

# 2026年AI视频生成行业报告： 国产爆款频出，行业加速步入“工业化商用”阶段 (精华版)

AI Video Generation industry

AI动画生成业界

概览标签：AI视频生成、Seedance 2.0、Vidu Q3、Kling 3.0

报告提供的任何内容（包括但不限于数据、文字、图表、图像等）均系头豹研究院独有的高度机密性文件（在报告中另行标明出处者除外）。未经头豹研究院事先书面许可，任何人不得以任何方式擅自复制、再造、传播、出版、引用、改编、汇编本报告内容，若有违反上述约定的行为发生，头豹研究院保留采取法律措施，追究相关人员责任的权利。头豹研究院开展的所有商业活动均使用“头豹研究院”或“头豹”的商号、商标，头豹研究院无任何前述名称之外的其他分支机构，也未授权或聘用其他任何第三方代表头豹研究院开展商业活动。

# 研究目的&摘要

## 01 | 研究目的

- 本报告系统梳理AI视频生成技术发展脉络及现状，对比主流模型能力与技术路径，并深入剖析AI视频生成市场规模、产业格局与行业应用现状。

## 02 | 研究问题

- AI视频生成模型发展现状如何？
- 目前主流AI视频生成模型有哪些？
- 全球AI视频生成市场规模为多少？不同行业占比如何？

## 01

### AI视频生成行业发展现状

- 技术方面：AI视频生成技术在分辨率、时长、物理真实模拟、主体一致性、镜头控制等多个方面实现显著突破，但面对复杂场景，均存在数量不稳定、交互出错失真等问题。商业化进程方面：AI视频生成模型C端主要面向专业内容创作者、短视频用户、泛娱乐群体等，通过月度订阅制收费；B端主要面向影视制作、广告营销等行业，通过API调用（按时长/Tokens等收费）保障收入稳健增长。

## 02

### 目前主流AI视频生成模型

- Seedance 2.0优势在于多模态叙事能力；Kling 3.0以智能分镜和主体锚定为核心；Sora 2依托社交化应用，侧重Cameos功能；Veo 3.1追求角色场景高度一致；Vidu Q3生成速度快；Gen 4.5强调“高精度、强可控”。

## 03

### AI视频生成市场规模及行业应用现状

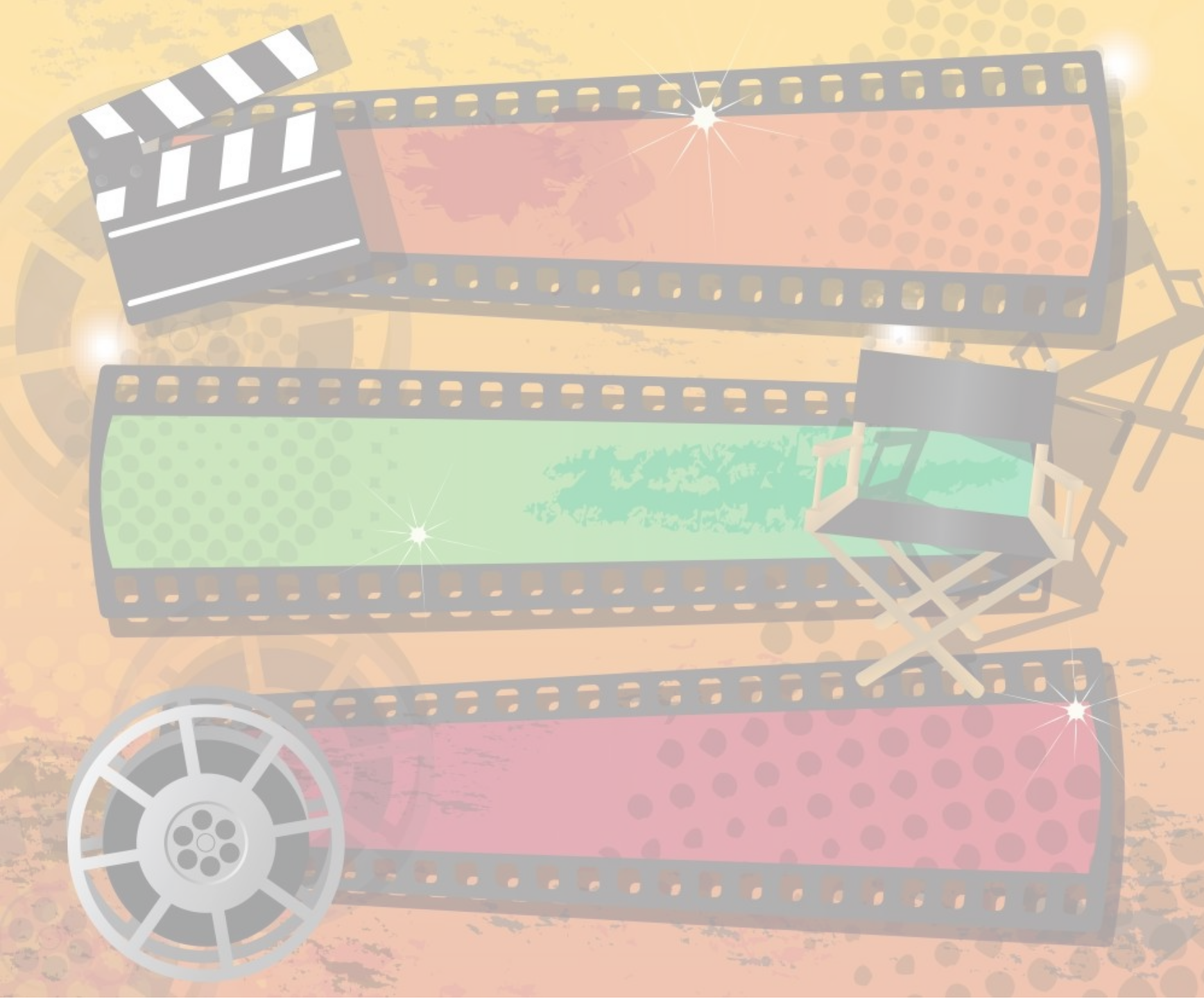
- 2025年，全球AI视频生成规模约86.8亿美元，预计2030年将增至519.3亿美元。分群体和行业看，目前行业以B端客户为主，约占65%-70%，其中营销广告、影视娱乐两大行业贡献70%以上的份额。



# 目录

## CONTENTS

◆ AI视频生成行业发展概览	-----	4
• AI视频生成所处赛道	-----	5
• AI视频生成典型方式	-----	6
• AI视频生成发展历程（技术）	-----	7
• AI视频生成发展历程（产品）	-----	8
◆ 主流AI视频生成模型及对比	-----	9
• AI视频生成参与者图谱	-----	10
• AI视频生成模型发展现状（技术）	-----	11
• AI视频生成模型发展现状（商业化进程）	-----	13
• AI视频生成模型发展现状（融资情况）	-----	14
• 主流AI视频生成模型	-----	15
• 主流AI视频生成模型对比	-----	19
◆ AI视频生成市场规模及行业应用现状	-----	22
• AI视频生成市场规模	-----	23
• AI视频生成行业应用	-----	24
• AI视频生成应用典型案例	-----	26
• AI视频生成发展趋势及挑战	-----	28
◆ 联系我们	-----	29
◆ 方法论与法律声明	-----	30



# 第一部分

## AI视频生成行业发展概览

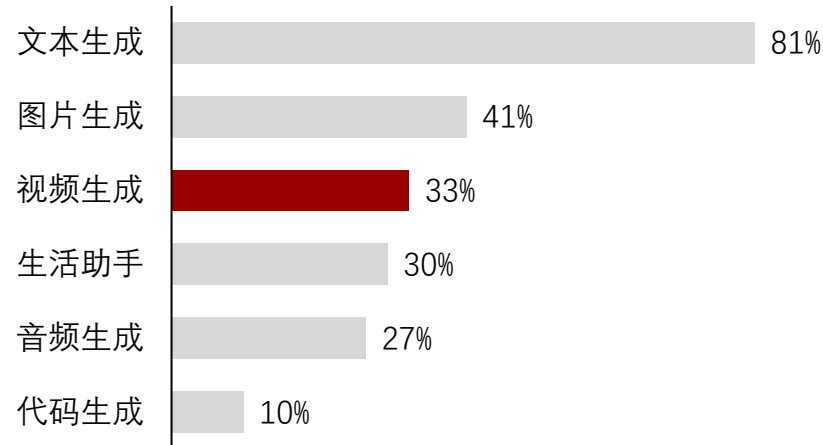
# AI视频生成所处赛道

目前，AIGC技术迈向多模态融合发展，进入全链路规模化应用爆发期，截至2025年6月，相关产品用户规模超5亿人，普及率跃升至36%+，AI视频生成作为重点赛道之一，用户使用率已达33%

## AIGC典型应用场景

类型	典型应用场景
文本生成	内容创作/续写：创作小说、剧本、诗歌、营销文案、演讲稿、邮件等，并进行续写与扩写
	信息处理/整合：文本摘要、信息提取、情感分析、自动翻译、生成结构化报告、构建知识图谱
	辅助写作/校对：文章润色、语法纠错、风格转换、内容推荐、辅助事实性核查
	对话式AI/智能助理：智能客服、聊天机器人、虚拟个人助理、AI情感陪伴、AI角色扮演
图片生成	代码生成/开发：自动编写/补全代码、修复错误、代码重构与迁移、生成测试用例、生成技术文档
	文生图/图生图：根据文字描述或者参考图片生成高分辨率的概念艺术、设计草图、产品原型图等
	图像编辑/修复：智能内容擦除/填充、图像扩展、智能修复（老照片上色、去噪、提升清晰度等）
视频生成	可控图像生成/编辑：根据草图、姿态、深度图等精控内容，风格转移、局部重绘、元素替换
	视频内容生成：根据文字/图片/音频生成短视频、广告片、动画片段等
	智能视频剪辑/处理：自动剪辑、智能匹配转场与特效、视频摘要生成、目标跟踪与移除
	视频增强与特效：视频超分辨率、插帧提升流畅度、风格化滤镜、人脸替换与美化等
音频生成	数字人/虚拟形象：生成并驱动超写实的数字人、虚拟主播，用于直播、客服、品牌代言、虚拟偶像等
	语音合成：将文本转换为自然流畅的人声，用于新闻播报、有声书、导航语音、虚拟主播配音
	声音克隆/转换：用少量样本复制特定任务声音，声音特征转换，跨语种声音克隆、个性化配音
跨模态生成	音乐/音效生成：自动作曲、编曲、作词，生成背景音乐、电影音效，人声与伴奏分离/重组
	视频内容理解/描述：图像或视频自动生成文字描述、标题或关键词，用于内容检索、辅助视障人士
	多模态交互：实现图文、音视频结合的复杂推理与对话问答
	3D模型生成：根据文本或2D图像生成3D模型，用于游戏开发、工业设计、VR/AR

用户使用AIGC产品的目的，2025.06



- AIGC指利用AI生成内容，涵盖文本生成、图片生成、视频生成、音频生成及跨模态生成等。目前，AIGC技术已从单模态迈向多模态融合发展，进入全链路规模化应用爆发期。截至2025年6月，中国AIGC相关产品用户规模突破5亿人，普及率由2024年的不足20%跃升至36%以上。其中，AI视频生成作为重点赛道之一，用户使用率达33%，典型应用场景包括视频内容生成、智能视频剪辑与增强、数字人/虚拟形象实时生成等。

# AI视频生成典型方式

AI视频生成包含文生视频（根据文本描述直接生成视频）、图生视频（结合图片与描述生成视频）、视频生视频（对已有视频进行更改）三类，目前主流国产AI视频生成模型多仅支持文本/图片输入

## AI视频生成的三种典型方式

方式	定义	案例
文生视频	基于文本提示生成符合描述的视频成片	<ul style="list-style-type: none"><li>例如：“生成一个短视频，主要场景为油菜花地，一个人在油菜花地奔跑，另一个人在帮忙拍视频”。</li></ul> 
图生视频	结合图片（关键帧）和文本提示生成符合描述的视频，或者将静态图片转为动态视频	<ul style="list-style-type: none"><li>例如：“基于输入图片生成视频，保持原图内容与结构，不新增元素，生成按顺序播放的故事画面，从第一格到最后一格依次展开，画面连贯、节奏稳定，按照图片中顺序播放”。</li></ul> 
视频生视频	对已有视频进行更改或增强（如视频中物体移除/增加、背景替换、调整视频光线）	<ul style="list-style-type: none"><li>例如：陆续上传一张大黄鸭的照片和一个视频，输入“将图片中的大黄鸭添加到视频中合适的位置上”。</li></ul> 

## 8大主流厂商AI视频生成工具占比，2025.12

输入模态	输出模态	占比*
文本/图片	视频	6.0%
文本/图片/视频	视频	2.7%
文本	视频	0.7%
文本/音频/视频	文本/音频/视频	0.7%

注：8大厂商指阿里、百度、百川、抖音、Deepseek、腾讯、月之暗面、智谱，合计共409个大模型，其中多模态模型占比为36.7%，\*占比指=8大模型厂商旗下不同【输入-输出】组合多模态大模型数量/8大厂商旗下多模态大模型合计数量。

□ AI视频生成主要包含文生视频（依据文本描述直接生成视频）、图生视频（结合图片与文本提示生成视频）、视频生视频（基于已有视频进行编辑与优化）三类。从阿里、腾讯等8家国产主流大模型厂商推出的产品来看，文生视频、图生视频为当前主流技术路线。

# AI视频生成发展历程（技术）

2014年以来，AI视频生成技术呈现“GAN（生成能力）→Transformer（长时序理解）→基于U-Net架构的Diffusion（更高质量生成）→Transformer + Diffusion的DiT融合架构（更强的可拓展性）”的演进路径

## AI视频生成技术发展历程

**技术架构：**2014年，Ian Goodfellow等提出一种由生成器和判别器组成的模型架构——**GAN**，为AI视频生成提供开创性架构；**2016年，C.Vondrick等提出第一个使用GAN生成视频的模型VGAN**；2018年，Karras等提出新的生成器架构，实现对合成图像的尺度控制，显著提高视频生成的质量和多样性；随后，研究者提出数十种基于GAN生成视频的模型，覆盖无条件的文生视频和有条件的语音、图片、视频引导生成视频等几乎所有应用场景

**架构特点：**生成质量高/快，适合图像生成，参数量小，较为轻便，但训练不稳定且训练成本高、模式崩溃、对超参数过于敏感

2016年左右

GAN

**技术架构：**2016年，开始关注自回归模型生成视频，并重点跟踪**Transformer**架构（按顺序生成数据，每次生成依赖于之前已生成内容），该架构中自注意力机制使得自回归模型可用于捕获视频、上下文等长距离依赖关系，显著提升模型对复杂内容的理解能力

**架构特点：**视频生成更为连贯自然，过程可控，概率模型清晰，适合序列数据，但生成速度慢、错误可能累积，长序列困难，计算成本通常较高

2020年左右

Transformer

2022年左右

**技术架构：**2021年起，**扩散模型**在图像生成领域的成功被复制到视频生成领域，通过对噪声逐步去噪来生成内容，其核心骨干网络普遍采用**U-Net**架构（基于SD逐帧生成），在生成质量、多样性和稳定性上全面超越GAN

**架构特点（U-Net+扩散模型）：**模型容易训练，但生成的视频内容一致性相对较差，长视频生成较困难，代表模型有Pika、Gen-2、MagicVideo-V2等

U-Net+扩散模型（Diffusion）

2024年之后

**技术架构：**2024年2月，Sora模型发布，将**扩散模型**和**Transformer**相结合（**DiT架构**），替代传统扩散模型中的U-Net架构，通过将视频时空内容切分为Patches，直接在去噪过程中利用Transformer的全局建模能力，实现卓越的可扩展性和长视频生成效果

**架构特点（Transformer+扩散模型）：**能够在不同模态之间进行有效学习与生成，更容易保证生成视频的长度和一致性，代表模型有Sora、清影、Pixverse V2、可灵、海螺AI等

Transformer+扩散模型（DiT）

# AI视频生成发展历程（产品）

从产品看，AI视频生成模型发展路径大致为：2024年，实现高清长视频突破，中国厂商快速崛起；2025年，多模态融合升级，多镜头叙事与物理模拟能力显著提升；2026年，电影级生成加速行业商业化落地

## AI视频生成产品发展历程

- Runway发布**第一代AI视频编辑模型Gen-1**，能对已有视频素材进行风格转换，3月推出第二代**视频生成模型Gen-2**，支持通过文字和图像生成最长18s的视频
- Pika发布Pika 1.0，能够生成和编辑3D动画、动漫、卡通和电影；Stability AI公司发布**Stable Video Diffusion**，支持基于原有**静止图像生成一段几秒钟的视频**
- OpenAI发布文生视频**Sora**，可直接输出长达60s的视频，支持复杂的多角度镜头，在时长、流畅度及逻辑性有显著优势，且**初步具备理解和模拟真实世界的能力**
- 生数科技发布中国首个长时长、高一致性、高动态性**视频大模型Vidu**，支持一键生成**16s、1080P**的高清视频和多镜头生成，能模拟真实物理世界，时空一致性高
- 快手发布**可灵大模型**，开放图生视频功能的同时推出**视频续写功能**，最长可生成约3分钟的视频，9月发布可灵1.5模型，新增“运动笔刷”功能，提升画面可控性

2023年2月

2023年11月

2024年2月

2024年4月

2024年6月

- Google发布文生视频**Veo 3**，支持在视频生成的同时能够完整生成音效、环境声和对话语音等声音内容，具备将音频与视觉内容精准匹配的能力，即**音画同步**。
- 字节发布**Seedance 1.0**视频生成模型，在**无缝多镜头叙事、稳定运动**等多个维度表现突出，且已在电商内容制作、游戏资产生成、影视预演等商业场景展开应用探索
- 阿里发布**Wan 2.6**系列，面向**专业影视制作和图像创作场景全面升级**，是中国首个**支持角色扮演功能**的视频生成模型，在音画同步等多方面实现突破
- 爱诗科技发布**PixVerse R1**，为全球首个支持1080P分辨率的**实时AI视频模型**，这不仅是生成速度的提升，更是AIGC从“预录制”向“实时动态生成”的范式转移
- **可灵3.0及Seedance 2.0**支持**电影级别视频生成**，大幅降低制作成本；昆仑万维发布**SkyReels V4**，首个**全能型模型**（支持多模态输入、音频生成、统一全场景任务）

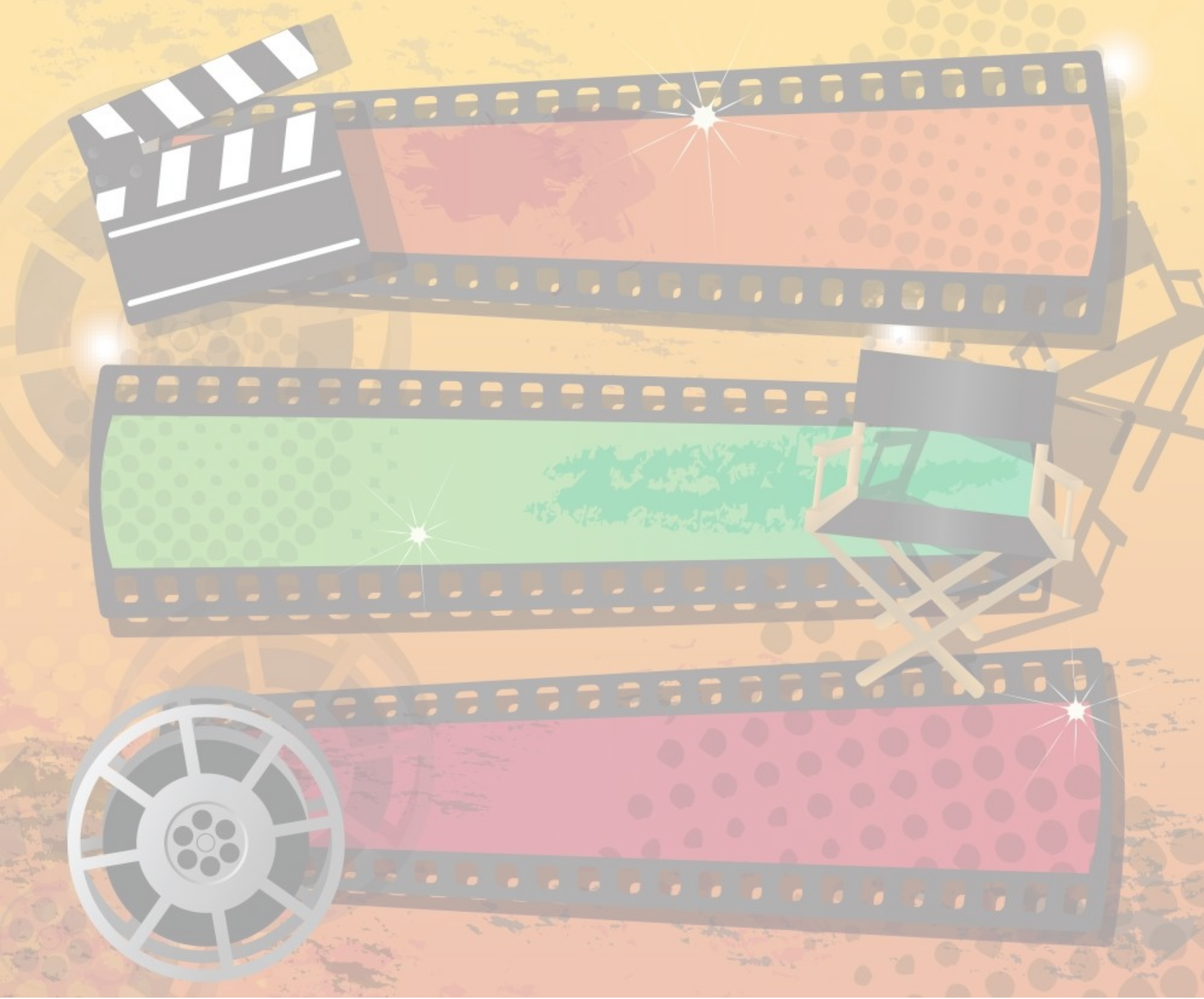
2025年5月

2025年6月

2025年12月

2026年1月

2026年2月



# 第二部分

## 主流AI视频生成模型及对比

# AI视频生成参与者图谱

现阶段，AI视频生成模型主要参与者包括：以OpenAI、Google、Runway、xAI为代表的海外厂商；以快手可灵、字节跳动、生数科技、爱诗科技、昆仑万维、阿里为代表的国产厂商

## AI视频生成参与者图谱

### 海外厂商

 OpenAI

 runway

 Google

 xAI

 Luma AI

 Lightricks

 stability.ai



 MOONVALLEY

### 国产厂商

 可灵AI

 ByteDance  
字节跳动

 生数  
ShengShu

 Tencent 腾讯

 阿里巴巴  
Alibaba.com

 爱诗 Asphere

 MINIMAX

 KUNLUN  
www.kunlun.com

 智谱·AI

 商汤  
sensetime

# AI视频生成模型发展现状（技术）

目前，AI视频生成技术在分辨率、时长、物理真实模拟、主体一致性、镜头控制等多个方面实现显著突破，但当对象较多、互动较为复杂时，各大视频生成模型均存在数量不稳定、交互出错失真等问题

## AI视频生成技术核心解决问题

01

### 视觉质量控制

- 确保每一帧都具有高质量细节和清晰度
- 避免局部畸变或不自然的视觉伪影
- 维持光照条件和色彩风格的一致性

04

### 精细动作与交互合理性

- 人物之间的自然交互（对话、肢体语言）
- 人物与物体的合理互动（拿起物品、推门）等
- 复杂动作序列的连贯表现

02

### 物理真实性模拟

- 确保生成内容符合现实物理规律
- 物体运动遵循惯性、重力等基本规律
- 复杂物理现象模拟及光影效果随时间自然过渡

05

### 场景转换的自然过渡

- 确保转场效果自然流畅
- 保持不同场景间的风格和氛围一致性
- 维持人物和关键元素在不同场景中的连贯性

03

### 时空一致性维持

- 保持视频连贯性，确保人物特征在所有帧一致
- 维持背景元素位置和外观的稳定性
- 随时间增加，一致性维持难度呈指数级增长

06

### 长时间叙事与情节控制

- 保持长视频情节的合理性和连贯性
- 确保角色行为遵循逻辑并符合设定的个性特征
- 在控制故事发展方向的同时保持自然流畅

# AI视频生成模型发展现状（技术）（接上页）

## AI视频生成模型技术突破

主要进展	说明
分辨率与时长双重突破	➤ 1080P分辨率已成为主流视频生成模型的标配，部分模型如Veo 3.1、Kling 3.0、Gen 4.5已支持4K画质。在视频时长上，Kling 3.0、Seedance 2.0等已支持15s视频，Vidu Q3已支持16s原生音视频直接输出，可支持更复杂的叙事逻辑
物理真实感与细节	➤ 新一代模型在理解和模拟物理世界上取得显著进步。例如：Sora 2的物理引擎能 <b>精确模拟重力与运动定律</b> （物体抛物线轨迹计算等）、 <b>光照与阴影</b> （反射与折射效果等）、 <b>物体交互</b> （刚体碰撞检测等）；在双人花滑场景中，Seedance 2.0能较好地演绎同步起跳、空中旋转、精准落冰等一系列高难度动作，同时在生成过程中保持对现实世界运动规律的 <b>遵循</b> ；Gen 4.5能模拟液体流动、材质纹理、头发摆动等真实物理细节
运动与镜头语言的精准掌控	➤ AI视频生成领域正从过去“抽卡”式随机生成，到如今 <b>支持多模态输入、精准资产引用和分镜编辑的工作台模式</b> ，模型能够更精准地理解文本指令，控制主体运动轨迹、镜头移动及景深变化。例如：Kling 3.0 <b>集成完善的资产管理与智能分镜控制面板</b> ，创作者可在侧边栏实时监控生成进度，并通过“智能分镜”开关对镜头的叙事结构进行微调；Seedance 2.0新增 <b>视频编辑能力</b> ，支持对指定片段、角色、动作或剧情进行定向修改
音画同步革命性突破	➤ 例如：SkyReel V4 <b>支持视频-音频联合生成</b> ，在生成画面的同时同步产出音效，无需任何后期音频合成；Seedance 2.0集成双声道立体声技术， <b>支持背景音乐、环境音效、人物解说等音频多轨并行输出</b> ，精准对齐画面节奏，同时，模型的音效设计高度自然，能真实还原磨砂玻璃轻刮声、毛绒之物揉搓声等；Vidu Q3 <b>实现口型和声音精准同步</b>
实时生成与交互技术	➤ 2025年12月，清华大学TSAIL实验室与生数科技联合研发的 <b>视频生成加速框架TurboDiffusion</b> 正式开源，这一成果将AI视频生成从传统“分钟级”耗时推进至“秒级”乃至“实时生成”的全新阶段；爱诗科技2026年1月发布全球首个支持最高1080P分辨率实时生成的模型 <b>PixVerse R1</b>
主体一致性	➤ 为保证角色一致性， <b>可通过上传参考图，用@语法锁定</b> ，模型会从参考图提取面部特征、发色、肤色、服装风格等，在生成时尽量保持一致

□ 当前，AI视频生成技术已在分辨率、生成时长、物理真实感、镜头控制、音画同步、生成效率及主体一致性等核心维度实现显著突破。例如：Kling 3.0支持4K、Vidu Q3支持16s原生音视频输出；Sora 2能精准模拟运动规律，在生成过程中保持对现实世界运动规律的遵循；Seedance 2.0支持文本、图片、视频、音频混合模态输入，可对指定片段、角色或剧情进行定向修改，同时支持背景音乐、环境音效、人物解说等音频多轨并行输出。尽管如此，主流视频生成模型在**复杂连续、多阶段物理现象仿真等高阶场景中仍面临明显挑战**。在复杂运动关联、多物体强交互等场景中，**普遍存在物体形变失真、实体穿透、互动逻辑错误等问题**，AI视频生成在特效密集镜头与高真实感强交互场景的生成能力，仍有较大优化与提升空间。

# AI视频生成模型发展现状（商业化进程）

在商业化布局方面，AI视频生成模型C端主要面向专业内容创作者、泛娱乐群体等，通过月度订阅制收费；B端主要面向影视制作、广告营销等行业，通过API调用（按时长/Tokens等收费）保障收入稳健增长

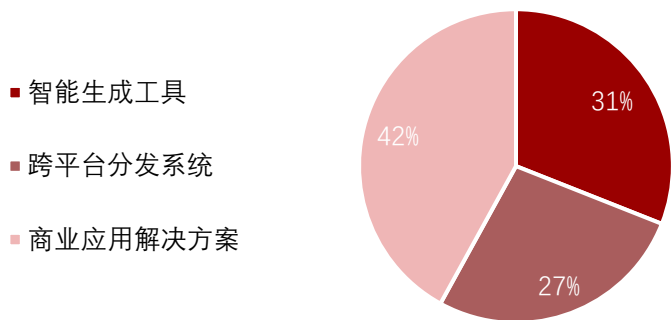
## AI视频生成模型收费模式



# AI视频生成模型发展现状（融资情况）

2025年下半年以来，AIGC视频赛道投融资显著增长，规模达30亿元，除生数科技、爱诗科技、Video Rebirth等AI视频生成模型企业外，以Nemo Video、Vattention为代表的视频制作Agent平台获资本关注

## 全球视频相关投融资分布



## 中国AIGC视频领域融资情况，2025.07-2026.03

获投企业	获投时间	融资轮次	投资金额	投资人
Motion (AI视频生成平台)	2025年7月	战略投资	数千万人民币	头部产业投资方
爱诗科技	2025年9月	B轮	6,000万美元	阿里巴巴、达晨财智、深创投等
生数科技	2025年9月	A轮	数亿人民币	博华资本、百度、建发新兴投资等
爱诗科技	2025年10月	B+轮	1亿元人民币	复星锐正、同创伟业、顺禧基金等
Nemo Video (视频制作Agent平台)	2025年10月	天使轮	数百万美元	锦秋基金、钛动科技、微光创投等
	2025年11月	Pre-A轮	近千万美元	IDG资本等
Sparki (视频创作与编辑Agent平台)	2025年11月	A轮	数百万美元	BAI资本领投
Video Rebirth (AI视频初创公司)	2025年11月	天使轮	5,000万美元	头部美元基金、互联网巨头、中韩老牌游戏上市公司等
爱诗科技	2026年1月	战略投资	1,420万美元	中国儒意
生数科技	2026年2月	A+轮	超6亿人民币	中关村科学城、星连资本、万兴科技、视觉中国等
Vattention (视频后期Agent平台)	2026年3月	种子轮	数百万美元	青锐资本、百度风投、常垒资本等
爱诗科技	2026年3月	C轮	3亿美元	鼎晖、中国儒意、三七互娱、亦庄国投、苏创投等20+国内外投资方

□ 近年来，视频生成赛道投融资规模呈现爆发式增长态势，尤其是2025年以来，随着商业化进程加速，相关赛道投融资热度达到顶峰，2025年全年全球视频相关投资规模较2024年增长83%，其中智能生成工具占比达31%、跨平台分发系统占27%，商业应用解决方案占42%。中国市场方面，一批新锐AIGC视频应用创业公司开始涌现并获资本热捧，近大半年（2025年7月-2026年3月）融资额达30亿元人民币，除生数科技、爱诗科技等头部AI视频生成模型企业外，还包括专注视频制作的Agent平台Nemo Video、Vattention和Sparki，以及致力于构建AI视频生成世界模型的Video Rebirth等。从投资方看，除专业风险投资资本外，还包括万兴科技、中国儒意、三七互娱等下游应用厂商。

来源：头豹研究院

# 主流AI视频生成模型

## 主流AI视频生成模型

模型	核心亮点
Seedance 2.0	① <b>复杂场景下更高可用率</b> ：在多主体交互和复杂运动场景中生成可用率达到业界SOTA水平；② <b>多模态能力显著增强</b> ：基于统一的多模态音视频联合架构训练，支持混合模态输入，允许用户同时输入9图+3视频+3音频+文本；③ <b>视频生成可控性大幅提升</b> ：模型的指令遵循与一致性表现全面提升，并支持稳定可控的视频延长、视频编辑；④ <b>深度支持工业级内容创作</b> ：模型支持15s高质量多镜头音视频输出，具有双声道音频能力，可实现极致拟真的试听效果，配合多模态参考和编辑能力，能大幅降低影视、广告、电商、游戏等场景的内容制作成本
Kling 3.0	① <b>智能分镜</b> ：告别单镜头堆砌，AI可深度理解剧本，自动调度景别与机位，一键生成电影感影像叙事；②在图生视频基础上，支持额外添加多图/视频主体，二次锚定画主角、道具与场景特征，锁定视觉核心；③支持中、英、日、韩、西多语种生成，以及四川话、粤语等各方言口音，多语混说口型神态自然流畅；④支持3-15s灵活时长，能容纳更复杂的叙事逻辑
Sora 2	① <b>物理模拟和音视频同步生成</b> ：能更真实地模拟运动、力学、人体动作，减少扭曲变形问题，首次支持语音、环境音、特效音同步生成，不仅能执行指令，还能基于物理规律预测后续动作；② <b>短时间内镜头叙事切换能力增强</b> ，支持15-25s视频生成，提升视频逻辑表达完整性；③ <b>内置数字分身</b> （用户通过创作自身虚拟形象，可将其一键置入AI生成的视频场景中）、二创等功能，并推出独立社交APP，创新商业模式
Veo 3.1	① <b>在多个功能中集成音频</b> ，在视频画面生成的同时，模型会自动根据画面内容生成环境音、动作声及氛围配乐，且能模拟出紧张的背景配乐和人物语言；②用户仅需提供首尾帧图片，模型即可生成平滑过渡的画像镜头；③ <b>支持输出最长8s的视频</b> ，光影效果较好，且生成速度较Sora 2快
Vidu Q3	① <b>支持16s音视频输出</b> ，强调声画同出、口型同步和多人对话，能实现口型和声音精准同步；② <b>多镜头自由切换</b> ：可根据剧情自动切换不同镜头；③支持中、英、日等多语种、多语言文字自然嵌入视频画面，无需后期逐帧跟踪、拼贴剪辑；④在全球“商业内容生成级”模型中保持 <b>最快生成速度</b> ，其处理速度较Sora 2提升10倍，较Veo 3 Fast和Grok-Imagine-Video快2倍
Gen 4.5	①实现极高的物理精度和视觉精确度；② <b>能处理多种美学风格</b> ，从照片级真实感和电影感到风格化动画，同时保持连贯的视觉语言；③支持原生音频生成和剪辑，且引入多镜头编辑功能，支持最长10s视频生成

# 主流AI视频生成模型：Seedance 2.0

一方面，Seedance 2.0采用双分支扩散变压器架构，实现视频音频原生协同生成和实时校准；另一方面，通过构建多模态全能参考系统，确保多镜头主体/叙事一致性，推动生成式AI视频走向工业化应用

## Seedance系列关键版本及Seedance 2.0关键创新点

版本	时间	输入模态	关键亮点	典型场景
Seedance 1.0 lite	2025年 5月	文生视频、图生视频	5s/10s, 480P/720P; 精准的指令遵循能力, 在多主体动作解析、镜头切换响应等方面具有优势, 丰富的影视级运镜技术, 兼顾效率、速度和性价比	电商广告、娱乐特效、影视制作、动态壁纸
Seedance 1.0 Pro	2025年 6月	文生视频、图生视频	1080P; 无缝多镜头叙事, 保持主体与风格高度连贯性; 精准的多动作与运镜控制; 超高运动稳定性与真实美感	影视分镜、电商广告、游戏
Seedance 1.5 Pro	2025年 12月	文生视频、图生视频	支持离线推理; 毫秒级音画同步, 原生支持音视频联合生成; 多人多语言自然对话; 影视叙事, 人物情绪/表情细腻	多语言电商广告、批量千人千面营销广告、影视/短剧/动漫领域部分场景成片级交付
Seedance 2.0	2026年 2月	统一多模态音视频合成: 文本+图片+视频+音频	支持2K; 多模态全能参考; 15s多镜头音视频; 强主体绑定; 可控性、可用率提升	工业级影视、社媒营销、电商广告、游戏内容生产平台

01

### 架构创新

- Seedance 2.0采用“视频+音频”双分支扩散变压器架构，打破“先画面、后音频”的串行生成模式，构建视觉与音频联合训练、同步生成的统一框架。两个分支共享统一多模态理解编码器，实现视频音频原生协同生成，从根源上解决音画不同步问题。同时，引入“跨分支校准模块”，能够实时校准视频音频的节奏、情绪、场景匹配度，确保生成的视频画面与音频完美契合，无需后期拼接或调整。
- 通过稀疏注意力机制优化长时序依赖，可稳定维持15s视频内容的连贯性，有效解决角色外观漂移、物理逻辑混乱等行业难题；同时采用“分层采样策略”，将视频生成分为“粗采样→细采样→优化采样”三个阶段，大幅提升采样效率及视频生成速度（较前代产品快30%），同时保证视频画面清晰度和流畅度。

02

### 控制创新

- Seedance 2.0构建全能参考系统，支持多类型、多数量参考素材输入，通过“@提及”语法要求参考其画面构图、镜头语言、动作节奏、音效特点等元素进行生成，甚至可以参考文字分镜内容，创作自由度大幅提升。即使面对包含大量角色互动和精准动作描述的复杂脚本，也能做到精准还原，并确保稳定的主体一致性。同时，模型具备一定编导思维，可自主规划镜头语言，并设计视觉呈现模板。
- “角色一致性约束模块”：采用“角色特征锚定技术”，在生成第一个镜头时，提取角色核心特征（面部轮廓、五官比例等）并进行锚定，后续镜头生成时，始终以锚定的角色为基础，避免出现特征偏差；采用“跨镜头注意力机制”，让模型生成每个镜头时，都能参考上一个镜头的角色特征，确保连贯性。

03

### 叙事创新

- 多镜头叙事能力是Seedance 2.0的标志性突破，模型可自动将单一叙事指令拆解成多场景、多镜头的连贯内容，确保人物、光影、风格、基调的整体一致性。一方面，模型会根据叙事节奏和镜头类型，自动生成淡入淡出、叠化、推拉摇移等过渡效果；另一方面，模型会根据用户文本提示中的叙事逻辑，确保每个镜头的内容都与上个镜头、下个镜头相互关联，形成完整的叙事链条。

# 主流AI视频生成模型：Kling 3.0

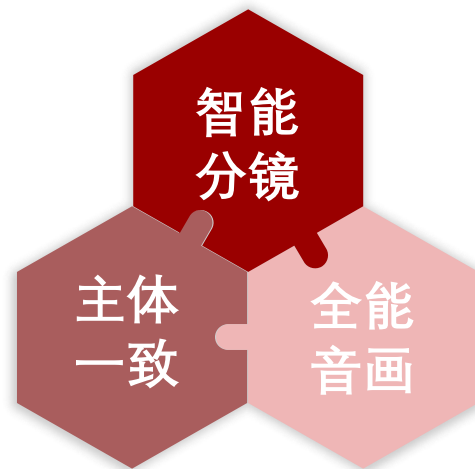
Kling 3.0关键突破在于：智能分镜系统可深度解读意图，自动调度机位和景别，确保叙事连续性；“图生视频+多图/视频主体参考”强化主体一致性；多语混说和精准映射，彻底解决指代混乱难题

## Kling系列关键版本及Kling 3.0关键能力

版本	时间	输入模式	关键亮点
Kling 1.5	2024年9月	文生视频、图生视频	分辨率从720P提升至 <b>1080P</b> ；“运动笔刷”功能可为图片中的元素 <b>指定运动轨迹</b> ，支持额外指定静止区域，提升可控性
Kling 2.0	2025年4月	文生视频、图生视频	<b>多模态编辑功能</b> ，支持输入图片等作为参考生成创意视频，可通过输入图片/文字，对生成的视频内容中元素进行增加、删减、替换等二次编辑和处理； <b>图生视频约占可灵AI视频创作量85%</b>
Kling 2.6	2025年12月	文生视频、图生视频	首次内置 <b>音频生成功能</b> ，支持中英文对白生成、歌唱演绎及环境音效同步输出（音画同出）
Kling 3.0	2026年2月	多模态输入输出高度统一，支持 <b>视频、多图素材</b> 等作为参考	<b>4K, 60fps原生输出</b> ； <b>原生智能分镜能力</b> ，对镜头精准掌控，可指定每个镜头时长、景别、运镜及叙事内容等，告别单镜头堆砌； <b>音画同步</b> ，支持多语种口型匹配、 <b>音色提取复刻</b> ；支持3-15s灵活时长

- 智能分镜系统：深度解读剧本意图，自动调度机位与景别，一键生成包括“正反打”对话、复杂转场在内的分镜，大幅降低后期调整成本。“**首尾帧精准控制**”解决多镜头角色变形、画面“崩脸”问题，**确保叙事连贯性**。
- 原生文字功能聚焦“**文字与画面原生融合**”，解决以往AI创作中文字模糊、乱码、违和感强等核心痛点。

- 主体参考：支持在图生视频基础上，**额外添加多图主体/视频主体**，对特定元素进行二次锚定，使得角色/场景等特征在复杂运镜中保持形象稳定，有效解决“主体崩坏/变形”痛点；支持**提取视频原声或为静态角色匹配音色**，配合精准的口型驱动，实现“本色出演”。



- 音画同步能力重磅升级，支持中、英、日、韩、西等多语种，四川话、粤语等多方言的音画同步，**多语混说口型神态自然**，毫无违和感，演绎流畅。多人同框下，想让谁说就由谁说，实现精准映射，**彻底解决指代混乱难题**，AI成为能精准执行导演意图的智能伙伴。

# 主流AI视频生成模型：Vidu Q3

Vidu Q3为首个支持16s声画同出的模型，支持全维度复杂动作迁移、丰富的镜头语言、精准切镜、多语言输出。该模型主要面向专业级影视制作场景，且已推出专为漫剧行业从零打造的AI漫剧解决方案

## Vidu系列关键版本及Vidu Q3关键能力

版本	时间	关键亮点
Vidu Q1	2025年3月	支持5s、1080P输出； <b>多主体细节可控</b> （尤其是动作、布局可控）； <b>AI音效功能</b> 实现一句话生成精准专属音效，且支持多种音效叠加；2025年7月上线 <b>参考生功能</b> ，最多支持7个主体同时输入并始终保持一致，可控性强
Vidu Q2	2025年9月	支持首尾帧、双出片模式及2-8s视频输出； <b>能理解并生成极其细腻的面部微表情</b> ，推动AI视频从“能动”到“会演”； <b>网页版、APP版及API同步上线</b> ，多渠道化；Vidu Q2参考生在生成质量更高视频内容的同时， <b>实际推理速度较Vidu Q1参考生提升近3倍</b> ；首次推出 <b>视频延长功能</b> ，突破AI视频时长限制，实现从“碎片镜头”到“完整故事”的跨越（旗舰版用户最长可将视频时长延长至5分钟）； <b>主要面向漫剧场景极致的性价比</b> ，适配大量团队现有 workflow，Vidu Q2 Pro参考生视频适合视频参考/编辑，方便创意复刻与修改
Vidu Q3	2026年1月	<b>主要面向专业级影视制作场景</b> （为服务于短剧、漫剧、影视剧等高频、强叙事的内容形态）；支持 <b>16s声画同出</b> 、丰富的镜头语言、精准切镜、多国文字渲染和多语言输出； <b>在全球“商业内容生成级”模型中保持最快生成速度</b> ，较Sora 2快10倍，比Veo 3 Fast和Grok-imagine-video快2倍；面向企业级市场上线 <b>超强动作同步Motion Sync功能</b> ，凭借 <b>对复杂动作的高精度复刻能力</b> ，将专业级动作迁移门槛大幅降低（用户仅需提供一张目标形象的参考图和一段动作源视频，即可生成一套高度拟真、流畅连贯的专业级复刻视频）

## 01

### 突破时长限制

Vidu Q3为**全球首个支持最长16s的音视频一体生成**，语音、旁白、对话、音效和音乐同步输出，口型精准对应，且模型支持**推拉摇移等多种经典镜头语言**和**根据脚本逻辑、场景分为自动完成流畅切镜**，用户可像导演一样精准控制每一帧的运镜节奏和视角切换，适配影视片段、宣传短片等专业场景。

## 02

### 精准复刻高复杂动作

超强动作同步Motion Sync功能基于更精准的动作控制技术，实现从原视频角色到生成角色的**全维度动作迁移**，确保即使是高速旋转、连续跳跃等复杂舞蹈动作，高难度、高精度的特效，生成视频的画面稳定性和细节表现力始终保持一致。该功能将过去需要专业动捕设备和后期团队数周工作的流程，压缩至秒级自动化生成。

## 03

### 为“剧”而生的专业模型

Vidu Q3主要面向专业级影视制作场景。2026年3月，正式推出**Vidu Q3模型AI漫剧解决方案**，专为漫剧行业从零打造，推出**非人形角色精准控制、多镜头连贯生成、提示词智能优化、配音口型同步**四大专属解决方案。同时，联合万兴科技推出中国首个**一站式AI漫剧创作平台**，以“模型+工具+场景”重构AI漫剧工业化标准。

# 主流AI视频生成模型对比（总体情况）

Seedance 2.0优势在于多模态叙事能力；Kling 3.0以智能分镜和主体锚定为核心；Sora 2依托社交化应用，侧重Cameos功能；Vevo 3.1追求角色场景高度一致；Vidu Q3生成速度快；Gen 4.5强调“高精度、强可控”

## 主流AI视频生成模型对比

维度	Seedance 2.0	Kling 3.0	Sora 2	Vevo 3.1	Vidu Q3	Gen 4.5
开发商	字节	快手	OpenAI	Google	生数科技	Runway
核心架构	双分支扩散变换器（叙事+视觉并行）	基于All-in-One理念，多模态输入输出高度统一	扩散模型+视频Transformer	层次化扩散模型架构，由多个专业化子模型协同工作	U-ViT架构（Transformer与扩散模型深度融合形成统一的架构）	自研时空混合Transformer架构，强化对时间维度的理解
分辨率/帧率	最高支持2K, 60fps（商用支持1080P, 24fps）	最高支持4K, 30fps	最高支持1080P, 60fps	最高支持4K, 24fps	最高支持1080P	最高支持4K, 60fps
输入模态	文本/图片/视频/音频（4模态混合）	文本/图片	文本为主，支持图片参考	文本/图片	文本/图片/视频/音频（多模态混合）	文本/图片/视频/音频，支持文/图/视频混合
核心定位及优势	原生多模态叙事：聚焦多镜头连贯叙事与音画同步生成，支持多模态参考输入	工业级视觉：以智能分镜与主体锚定为核心，强调影视级视觉与一致性问题解决	社交化应用：依托社交平台载体，强调物理真实感和Cameos功能，简化创作流程	以多图参考与4K超分为突破口，追求角色与场景的高度一致性	“为剧而生”，专注于叙事完整性与生成速度	“高精度、强可控”：实时交互编辑以及风格化效果
商业与生态	按token计费，含视频输入价格为28元/百万tokens，不含视频输入价格为46元/百万tokens	月活用户突破1,200万，2025年，可灵全年收入预计达1.4亿美元	免费邀请制、订阅收费（\$20-\$200/月）及面向开发者API按秒计费（\$0.1/s-\$0.5/s）	订阅收费，\$7.99-\$249.99/月；API服务，标准版\$0.4/s，快速版为\$0.15/s	订阅收费，¥69-¥699/月	订阅收费，\$15-\$95/月

# 主流AI视频生成模型对比（C端收费对比）

AI视频生成模型C端以订阅模式为主，大多采用“免费试用+多档订阅+积分充值”的服务架构。从定价看，海外模型基础版约15-20美元/月，国产模型基础版集中在60-80元/月，具备一定性价比优势

## 主流AI视频生成模型C端订阅费

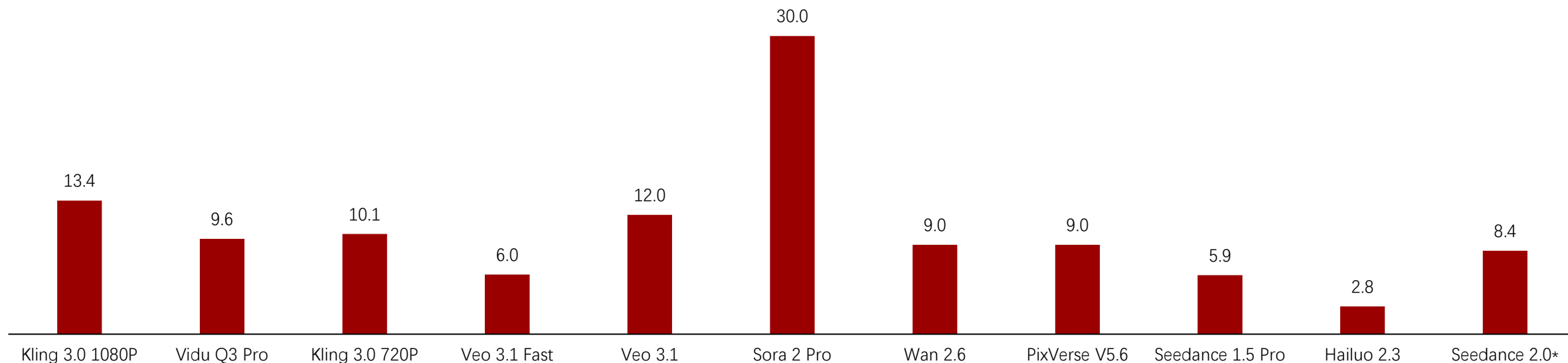
模型	Seedance 2.0	Kling 3.0	Sora 2	Veo 3.1	Vidu Q3	Gen 4.5
收费模式	免费+多档订阅+积分充值	免费+多档订阅+积分充值	免费+多档订阅	免费+多档订阅+积分充值	免费+多档订阅+积分充值	免费+多档订阅（积分兑换）
订阅价格	基础：¥79/月；标准：¥239/月；高级：¥649/月（连续包月6折，连续包年5折）	黄金：¥66/月；铂金：¥266/月；钻石：¥666/月，黑金：¥1,314/月（连续包月87折，连续包季77折）	Plus：\$20/月；Pro：\$200/月	Plus：\$7.99/月；Pro：\$19.99/月；Ultra：\$249.99/月（连续包月享受折扣）	标准版：¥69/月；专业版：¥259/月；旗舰版：¥699/月（连续包月最低8.5折，按年最低7折），企业版¥9,999起	Standard：\$15/月；Pro：\$35/月；Unlimited：\$95/月（按年包月提供8折优惠），提供企业定制服务
订阅使用权限	免费登录得积分，基础1,080分/月，标准4k分/月，高级15K分/月，最多生成1,500个视频	免费登录得积分，黄金660分/月，约33个720P的视频，铂金3K分/月，钻石8K分/月，黑金26K分/月，约1,300个720P视频	免费：约12次/天（邀请制）；Plus：50-100次/天，1080P；Pro：无限生成	免费版50分/天，plus版200分/月，pro版1K分/月，Ultra版25K分/月，1K分对应Ve03.1 Fast模型生成约20部，Quality模型生成约10部	登录免费送分，标准版800分/月，专业版4K分/月，旗舰版8K分/月，每日可提交200个视频，以Q1模型为例，一个5s，1080P的视频20分	免费版一次性得125分，仅可使用前代产品；Standard版625分/月，对应约25s Gen 4.5视频；pro版2,250分/月，对应约90s Gen 4.5视频；Unlimited版无限生成
增值服务	提供积分充值服务，约10分/元，充值增多享有不同程度折扣	提供积分充值服务，约10分/元，充值越多，加赠2分/元	——	提供积分充值服务选择，约100积分/美元	提供积分充值服务	——
其他	接入即梦，可使用即梦系列产品	可使用可灵AI系列产品	内嵌于ChatGPT会员体系	内嵌于Gemini会员体系，未单独提供服务	可使用前代产品	可使用前代产品

# 主流AI视频生成模型对比（B端收费对比）

在B端，主流AI视频生成模型厂商正尝试开放API服务，其中Sora 2 Pro成本最高，Kling 3.0和Vevo 3.1其次，Vidu Q3、Wan 2.6等其他国产模型价格偏低，Seedance 2.0综合成本约\$0.14/s，但暂未完全开放API接入

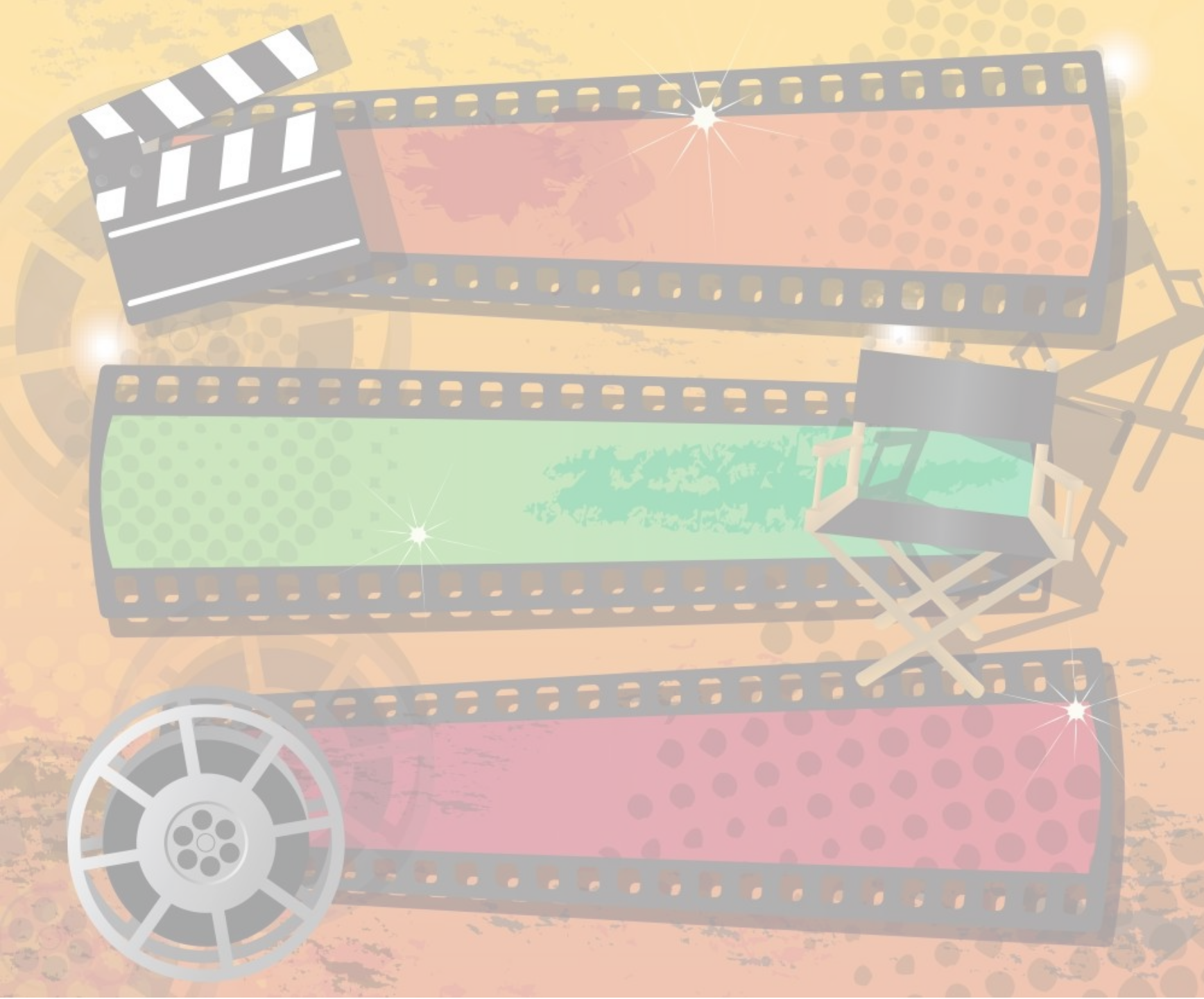
## 主流AI视频生成模型API调用价格（默认设置下，生成1分钟1080P视频的成本）

单位：美元/分钟



\*根据API易测算，Seedance 2.0生成10s、1080P的视频，价格约10元人民币，以此大致估算其生成1min、1080P的视频约8.4美元。

- 当前AI视频生成厂商面向B端客户，主要以开放API接口、按秒/按量计费为核心商业化模式。根据Artificial Analysis测算数据，在各大模型默认设置下，生成1分钟1080P的视频（不含音频）的成本呈现梯队差异，Sora 2 Pro的成本高达30美元，主打高保真、强物理一致性的顶级内容创作，其次为Kling 3.0 1080P和Vevo 3.1，价格分别为13.4美元和12美元，Vidu Q3、Wan2.6、Seedance 1.5 Pro等其他国产模型的价格均低于10美元。Seedance 2.0作为最新模型代表，暂未完全开放API接入，其定价标准为：含视频输入（视频编辑功能）为28元/百万tokens，不含视频输入（纯生视频）为46元/百万tokens，综合成本约0.14美元/秒，生成10s、分辨率为1080P的视频价格仅约1.4美元。



# 第三部分

AI视频生成市场规模及行业应用现状

# AI视频生成市场规模

2025年，全球AI视频生成规模约86.8亿美元，预计2030年将增至519.3亿美元。分群体和行业看，目前行业以B端客户为主，约占65%-70%，其中营销广告、影视娱乐两大行业贡献70%以上的份额

全球AI视频生成市场规模（单位：亿美元），2024-2030

CAGR: 43.0%



- 报告完整版/高清图表或更多报告：请登录 [www.leadleo.com](http://www.leadleo.com)
- 如需进行品牌植入、数据商用、报告调研等商务需求，欢迎与我们联系  
联系邮箱：[service@leadleo.com](mailto:service@leadleo.com)

Yoy: 100.9%

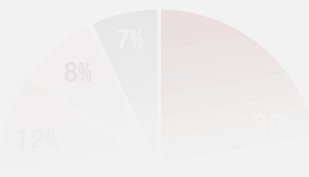
43.2

154.5

247.2

341.1

20%

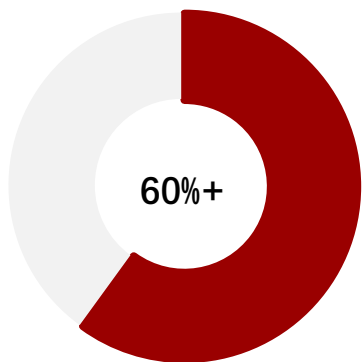
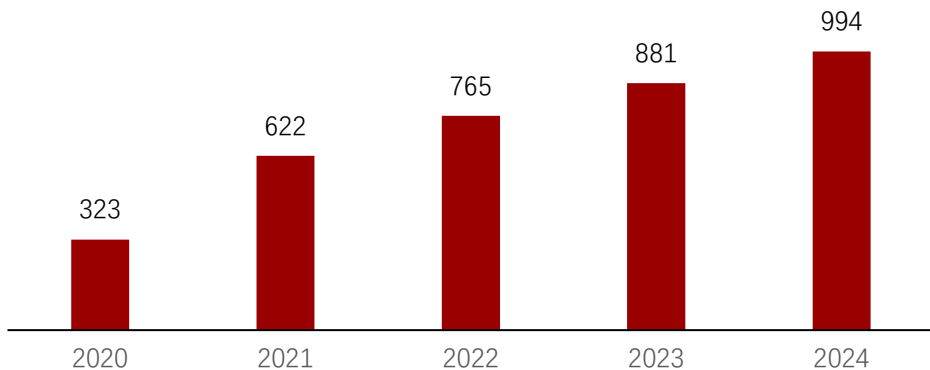


# AI视频生成行业应用：营销广告业

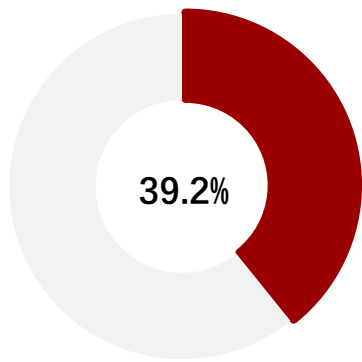
广告营销是AI视频生成短期内最直接变现场景之一：全球短视频广告年花费近千亿美元，AI视频生成技术从生产端大幅降本增效、投放端显著提升广告效果，助力行业营销即时化、个性化、规模化生产落地

## 全球短视频广告花费

单位：亿美元



AI工具可降低超60%的广告生产成本



AI生成视频的冷启动期点击率提升39.2%

来源：头豹研究院

### AI视频生成在广告业的应用场景

- 输入产品图+品牌调性参考+文案，一键生成多平台适配的短视频。智能分镜技术能够多角度、场景化地呈现产品特性，让消费者更直观地了解商品细节与使用效果。
- AI数字人被用于商品讲解、多语种播报等标准化带货场景。此外，在直播与店铺场景中，利用图生视频驱动“AI动态橱窗秀”、“AI穿搭组合购”等，有效提升直播间停留和转化效率。
- AI视频生成模型让视频广告即时化、个性化生产成为可能，商家只需上传商品图片，就可针对不同人群定制风格，同时可批量产出创意变体，快速开展A/B测试，大幅提升投放ROI。

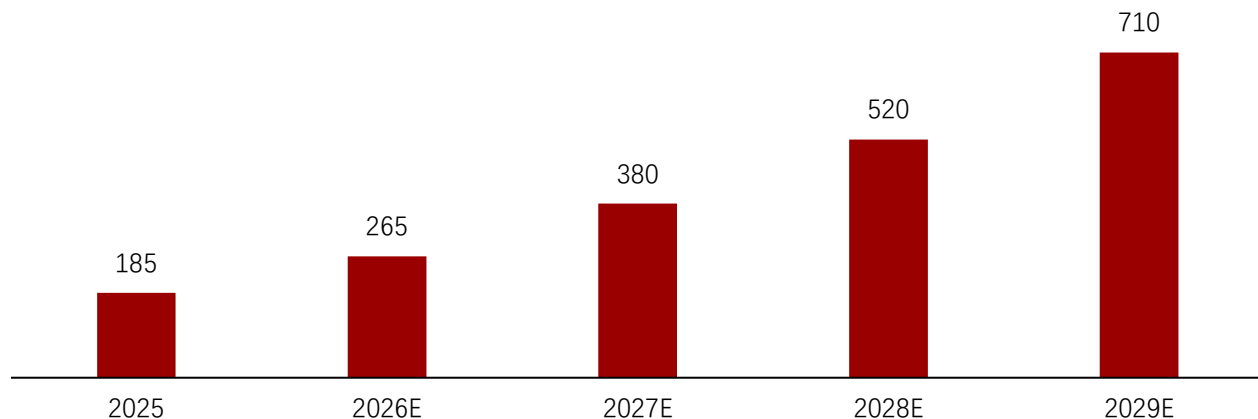
目前，全球短视频广告年花费已接近千亿美元，短视频已然成为核心的新产品发现引擎，近50%的在线购物者会在访问店铺前寻找与产品相关的视频。随着AI视频生成技术的快速迭代，广告营销成为行业短期内最直接变现场景。在生产端，AI工具实现全流程提效：只需上传产品图、品牌调性参考与文案，即可一键生成适配抖音、小红书、快手等多平台的短视频。配合AI数字人多语种实时讲解、A/B测试等，广告创作效率大幅提升，提案通过率翻倍，单条广告制作成本整体下降超60%。在投放端，AI生成广告的互动率显著高于传统广告。例如：AI生成视频的冷启动期点击率提升近40%；使用AI工具后，部分快消品牌广告点击率较传统素材提升2.5倍。

# AI视频生成行业应用：影视业

一方面，Seedance 2.0推动AI视频生成迈向专业级工业化内容生产，可促使影视制作成本下降超50%，生产周期压缩近60%；另一方面，多模态参考系统、长镜头连贯性等能力突破，极大提升AI视频的可用率

## 中国AI+影视市场规模

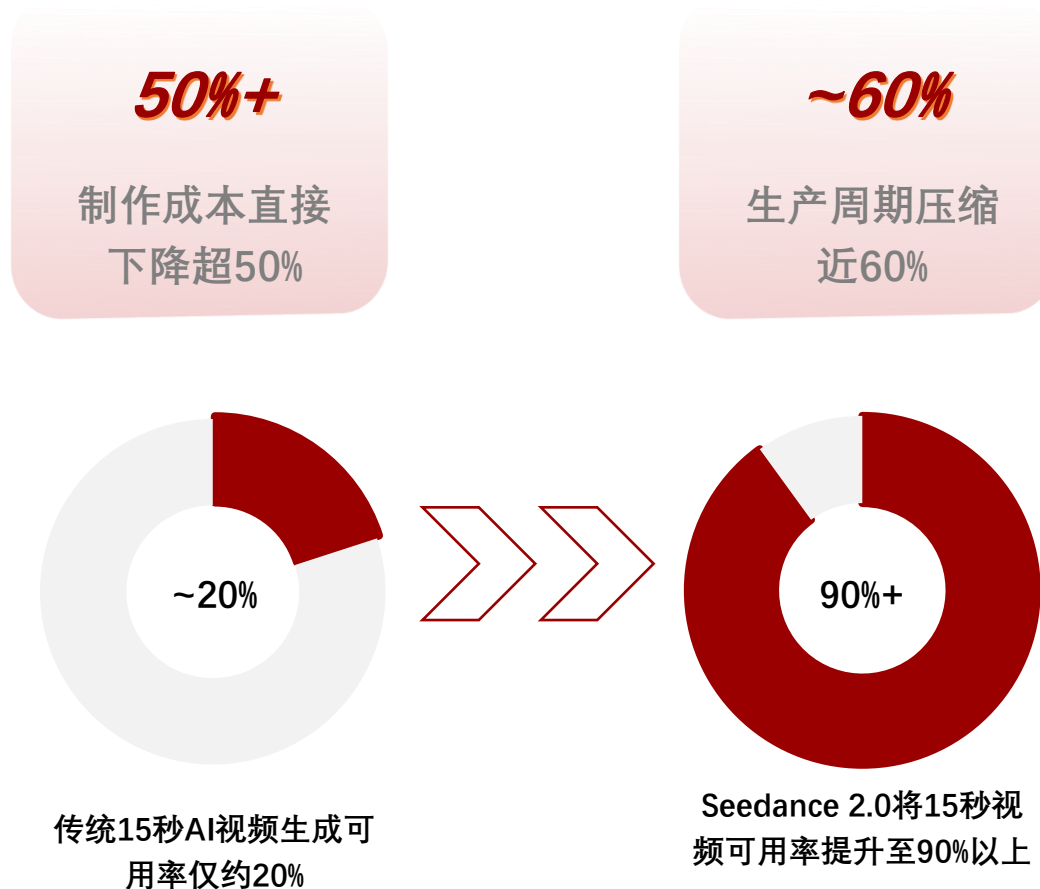
单位：亿元



□ 当前，中国影视行业正处于从数字化向智能化深度转型的关键节点。2025年，中国AI影视市场规模已达185亿元，预计2029年将突破700亿元，行业增长空间显著。2026年，以Seedance 2.0为代表的新一代AI视频生成模型，正在从底层重塑影视工业化的成本与效率逻辑。一方面，Seedance 2.0推动AI视频生成从“单模态画面生成”迈向“多模态视听一体化生产”，实现从简单素材制作到专业级工业化内容生产的跨越。依托该模型，以往需专业团队耗时数月、高成本投入的电影级视听效果，普通创作者即可在短时间内完成，制作成本下降超50%，生产周期压缩近60%。另一方面，模型在多模态参考系统、长镜头连贯性等关键能力实现突破，大幅改善传统AI视频生成画面不稳定、一致性差等痛点。数据显示，传统AI视频生成内容可用率仅约20%，而Seedance 2.0将15s视频可用率提升至90%以上，有效降低废片率和试错成本，进一步释放影视创作生产力。

来源：头豹研究院

## Seedance 2.0对影视行业的影响



# AI视频生成应用典型案例（1/2）

作为全球首个AI原生营销视频多智能体平台，inSai Hilight通过理解与洞察、创意与结构、执行与成片的三层架构，实现商品链接/图片秒变100%原创、可直接用于投流的AI营销视频，创作效率得到数十倍提升

## inSai Hilight（AI原生营销视频多智能体平台）核心能力

**核心能力一：智能成片，商品链接秒变营销视频。** 仅需输入商品链接或产品图片，inSai Hilight即可结合全球创意知识库，自动拆解产品卖点、生成创意脚本和规划分镜，全程无需人工介入，即可生成高质量、可直接用于TikTok等平台投流的营销视频，大幅降低创意、脚本、拍摄、剪辑等一系列繁琐流程和高昂成本。

**核心能力二：数字分身，生成匹配专属形象。** 支持通过图片或视频定制专属数字人形象和声音克隆，亦可从平台库按性别、年龄、国籍、姿态等维度灵活选用。inSai Hilight融合文本、图片、音视频、数字人多模态模型，实现跨帧一致性优化，还构建负责智能自检的Agent，在视频片段生成后再进行校验，确保最后输出结果的准确性。

**核心能力三：无缝衔接创意，突破素材产能瓶颈。** inSai Hilight提供“一键换装、换模特、换背景、鞋子上脚、商品动态展示”等灵活、实用、丰富、便捷的AI素材生成功能，覆盖商品展示、场景延展、细节特写等多元营销场景。所有素材均由AI原创生成，无版权风险、无需实拍，在保障视频质量的同时实现规模化产出。

**核心能力四：统一管理全局素材，实现高效复用。** inSai Hilight自动归档用户生成的所有图片、音频、视频素材，并按项目进行可视化分类管理。同时，inSai Hilight还支持团队权限配置、版本追踪，方便用户随时查阅、下载、复制生成。让每一次创作都成为可积累、可复利、可变现的数字资产。

理解与洞察层：把“需求+素材”变成“可执行的营销指令”

Brief理解Agent      素材理解Agent      信息总结Agent      趋势洞察Agent



创意与结构层：把“好想法”变成“能被执行的视频结构”

创意生成Agent      剧本策划Agent      素材匹配Agent      素材增强Agent



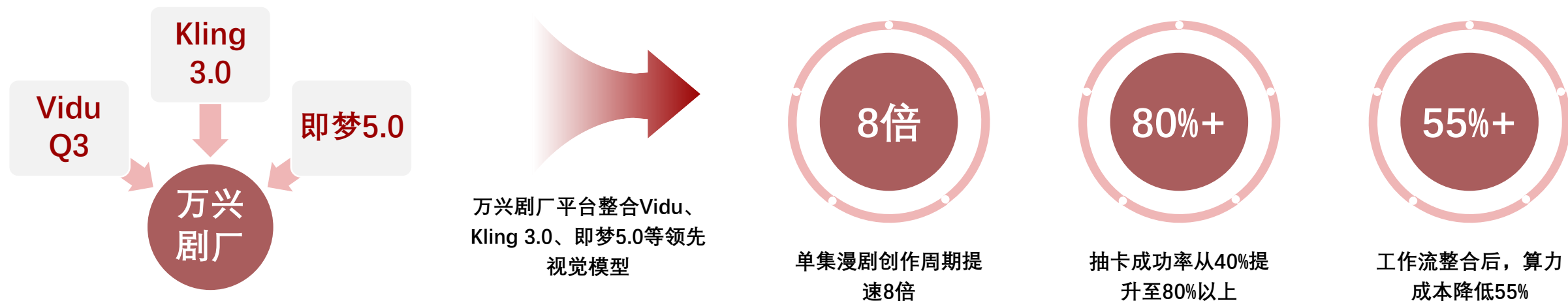
执行与成片层：把“结构化方案”变成“可投放的视频资产”

剪辑执行Agent      成片生成Agent      质检复盘Agent

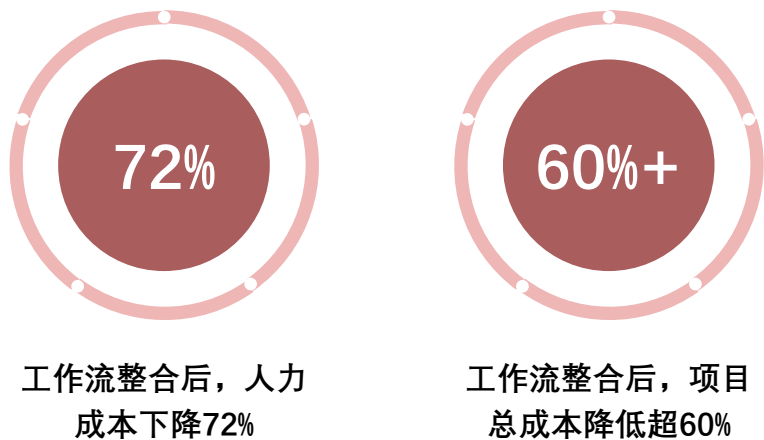
## AI视频生成应用典型案例（2/2）

万兴剧厂采用集成Vidu、Kling等主流视频生成模型的“多模驱动”架构，通过智能匹配最优模型、运用提示词工程优化效果、搭建知识库实现资产封装与复用等方式整合 workflow，使得项目综合成本下降超60%

### 万兴剧厂整合视频生成模型的效果



- 影视项目创作通常涵盖剧本拆解、分镜设计、角色设定、视频生成、后期剪辑等全流程环节，单一模型难以支撑复杂剧情的工业化制作需求，而频繁在多工具间切换也会显著拉低整体效率。万兴剧厂平台采用“多模驱动”架构，集成Vidu、Kling 3.0、即梦5.0等主流视觉生成模型（公司已于2026年2月正式落地Vidu Q3图生视频能力，并同步推进Seedance 2.0接入）：通过按题材类型智能匹配最优生成模型、运用提示词工程优化生成效果、引入“知识库”优化资产复用（如封装后的特定构图提示或运镜方式，使用时可直接调用）等方式整合优化全流程 workflow，使得单集漫剧创作周期提速8倍，抽卡成功率从40%提升至80%以上，算力成本降低55%，人力成本下降72%，项目总成本降低超60%。



# AI视频生成发展趋势及挑战

【1】发展趋势：模型易用性与轻量化并行推进，高分辨率、时序稳定的垂直领域专业视频生成模型将加速落地。【2】发展挑战：版权法规缺位、伦理规范不足，或将持续制约行业商业化落地进程

## AI视频生成发展趋势及挑战



### 发展趋势

#### 01 易用性和轻量化同步推进

现阶段文生、图生、视频生视频多为独立子模型，未来将向**统一多模态架构融合**（在统一的、单一框架下兼容文/图/视频生成与编辑），且正推进视频生成与大语言模型原生绑定，强化提示词理解、叙事规划与时序逻辑推理，易用性提升；**视频生成模型正从云端大参数量化走向轻量化和高效化**，短视频生成、视频编辑等高频场景将走向PC端、移动端本地离线部署。

#### 02 专业垂直领域模型推出

例如：（1）生数科技推出Vidu Q3模型AI漫剧解决方案，从底层算法到功能设计，全面围绕漫剧内容生产需求展开；（2）专为电商出海营销的AI原生视频多智能体平台inSai Hilight的推出。

#### 03 视频分辨率及技术稳定性/一致性持续提升



### 落地挑战

#### 01 版权问题

当前视频生成与多模态大模型训练高度依赖互联网公开视频素材，全球范围内尚未形成针对AI训练数据使用、视听内容/声音版权等的统一明确法规，由此引发的侵权纠纷频发（如Seedance 2.0高逼真生成能力引发训练数据来源合规性的质疑），**版权边界模糊、维权成本高企等问题日益突出**。加快构建覆盖事前数据合规、事中生成管控、事后侵权溯源处置的全流程治理体系，已成为行业健康发展的迫切需求。

#### 02 伦理道德规范问题

例如：Seedance 2.0的面部克隆与零样本跨模态声音合成能力，突破传统深度伪造技术限制，仅需单张面部照片即可生成音画高度同步、逼真度极高、肉眼难以辨识的声像内容，这大幅降低身份伪造、虚假诈骗的技术门槛，极易被用于实施精准诈骗，直接威胁公民财产安全与人格权益，对社会信任与公共秩序构成严峻挑战。

# 头豹业务合作

## 会员账号

可阅读全部原创报告和百万数据，提供PC及移动端，方便触达平台内容

## 定制报告/词条

行企研究多模态搜索引擎及数据库，募投可研、尽调、IRPR等研究咨询

## 定制白皮书

对产业及细分行业进行现状梳理和趋势洞察，输出全局观深度研究报告

## 报告作者



袁栩聪  
首席分析师



付淑芳  
行业分析师

• [service@leadleo.com](mailto:service@leadleo.com)

## 招股书引用

研究覆盖国民经济19+核心产业，内容可授权引用至上市文件、年报

## 市场地位确认

对客户竞争优势进行评估和证明，助力企业价值提升及品牌影响力传播

## 行研训练营

依托完善行业研究体系，帮助学生掌握行业研究能力，丰富简历履历

## 业务咨询

- 客服电话：400-072-5588
- 官方网站：[www.leadleo.com](http://www.leadleo.com)

## 深圳办公室

广东省深圳市南山区粤海街道  
华润置地大厦E座4105室  
邮编：518057

## 上海办公室

上海市静安区南京西1717号  
会德丰国际广场 2701室  
邮编：200040

## 南京办公室

江苏省南京市栖霞区经济开发  
区兴智科技园B栋401  
邮编：210046



# 方法论

- ◆ 头豹研究院布局中国市场，深入研究19大行业，532个垂直行业的市场变化，已经积累了近100万行业研究样本，完成近10,000多个独立的研究咨询项目。
- ◆ 研究院依托中国活跃的经济环境，研究内容覆盖整个行业的发展周期，伴随着行业中企业的创立，发展，扩张，到企业走向上市及上市后的成熟期，研究院的各行业研究员探索和评估行业中多变的产业模式，企业的商业模式和运营模式，以专业的视野解读行业的沿革。
- ◆ 研究院融合传统与新型的研究方法，采用自主研发的算法，结合行业交叉的大数据，以多元化的调研方法，挖掘定量数据背后的逻辑，分析定性内容背后的观点，客观和真实地阐述行业的现状，前瞻性地预测行业未来的发展趋势，在研究院的每一份研究报告中，完整地呈现行业的过去，现在和未来。
- ◆ 研究院密切关注行业发展最新动向，报告内容及数据会随着行业发展、技术革新、竞争格局变化、政策法规颁布、市场调研深入，保持不断更新与优化。
- ◆ 研究院秉承匠心研究，砥砺前行的宗旨，从战略的角度分析行业，从执行的层面阅读行业，为每一个行业的报告阅读者提供值得品鉴的研究报告。

# 法律声明

- ◆ 本报告著作权归头豹所有，未经书面许可，任何机构或个人不得以任何形式翻版、复刻、发表或引用。若征得头豹同意进行引用、刊发的，需在允许的范围内使用，并注明出处为“头豹研究院”，且不得对本报告进行任何有悖原意的引用、删节或修改。
- ◆ 本报告分析师具有专业研究能力，保证报告数据均来自合法合规渠道，观点产出及数据分析基于分析师对行业的客观理解，本报告不受任何第三方授意或影响。
- ◆ 本报告所涉及的观点或信息仅供参考，不构成任何证券或基金投资建议。本报告仅在相关法律许可的情况下发放，并仅为提供信息而发放，概不构成任何广告或证券研究报告。在法律许可的情况下，头豹可能会为报告中提及的企业提供或争取提供投融资或咨询等相关服务。
- ◆ 本报告的部分信息来源于公开资料，头豹对该等信息的准确性、完整性或可靠性不做任何保证。本报告所载的资料、意见及推测仅反映头豹于发布本报告当日的判断，过往报告中的描述不应作为日后的表现依据。在不同时期，头豹可发出与本报告所载资料、意见及推测不一致的报告或文章。头豹均不保证本报告所含信息保持在最新状态。同时，头豹对本报告所含信息可在不发出通知的情形下做出修改，读者应当自行关注相应的更新或修改。任何机构或个人应对其利用本报告的数据、分析、研究、部分或者全部内容所进行的一切活动负责并承担该等活动所导致的任何损失或伤害。

