

AI VIDEO GENERATION EVALUATION

# AI 视频能力评测汇报



王娟



2026.03

01

# 项目背景

PROJECT BACKGROUND



# 产品选择

PRODUCT SELECTION



**Sora**

OpenAI



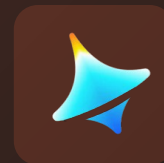
**Veo**

Google



**可灵**

Kling



**即梦**

Jimeng



**海螺**

Hailuo



**Luma**

Luma AI



**万相**

Wan 2.6



**Vidu**

Vidu

● 共 8 款主流 AI 视频生成产品参与评测

# 模型的选择与优势

## MODEL SELECTION & ADVANTAGES

软件名称	版本号	内部模型	发布时间	核心特性
Sora	Sora 1.0 (预览)	Sora 1.0	2024年2月	首次亮相, 支持最长 60 秒视频生成, 引发行业震动但未公开发布
	Sora 2	Sora 2	2025年9月30日	正式公开发布, 支持物理精确性、道具度和可控性大幅提升, 支持同步对话及音效生成
Veo	Veo 1	Veo 1	2024年5月	Google 首款视频生成模型, 支持文生视频, 720p 分辨率, 视频长度数秒
	Veo 2	Veo 2	2024年12月	提升真实感和电影感, 支持 8-10 秒视频, 改善物理模拟、引入安全过滤
	Veo 3	Veo 3	2025年5月	首次集成同步音频 (对话、音效、环境音), 支持 4-8 秒 1080p 视频, 推出 Flow 编辑器
	Veo 3 Fast	Veo 3-fast	2025年7月	加速版本, 优化速度和成本, 适用于快速原型制作
	Veo 3.1	Veo 3.1	2025年10月14日	扩展至 60 秒视频长度, 增强真实感, 优化音频功能。新增多参考图像融合、首尾帧过渡等功能
	Veo 3.1 Fast	Veo 3.1-fast	2025年9月	3.1 快速版本、平衡速度与质量
Luma AI	v1.0 (Preview)	Dream Machine v1	2024年6月12日	首个公众版本, 免费开放, 支持文生 / 图生视频、生成 5 秒片段, 具备基础物理模拟能力, 引发全民 AI 视频热潮。
	v1.5	Dream Machine v1.5	2024年8月20日	大幅提升写实度, 增强了对提示词的理解能力和运动一致性、生成速度提升, 并在角色保持方面有明显改进。
	v1.6	Dream Machine v1.6	2024年9月4日	引入运镜控制, 新增 Camera Motion 功能, 支持通过文本指令 (如 "pan left", "zoom in") 精确控制摄像机推拉摇移。
	Ray 2	Ray 2	2025年1月15日	架构大升级, 算力提升 10 倍、支持原生 1080p 分辨率, 视频时长延长至 10 秒 (可 Extend 至 30 秒), 物理规律模拟更加真实。
	Ray 2 Modify	Ray 2-modify	2025年6月	视频重构功能, 新增 "Modify Video" 能力, 允许用户通过文本指令修改视频中的物体, 环境或风格, 而无需复杂的遮罩。
	Ray 3	Ray 3	2025年9月18日	首个 "推理型" 视频模型, 具备视觉推理能力 (能理解意图并自我修正), 支持原生 16bit HDR 输出, 新增 "单图控制" 和 "Draft Mode" 快速预览模式。
	Ray 3 Modify	Ray 3-modify	2025年12月18日	专业级编辑模型, 专注于视频后期, 支持精确关键帧控制、角色参考 (Character Reference) 锁定角色一致性, 以及场景重塑功能。
	Ray 3.14	Ray-3.14	2026年1月26日	消除 "画质 - 速度 - 成本" 不可能三角; 4 倍速生成, 原生 1080p 输出, 单秒成本降低 3 倍。
Wan (万相)	Wan 1.0	Wan-1.0	2024年3月	阿里通义万相首个视频生成模型
	Wan 2.0	Wan-2.0	2024年8月	提升画质和指令理解能力
	Wan 2.1	Wan-2.1	2024年10月	优化生成稳定性、支持 ComfyUI 本地部署
	Wan 2.2 Pusa V1.0	Wan-2.2-pusa	2024年11月	LoRA 微调模型, 采用双 ControlNet 架构、支持普帧生成、角色一致性控制
	Wan 2.5	Wan-2.5	2025年6月	架构优化, 提升多场景适应性

# 模型的选择与优势

## MODEL SELECTION & ADVANTAGES

软件名称	版本号	内部模型	发布时间	核心特性
Kling (可灵)	Kling 1.0	Kling-1.0	2024 年 6 月 6 日	快手首款视频生成模型正式上线, 支持文生视辉和图生视辉
	Kling 1.5	Kling-1.5	2024 年 9 月	提升画质和运动表现, 优化生成稳定性
	Kling 1.6 (高画质版)	Kling-1.6-hq	2024 年 11 月	推出高画质版本, 进一步提升视觉质量
	Kling 2.0	Kling-2.0	2025 年初	架构升级, 支持更长视频时长
	Kling 2.1	Kling-2.1	2025 年 3 月	优化指令遵循能力
	Kling 2.5 Turbo	Kling-2.5-turbo	2025 年 9 月 19 日	重新定义最强性价比, 提升生成速度和价格竞争力
	Kling 2.6	Kling-2.6	2025 年 11 月	支持灵动画布、多模态编辑功能
	可灵 O1	Kling-O1	2025 年 12 月	大一统模型, 融合参考生视频、文生视频、首尾帧生视频、视频编辑等多任务
	Kling 3.2	Kling-3.2	2026 年 3 月 6 日	极致性价比提升; 优化了指令遵循度 (Prompt Adherence), 强化了运动笔刷 (Motion Brush) 的精确度。
Hailuo (海螺)	Video-01	Hailuo-01	2024 年 8 月 31 日	MiniMax 首款视频生成模型、支持 720p/25fps 高清视频。具备电影感镜头移动
	Hailuo 02	Hailuo-02	2025 年 6 月 18 日	NCR 架构创新, 参数量 3 倍、训练数据 4 倍, 支持原生 1080p 生成, 极限物理表现, SOTA 指令遵循
	Hailuo 2.3	Hailuo-2.3	2025 年 10 月 28 日	进一步升级动态表现力, 画面更真实稳定, 肢体动作、表情变化表现更佳
Jimeng (即梦)	视频 1.2	PixelDance 1.2	2024 年 7 月	来自 AI Lab 团队。早期视频生成尝试
	PixelDance (视频 P2.0 Pro)	PixelDance-2.0-pro	2024 年 9 月 24 日	豆包视频生成模型, 支持多主体互动、极致运镜 (推拉摇移、360 度环绕)
	视频 (视频 S2.0)	Seaweed-2.0	2024 年 11 月 15 日	与 PixelDance 同期发布的另一款视频生成模型, 面向公众免费开放
	Seedance 1.0 (视频 3.0)	Seedance-1.0	2025 年 2 月	全新视频生成模型, 整合前代优势
	视频 3.5 Pro	Seedance-3.5-pro	2025 年 5 月	当前主推版本、能力全面提升
	Seedance 2.0	Seedance 2.0	2026 年 2 月 7 日	行业突破: 首个“音视频双支路”统一架构, 视频与音频同步生成, 可用率高达 90%, 支持多模态 12 路参考输入。
Vidu	Vidu 1.0	Vidu-1.0	2024 年 4 月	生数科技首款视频生成模型, 中国首个月自研 Diffusion+Transformer 架构
	Vidu 1.5	Vidu-1.5	2024 年 8 月	提升生成速度和画质稳定性
	Vidu Q1	Vidu-Q1	2025 年 7 月	首款 AI 生图模型, 仅用 2 个月完成研发, 超越众多成熟模型
				实现从“生成视频”到“生成演技” 在极致表情变化、推拉运镜、语义理解和生成速度方面取得重大进

02

# 评估维度与标准

EVALUATION DIMENSIONS & CRITERIA



# AI 视频生成

AI VIDEO GENERATION



文生视频

Text-to-Video

Text Prompt → Video



图生视频

Image-to-Video

Image + Prompt → Video

核心问题

# 让我们先确定一个问题

---

一个好的 AI 视频要具备什么？



# 好视频的标准

STANDARDS FOR GOOD VIDEO

文生视频

图生视频



美学

Aesthetics



画质

Image Quality



语义遵循

Semantic Adherence



物理逻辑

Physics Logic



时间一致

Temporal Consistency



场景一致

Scene Consistency



人物一致

Character Consistency



运动逻辑

Motion Logic

💡 不管是文生视频还是图生视频，大家对好的视频的理解维度都是一样的，包括但不限于**美学、画质、物理逻辑、运动逻辑、时间一致、场景一致、空间一致、人物一致、语义遵循**等

# 模型评估维度

## MODEL EVALUATION DIMENSIONS

### 01 美学

衡量生成内容的艺术价值，关注整体美感、光影布局、色彩搭配以及构图的专业度。优秀模型应能产出具有电影质感、视觉冲击力且审美在线的画面，而非平淡无奇的记录式图像。

### 02 画质

考察生成内容的清晰度，重点检查画面是否清晰、锐化是否严重。同时，扣分项包括画面模糊、过度锐化、明显的马赛克或由于生成缺陷导致的空白或阴影。

### 03 语义遵循

验证模型是否“听懂”了指令。评估生成的视频是否准确包含了用户描述的所有关键要素（主体、环境、颜色、数量等），以及是否正确执行了指令。

### 04 物理与运动逻辑

检测动态生成的真实感。重点观察物体运动是否符合重力与惯性、流体运动是否自然，特别对于人物/动物，需评估肢体结构是否在运动中保持正常，避免出现四肢扭曲、穿模。

### 05 时间与角色一致性

评估在角色的身份（长相、衣着）是否保持稳定、不会“变脸”或“换装”。同时检查背景和物体是否在与镜头之间保持稳定，不存在画面闪烁、抖动或物体随机变形的情况。



**评估方法：**选择 1-5 分打分法，并建立严格的评分标准。从 1. 语义遵循度、2. 物理与运动逻辑、3. 时间与角色一致性、4. 美学、5. 画质五个维度进行打分

# 模型评估维度

MODEL EVALUATION DIMENSIONS

## 1-5 分严苛打分体系



### 总原则：拒模糊打分，建立严格的“及格线”

评分不是简单的“好”或“坏”，而是需要建立明确的分级标准。每个分数段都有具体的定义和评判依据，确保评分的客观性和可复现性。拒绝“差不多就行”的模糊态度，用严苛的标准推动模型能力的真实展现。

# 模型评估维度

MODEL EVALUATION DIMENSIONS

## 1-5分 严苛打分体系

总原则：拒模糊打分，建立严格的“及格线”

	分值	描述	Prompt	案例
语义遵循	1	<b>关联性断裂:</b> <ul style="list-style-type: none"><li>生成的内容与提示词毫无关系，或者完全忽略了指令中的核心主体。例如要求“外太空”，结果生成了“森林”。</li></ul>	帮我生成一个带有钢琴的房间	
	2	<b>部分关键词触发</b> <ul style="list-style-type: none"><li>视频中出现了指令提到的某些元素，但它们之间的关系或整体场景完全不对。画面逻辑与指令意图严重脱节、更像是AI的“自由发挥”。</li><li>如：提示词“一只狐狸在结冰的湖面上轻盈地跳跃”但只有狐狸只走路没有跳跃</li></ul>	一只狐狸在结冰的湖面上轻盈地跳跃	
	3	<b>要素缺失/偏差</b> <ul style="list-style-type: none"><li>主体正确，但动作或环境出现偏差。例如指令要求“猫在奔跑”，结果生成了“猫在睡觉”；或者多个主体中漏掉了一个。</li></ul>	在暴雨倾盆的夜晚。一位送餐员（尽管天气恶劣，但他依然面带微笑）骑着电动车穿过霓虹闪烁的街道	
	4	<b>主要要素准确</b> <ul style="list-style-type: none"><li>视频捕捉到了指令的核心主体，可能在极其细微的次要描述（如某个特定的小装饰品颜色）上有所忽略，但不影响整体叙事和核心动作。</li></ul>	一个穿着白衬衫的小男孩，在公园的沙坑里用蓝色的铲子堆城堡，他的身边有一只黑色的狗	
	5	<b>全要素达成</b> <ul style="list-style-type: none"><li>完美呈现了Prompt中所有的主体、动作，环境细节及艺术风格。对复杂的方位词（如“左边”、“背景中”）和程度词（如“缓缓地”、“剧烈地”）有精准响应。</li></ul>	一只戴着红色领结的泰迪熊，坐在一张摆满复古书籍和一杯冒着热气的咖啡的木桌前，用羽毛笔认真地写信，窗外有雪花缓缓飘落。	

	分值	描述	案例
时间与角色一致性	1	<b>逻辑崩塌:</b> <ol style="list-style-type: none"><li>视频每一帧看起来都像不同的人；</li><li>主体在运动过程中突然变成另一种物体（如人变马）；</li><li>空间结构完全乱套。</li></ol>	
	2	<b>视觉混乱</b> <ul style="list-style-type: none"><li>视频中出现了指令提到的某些元素，但它们之间的关系或整体场景完全不对。画面逻辑与指令意图严重脱节、更像是AI的“自由发挥”。</li><li>如：提示词“一只狐狸在结冰的湖面上轻盈地跳跃”但只有狐狸只走路没有跳跃</li></ul>	
	3	<b>轻微“AI感”</b> <ol style="list-style-type: none"><li>人增长相大致相同，但仔细会发现“微变形”现象（如眼睛特征消失、脸型轻度变形）；</li><li>衣服的过渡可能在跳动；</li><li>背景物体在镜头移动时有轻微的漂移或重组。</li></ol>	
	4	<b>极其细微偏移</b> <ul style="list-style-type: none"><li>角色特征保持得很好，仅在玻璃角度（如完全侧脸）或快速运动时，细节（如耳环样式、材料纹理）有几乎察觉不到的轻微波动。</li></ul>	
	5	<b>完全一致</b> <ul style="list-style-type: none"><li>角色/物体的五官、发型、疤痕等特征在转头或大动作时毫无变化；短装材质和花纹全程固定；背景建筑或道具位置绝对稳定。</li></ul>	

# 模型评估维度

MODEL EVALUATION DIMENSIONS

## 1-5分 严苛打分体系

总原则：拒模糊打分，建立严格的“及格线”

分值	描述	案例
1	<b>严重违背物理定律：</b> <ol style="list-style-type: none"><li>完全无视重力：物体凭空悬浮、向上掉落、或静止在半空；</li><li>穿模/融合：人物直接穿过墙壁、桌椅，或身体部位相互穿透；</li><li>肢体扭曲：关节反向弯曲、四肢像面条一样随意拉伸变形；</li><li>瞬间位移：物体在没有过渡的情况下突然跳跃位置。</li></ol>	
2	<b>明显运动异常：</b> <ol style="list-style-type: none"><li>惯性缺失：急停急起没有缓冲，像被无形的手操控；</li><li>流体错误：水流像固体一样流动，烟雾像石头一样硬，火焰不随风动；</li><li>步态怪异：人物走路像滑行（月球漫步）、脚不沾地、或同手同脚严重；</li></ol>	
3	<b>基本合理但有瑕疵：</b> <ol style="list-style-type: none"><li>大动态尚可，微操失误：整体跑步姿势正常，但摆臂幅度不协调；</li><li>接触面轻微漂浮：鞋底与地面接触瞬间有轻微滑动（Skating），但未完全离地；</li><li>布料/头发解算简单：衣物随风摆动，但方向偶尔与风向不一致，或摆动频率单一；</li></ol>	
4	<b>自然流畅，细节逼真：</b> <ol style="list-style-type: none"><li>惯性与动量准确：加速、减速、转弯都有合理的身体倾斜和缓冲；</li><li>复杂的次级运动：头发、衣角、配饰随主运动产生自然的滞后和跟随效果；</li><li>精准的碰撞反馈：物体撞击后有形变、反弹或破碎，且符合材质特性；</li></ol>	
5	<b>完美拟真，电影级物理解算：</b> <ol style="list-style-type: none"><li>微观物理精确：肌肉收缩带动骨骼运动，皮肤挤压变形真实；</li><li>复杂交互零失误：多物体同时碰撞、堆叠、滚动，逻辑严密无穿模；</li><li>环境互动完美：脚印深浅符合体重与地面硬度，风吹草动呈现波浪式传导；</li></ol>	

分值	描述	案例
1	<b>视觉灾难：</b> <ol style="list-style-type: none"><li>画面严重失真、回血，充满马赛克；</li><li>色彩杂乱无章，无法分辨画面主体；</li><li>画面不忍直视，美感及低。</li></ol>	
2	<b>粗糙不型：</b> <ol style="list-style-type: none"><li>画面有明显的模糊、绿点或色块；</li><li>色彩失真（如皮肤发黄、天空过曝）；</li><li>构图混乱，主体位置尴尬或稀疏性；</li><li>易面美感知强随着（应为“易产生审美疲劳”）</li></ol>	
3	<b>平庸写实：</b> <ol style="list-style-type: none"><li>画面清晰度尚可，但缺乏艺术处理；</li><li>色彩可能偏灰或饱和度过高，构图类似随手抓拍，没有明显的视觉重点；</li><li>普通通量水平（应为“普通量产水平”）</li></ol>	
4	<b>精美自然：</b> <ol style="list-style-type: none"><li>画面清晰、干净；</li><li>光影逻辑合理，色彩搭配和谐；</li><li>构图专业，符合大众审美，没有明显的废筐杂乱感（应为“废片杂乱感”）；</li><li>稍微带有AI感，精美的真实电影拍摄质感。</li></ol>	
5	<b>无可挑剔：</b> <ol style="list-style-type: none"><li>分辨率极高，质感细腻（皮肤纹理、发丝清晰）；</li><li>构图极具艺术感（如完美的黄金分割、极简主义等）；</li><li>光影处理达到顶级电影灯光水准；</li><li>【UHD*】高水平的光影处理：用符合故事逻辑的明暗、色温与层次，平衡真实感和风格化，精准烘托情绪并引导观众视线；</li><li>毫无AI感，媲美的真实电影拍摄质感。</li></ol>	

分值	描述	案例
1	<b>高度失真：</b> <ol style="list-style-type: none"><li>浓厚马赛克或严重的像素化。（“像素化”指画面被拆成彩色方块）</li><li>画面主体模糊到难以辨认，或充满了大量的数学伪影。</li></ol>	
2	<b>明显模糊：</b> <ol style="list-style-type: none"><li>画面有明显的颗粒、噪点或色块；</li><li>细节丢失严重，主体边缘出现模糊的重影或光晕。</li></ol>	
3	<b>普通视频质感：</b> <ol style="list-style-type: none"><li>画面清晰度足够看清内容，但缺乏细节深度【如画面关系（明暗交界线）、建筑刚、人物反映画面等等】；</li><li>存在轻微的AI渣味地（水印效果），边缘不够锐利。</li></ol>	
4	<b>优质 1080P/2K 质感：</b> <ol style="list-style-type: none"><li>细节饱满，主体清晰；</li><li>但在远景或极复杂纹理（如密集的树叶）处可能存在轻微的软化。</li></ol>	
5	<b>原生 4K 质感：</b> <ol style="list-style-type: none"><li>纹理强度锐利，如瞳孔细节、皮肤毛孔、布料纤维清晰可见；</li><li>完全没有编码杂讯或人工锐化痕迹。</li></ol>	

03

# 文生视频

TEXT-TO-VIDEO




评估方法

# 文生视频评估标准

通过【提示词难度】来模拟

不同用户 / 不同指令下的生成上限

 简单难度

 中等难度

 困难难度

# 文生视频评估标准

TEXT-TO-VIDEO EVALUATION CRITERIA



## 简单难度

Easy

### 核心要素：

单一主体 + 简单动作 + 常规场景

### 字数限制：

5-30 字

### 主要考察点：

基础语义遵循无幻觉

示例："一个破旧的机器人坐在天台上晒太阳"



## 中等难度

Medium

### 核心要素：

渲染风格 + 多主体互动 + 具体场景 + 主体细节描述 + 动态交互

### 字数限制：

30-100 字

### 主要考察点：

风格一致性、多主体位置关系、光影细节

示例："赛博朋克风格，低角度仰拍，背景是一个寒夜雨后的城市。一个侦探在雨中递给机器人一把伞，霓虹灯倒影清晰"



**评估逻辑：**通过递增的提示词复杂度，全面测试模型在不同场景下的生成能力和上限表现

# 文生视频评估标准

TEXT-TO-VIDEO EVALUATION CRITERIA



## 困难难度

Hard Difficulty

### ★核心要素

详细风格 + 详细分镜 + 宏大场景 + 动态光影 + 物理特效

### 🔗字数限制

100 字以上

### 📍主要考察点

长文本逻辑保持、时间一致性、物理规律 ( 粒子 / 流体 )、运镜控制

" 风格：史诗级奇幻电影风格，融合了高饱和度的魔法粒子特效与写实的自然环境。画面具有极高的视觉冲击力，冷色调的冰雪世界与诡谲的紫色迷雾形成鲜明对比。整体呈现出一种现代大片级的 CG 质感，伴随电影感的动态模糊和宏大的色彩分级。 "

### “ Prompt 示例

0 - 3.5s 在一片荒凉且乱石嶙峋的荒野中，一位满头银发的中国修仙长者傲然挺立。他的双眼闪烁着幽邃的蓝色光芒，长发在刺骨的寒风中狂乱舞动。随着他缓缓抬起手臂，周围的空气开始凝结，无数冰晶与雪花围绕着他剧烈旋转，形成一股强大的寒冰漩涡。

细节：- 长者眼神凝重，双眼迸发出耀眼的蓝光。- 寒风吹动他的白色长袍和发丝，展现出极佳的物理动态。- 他伸出右手，掌心汇聚起强大的冰霜能量。3.5s - 结束 迅速拉远切换至航拍视角，一头巨大的寒冰中国龙从长者身后的地表破土而出。中国龙通体由晶莹剔透的坚冰构成，体型遮天蔽日，它对着苍穹发出震耳欲聋的咆哮，冰冷的吐息化作漫天飞雪，震撼着整片大地。

细节：寒冰中国龙巨大的头颅冲破地面，带起无数冰块飞溅。- 寒冰中国龙完全舒展身体，遮蔽了阳光，投下巨大的阴影。- 它张开巨口向天空咆哮，口中喷出白色的寒冰雾气。- 镜头环绕寒冰中国龙进行动态追踪，展现其宏伟的体量。”

# 模型评估框架

MODEL EVALUATION FRAMEWORK

模型	任务	美学与画质	语义遵循度	物理与运动逻辑	时间与角色一致性	总分
Sora	简单 prompt					
Veo	简单 prompt					
Luma	简单 prompt					
Kling (可灵)	简单 prompt					
Hailuo (海螺)	简单 prompt					
Jimeng (即梦)	简单 prompt					
Wan 2.6 (万相)	简单 prompt					
Vidu	简单 prompt					

模型	任务	美学与画质	语义遵循度	物理与运动逻辑	时间与角色一致性	总分
Sora	中等 prompt					
Veo	中等 prompt					
Luma	中等 prompt					
Kling (可灵)	中等 prompt					
Hailuo (海螺)	中等 prompt					
Jimeng (即梦)	中等 prompt					
Wan 2.6 (万相)	中等 prompt					
Vidu	中等 prompt					

模型	任务	美学与画质	语义遵循度	物理与运动逻辑	时间与角色一致性	总分
Sora	困难 prompt					
Veo	困难 prompt					
Luma	困难 prompt					
Kling (可灵)	困难 prompt					
Hailuo (海螺)	困难 prompt					
Jimeng (即梦)	困难 prompt					
Wan 2.6 (万相)	困难 prompt					
Vidu	困难 prompt					

看看视频效果吧



简单难度

Easy



中等难度

Medium

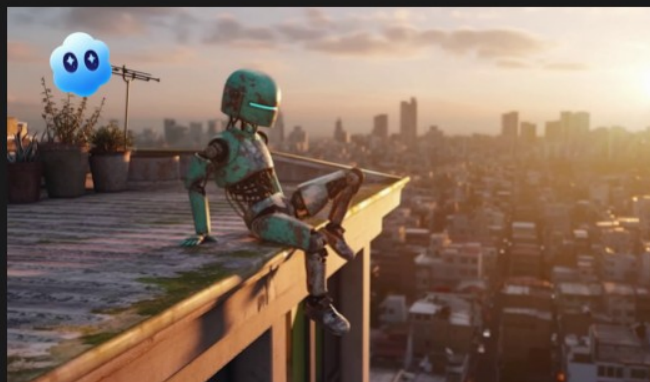
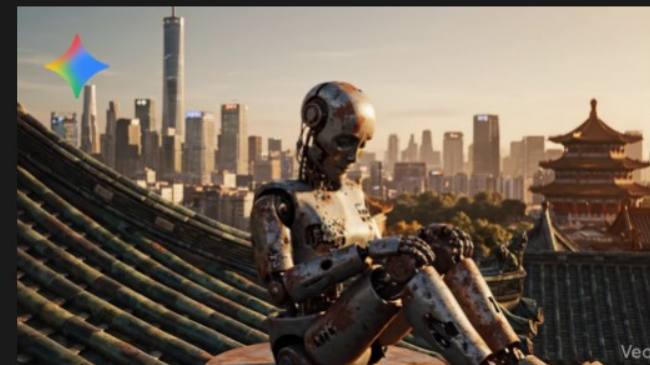
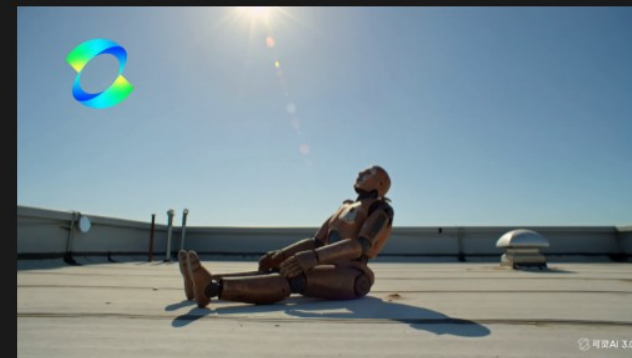


困难难度

Hard

# 简单-视频效果

EASY DIFFICULTY - VIDEO RESULTS



# 简单难度排名

EASY DIFFICULTY RANKING



评判逻辑：所有人总分相加

A= 难度	模型	A= 评委	美学	语义	物理	一致性	创意	# 总分	平均分
简单	Veo	评委1	4	3	5	5	4	21	4.2
简单	Veo	评委2	5	4	5	4	3	21	4.2
简单	Veo	评委3	5	4	4	4	4	21	4.2
简单	Veo	评委4	5	3	4	5	4	21	4.2
简单	Veo	评委5	4	5	3	4	5	21	4.2
简单	Veo	评委6	4	5	4	4	3	20	4.0
简单	Veo	评委7	5	4	4	4	3	20	4.0
简单	Veo	评委8	4	5	3	4	4	20	4.0
简单	Kling	评委1	4	4	5	5	2	20	4.0
简单	Kling	评委2	5	4	4	5	2	20	4.0
简单	Kling	评委3	3	5	3	5	4	20	4.0
简单	Kling	评委4	5	3	4	4	4	20	4.0
简单	Kling	评委5	5	3	5	5	2	20	4.0
简单	Kling	评委6	5	3	4	4	3	19	3.8
简单	Kling	评委7	4	4	4	5	2	19	3.8
简单	Kling	评委8	4	4	5	3	3	19	3.8
简单	Jimeng	评委1	5	3	4	5	3	20	4.0
简单	Jimeng	评委2	4	4	4	5	2	19	3.8
简单	Jimeng	评委3	5	4	2	4	4	19	3.8
简单	Jimeng	评委4	3	4	3	4	5	19	3.8
简单	Jimeng	评委5	4	4	4	4	3	19	3.8
简单	Jimeng	评委6	5	3	4	4	3	19	3.8
简单	Jimeng	评委7	4	5	3	4	3	19	3.8
简单	Jimeng	评委8	4	4	3	5	4	19	3.8
简单	Wan2.6	评委1	4	5	3	4	4	20	4.0
简单	Wan2.6	评委2	4	4	4	4	3	19	3.8
简单	Wan2.6	评委3	5	4	4	5	2	20	4.0
简单	Wan2.6	评委4	3	4	3	5	4	19	3.8
简单	Wan2.6	评委5	3	5	4	4	3	19	3.8
简单	Wan2.6	评委6	4	4	4	4	3	19	3.8
简单	Wan2.6	评委7	4	4	3	5	4	20	4.0
简单	Wan2.6	评委8	4	5	4	4	3	20	4.0
简单	Hailuo	评委1	4	4	4	5	3	20	4.0
简单	Hailuo	评委2	4	3	4	5	4	20	4.0
简单	Hailuo	评委3	3	4	3	5	4	19	3.8

# 中等-视频效果

MEDIUM DIFFICULTY - VIDEO RESULTS



# 中等难度排名

EASY DIFFICULTY RANKING

评判逻辑：所有人总分相加



## 评测结论 - 即梦

中等难度结论：综合评判即梦效果最佳

难度	模型	评委	美学	语义	物理	一致性	创意	总分	平均分
中等	Jimeng	评委1	4	3	4	5	4	20	4.0
中等	Jimeng	评委2	4	4	5	4	3	20	4.0
中等	Jimeng	评委3	4	5	4	3	3	19	3.8
中等	Jimeng	评委4	5	5	2	3	4	19	3.8
中等	Jimeng	评委5	4	5	4	3	3	19	3.8
中等	Jimeng	评委6	3	5	4	3	4	19	3.8
中等	Jimeng	评委7	5	4	3	4	3	19	3.8
中等	Jimeng	评委8	4	5	2	4	4	19	3.8
中等	Veo	评委1	4	3	5	5	3	20	4.0
中等	Veo	评委2	4	3	4	5	3	19	3.8
中等	Veo	评委3	4	2	5	5	3	19	3.8
中等	Veo	评委4	5	1	4	5	4	19	3.8
中等	Veo	评委5	3	4	5	4	3	19	3.8
中等	Veo	评委6	4	4	5	3	3	19	3.8
中等	Veo	评委7	4	4	3	4	4	19	3.8
中等	Veo	评委8	5	4	4	5	1	19	3.8
中等	Kling	评委1	3	4	5	4	3	19	3.8
中等	Kling	评委2	3	5	5	3	3	19	3.8
中等	Kling	评委3	3	5	5	3	3	19	3.8
中等	Kling	评委4	3	5	4	3	4	19	3.8
中等	Kling	评委5	4	5	3	3	4	19	3.8
中等	Kling	评委6	3	5	3	4	4	19	3.8
中等	Kling	评委7	4	4	4	3	3	18	3.6
中等	Kling	评委8	3	4	4	3	4	18	3.6
中等	Wan2.6	评委1	5	3	4	3	4	19	3.8
中等	Wan2.6	评委2	3	3	5	3	5	19	3.8
中等	Wan2.6	评委3	4	3	3	5	4	19	3.8
中等	Wan2.6	评委4	4	3	3	4	4	18	3.6
中等	Wan2.6	评委5	4	4	3	5	2	18	3.6
中等	Wan2.6	评委6	3	3	4	3	5	18	3.6
中等	Wan2.6	评委7	3	5	2	3	5	18	3.6
中等	Wan2.6	评委8	3	4	4	5	2	18	3.6
中等	Hailuo	评委1	3	4	3	3	5	18	3.6
中等	Hailuo	评委2	4	3	3	4	4	18	3.6

# 困难-视频效果

HARD DIFFICULTY - VIDEO RESULTS



# 困难难度排名

EASY DIFFICULTY RANKING

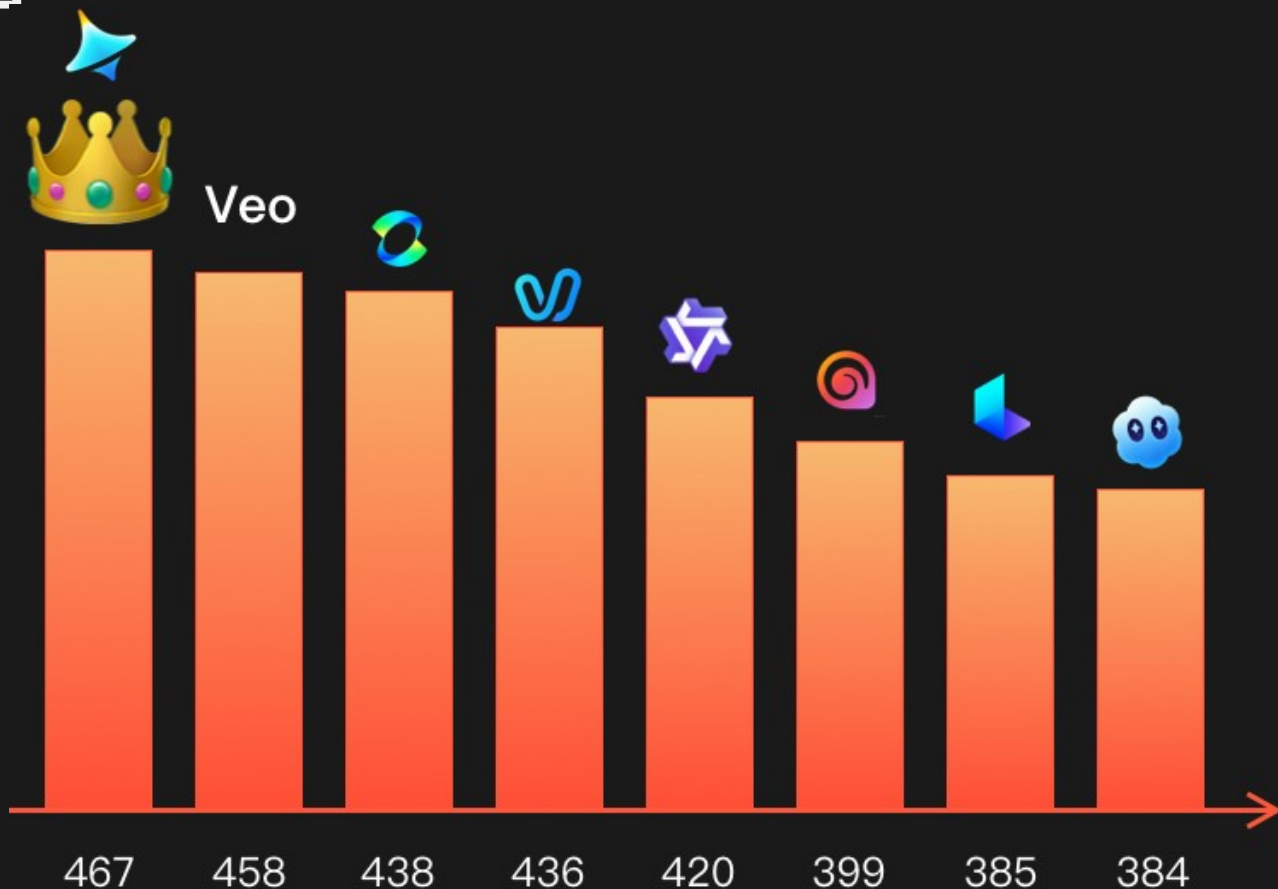
评判逻辑：所有人总分相加



**评测结论 - 即梦**  
 困难难度结论：**综合评判即梦效果最佳**

难度	模型	评委	美学	语义	物理	一致性	创意	总分	平均分
困难	Jimeng	评委1	4	4	3	4	4	19	3.8
困难	Jimeng	评委2	4	3	4	5	3	19	3.8
困难	Jimeng	评委3	3	4	3	4	4	18	3.6
困难	Jimeng	评委4	4	4	2	4	4	18	3.6
困难	Jimeng	评委5	3	5	2	3	4	17	3.4
困难	Jimeng	评委6	4	4	3	3	3	17	3.4
困难	Jimeng	评委7	3	4	3	4	3	17	3.4
困难	Jimeng	评委8	4	3	2	4	4	17	3.4
困难	Veo	评委1	4	3	4	4	3	18	3.6
困难	Veo	评委2	3	3	4	4	4	18	3.6
困难	Veo	评委3	4	2	4	5	3	18	3.6
困难	Veo	评委4	3	3	3	4	4	17	3.4
困难	Veo	评委5	4	2	3	4	4	17	3.4
困难	Veo	评委6	3	3	4	3	3	16	3.2
困难	Veo	评委7	4	2	3	4	3	16	3.2
困难	Veo	评委8	3	3	2	4	4	16	3.2
困难	Kling	评委1	3	3	4	3	4	17	3.4
困难	Kling	评委2	3	4	3	3	4	17	3.4
困难	Kling	评委3	4	3	3	4	3	17	3.4
困难	Kling	评委4	3	4	2	3	4	16	3.2
困难	Kling	评委5	3	3	3	4	3	16	3.2
困难	Kling	评委6	4	2	3	3	4	16	3.2
困难	Kling	评委7	3	3	2	4	3	15	3.0
困难	Kling	评委8	2	4	3	3	3	15	3.0
困难	Wan2.6	评委1	3	3	3	4	4	17	3.4
困难	Wan2.6	评委2	4	2	3	3	4	16	3.2
困难	Wan2.6	评委3	3	3	2	4	4	16	3.2
困难	Wan2.6	评委4	3	2	3	4	3	15	3.0
困难	Wan2.6	评委5	4	2	2	3	4	15	3.0
困难	Wan2.6	评委6	3	3	1	4	4	15	3.0
困难	Wan2.6	评委7	2	3	2	4	4	15	3.0
困难	Wan2.6	评委8	3	2	3	3	3	14	2.8
困难	Hailuo	评委1	3	2	3	4	4	16	3.2
困难	Hailuo	评委2	3	3	2	4	3	15	3.0

## 文生视频综合排名



在文生视频场景中，**即梦凭借维度均衡性成为综合首选**。各个产品在文生视频方面的能力**差异主要体现在语义遵循度与物理和运动逻辑上**。其他排名靠前的模型在各种难度的 Prompt 下能够实现高度的语义对齐，物理和运动逻辑基本符合物理引擎和自然规律；时间和角色一致性这方面也有很好的保持。相比之下，排名靠后的产品虽然在单帧质量或局部美学方面具有一定的可用性，但在**高难度的 Prompt 生成过程中更容易出现语义不遵循、动作不自然以及脱离物理规律的情况**，这反映出它们在复杂场景下控制整体结构的能力有限。总体而言，文生视频模型能力的高低更多取决于其对 prompt 的遵循并输入以及对于物理与运动规律的理解程度

# 文生视频评测总结



## 简单难度

- 1 Veo 综合最优
- 2 可灵、即梦 紧随其后
- 3 万相、海螺 表现稳定



## 中等难度

- 1 即梦 可控性最强
- 2 Veo、可灵 一致性佳
- 3 万相、海螺 美学突出



## 困难难度

- 1 即梦 复杂场景王者
- 2 Veo 物理逻辑优秀
- 3 可灵、万相 基本稳定

📈 **总体趋势：**随着难度提升，各模型差距拉大，即梦在复杂场景下优势更加明显

# 文生视频评测完成



三大难度级别评测  
8款主流产品全面对比

→ 接下来进入 **图生视频** 评测环节

04

# 图生视频

IMAGE-TO-VIDEO



评估方法

# 图生视频评估标准

通过【画面难度】来模拟

复杂情况下的生成上限

 单人 + 低动态

 单人 + 高动态

 多人 + 低动态

 多人 + 高动态

# 图生视频评估标准

IMAGE-TO-VIDEO EVALUATION CRITERIA



## 难度 1: 单人 + 低动态

Single Person + Low Motion

### 核心要素:

单人 (简单) 场景 + 低动态范围。

### 考察重点:

人物特征保持、基础运动生成、背景稳定性

示例: " 一个低角度跟踪镜头跟随年轻男子自信地向前走过城市广场, 镜头从侧后方跟随他移动, 他自然地环顾四周观察建筑, 夹克随着动作飘动, 流畅的斯坦尼康式跟拍, 现代建筑从旁边经过, 电影行走场景 "



## 难度 2: 单人 + 高动态

Single Person + High Motion

### 核心要素:

单个 (简单) 场景 + 高动态范围。

### 考察重点:

肢体结构保持、运动流畅性、物理逻辑合理性

示例: " 一个年轻人站在火车站台上, 黄色地铁列车从左到右呼啸而过带有运动模糊, 他的头发和夹克在风中轻轻飘动, 他微微转头跟随列车, 环境站台灯光, 真实运动, 电影感 "

# 图生视频评估标准

IMAGE-TO-VIDEO EVALUATION CRITERIA



## 难度 3: 多人 + 低动态

Multiple People + Low Motion

### 核心要素:

多人 (简单) 场景 + 低动态范围。

示例: " 电影级 CG 画面, 两个精致的精灵少女坐在发光的森林湖畔。镜头从她们的背影缓慢侧绕至侧脸。一个精灵正伸出手, 指尖停着一只半透明的发光蝴蝶, 蝴蝶翅膀扇动产生粒子微光, 照亮了她皮肤上的毛孔和细微纹理。另一个精灵转头看向伙伴, 露出温柔的微笑。她们身上的薄纱长裙随微风轻轻起伏, 背景是景深虚化的萤火虫森林。8k 分辨率, 虚幻引擎 5 风格, 柔和的丁达尔效应。 "



## 难度 4: 多人 + 高动态

Multiple People + High Motion

### 核心要素:

多人 (简单) 场景 + 高动态范围。

示例: " 行车手在森林道路上高速骑行, 背景和前景有电影级运动模糊, 充满活力的 HDR 色彩分级呈现丰富绿色和明亮天空, 动态速度效果, 镜头从身后略偏高角度跟随, 树木模糊成运动条纹, 影子清晰投射在道路上, 明亮高光照亮云朵和叶子, 高动态范围, 强烈运动能量, 专业户外动作视频 "

# 模型评估框架

## MODEL EVALUATION FRAMEWORK

模型	任务	美学与画质	语义遵循度	物理与运动逻辑	时间与角色一致性	总分
Sora	难度1:单人+低动态					
Veo	难度1:单人+低动态					
Luma	难度1:单人+低动态					
Kling (可灵)	难度1:单人+低动态					
Hailuo (海螺)	难度1:单人+低动态					
Jimeng (即梦)	难度1:单人+低动态					
Wan 2.6 (万相)	难度1:单人+低动态					
Vidu	难度1:单人+低动态					

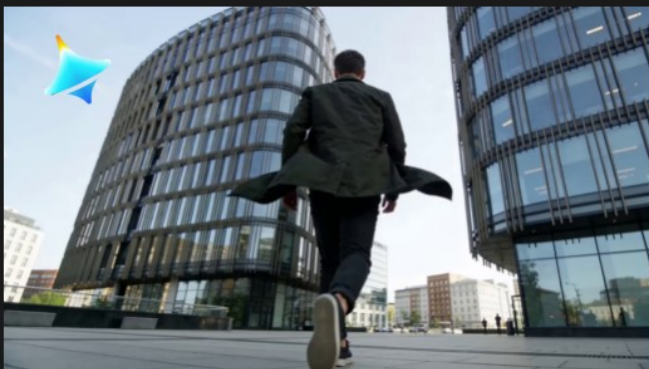
模型	任务	美学与画质	语义遵循度	物理与运动逻辑	时间与角色一致性	总分
Sora	难度3:多人+低动态					
Veo	难度3:多人+低动态					
Luma	难度3:多人+低动态					
Kling (可灵)	难度3:多人+低动态					
Hailuo (海螺)	难度3:多人+低动态					
Jimeng (即梦)	难度3:多人+低动态					
Wan 2.6 (万相)	难度3:多人+低动态					
Vidu	难度3:多人+低动态					

模型	任务	美学与画质	语义遵循度	物理与运动逻辑	时间与角色一致性	总分
Sora	难度2:单人+高动态					
Veo	难度2:单人+高动态					
Luma	难度2:单人+高动态					
Kling (可灵)	难度2:单人+高动态					
Hailuo (海螺)	难度2:单人+高动态					
Jimeng (即梦)	难度2:单人+高动态					
Wan 2.6 (万相)	难度2:单人+高动态					
Vidu	难度2:单人+高动态					

模型	任务	美学与画质	语义遵循度	物理与运动逻辑	时间与角色一致性	总分
Sora	难度4:单、多人+高动态					
Veo	难度4:单、多人+高动态					
Luma	难度4:单、多人+高动态					
Kling (可灵)	难度4:单、多人+高动态					
Hailuo (海螺)	难度4:单、多人+高动态					
Jimeng (即梦)	难度4:单、多人+高动态					
Wan 2.6 (万相)	难度4:单、多人+高动态					
Vidu	难度4:单、多人+高动态					

# 难度1-视频效果

HARD DIFFICULTY - VIDEO RESULTS



# 难度 1- 排名

EASY DIFFICULTY RANKING

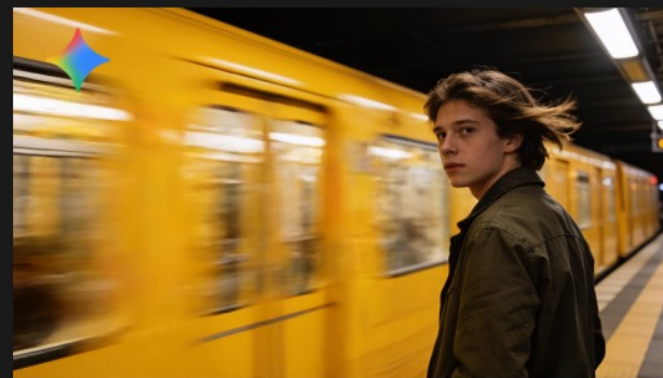
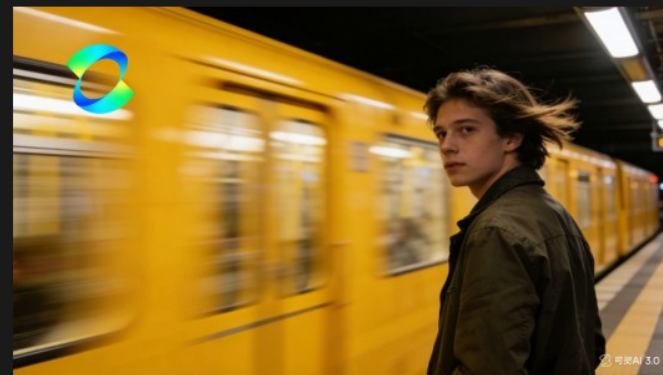


评判逻辑：所有人总分相加

A= 难度	模型	A= 评委	美学	语义	物理	一致性	创意	# 总分	平均分
难度1	Kling	评委1	5	5	4	4	3	21	4.2
难度1	Kling	评委2	4	5	5	4	3	21	4.2
难度1	Kling	评委3	5	4	4	5	3	21	4.2
难度1	Kling	评委4	4	5	4	4	4	21	4.2
难度1	Kling	评委5	5	4	5	3	3	20	4.0
难度1	Kling	评委6	4	4	4	5	3	20	4.0
难度1	Kling	评委7	5	5	3	4	3	20	4.0
难度1	Kling	评委8	4	4	5	5	2	20	4.0
难度1	Wan2.6	评委1	5	4	4	4	2	19	3.8
难度1	Wan2.6	评委2	4	5	4	4	2	19	3.8
难度1	Wan2.6	评委3	5	4	4	5	1	19	3.8
难度1	Wan2.6	评委4	4	5	4	4	2	19	3.8
难度1	Wan2.6	评委5	5	4	5	3	2	19	3.8
难度1	Wan2.6	评委6	4	4	4	5	2	19	3.8
难度1	Wan2.6	评委7	5	5	3	3	2	18	3.6
难度1	Wan2.6	评委8	4	4	5	3	2	18	3.6
难度1	Ve0	评委1	5	4	4	4	2	19	3.8
难度1	Ve0	评委2	4	5	4	4	2	19	3.8
难度1	Ve0	评委3	5	4	4	5	1	19	3.8
难度1	Ve0	评委4	4	5	4	4	2	19	3.8
难度1	Ve0	评委5	5	4	5	3	2	19	3.8
难度1	Ve0	评委6	4	4	4	4	2	18	3.6
难度1	Ve0	评委7	5	5	3	3	2	18	3.6
难度1	Ve0	评委8	4	4	5	3	2	18	3.6
难度1	Hailuo	评委1	5	4	4	4	2	19	3.8
难度1	Hailuo	评委2	4	5	4	4	2	19	3.8
难度1	Hailuo	评委3	5	4	4	5	1	19	3.8
难度1	Hailuo	评委4	4	5	4	4	2	19	3.8
难度1	Hailuo	评委5	4	4	4	4	2	18	3.6
难度1	Hailuo	评委6	5	4	5	2	2	18	3.6
难度1	Hailuo	评委7	4	5	4	3	2	18	3.6
难度1	Hailuo	评委8	5	4	4	3	2	18	3.6
难度1	Luma	评委1	5	4	4	4	2	19	3.8
难度1	Luma	评委2	4	5	4	4	2	19	3.8

# 难度2-视频效果

HARD DIFFICULTY - VIDEO RESULTS



# 难度 2- 排名

EASY DIFFICULTY RANKING



评判逻辑：所有人总分相加

A= 难度	模型	A= 评委	美学	语义	物理	一致性	创意	# 总分	平均分
难度1	Jimeng	评委1	2	2	2	2	3	11	2.2
难度1	Jimeng	评委2	3	2	2	2	2	11	2.2
难度1	Jimeng	评委3	2	3	2	2	2	11	2.2
难度1	Jimeng	评委4	2	2	3	2	2	11	2.2
难度1	Jimeng	评委5	2	2	2	3	2	11	2.2
难度1	Jimeng	评委6	3	2	2	2	2	11	2.2
难度1	Jimeng	评委7	2	3	2	2	2	11	2.2
难度1	Jimeng	评委8	2	2	3	2	2	11	2.2
难度1	Veo	评委1	5	5	4	4	4	22	4.4
难度1	Veo	评委2	4	5	5	4	4	22	4.4
难度1	Veo	评委3	5	4	4	5	4	22	4.4
难度1	Veo	评委4	4	5	4	4	5	22	4.4
难度1	Veo	评委5	5	4	5	4	4	22	4.4
难度1	Veo	评委6	4	4	4	5	5	22	4.4
难度1	Veo	评委7	5	5	3	4	5	22	4.4
难度1	Veo	评委8	4	4	5	5	4	22	4.4
难度1	Kling	评委1	4	4	4	4	3	19	3.8
难度1	Kling	评委2	4	4	4	3	4	19	3.8
难度1	Kling	评委3	4	4	3	4	4	19	3.8
难度1	Kling	评委4	4	3	4	4	4	19	3.8
难度1	Kling	评委5	4	4	4	4	4	20	4.0
难度1	Kling	评委6	4	4	4	4	4	20	4.0
难度1	Kling	评委7	4	4	4	4	4	20	4.0
难度1	Kling	评委8	4	4	4	4	4	20	4.0
难度1	Wan2.6	评委1	5	4	4	4	4	21	4.2
难度1	Wan2.6	评委2	4	5	4	4	4	21	4.2
难度1	Wan2.6	评委3	4	4	5	4	4	21	4.2
难度1	Wan2.6	评委4	4	4	4	5	4	21	4.2
难度1	Wan2.6	评委5	4	4	4	4	5	21	4.2

# 难度3-视频效果

HARD DIFFICULTY - VIDEO RESULTS



# 难度 3- 排名

EASY DIFFICULTY RANKING



**评测结论 -Kling**  
 高级难度结论：综合评判 Kling 效果最佳

评判逻辑：所有人总分相加

A: 难度	模型	A: 评委	美学	语义	物理	一致性	创意	总分	平均分
难度 3	Kling	评委 1	5	5	4	4	2	20	4.0
难度 3	Kling	评委 2	4	5	5	4	2	20	4.0
难度 3	Kling	评委 3	5	4	4	5	2	20	4.0
难度 3	Kling	评委 4	4	5	4	4	3	20	4.0
难度 3	Kling	评委 5	5	4	5	3	3	20	4.0
难度 3	Kling	评委 6	4	4	4	5	3	20	4.0
难度 3	Kling	评委 7	5	5	3	4	2	19	3.8
难度 3	Kling	评委 8	4	4	5	5	1	19	3.8
难度 3	Wan2.6	评委 1	5	4	4	4	2	19	3.8
难度 3	Wan2.6	评委 2	4	5	4	4	2	19	3.8
难度 3	Wan2.6	评委 3	5	4	4	5	1	19	3.8
难度 3	Wan2.6	评委 4	4	5	4	4	2	19	3.8
难度 3	Wan2.6	评委 5	4	4	4	4	2	18	3.6
难度 3	Wan2.6	评委 6	5	4	5	2	2	18	3.6
难度 3	Wan2.6	评委 7	4	5	4	3	2	18	3.6
难度 3	Wan2.6	评委 8	5	4	4	3	2	18	3.6
难度 3	Ve0	评委 1	5	4	4	3	2	18	3.6
难度 3	Ve0	评委 2	4	5	4	3	2	18	3.6
难度 3	Ve0	评委 3	5	4	4	3	2	18	3.6
难度 3	Ve0	评委 4	4	5	4	3	2	18	3.6
难度 3	Ve0	评委 5	5	4	5	2	2	18	3.6
难度 3	Ve0	评委 6	4	4	4	3	2	17	3.4
难度 3	Ve0	评委 7	5	5	3	2	2	17	3.4
难度 3	Ve0	评委 8	4	4	5	2	2	17	3.4
难度 3	Hailuo	评委 1	5	4	4	2	2	17	3.4
难度 3	Hailuo	评委 2	4	5	4	2	2	17	3.4
难度 3	Hailuo	评委 3	5	4	4	2	2	17	3.4
难度 3	Hailuo	评委 4	4	5	4	2	2	17	3.4
难度 3	Hailuo	评委 5	4	4	4	2	2	16	3.2
难度 3	Hailuo	评委 6	5	4	5	2	2	16	3.2
难度 3	Hailuo	评委 7	4	5	4	2	2	16	3.2
难度 3	Hailuo	评委 8	5	4	4	2	2	16	3.2
难度 3	Luma	评委 1	4	4	4	2	2	16	3.2
难度 3	Luma	评委 2	5	4	3	2	2	16	3.2

# 难度4-视频效果

HARD DIFFICULTY - VIDEO RESULTS



# 难度 4- 排名

EASY DIFFICULTY RANKING

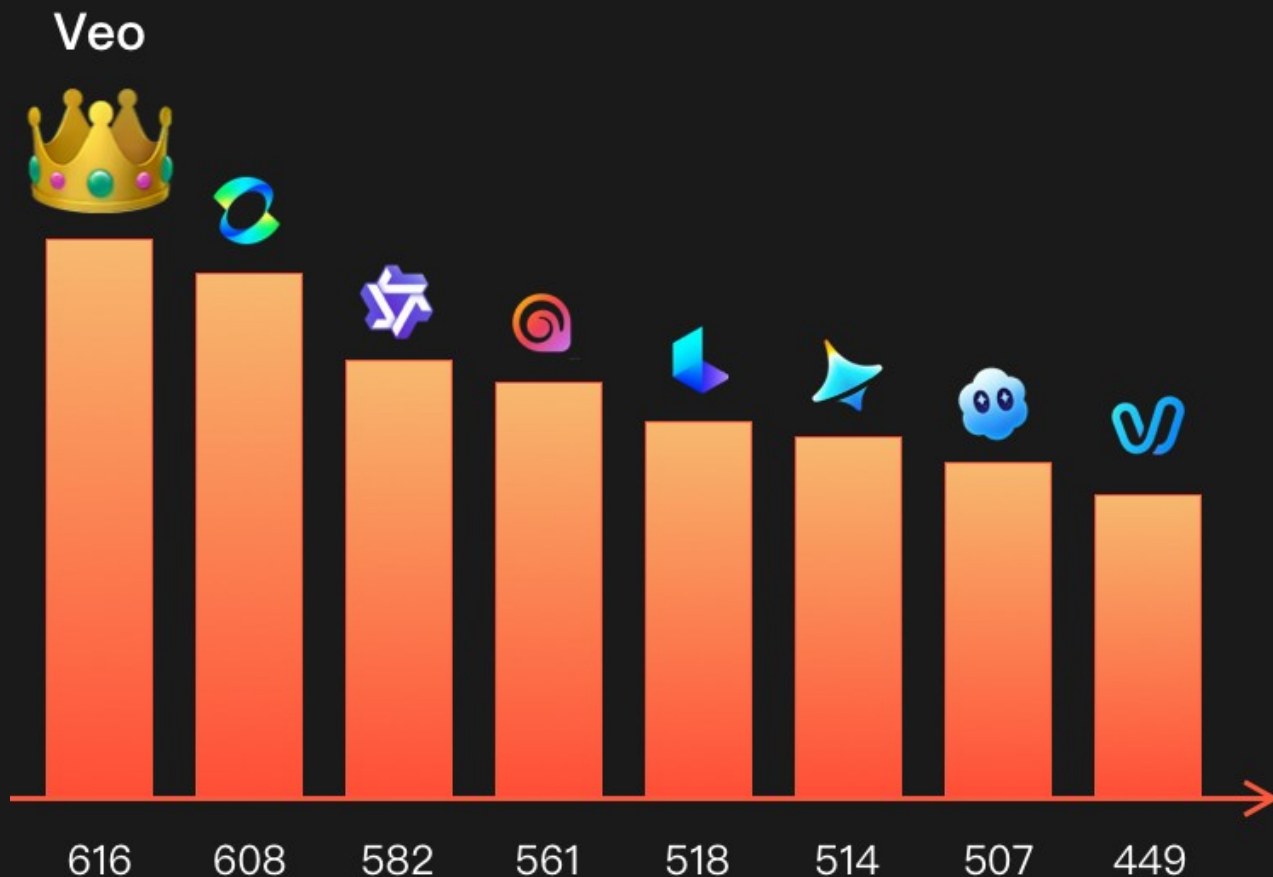


**评测结论 -Veo**  
 高级难度结论：**综合评判 Veo 效果最佳**

评判逻辑：所有人总分相加

难度	模型	评委	美学	语义	物理	一致性	创意	总分	平均分
难度 4	Veo	评委 1	5	5	4	4	2	20	4.0
难度 4	Veo	评委 2	4	5	5	4	2	20	4.0
难度 4	Veo	评委 3	5	4	4	5	2	20	4.0
难度 4	Veo	评委 4	4	5	4	4	3	20	4.0
难度 4	Veo	评委 5	5	4	5	3	3	20	4.0
难度 4	Veo	评委 6	4	4	4	5	3	20	4.0
难度 4	Veo	评委 7	5	5	3	4	3	20	4.0
难度 4	Veo	评委 8	4	4	5	5	2	20	4.0
难度 4	Wan2.6	评委 1	5	4	4	4	2	19	3.8
难度 4	Wan2.6	评委 2	4	5	4	4	2	19	3.8
难度 4	Wan2.6	评委 3	5	4	4	5	1	19	3.8
难度 4	Wan2.6	评委 4	4	5	4	4	2	19	3.8
难度 4	Wan2.6	评委 5	4	4	4	4	2	18	3.6
难度 4	Wan2.6	评委 6	5	4	5	2	2	18	3.6
难度 4	Wan2.6	评委 7	4	5	4	3	2	18	3.6
难度 4	Wan2.6	评委 8	5	4	4	3	2	18	3.6
难度 4	Kling	评委 1	5	4	4	3	2	18	3.6
难度 4	Kling	评委 2	4	5	4	3	2	18	3.6
难度 4	Kling	评委 3	5	4	4	3	2	18	3.6
难度 4	Kling	评委 4	4	5	4	3	2	18	3.6
难度 4	Kling	评委 5	4	4	4	3	2	17	3.4
难度 4	Kling	评委 6	5	4	3	3	2	17	3.4
难度 4	Kling	评委 7	4	5	4	2	2	17	3.4
难度 4	Kling	评委 8	5	4	3	3	2	17	3.4
难度 4	Hailuo	评委 1	5	4	4	3	2	18	3.6
难度 4	Hailuo	评委 2	4	5	4	3	2	18	3.6
难度 4	Hailuo	评委 3	5	4	4	3	2	18	3.6
难度 4	Hailuo	评委 4	4	4	4	3	2	17	3.4
难度 4	Hailuo	评委 5	5	4	3	3	2	17	3.4
难度 4	Hailuo	评委 6	4	5	4	2	2	17	3.4
难度 4	Hailuo	评委 7	5	4	3	3	2	17	3.4
难度 4	Hailuo	评委 8	4	4	4	3	2	17	3.4
难度 4	Sora	评委 1	5	4	4	3	2	18	3.6
难度 4	Sora	评委 2	4	5	4	3	2	18	3.6

## 图生视频综合排名



在图生视频评测中，**Ve**o 综合性能稳定，复杂场景建模能力出色，占据图生视频第一。各个产品的图生视频差异主要体现在语义遵循度、物理和运动逻辑、时间和人物一致性方面。排名靠前的模型能够有效保持输入图像的主体结构、风格特征和空间关系，并在此基础上扩展出连贯的动态效果，其生成的视频在运动逻辑、镜头衔接及画质稳定性方面表现较为成熟。而排名靠后的产品在动态延展过程中更容易出现细节丢失、物体凭空出现或消失以及人物运动较为僵硬。这表明排名靠后的产品在将静态视觉信息转化为稳定时间序列方面仍存在瓶颈。总体看，图生视频模型能力的高低更多取决于对输入图像结构的保持能力以及对物理运动规律的理解程度

05

# 总结论

CONCLUSION



# 头部产品表现分析

MODEL EVALUATION FRAMEWORK

即梦

## 文生视频 场景冠军



核心能力与技术亮点

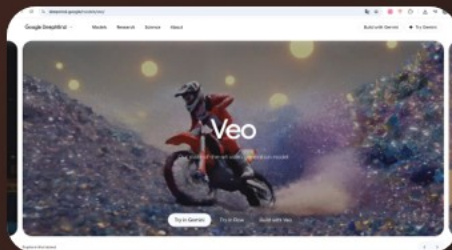
- 强大的复杂文本指令解析能力  
生成结果与Prompt目标高度一致
- 优秀的时间与角色一致性  
时间推移和多角色场景中保持稳定的身份和连续性动作
- 自然的物理与运动逻辑  
动作衔接自然，符合基本物理引擎，错误率低
- 优秀的画质与美学表现  
整体图像完成度高，风格统一，视觉层次清晰

### 优势场景

文生视频、复杂文本指令解析、时间与角色一致性

Veio

## 图生视频 场景冠军



核心能力与技术亮点

- 强大的复杂文本指令解析能力  
精准捕捉文本中的场景描述与物体属性
- 优秀的时间与角色一致性  
在各个难度的Prompt中，生成视频的稳定性优于大多数产品
- 合理的物理与运动逻辑  
物体轨迹合理，镜头切换自然，空间连续性强
- 出色的画质与美学表现  
清晰度和细节还原能力出色，图像噪点控制良好

### 优势场景

图生视频、画质还原、语义通顺、场景稳定性

# 中等表现产品分析

MODEL EVALUATION FRAMEWORK

## 可灵

文生视频3名 & 图生视频2名

- ✔ 优势：语义符合度高，文本与图像映射关系清晰
- ✔ 优势：角色身份连续性尚可，基本叙事结构完整

- ⚠ 局限：物理和运动逻辑波动较大，部分场景存在违背常识的动作
- ⚠ 局限：色彩搭配与构图合理性不足

## 万相

文生视频5名 & 图生视频3名

- ✔ 优势：美学表现力强，构图和色彩完成度高
- ✔ 优势：细节清晰，视觉吸引力出色

- ⚠ 局限：复杂Prompt中语义精确度不高，“好看但不完全相关”
- ⚠ 局限：不适合复杂图像或高精度动态化的场景

## 海螺

文生视频6名 & 图生视频4名

- ✔ 优势：图生视频还原的专项强者，语义遵循度准确
- ✔ 优势：准确理解图像中的核心元素与空间关系

- ⚠ 局限：图像质量和美感较为保守，整体冲击力不足
- ⚠ 局限：物理与运动逻辑表现一般，大体符合常识

## Vidu

文生视频4名 & 图生视频7名

- ✔ 优势：语义遵循度总体可靠，基本涵盖所有关键指令
- ✔ 优势：在图生视频中能深度解析图片中的隐含语义

- ⚠ 局限：美学呈现和整体风格的视觉设计感较弱
- ⚠ 局限：复杂动作场景中的物理逻辑稍显僵硬

# 尾部产品问题诊断

BOTTOM-TIER PRODUCTS DIAGNOSIS

## Luma

### ● 多维度短板

语义遵循度、时间与角色一致性和物理与运动逻辑波动较大  
存在人物前后不一致、物体不符合物理逻辑的现象  
物体凭空出现消失的情况

### ● 画质问题

画质分辨率低  
色彩搭配与构图合理性较差

### ● 竞争力评估

综合竞争力较低，在多维评估中表现不佳，难以满足生产级任务需求。

## Sora

### ● 潜力与限制

理论能力上限较高，但在当前表现受限  
在理想条件下可生成结构完整的视频结果

### ● 时序与角色一致性

物理与运动逻辑  
物体凭空出现，半空悬停的情况

### ● 功能缺陷

在图生视频方面不支持真人图像类的视频生成，限制了应用场景。

 **改进空间分析：**物理引擎模拟能力需要加强，提高对真实世界物理规律的模拟准确性；语义理解能力有待提升，特别是在复杂指令解析和文化语境理解方面

# 行业共性问题剖析

INDUSTRY COMMON ISSUES ANALYSIS

## ✘ 物理逻辑缺陷

Sora、Luma、可灵、即梦  
在处理中等难度文生视频时，普遍存在物理与运动逻辑方面的问题

### 主要表现

- 物体凭空悬浮或消失
- 运动轨迹不合理
- 动作不符合常识

### 典型案例

- ⚠ 伞的打开方式不合理，违背基本物理规律
- ⚠ 物体从空中无故出现或消失

## ✘ 语义偏离

Sora、可灵、Luma、海螺、Wan、Vidu 等  
在多主体任务或特定文化符号生成中表现出语义理解偏离

### 主要表现

- 文化符号理解错误
- 多主体关系处理不当
- 指令解析存在系统性偏差

### 典型案例

- ⚠ Sora和可灵在生成“中国龙”时，错误地生成了“西方龙”
- ⚠ Luma、可灵、海螺、Wan、Vidu在生成难度二的视频时，对于多主体中的列车行驶方向判定错误



**结论：**当前视频生成模型在物理逻辑和语义理解方面仍存在系统性缺陷，需要在基础模型层面进行突破

# I 技术选型策略建议

## TECHNOLOGY SELECTION STRATEGY

**i** 基于八大视频生成产品在文生视频与图生视频两大核心场景的量化评估结果，我们提出以下技术选型建议，以帮助用户根据具体需求进行精准选型



### 主力生产力组合

#### 推荐组合：Veo + 即梦

- > **Veo:** 图生视频场景综合最优
- > **即梦:** 文生视频场景表现第一
- > 两者结合可覆盖主流视频生成需求



### 专项补充方案

#### 特定场景补充：海螺

- > 图生视频领域画质还原专项强者
- > 对画质还原和语义遵循有较高要求
- > 可在 Veo 表现不佳时提供有效补充



### 规避建议

#### 暂不建议用于生产级任务

- > **Sora:** 无法生成真人参考视频
- > **Luma:** 多维度短板，综合竞争力低
- > 更适合探索性测试而非实际应用

# 总结

## SUMMARY



**Vidu**  
Vidu Q2

### 优势

多镜头主体一致性最强

### 局限

目前画面内多为一镜到底，不会自动切镜



**Sora**  
Sora2

### 优势

1. 美学风格表现最佳
2. 镜头语言表现最佳
3. 主体一致性排第二

### 局限

画质较差



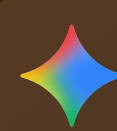
**即梦**  
seedream2.0

### 优势

可控性之王，成熟的运动笔刷、运镜控制，最像“生产力工具”的产品

### 局限

1. 目前画面内多为一镜到底
2. 多主体参照需大量抽卡



**Veo**  
Veo 3.1

### 优势

1. 人物动作逻辑感最强
2. 镜头语言连贯性极具导演思维

### 局限

1. 目前画面内多为一镜到底
2. 多主体会愿需大量抽卡



文生视频冠军  
即梦 (Jimeng)



图生视频冠军  
Veo



最佳组合  
Veo + 即梦



# Thanks



感谢观看

