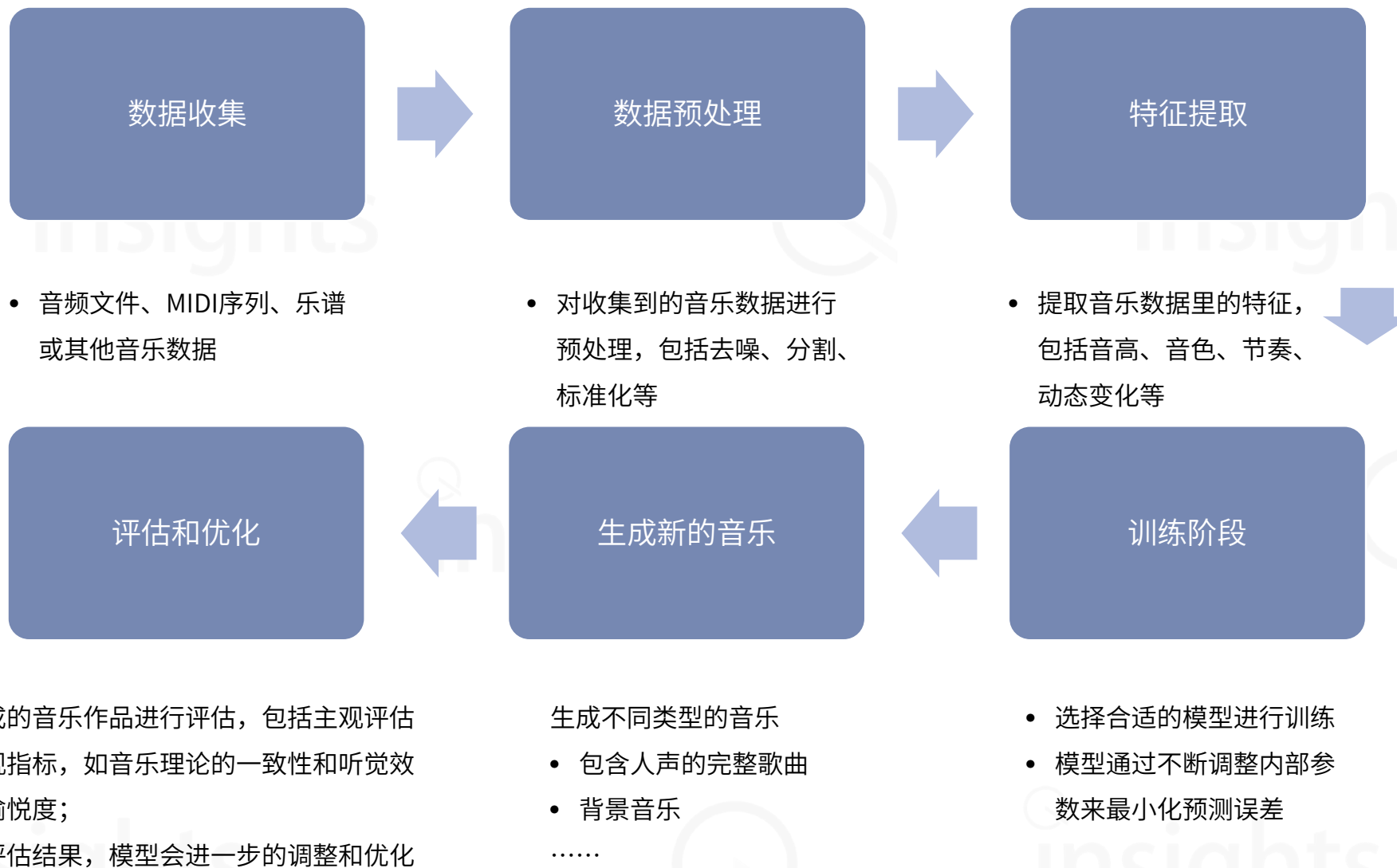
2016

2020

2023

2024

# AI音乐生成的基本流程



## 两种技术路线

音乐符号模型打开AI音乐生成的大门，音频模型找到AI落地音乐应用的方式

符号模型的研究要早于音频模型的研究，但符号模型生成的音乐在最终效果上存在明显的机械、不自然特征，因此在应用侧难以落地。而音频路线在近两年取得了突破性进展，生成的作品更加流畅自然。因此，这条路线成为目前业内更受欢迎的一条路线。

此外，音频模型路线因其生成音乐的完整性更高，且能够直接生成歌曲，对于用户来说可以做到零门槛创作，在商业化上也更具潜力。

音频生成模型

- 原理  
直接用海量音频数据训练模型，使用深度神经网络和VAE或Transformer等方法，端到端直接生成音乐

- 局限  
算力成本高昂；  
音频生成的侵权风险相较于符号生成更大（后者学习乐理规则，前者直接模仿成品音乐）

符号生成模型

- 原理  
提取出音频中的各类音乐信息，如歌词、旋律、乐器等，并对这些信息进行标注，拆分成音乐的各个环节来生成词、曲等

- 局限  
生成的是音乐的不同部分而非成品音乐，需要通过传统的音乐制作流程，将各部分融合起来，制作成音频内容

### 音频+音乐符号协同生成

- 可根据文本直接生成音频
- 乐谱生成，并且支持对已生成的乐谱进行自动编辑操作
- 案例：  
腾讯的「琴乐大模型」，同时支持文本生成音乐，以及乐谱生成

# 音频模型的代表

音频路线成为研究热点，为应用爆发奠定技术基础

MusicLM在跨模态创作和个性化音乐生成上表现出色，而MusicGen则在音乐质量和风格适应性方面具有优势。MusicLM的出现标志着音乐生成技术向更高级别的艺术创作和智能化方向发展，而MusicGen则代表了音乐生成技术在音乐制作和质量方面的进一步提升。

MusicLM是一个基于Transformer架构的模型。

MusicLM的一个显著特点是能够生成多音轨作品，生成的音乐更具艺术性和个性化。

MusicGen同样基于Transformer架构，但更侧重于生成高质量的音乐样本。

MusicGen能够生成具有高度复杂性和多样性的音乐作品，同时保持音乐的结构和风格一致性，生成的音乐听起来更加自然。

谷歌  
MusicLM

多模态能力

结合了文本和音乐两种模态，能够实现跨模态的创意表达

长距离依赖处理

模型能够捕捉音乐序列中的长距离依赖关系，生成连贯的音乐作品

上下文理解

通过文本描述，模型能够更好地理解用户的需求和创作意图

Meta  
MusicGen

高质量音乐生成

模型能够生成具有丰富和声和旋律的音乐，接近专业音乐作品的水准

风格适应性

能够学习和模仿各种音乐风格，从而生成符合特定风格的音乐作品

高效生成

通过优化模型结构和训练过程，MusicGen能够快速生成音乐，提高创作效率



02

产业篇

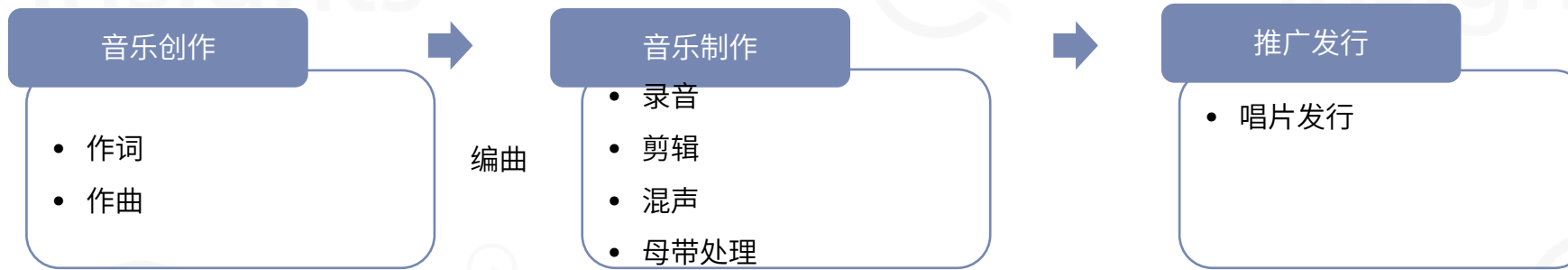
# AI改变音乐产业

## 音乐制作流程简化，提供创作新思路

音乐产业的核心动力来自于创作部分，属于音乐从0到1的阶段；而音乐制作是音乐从1到100的过程，包含大量的音乐工程。

音乐产业在进入工业化阶段后，存在供过于求，创新力不足等问题。

AI音乐生成的出现，为音乐产业带来新的创作思路，让越来越多的人加入到音乐创作当中。此外，AI对于音乐产业的一大变革在于音乐流程的极大简化。原先不同步骤的工序需要不同的人/团队来完成。AI出现后，一个人可以完成所有工序。其中，生成式AI更多运用在创作部分，在制作和发行流程中，也会加入其他AI技术。本报告主要聚焦于生成式AI对音乐产业中部分环节的变革。



传统方式

通过学习大量乐理知识来形成对音乐的理解，花费大量时间来构思和设计音乐结构

AI

通过学习海量音乐数据来掌握音乐的基本规律，并进行创作

音频工程师、母带工程师等对音乐进行处理

作为自动化音频处理工具，保证音乐最终的质量

由唱片公司代理，承包艺人的宣发工作

艺人可以在流媒体平台发布自己的专辑或单曲，平台为用户匹配适合的歌曲风格，并提供个性化音乐推荐

# AI音乐现状

## AI制作走向成熟，AI创作处于发展上升期

AI在音乐制作中的应用已经相对成熟，在母带处理、混音等工程含量较高的环节已经的到很好的应用，如Landr通过AI进行母带处理并由此形成稳定的商业模式。然而，在音乐创作中，AI的应用处在早期阶段。Suno属于这个赛道上第一家出圈的公司。昆仑万维、网易天音等是中国市场中的领先者。

这份报告在产品侧主要聚焦于生成式AI在音乐创作上的应用。

音乐创作

在音乐创作中，AI不仅需要理解和模仿复杂的音乐结构、和声、旋律、节奏等元素，还需要让人声和乐器完美结合

具体体现

需要音乐基础

AI编曲

创作各种风格的编曲，提供音乐编排等建议

AI作曲

根据用户提供的参数自动生成旋律、和弦等

无需音乐基础

AI作词

选择场景、心情、环境等描述情感/氛围的关键词，或输入创作灵感，AI一键生成歌词

一键生成

输入提示词生成音乐，大致分为两类：  
1) 一键生成完整音乐  
2) 一键生成完整音轨

音乐制作

音频信号处理

自动去除音频中的杂音、噪音，音频压缩等

混音处理

根据创作者的要求，自动平衡轨道间的音量，调整均衡器设置

母带处理

自动检测音乐特点，并自动应用适当的母带处理效果

## AI音乐创作

## AI音乐制作

一键生成

SUNO udio Stable Audio

Mubert CASSETTE

天工 天音 BGM猫 MUSICHERO

海绵音乐 万兴喵影

AI作曲

amper Aiva boomy

soundful Loudly

SOUNDRAW ecret music

AI作词

sudowrite moises 歌

音频处理

audio descriipt

audioshake Indie

混音处理

IZOTOPE RipX DAW

母带处理

LANDR MASTERED

AI Mastering

BandLab Mastering

# 一键生成类产品

产品密集发布，初步实现「人人皆可创作」

今年3月，Suno发布V3模型。用户只需给出风格和主题词，几秒钟便可以生成一首两分钟的原創歌曲。一个月后，另一款有相似功能的产品Udio也正式推向市场。

「一键生成」成为了AI音乐市场最热的话题，中国市场中网易、昆仑万维、腾讯等公司也推出了音乐生成产品。生成式AI在音乐领域，开启了「人人皆可创作」的时代。产品的可观表现，也激发了资本对于生成式AI落地的信心。Suno在5月22日完成A轮1.25亿美元融资，Udio在4月正式上线之际也宣布完成1000万美元的种子轮融资。

海外产品	产品是什么	产品特点
Suno	由AI驱动的音频和音乐生成工具	高品质、多语言、完整性高的歌曲生成
Udio	由AI驱动的音乐创作和分享平台	在合成人声中捕捉情感的能力
Stable Audio	AI生成音乐平台	生成长度可调节
CassetteAI	AI驱动的音乐生成平台	一键生成完整音轨

本土产品	产品是什么	产品特点
网易天音	使用AI进行编曲、作词或一键生成音乐	支持一键生成、AI作曲/AI作词
天工SkyMusic	一键生成音乐平台	作品的情感表达力丰富，人声逼真
海绵音乐	字节跳动推出的免费AI音乐创作和分享平台	音乐风格更符合国人喜好
BGM猫	灵动音推出的一键生成背景音乐产品	音乐高能点可调节

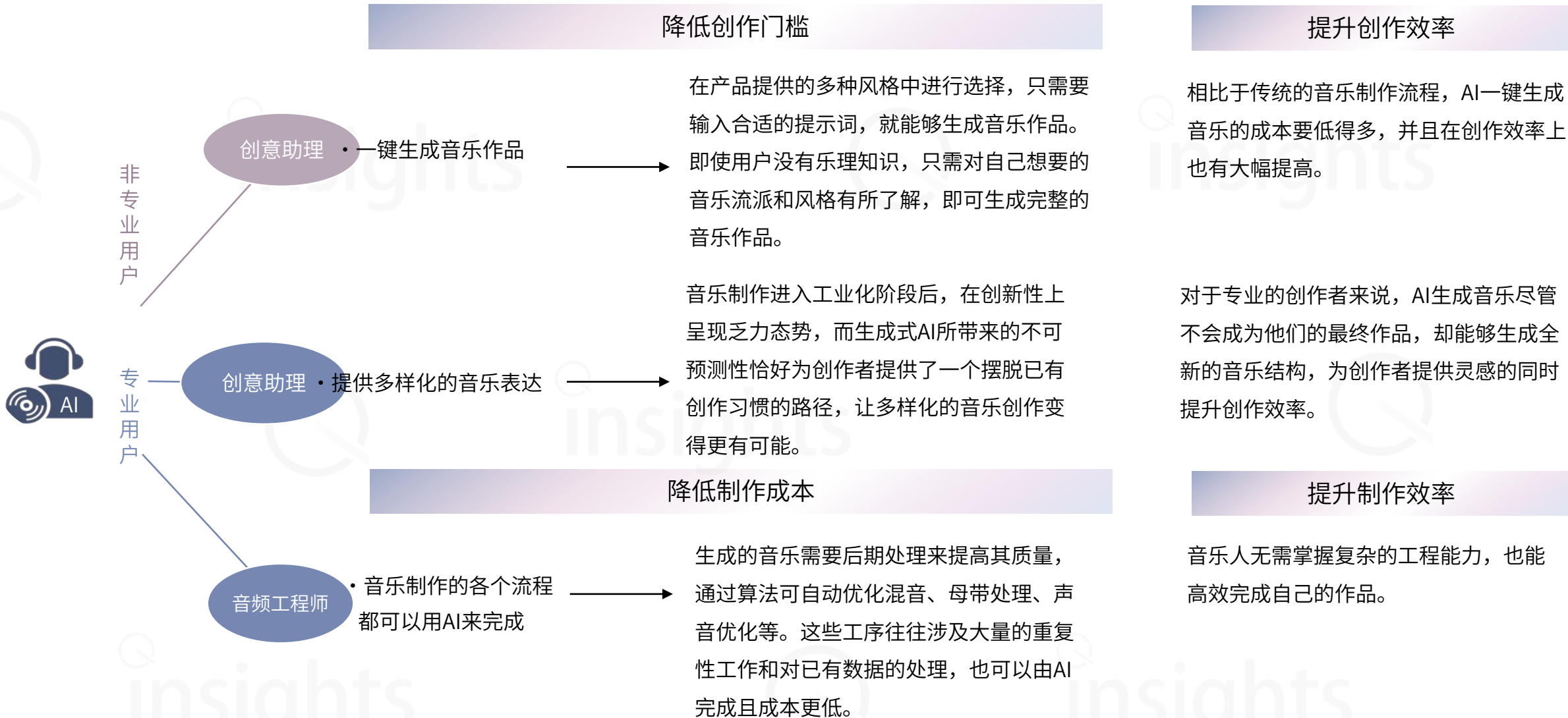
Suno最新发布的V3.5，已经能够完成4分钟的音乐创作，在歌曲的完整度和情感表现上，都较V3.0版本有所提升。

在中文歌曲的呈现上，Suno等海外产品在人声上会出现明显的瑕疵（如声音不够稳），而中国本土的音乐生成软件不论是对中文歌的旋律、节奏等的把控，还是对歌词的理解上，都有更好的表现力。也因此，本土产品在生成音乐的曲风上更加符合本土市场需求。

# 一键生成类产品对音乐人群的作用

专业人士的「降本增效」工具，爱好者的音乐入门产品

对于非专业人士来说，AI生成音乐主要解决了缺乏乐理的难题。而对于专业人士来说，目前AI更多是在制作环节帮助音乐人更高效地完成工作。



# 应用场景-影视

## 短剧市场将率先接受AI音乐成品，电影市场中AI仅用于音乐生产的部分环节

AI音乐在影视上的应用可大致分为两个方向：电影配乐和短剧音乐。电影配乐既是观众情绪的助推器，也是电影剧情的“提示词”，这意味着AI想要达到影视配乐的标准，需要对电影从整体和细节上都准确把握。目前,生成式AI类产品尚无法做到这点。在电影配乐环节，生成式AI更多地是充当灵感激发和效率提升的作用，在创作环节还是由作曲家来把控。而在短剧领域，对于音乐的要求更简单，更加注重音乐传播度，音乐首先服务于传播效果。因此，在音乐的选择上也偏向于有记忆点，旋律简单易于传唱的类型。生成式AI已经能够很好地完成这种类型的音乐创作。正因为这样的适配度，使得AI音乐率先落地在短剧领域。

### 电影配乐

- 保持创作一致性

电影配乐需要导演与作曲家在拍摄过程中保持充分沟通。作曲家在拍摄过程中需要制作小样（demo）给到导演。传统的方式是导演与作曲家沟通音乐诉求，由后者使用MIDI技术制作小样。由于MIDI制作需要耗费一定时间，因此难以快速呈现音乐与视觉同步后的效果。使用AI音乐工具，作曲家能够根据导演诉求快速生成音乐小样来做视听同步测试，并及时调整，确保音乐跟电影在创作上的一致性。

- 提升基础工作的效率

大型电影中的配乐通常涉及交响乐，由作曲家带领团队分工协作。AI音乐工具能够完成一些基础工作，如乐器的组合、和声写作。谷歌DeepMind研发的模型Lyria能够将输入的音频转化成不同风格并加入伴奏，从而让作曲家快速甄别出最适合演奏的乐器。

### 短剧音乐

- 供给端角度

一部短剧的制作周期通常在几周左右，意味着跟剧集相关的所有内容都要在这个时间内完成，包括音乐制作。此外，由于短剧通常以小制作低成本的方式投入，可能没有足够的预算购买版权。因此，通过AI音乐平台创作与剧情相匹配的歌曲也成为短剧的选择。

- 市场需求侧

近两年短剧市场呈井喷式增长，仅2023年前8月，全国备案拍摄的微短剧有3574部。据机构数据显示，2023年国内微短剧市场规模373.9亿元。短剧市场的繁荣，也将给AI音乐带来更多市场空间。

- 用户需求侧

短剧的观众对于整体内容水平的包容度更高，AI音乐已经能够达到为短剧配乐的水平。

# 应用场景-游戏

## AI音效制作较为成熟，配乐生成最先用在休闲类游戏

音乐、音效、语言是游戏声音设计中的三个基本元素。这份报告主要讨论生成式AI对游戏中音乐和音效的影响。音效可简单分为单音轨音效和多音轨音效，前者在生成上难度最低。在音乐生成上，主要有角色定制曲和游戏主题曲两类，二者在音乐风格上有较大差异。游戏主题曲的商业价值和品牌影响力更高，其对音乐的要求更加严格，创作和制作过程也更加复杂，因此也是生成难度最高的。

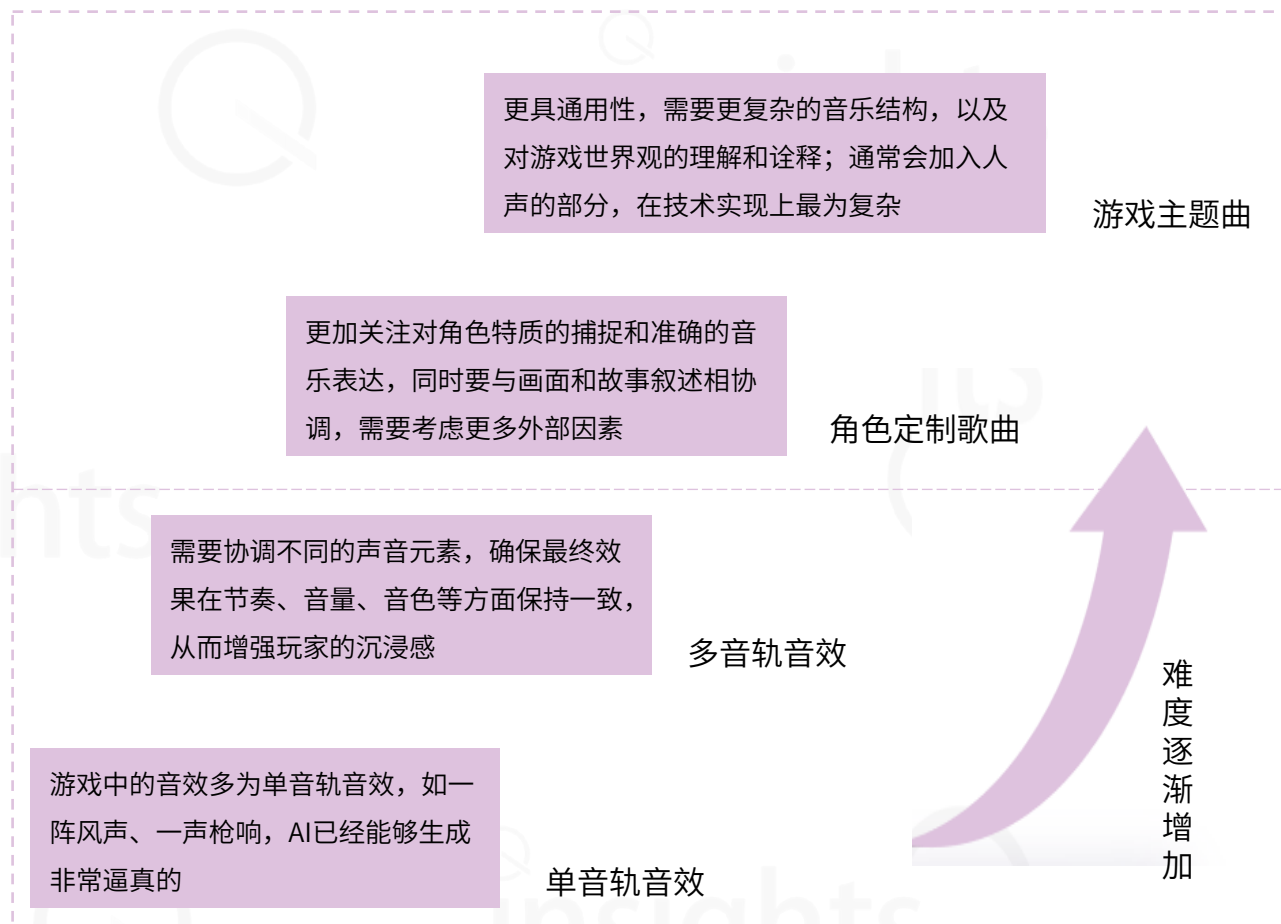
### 配乐生成

精品游戏中，配乐已经成为稳固世界观、建立游戏与玩家之间纽带的重要载体。游戏开发商会对配乐有非常高的要求，目前AI尚且无法取代，仅局限于对创作者产生一定的灵感激发作用。而对于休闲游戏，如消消乐这类低成本、轻叙事的游戏来说，对配乐没有特别高的需求且更注重成本控制。在这个细分领域中，开发者通常会向商业曲库购买音轨。有了AI音乐生成后，游戏开发者可以通过更低廉的成本来完成配乐。

### 音效生成

音效合成通常是由调音师创作demo，时间长容易出现灵感枯竭的情况。通过AI生成音效，能够为调音师提供不同风格demo的形式，从而保持创作能力在线。

AI音效生成目前在游戏领域已有一定应用，如OptimizerAI，能够根据提示词生成适合多场景的音效。



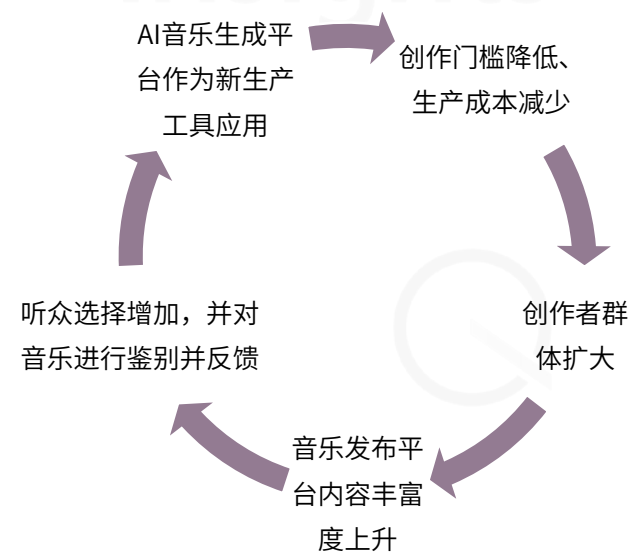
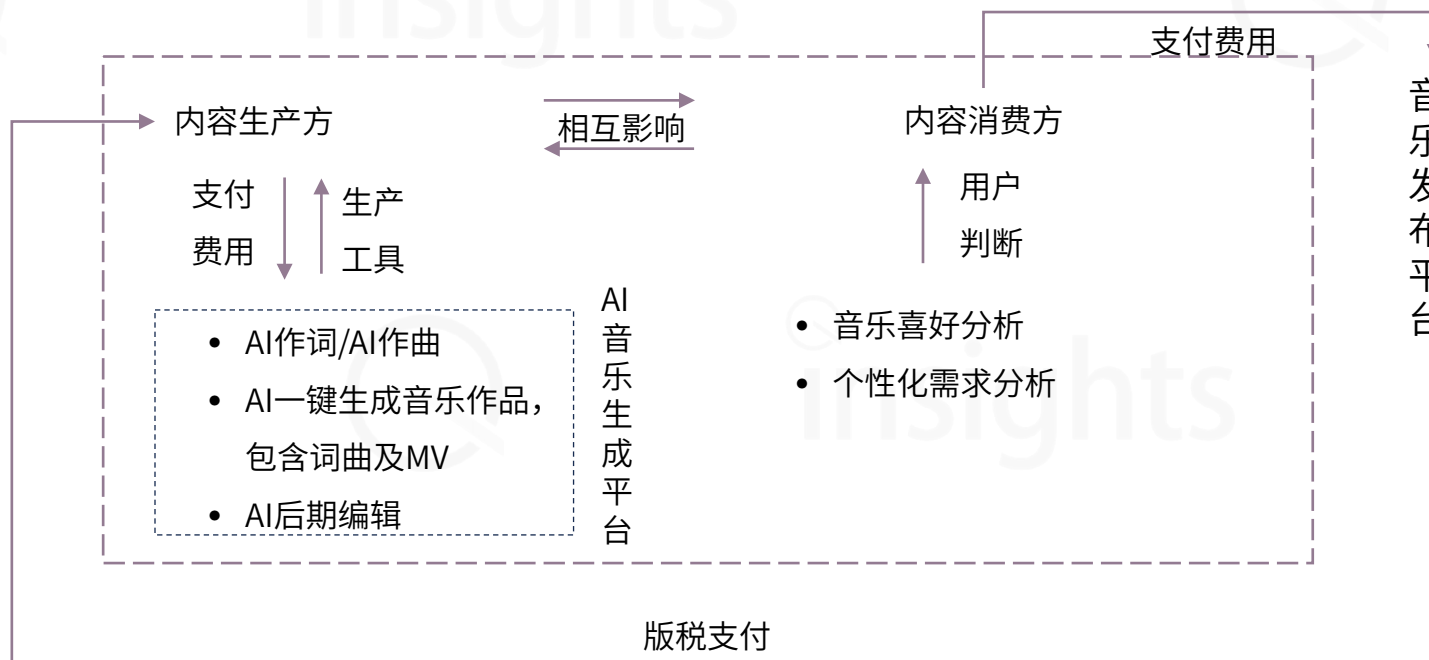


# 商业模式

流媒体平台或成为商业化中最确定的一方，传统的音乐工程在此轮变革中是受冲击最大的一方

目前，AI生成音乐能够触达的是中低端音乐市场，也就是长尾部分。在这个市场中，平台模式是最合适的商业化手段。这个阶段，付费群体主要是C端用户。个人创作者通过音乐生成软件创作音乐，在网易云音乐、Spotify等平台发布并收取版费。

此外，生成式AI工具的出现，为音乐产业带来了新的生产工具，也因此为音乐产业带来变化。首先，音乐生成软件的出现，将创作者的范围进一步扩大。中低端音乐市场的竞争加剧，高端市场受到的冲击较小。其次，音乐生成软件极大便利了创作者的创作流程，处在中间环节的工程化部分的成本得到大幅下降，这部分的工作也将减少。最后，音乐生成软件公司将会在软件的基础上，搭建音乐分享平台，以此来保持网站的访问量和使用量。



音乐生成平台和音乐发布平台属于两个不同的平台方。  
 音乐生成平台商通过会员制向创作者收取费用；  
 创作者将生成好的音乐在流媒体平台发布，可赚取版费。

AI音乐生成产品是一类新生产工具，能够帮助创作者降低创作成本，同时增加流媒体平台的作品丰富度。此外，AI音乐生成产品如Suno为音乐市场带来最大的变化是将创作者扩大为大众群体，实现人人皆可创作音乐。

# 音乐生成面临的挑战

数据是技术迭代和商业化过程中最关键的一环，情感把控是产品迭代的关键

尽管音乐生成产品已经有可观的表现，但仍然面临各方挑战。量子位智库从技术、音乐属性、商业三个层面对音乐生成类产品面临的挑战进行分析。

## 技术层面

### 数据收集

- 高质量音频数据有限，导致生成音乐的品质不稳定

### 数据处理

- 为了捕捉到声音的所有细节，音频通常以高采样率进行录制。高采样率导致音频数据量巨大，增加超长序列建模难度。生成 30s 的音乐需要生成超过一百万个采样点，而在文生文大模型中，十万个 token 的上下文支持已经属于超长上下文的范畴
- 音频数据具有时间维度上的强关联，且难以用文字精准描述

### 精细化控制

- 生成的音乐不满意只能重新生成。对于Suno这类一键生成音乐的产品来说，无法支持修改。而更专业的AI音乐生成工具，在操作性上对非专业人士并不友好

## 音乐属性层面

### 人类听感

- 人类的听觉系统异常敏感，对细节要求很苛刻，一旦AI音乐作品中出现不和谐的音符或人声，会影响整首歌的听感

### 情感表达

- 音乐是节奏、和声、音色、歌词的混合体，是一种综合的复杂感觉，很难用文字精准描述，提示词无法充分表达出创作者需要的情感/情绪

## 商业层面

### 版权归属

- 当有AI生成的音乐作品与现有作品相似时，难以界定AI的原创性以及作品的版权归属

03

展望篇

# 未来展望

## 解决情感表达匮乏是音乐生成类产品商业化的关键，用户增长依赖AI音乐的社交属性

目前，音乐生成产品在情感理解和表达上相对简单，无法完成更高水准的音乐创作。一方面是当前的模型在情感的理解上有局限。未来，跨模态理解有望帮助AI模型更好地理解人类情感。此外，AI音乐也将通过强社交属性打开音乐市场，获得更多受众。

### 音乐大模型的发展

多模态音乐模型是未来音乐生成走向音乐高端市场的必要技术支撑。目前，音乐生成仅能够满足中低端音乐市场的部分需求，而这部分市场通常是以量取胜。对于普通创作者来说，通过AI生成音乐赚钱往往是薄利多销的手段。

想要在高端市场中占有一定的份额，需要在生成音乐的质量上做进一步提升。一方面，需要让生成的音乐在局部修改后，还能够保持整体的一致性。目前音乐生成的大部分产品不支持编辑功能，如果第一遍没有达到效果，修改提示词后生成的音乐同样不可控。在B端场景中，如果客户对生成作品有新的需求，无法在现有基础上进行更改，这也阻碍了音乐生成类产品在B端的发展。

另一方面，音乐模型需要做到多模态理解，通过结合其他艺术形式，如文学（字）、视觉艺术等，进行多模态学习，从而更全面地理解和表达情感。

### AI音乐市场发展

#### 增加社交属性

AI音乐+社交的组合带来“人人皆可创作”的效应，使其成为天然的社交高地。类比于短视频时代的视频创作，当创作门槛被拉低后，每个人都能够成为创作者和分享者。创作平台不仅承担了发布渠道的功能，也衍生出分享和社交的功能。Suno创始人对于音乐生成的愿景是“让音乐成为一种社交载体，所有人随时随地制作音乐并且分享出来”。

#### 音视频结合

目前，已经有公司将音频+视频的模式落地在MV制作中，如月之暗面面向海外市场的Noisee，一款AI音乐生成视频产品。用户需要提供一段音乐，外加关键词即可生成一段MV。科大讯飞等公司也在这个方向上进行产品化尝试。音乐+视频的模式，有望打造出AI文娱类产品的一款爆品。

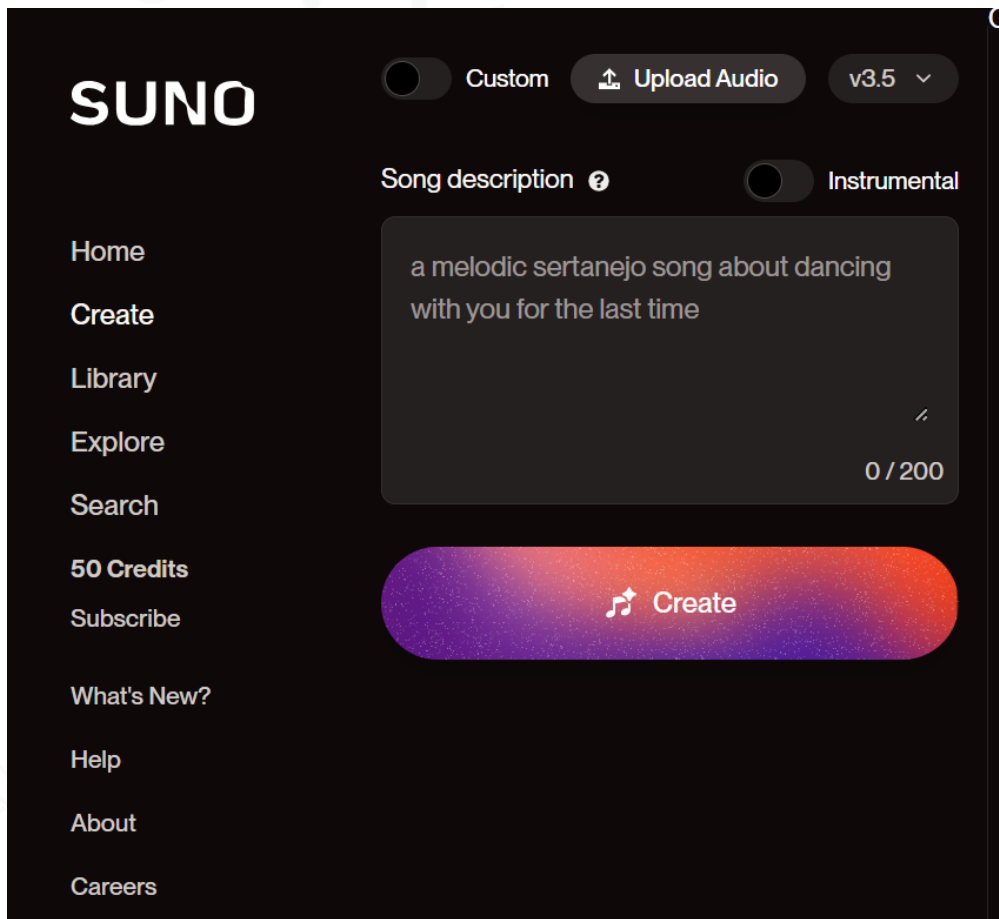
04

产品案例

2024年5月，Suno发布V3.5版本，其中最明显的改进是能够生成4分钟的歌曲。在今年3月，SunoV3发布引发了相关行业的广泛关注。用户只需输入提示词，即可生成包含人声演唱的完整音乐作品。Suno创始人将这款产品的主要用户群体定位在想要创作歌曲的音乐爱好者上，用户在没有乐理知识的情况下，通过用文字描述音乐风格及主题的方式（或上传音频文件），来生成完整作品。

目前，Suno的用户已经达到千万级别，公司的愿景是让每个人都能成为音乐创作者。

## 产品界面



## 技术迭代

- 负责创建人声和旋律

### Bark

2023年4月，初创团队推出基于Transformer的文本到音频的开源模型Bark，不仅能够生成多语言音频，还可以生成简单的音效。



- 负责音乐和音效

### Chirp

2023年9月推出V1版本，支持多语言转音乐，甚至在歌词中切换语言。



### Suno

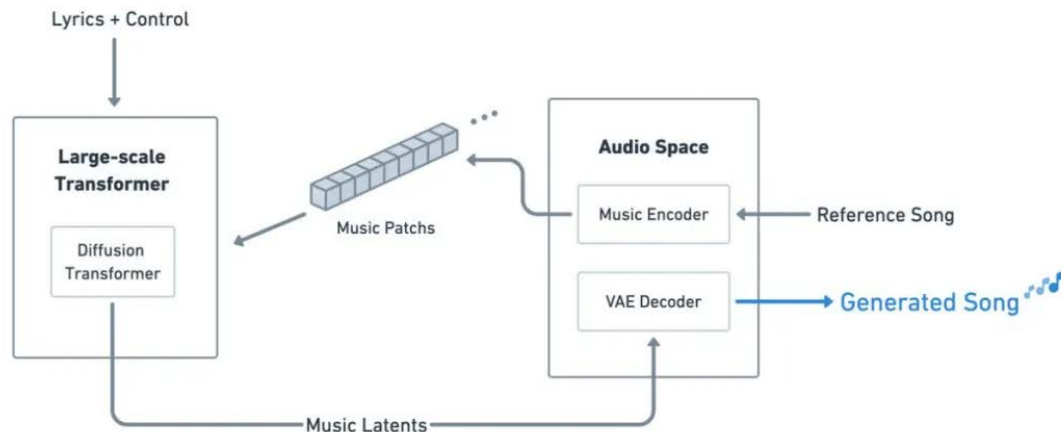
一键生成完整音乐

昆仑万维基于「天工3.0」打造出AI音乐生成大模型「天工SkyMusic」。打开天工APP，选择AI音乐板块，输入歌名和歌词，选择想要参考的音频，即可生成完整歌曲。在歌词部分，也可以选择由AI来创作的方式。产品亮点在于中文的人声清晰且稳定性高（对比海外产品的中文发音会有明显的抖动），同时还支持粤语等地方方言的发声。

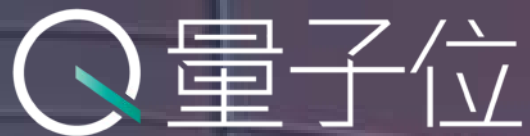
产品界面



技术架构



- 采用音乐音频领域类 Sora 模型架构
- 用户输入的参考音乐被拆解为不同的Music Patches，由Large-scale Transformer负责谱曲，学习Music Patches的上下文依赖关系，同时完成音乐可控性；
- Diffusion Transformer负责演唱部分，通过LDM(Latent Diffusion Model)让 Music Patches被还原成高品质音频



# 量子位智库 insights

## 关于量子位智库：

量子位旗下科技创新产业链接平台。致力于提供前沿科技和技术创新领域产学研体系化研究。

面向前沿AI&计算机，生物计算，量子技术及健康医疗等领域最新技术创新进展，提供系统化报告和认知。

通过媒体、社群和线下活动，基于专题技术报道及报告、专项交流会等形式，帮助决策者更早掌握创新风向。

## 关于量子位：

量子位（QbitAI），专注人工智能领域及前沿科技领域的产业服务平台。

全网订阅超过500万用户，在今日头条、知乎、百家号及各大科技信息平台量子位排名均为科技领域TOP10，内容每天可覆盖数百万人工智能、科技领域从业者。



关注量子位智库公众号



微信号：Qbitbot020  
量子位智库小助手

### 分析师

丁乔  
dq5142249297（微信）

### 智库负责人

李根  
ligen603（微信）

### 商务合作

赵萌  
13343397239（手机&微信）