

北京大学
PEKING UNIVERSITY

DeepSeek内部研讨系列

AI视频生成技术原理与行业应用

AI肖睿团队

(段永华、张惠军)

2025年11月25日

- 北大青鸟人工智能研究院
- 北大计算机学院元宇宙技术研究所
- 北大教育学院学习科学实验室

摘要

一、本讲座面向视觉内容创作者、AI技术爱好者、创意从业人员，学术研究人员、相关教育工作者和学生，旨在全面解析AI视频生成技术的技术原理与行业应用。我们聚焦前沿模型架构、厂商发展现状与行业应用案例，力求兼顾技术深度与实用价值。报告深入分析各主流厂商技术能力与产品定位，剖析不同行业的应用场景与工作流程。为创作者、决策者与开发者提供从技术理解到商业落地的系统指南，助力把握创新机遇与实施路径。

二、讲座涵盖以下几个模块：

1. **AI视频概览：**介绍 AI 和 AI 视频的基本概念、分类、行业发展现状，建立对这一技术领域的整体认识。
2. **技术原理与架构：**详细解析 AI 视频从训练到生成的完整过程、基础技术架构、训练数据，以及从基础架构到商业模型的演进路径。
3. **能力发展与突破：**分析主流厂商的技术能力、AI 视频技术的评估标准，以及市场上主要的 AI 视频工具和模型的比较。
4. **行业应用案例与价值：**全面展示 AI 视频在影视娱乐、短视频营销、教育培训、医疗健康等多个行业的具体应用案例和价值。
5. **工具/平台使用推荐：**可灵、Runway、即梦等10款国内外主流AI视频平台，提供基于应用场景、易用性分级及成本梯度的多维选型建议。

三、在技术学习的道路上，优质学习资源至关重要。AI基本概念和原理部分，推荐大家参考《人工智能通识教程（微课版）》这本系统全面的入门教材，结合B站“思睿观通”栏目的配套视频进行学习。此外，欢迎加入ai.kgc.cn社区，以及“AI肖睿团队”的视频号和微信号，与志同道合的AI爱好者交流经验、分享心得。

AI视频生成技术原理与行业应用

目 录

CONTENTS

01 AI视频概览

02 技术原理与架构

03 能力发展与突破

04 行业应用案例与价值

05 工具选择推荐



PART 01

AI视频概览

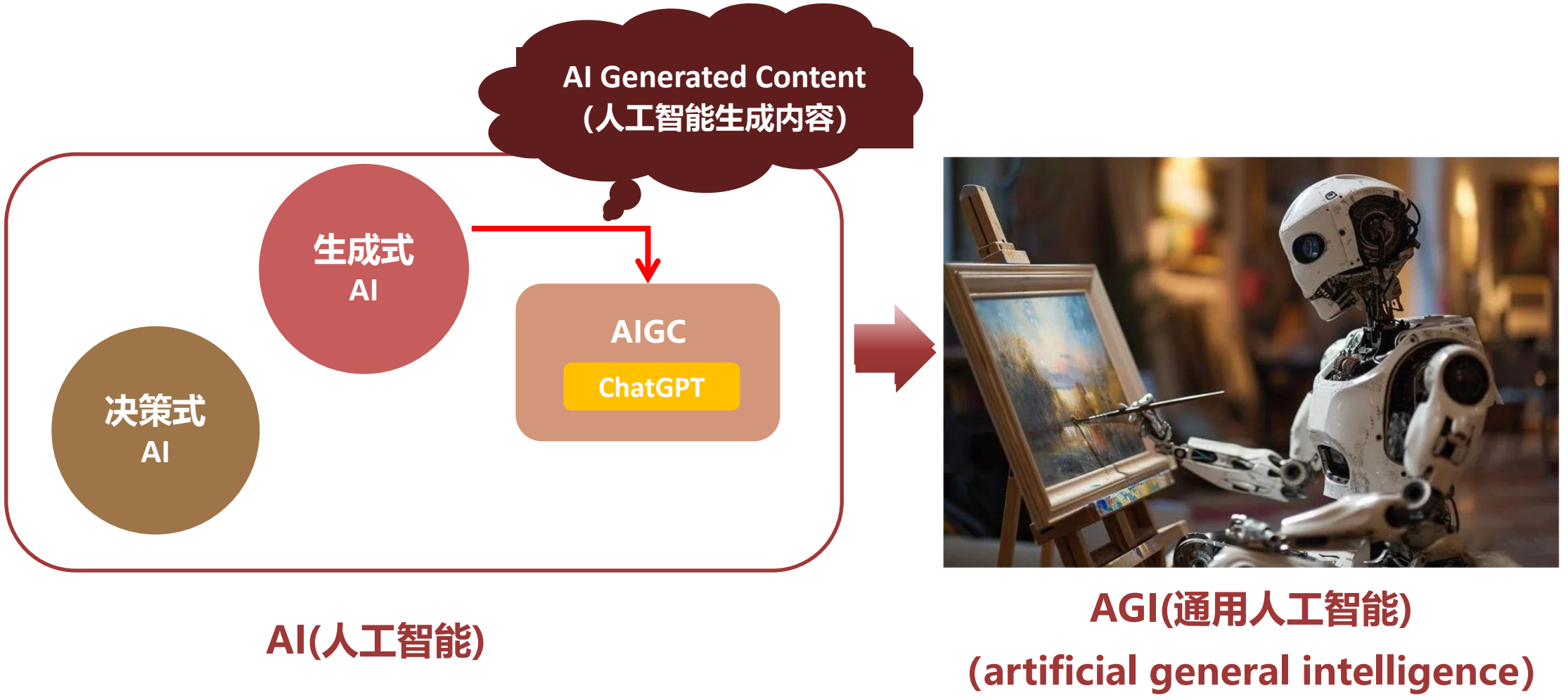
1. 认识AI

—— AI · AGI · 基于学习的AI · GenAI · AIGC

2. 认识AI视频

——定义 · 任务分类 · 发展 · 主流工具

认识AI，从厘清“名词”开始



1. 狭义AI 与 通用AI

按智能水平/范围划分，AI分为**狭义 AI** 和**通用AI**，狭义AI是当今最常见的 AI 类型，专注于特定任务；而通用AI，也称为 AGI，可以像人类一样思考、学习和运用知识。

狭义AI

- 通常被称为弱人工智能或特定人工智能
- 不具有意识或高级认知功能
- 仅限于解决特定、熟悉的任务
- 价值：提高效率、提高准确性、降低成本

AlphaGo



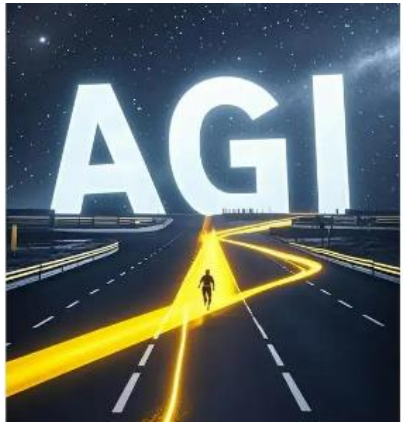


可灵AI

deepseek

通用AI

- 模仿人类智能，被称为强人工智能
- 假设具备意识和认知推理能力
- 能解决各种不熟悉的任务
- 尚未在实践中实现



2. 基于规则的AI 与 基于学习的AI

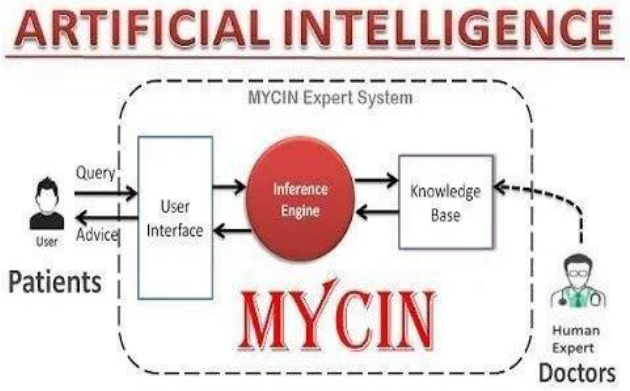
在**构建方式和范式**上，早期的AI范式是**基于规则的AI**，又称为“符号AI（symbolic AI）”，依赖于由人类专家预先编写的显式逻辑规则进行决策；而**基于学习的AI**则是通过分析海量数据来自动学习和归纳模式，是现代AI的主流范式。

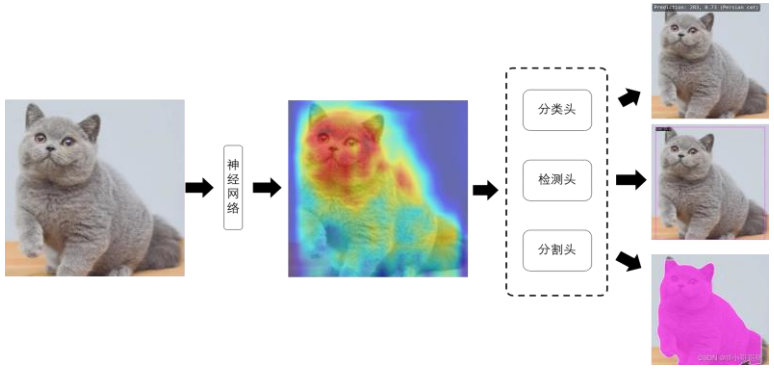


基于规则的AI

- 遵循预定义的规则和逻辑操作
- 决策基于人类专家提供的明确指示进行
- 易于理解和控制，但缺乏灵活性
- 无法适应或从新数据或经验中学习

例如：专家系统、传统聊天机器人







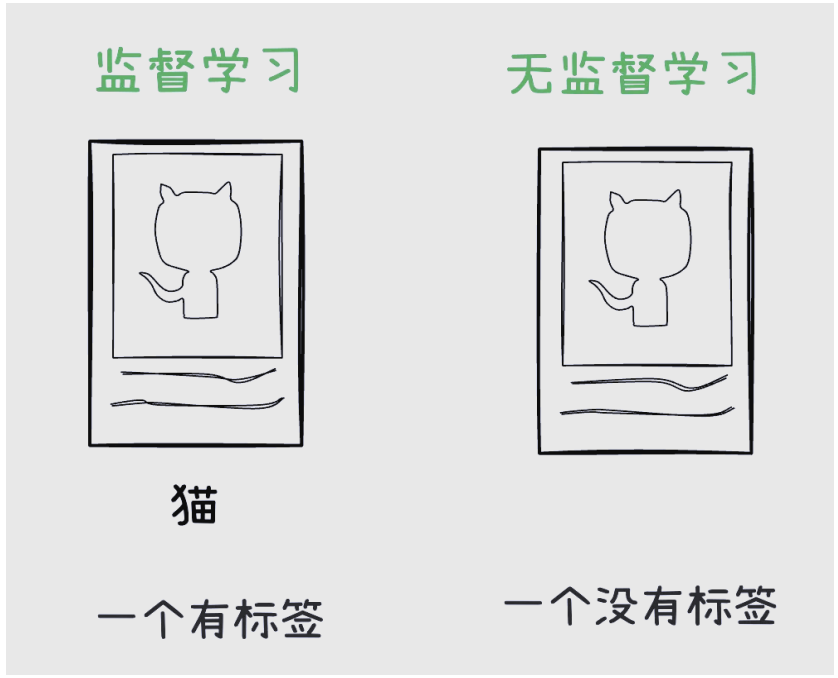
基于学习的AI

- 从数据中学习
- 利用算法识别模式并做出决策
- 灵活，可随着时间改进，但需要大量数据
- 能够处理复杂、不断变化的情况

例如：图像识别、自然语言处理、预测分析

3. 监督学习 与 无监督学习

按**学习和训练方式**，分为**监督学习**和**无监督学习**。监督学习是利用带有明确标签的“正确答案”数据进行训练，而无监督学习则是在没有标签的数据中自主发现其内在的结构与模式。



监督学习：从带有标签的训练数据中学习

算法通过分析已知的输入-输出对来学习映射函数，目标是对新的、未见过的输入数据进行准确预测。

无监督学习：在未标记的数据中发现模式

系统试图从数据的内在结构中发现模式或规律，而不依赖于预先定义的输出标签。

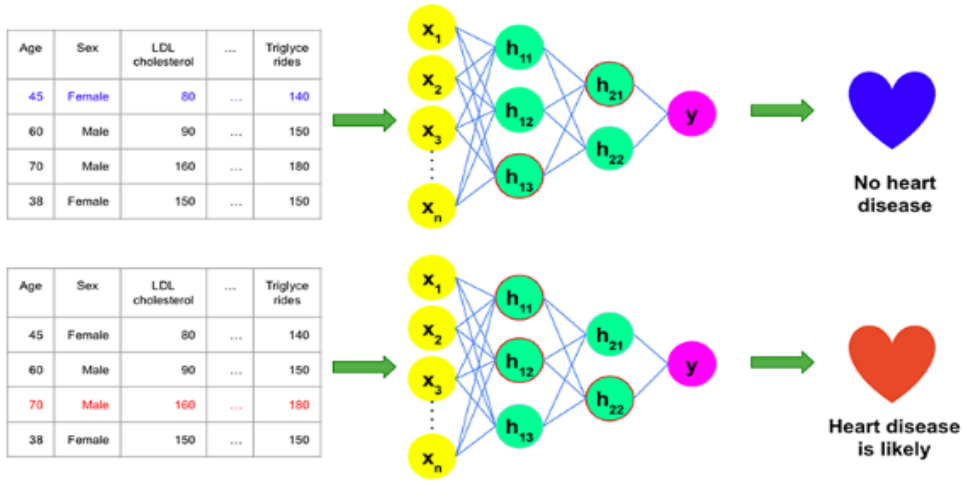
特点	监督学习	无监督学习
数据标签	需要带标签的训练数据	不需要带标签的数据
学习过程	学习输入与输出之间的映射关系	发现数据内在的结构和模式
训练目标	最小化预测错误	找出数据中的隐藏结构
应用场景	分类、回归问题	聚类、降维、关联规则学习
数据准备	需要大量人工标注	无需人工标注
应用例子	垃圾邮件过滤、图像识别、疾病诊断	客户细分、异常检测、推荐系统

4. 决策式AI VS 生成式AI

在基于学习的AI范式下，按**AI要完成的核心任务**分为**决策式AI**与**生成式AI**，决策式AI专注于对输入数据进行识别、判断和预测；生成式AI的目标是根据学习到的模式创造和生成全新的、原创的内容。

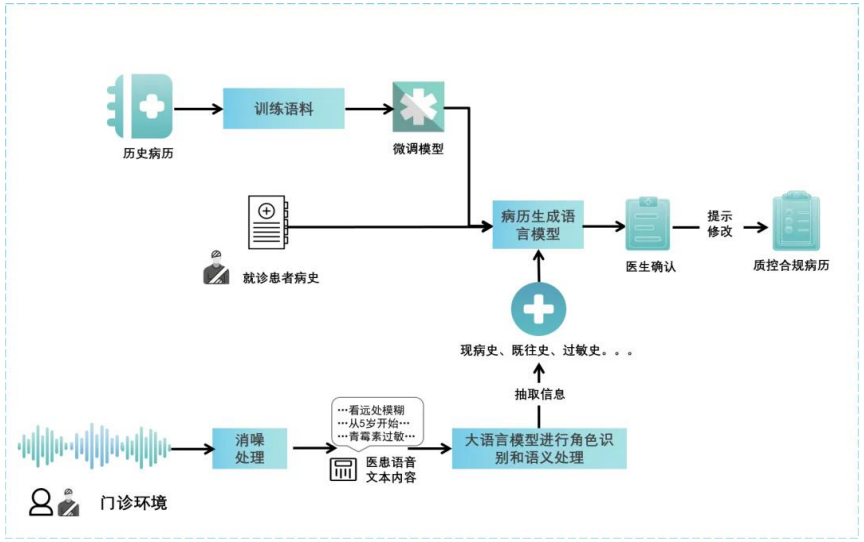
决策式AI：专注于分析现有数据，以对未来事件或趋势做出预测，常被应用于金融、医疗保健和营销等领域。决策式AI对于需要进行预测的决策过程至关重要。

例如，一个决策式AI模型可能会分析历史销售数据以预测未来销售，或者它可能会使用患者数据来预测出现某些医疗状况的可能性。



生成式AI：能够根据所接受的训练数据生成新的内容。它可以生成图像、文本、音乐等，被广泛应用于创意产业、内容生成，甚至是为各种应用创建合成数据。

例如：复旦大学附属眼耳鼻喉科医院推出近视儿童 AI 生成式病历，医生只需专注于与患者的沟通，系统后台即可完成语音识别、关键词抓取、语义理解及病历生成的全过程。



AIGC—生成式AI在内容创作领域的多模态发展

北京大学 PEKING UNIVERSITY

AIGC（AI Generated Content），指人工智能生成内容，广泛应用于文本生成、音频生成、图像生成、视频生成及跨模态生成等。

类型	应用场景
文本生成	1. **内容创作与续写：** 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
	2. 信息处理与整合： 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
	3. 辅助写作与校对： 文章润色、语法纠错、风格转换、内容推荐、辅助事实性核查
	4. 对话式AI与智能助理： 智能客服、聊天机器人、虚拟个人助理(VPA)、情感陪伴、AI角色扮演(RPG)
	5. 代码生成与开发： 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
图像生成	**1. 文生图：** 根据文字描述生成高分辨率的概念艺术、设计草图、产品原型图、广告营销素材
	2. 图像编辑与修复： 智能内容擦除与填充、图像扩展、智能修复（老照片上色、去噪、提升清晰度）
	3. 可控图像生成与编辑：根据草图、姿态、深度图等精控内容，风格迁移、局部重绘、元素替换
视频生成	**1. 视频内容生成：** 根据文字或图片生成短视频、广告片、动画片段等
	2. 智能视频剪辑与处理： 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
	3. 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
	4. 数字人与虚拟形象： 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
音频生成	**1. 语音合成 (TTS)：** 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
	2. 声音克隆与转换： 用少量样本复制特定人物声音，声音特征转换，跨语种声音克隆、个性化配音
	3. 音乐与音效生成： 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
跨模态生成	**1. 视觉内容理解与描述：** 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
	2. 多模态交互： 实现图文、音视频结合的复杂推理与对话问答
	3. 3D模型生成： 根据文本或2D图像生成3D模型，应用于游戏开发、工业设计、虚拟现实（VR/AR）

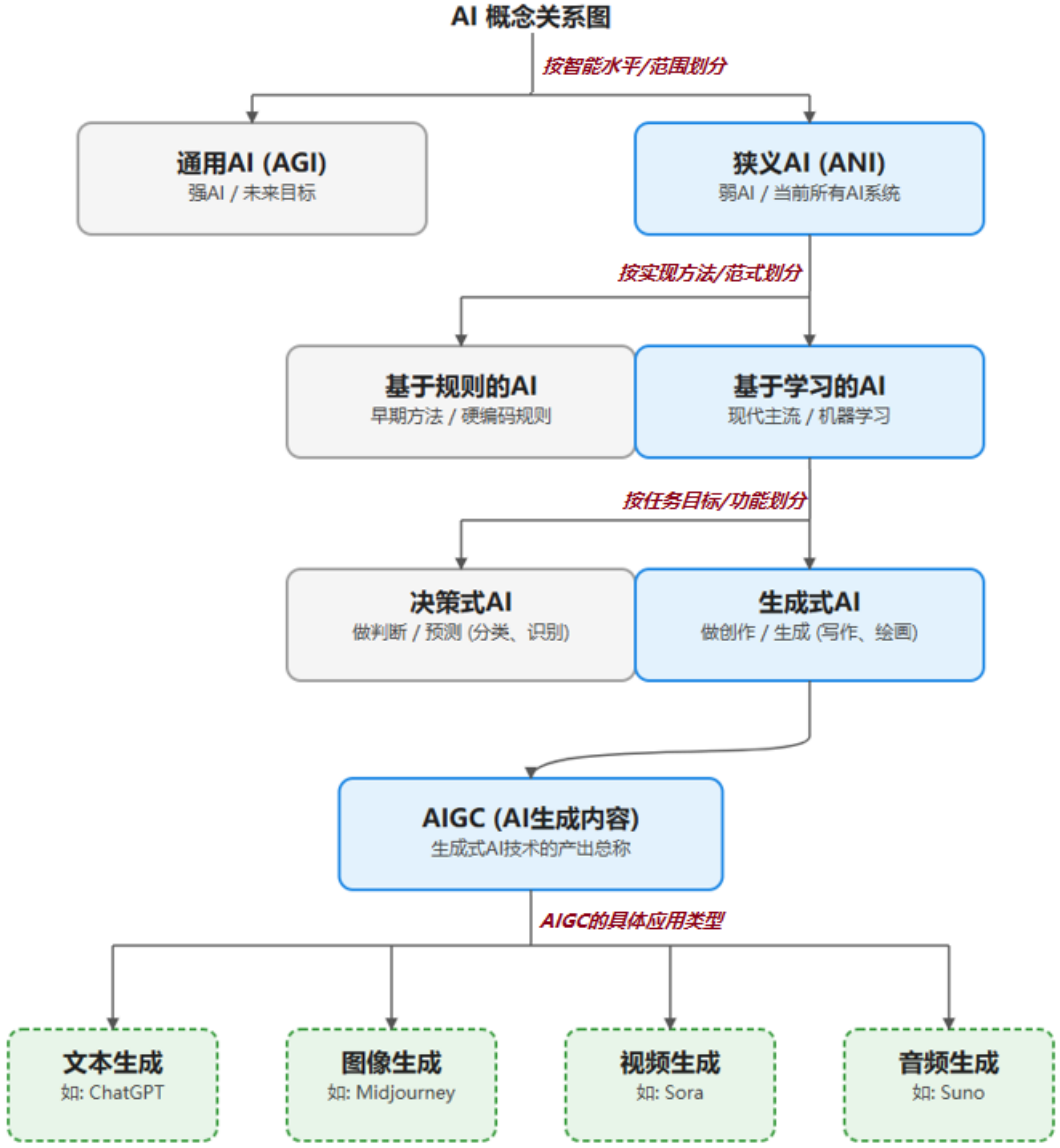








「AI相关概念」小结



1. 通用AI（AGI）是拥有与人类相当的全面智慧的终极理想，目前尚未实现。今天所讨论和使用的一切AI，都属于狭义AI 的范畴，它们是为解决特定任务而设计的工具。
2. 构建AI的方法，早期依赖基于规则的AI，像一本写好的“说明书”；而现代AI的主流是基于学习的AI，它能像学生一样从海量数据中自主学习规律，是后续高级功能的基础。
3. 基于学习的AI根据其任务目标，又可分为两大核心能力。决策式AI 像在做“判断题”，负责识别和分类（如人脸识别）；而生成式AI 则像在做“创作题”，负责创造全新的原创内容（如写文章、绘画）。
4. 生成式AI创造生成内容简称为 AIGC，文本生成、图像生成、视频生成、音频生成等都是AIGC在不同媒介上的具体表现形式。

AI视频生成是基于学习的AI，属于生成式AI范畴，而其产出的动态影像内容，是目前 AIGC 中技术最复杂、发展最迅速的分支之一。本报告将聚焦于此，深入探讨AI视频技术的核心原理、当前进展、行业应用与与实践以及未来前景展望。



PART 01

AI视频概览

1. 认识AI

—— AI · AGI · 基于学习的AI · GenAI · AIGC

2. 认识AI视频

——定义 · 任务分类 · 发展 · 主流工具

AI视频生成—重塑视频内容创作范式

2024年初，OpenAI发布的Sora模型凭借一分钟长度、高清画质及物理世界连贯性理解，重新定义了AI视频生成能力。随后，多个模型相继问世，展现相近或更高的技术水平。AI视频正从实验性"技术炫技"阶段迅速发展为影视预演、广告创意、短视频制作和在线教育的实用工具。尽管在可控性、逻辑一致性和成本方面仍存挑战，但行业变革的临界点已清晰可见。



2024年初Sora生成的60秒视频《行走在街头的女人》



可灵AI基于文本提示词生成的视频



RunwayML技术应用于电影《瞬息全宇宙》特效

AI视频的概念及任务分类

北京大学 PEKING UNIVERSITY

AI视频指利用人工智能（特别是机器学习和计算机视觉）来生成、编辑、分析或增强视频内容。

视频生成

重点讨论

从非视频模态（如文本、图像、音频）信息中，从零开始或大幅度创造全新的视频序列。

- **基于文本生成视频**
- **基于图像生成视频**（单图动画化、多图序列故事化转换）
- **基于音频生成视频**（音景视觉化、音乐视频自动生成）
- **3D/世界模型生成**（文本/图像到3D场景、3D场景动画生成、2D到3D视频转换）
- **数字人生成**（文本驱动表情/动作、对口型、克隆、形象定制）

视频编辑

在现有视频的基础上，通过AI进行内容、风格、时序或属性的修改与提升。

- **内容修改**（视频修复/填充、视频中物体替换、视频画面扩展）
- **时序编辑**（视频插帧实现流畅慢动作、视频变速/重定时）
- **视觉属性编辑**（风格化转换、重着色/调色、光照重置、特效添加）
- **视觉质量增强**（超分辨率提升、去噪/去模糊处理、视频防抖优化

视频理解

让AI模仿人类的视觉和认知能力，解析、描述和推理视频内容。

- **视频高层语义分析**（视频分类标签、行为识别、事件检测、场景分割）
- **视频内容摘要与交互**（关键内容摘要生成、自动字幕/描述生成、视频问答交互、内容检索）
- **视频时空感知**（目标检测与追踪、人体姿态估计与追踪）
- **多模态理解**（视觉-语言对齐、音视频事件定位、跨模态内容检索）

AI视频生成三种典型方式

01

文本到视频

基于文本提示生成符合描述的视频

“日出，镜头光晕清冷风，一个五官精致的年轻中国女子，长长的头发被风吹乱，头发丝飘散在脸上，穿着夏装，背景海边沙滩”



02

图像到视频

将静态图像转换为动态视频



一个身穿中国古代红色服饰的女生，在一个未来世界的街道上行走

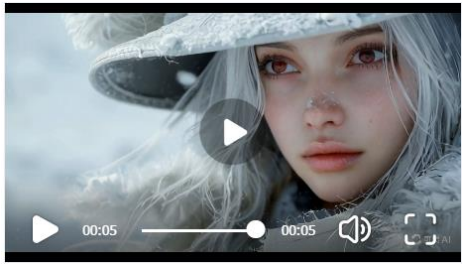


严格来说，属于视频编辑范畴

03

视频到视频

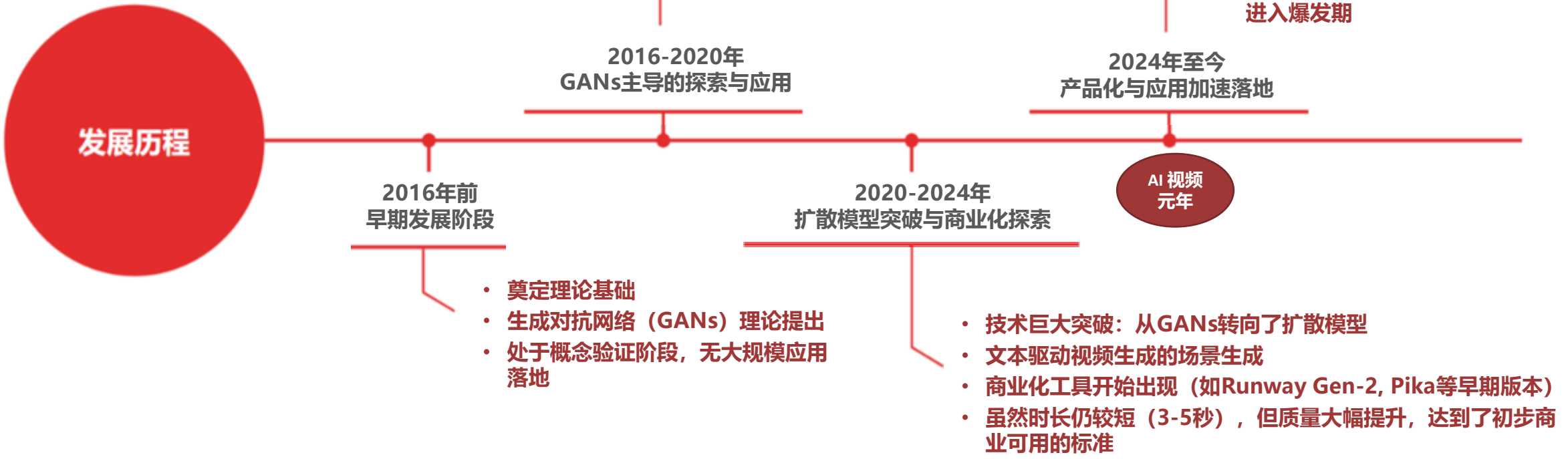
对已有视频进行增强或转换



一个女人站在雪地里，抬起右手摸了一下帽檐



AI视频生成的历史演进



当前主流的AI视频生成平台与工具

厂商名称	工具名称	发布时间	特 点
OpenAI	Sora	2024年	24年2月首次向外展示60秒演示视频引起轰动，开启AI视频元年，24年12月正式发布，对物理世界有深刻理解，能生成复杂场景、多角度镜头和富有情感的角色，已集成进ChatGPT，可生成20秒视频。
快手	可灵 AI	2024年	长时长与高画质：支持生成长达2分钟、1080p、30fps的视频。采用类似Sora的DiT架构，能模拟真实物理世界，迭代迅速，Kling 2.5 Turbo（2025年9月）成本降低30%，目前全球用户已突破4500万。
字节	即梦AI	2024年	画面审美、光影质感及中文语义理解上表现卓越。背靠抖音与剪映生态，打通了从“AI生成”到“剪辑发布”的全链路，国内唯一一个具有内容消费属性的社区，月活5400万，25年10月网站1000万访问量。
Minimax	海螺AI	2024年	一个类似ChatGPT的综合AI助手，视频生成是其多模态能力的一部分。在一个平台内提供文本、语音、图像、视频等多种AI能力。具备基础的文生视频功能，更侧重于多模态能力的融合体验。
爱诗科技	PixVerse	2024年	24年1月（早于Sora）发布V1版本，免费额度较高，模型迭代迅速，尤其在动漫和艺术风格上表现出色。因各种视频特效模版实现破圈，再靠模版在社交媒体平台上的传播来吸引更多用户，目前APP月活2300万。
Runway	Runway	2023年	最早AI视频商业化的公司之一，提供完整的"AI魔法工具"套件，包括运动笔刷、镜头控制、视频修复等专业级编辑功能。产品迭代迅速，最新Gen-4模型在运动、物理真实感和一致性上表现突出。
Pika Labs	Pika	2023年	在创意社区非常流行，以其出色的艺术和电影感著称。支持文生视频、图生视频，并提供局部修改、扩展画布、修改视频内容等强大编辑功能。早期通过Discord积累了大量用户和反馈，产品迭代紧贴用户需求。
生数科技	Vidu	2024年	国内首个发布的类Sora架构（U-ViT）的文生视频大模型。能够一键生成16秒的1080p高清视频，时空连贯性好，支持1-3张参考图的多元素交互控制。对中国文化有较好理解。
Luma Labs, Inc.	Luma AI	2024年	其视频模型Dream Machine生成速度快，视频质量极高，物理一致性和角色连贯性表现优异。以其3D高斯溅射和NeRF（神经辐射场）技术闻名，Ray3（2025年9月）号称是首个能"思考和推理"的视频AI模型。
阿里	通义万相	2024年	集成的WAN模型被誉为“开源界的Sora”。能生成1080p高清、光影逼真且运动幅度巨大的视频。其最大差异化优势在于“模型开源”策略，支持本地部署，不仅生成质量处于第一梯队，更极大地降低了开发者的使用门槛。

通用大模型平台的视频生成功能

除了垂直视频生成平台，越来越多的通用大模型已通过集成顶尖的视频模型，获得了非常强大的视频生成能力。

平台/厂商	通用大模型产品	核心特点与策略
OpenAI	ChatGPT	OpenAI将Sora集成到ChatGPT中，用户可以在一个对话界面内，通过自然语言完成从文本构思、图像生成到高质量视频创作的全流程。使ChatGPT从一个语言助手，成为一个全能的创意和生产力平台。
Google	Gemini (生态)	Gemini是谷歌的多模态大脑，而Veo是其专门用于视频生成的“利器”。谷歌还将Veo植入YouTube、Google Photos (动态回忆)以及其云平台Vertex AI中，让用户和开发者在现有产品中直接使用顶级的视频生成能力，核心是赋能现有生态。新的Veo 3模型还增加了对话、环境音效以及其他音频效果的支持。
Meta	Meta AI	Meta AI被集成在Facebook, Instagram, WhatsApp等社交应用中。其视频生成模型（如Emu Video）让用户能快速生成创意短视频等。其长远目标是为未来的元宇宙/AR眼镜提供实时、动态内容生成能力。
阿里巴巴	通义千问	阿里通义大模型家族中，负责图像/视频生成的是“通义万相”。它不仅在“通义千问”APP中提供给普通用户使用，更重要的是通过阿里云向企业客户提供API服务。企业可以将其视频生成能力集成到自己的电商、广告、传媒等业务流程中，体现了其强烈的产业服务属性。
字节跳动	豆包	字节跳动拥有抖音/TikTok和剪映这两大视频“主战场”。其大模型生成的视频内容可以与剪映的编辑能力无缝衔接，并直接在抖音/TikTok上分发，形成“AI生成 -> AI剪辑 -> AI分发”的完美闭环。这是最贴近C端视频消费场景的策略。

「AI视频概览」要点总结

明晰AI视频的“出身”与“定位”

- 厘清了AI的核心概念谱系，明确了AI视频生成技术是当前狭义AI (ANI) 范畴下，基于学习的AI 中的一个分支。
- AI视频功能上属于生成式AI (Generative AI)，其产出物是AIGC 的重要组成部分，代表了AI从“分析预测”到“创造生成”的关键跃迁。这一认知是理解其技术价值与潜力的基石。

掌握AI视频的“功能”与“形态”

- 深入理解AI视频的内核，明晰其本质——算法与数据驱动的动态影像创造。
- AI视频不仅仅是生成视频，其有三大核心任务分类（视频生成、视频编辑、视频理解），本报告侧重AI视频生成。
- AI视频生成有三种典型方式——文生视频、图生视频、视频到视频。

洞悉AI视频的“历程”与“格局”

- 从GANs理论提出，到扩散模型的广泛应用，再到以Sora为代表的DiT模型出现，AI视频技术已迈入"AI视频元年"。
- 目前国内外头部科技公司纷纷布局，OpenAI的Sora、快手的可灵AI、Runway的Gen-3等产品各具特色，能力从秒级短片扩展到分钟级高清视频，能力不断迭代增强，逐步进入产品商业化时代



PART 02

AI视频技术原理与架构

1. 基础技术范式与模型架构

——基础范式·模型架构·演进路径

2. 从基础生成架构到商业模型

3. AI视频生成的燃料：训练数据

从模型训练到创意实现，AI视频的生成之路

模型

模型生成阶段 ← → 模型使用阶段

带文本描述的视频
带文本描述的图片
无标签视频



01 数据收集

构建模型的“素材库”

获取原始数据，包括文本、图像或视频片段，构建规模庞大、内容多样、反映真实物理规律的数据集。

02 数据预处理

将现实世界编码为“视觉词元”

将原始数据清洗、标准化，并转换为模型能理解的数学语言，以便于训练

1. 清洗：去重、过滤低质量内容
2. 标准化：统一分辨率和帧率。
3. 数据编码

03 模型训练

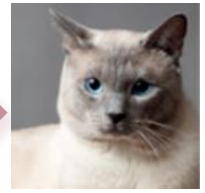
教会AI从“混沌”创造“秩序”

使用深度学习算法训练出视频生成模型

04 视频生成

根据指令开始“造梦”

根据用户输入提示 ，模型循环多次，逐步去噪，生成清晰视频画面。



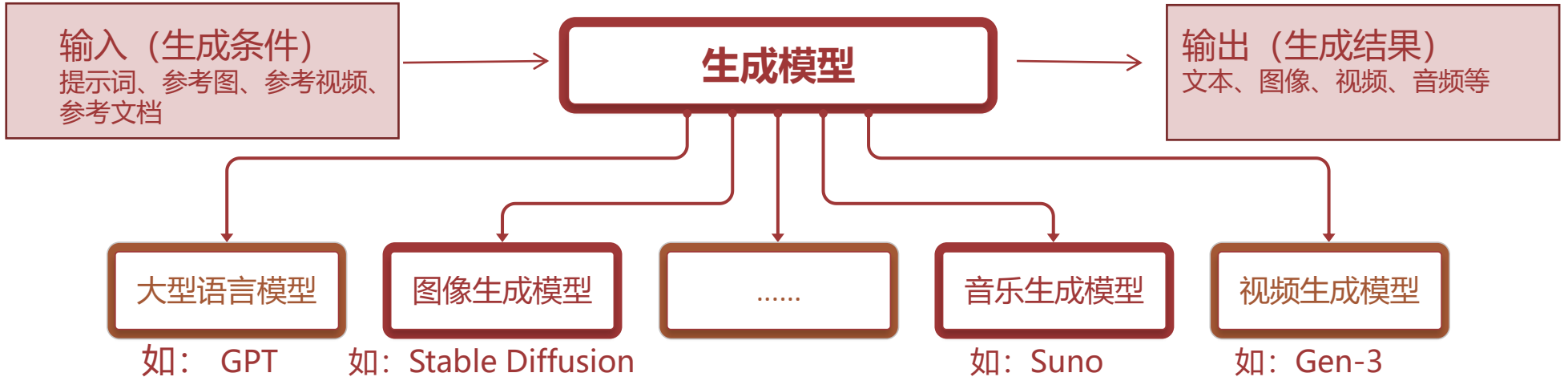
什么是“模型”？

“模型”是一个计算机程序或数字文件，它是一个为了完成特定智能任务（如识别图像、翻译语言）而被创造出来的、基于数据和算法的系统。它通过在海量数据集上进行“训练”，学会了识别特定模式，并做出预测或生成新内容的能力。

特点：

1. 是“学习”而非“编程”出来的
2. 由算法和数据共同构成，数据是模型的“原料”
3. 功能是“输入到输出”的映射
4. 目标是预测和决策，生成模型属于预测

特性	传统程序	AI模型
逻辑来源	由人明确编写的固定规则	从数据中自动学习到的模式和概率
处理方式	基于确定的指令执行	基于统计和推理进行预测
面对新情况	只能处理预设范围内的情况	具备对未知数据的泛化和推断能力
核心构成	硬编码的逻辑和算法	经过训练的参数和模型结构



理解视频生成技术，从基础生成技术范式开始

北京大学 PEKING UNIVERSITY

对比项目	变分自编码器(VAE)	生成对抗网络(GAN)	自回归模型(Autoregressive)	扩散模型(Diffusion)
基本概念	通过编码器将数据映射到潜在空间，再由解码器重构数据，同时优化重构误差和潜在空间分布	由生成器和判别器构成的对抗学习框架，通过博弈训练生成逼真样本	按顺序生成数据，每次生成依赖于之前已生成的内容	通过逐步给数据添加噪声然后学习反向去噪的过程来生成数据
优点	- 有明确的潜在表示 - 训练相对稳定 - 可用于特征学习 - 生成速度快	- 生成质量高 - 生成速度快 - 适合图像生成	- 生成过程可控 - 概率模型清晰 - 适合序列数据	- 训练稳定 - 多样性好 - 可控性强
缺点	- 生成质量较低 - 模糊问题 - 重构与KL散度平衡困难	- 训练不稳定 - 模式崩溃问题 - 难以评估	- 生成速度较慢 - 错误可能累积 - 长序列困难	- 采样速度慢 - 计算资源消耗大 - 理论复杂
典型模型	β-VAE：更好的解耦表示 VQ-VAE：向量量化VAE NVAE：层级VAE架构	DCGAN: 深度卷积GAN StyleGAN: 风格可控图像生成 CycleGAN: 图像风格转换	GPT系列: 文本生成 PixelCNN: 图像生成 WaveNet: 音频生成	DDPM: 基础扩散模型 Stable Diffusion: 潜空间扩散 DALL-E 2/3: 文本到图像
应用场景	图像生成、特征学习、异常检测、数据压缩	图像生成、图像增强、风格转换、艺术创作	自然语言生成、代码补全、音乐创作、视频预测	图像生成、超分辨率重建、图像编辑、3D模型生成

视频生成技术范式演进路径

北京大学 PEKING UNIVERSITY

2016年前
早期发展

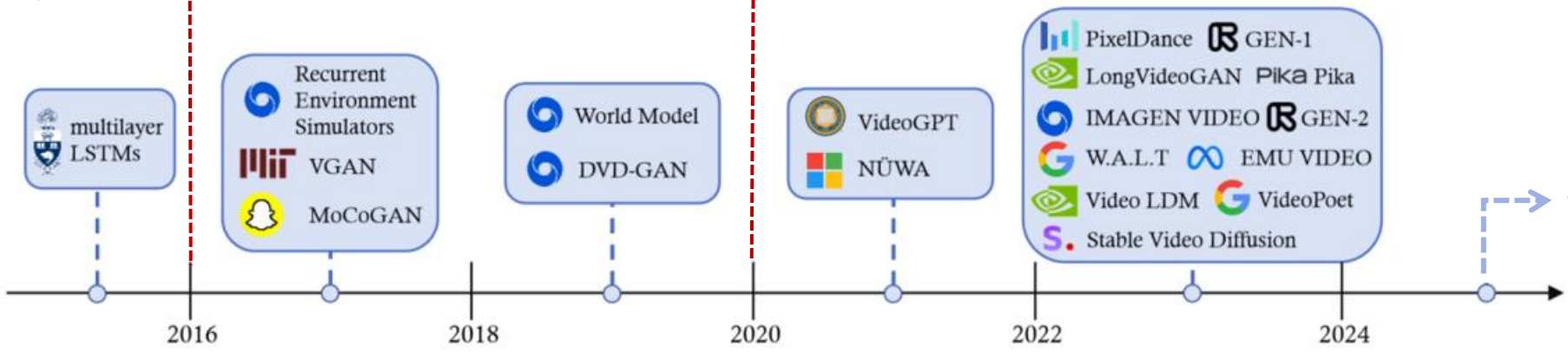
- 通过图像拼接或简单变换来模拟视频效果。

2016-2020年
GAN/VAE生成阶段

- 模型参数量小，较轻便，所以更加擅长对单个或多个对象类进行建模。

2020-至今
自回归模型及扩散模型生成阶段

- 在生成质量上优于GAN，逐步应用于高质量视频生成。
- 通过逐步去噪生成视频帧，具有稳定性强、画质高的特点。



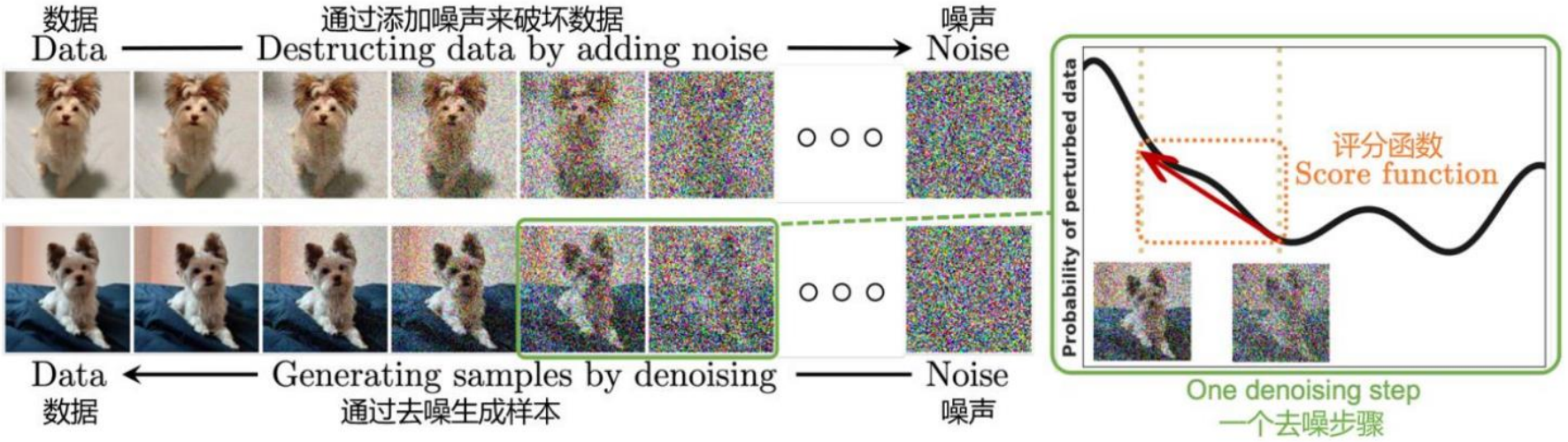
Sora：**Sora 2**
即梦AI：**Seedance 1.0**
Luma AI：**Ray 3**
Runway：**Gen-4**
Vidu：**Vidu Q2**
可灵AI：**kLing 2.5**
Pika Labs：**Pika 2.5**
PixVerse：**Model V5**
Meta：**Movie Gen**
海螺AI：**Hailuo 2.3**
Google：**Veo 3.1**
通义万相：**WAN 2.5**
腾讯混元：**HunyuanVideo**

技术范式演进过程中涌现的一系列代表性视频生成模型。

Ref: Liu Y, Zhang K, Li Y, et al. Sora: A Review on Background, Technology, Limitations, and Opportunities of Large Vision Models[J]. arXiv preprint arXiv:2402.17177, 2024..

为什么是Diffusion？—扩散模型的原理性优势

原理： 通过连续向数据添加随机噪声，直到得到一个纯高斯噪声数据，然后再学习逆扩散的过程，经过反向降噪推断来生成图像，通过系统地扰动数据中的分布，再恢复数据分布，逐步优化过程。



扩散模型的原理性优势：

稳定性： 相较于GAN难以收敛的对抗式训练，扩散模型以其目标明确、过程稳定的去噪任务，从根本上缓解了训练不稳定和模式崩溃等核心痛点。

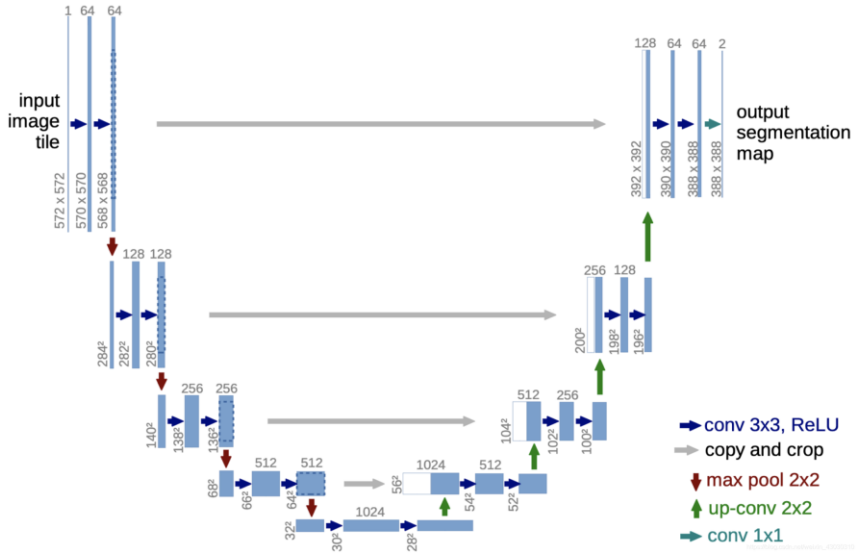
多样性： 在生成结果上，它不仅在生成结果的多样性上超越了 GAN，还克服了自回归模型难以维持全局一致性的缺陷，实现了保真度与多样性的双重领先。

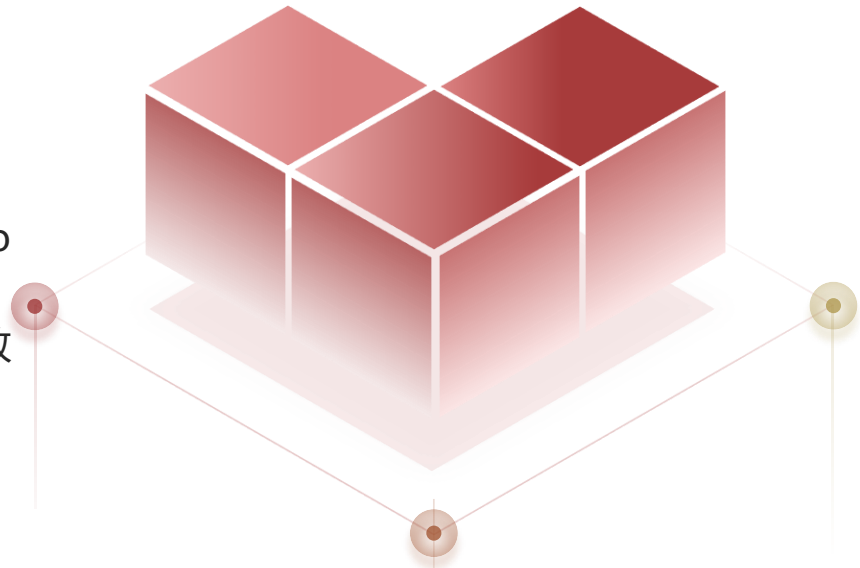
可控性： 凭借无分类器引导（CFG）带来了前所未有的文本可控性，从而在文生图等应用中取得突破性成功，一举成为主流范式。

扩散模型的两条架构路线

U-Net

- 基于SD逐帧生成
- 代表模型：Pika、Gen-2、Stable Video Difusion、MagicVideo-V2等
- 模型容易训练，然而生成的视频内容一致性要差一些，长视频生成存在困难。





混合

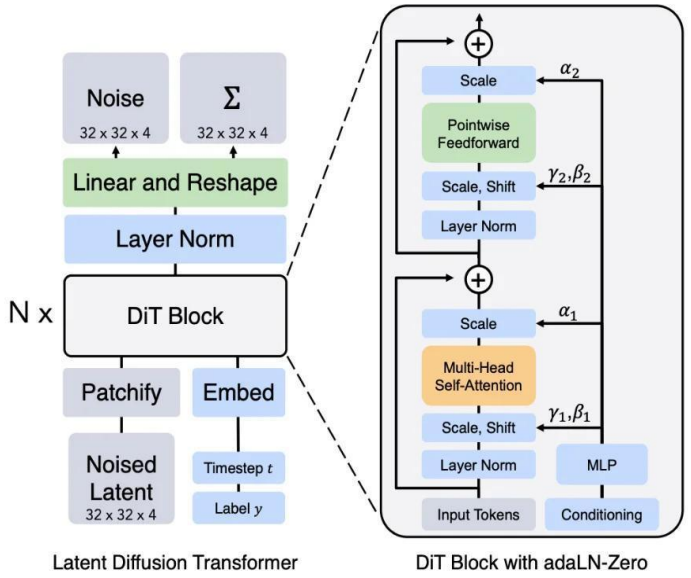
融合了U-Net结构和Transformer优势的混合架构

- U_ViT：代表模型 Vidu
- STUNet：代表模型 Google Lumiere

当前主流

DiT: Diffusion+Transformer

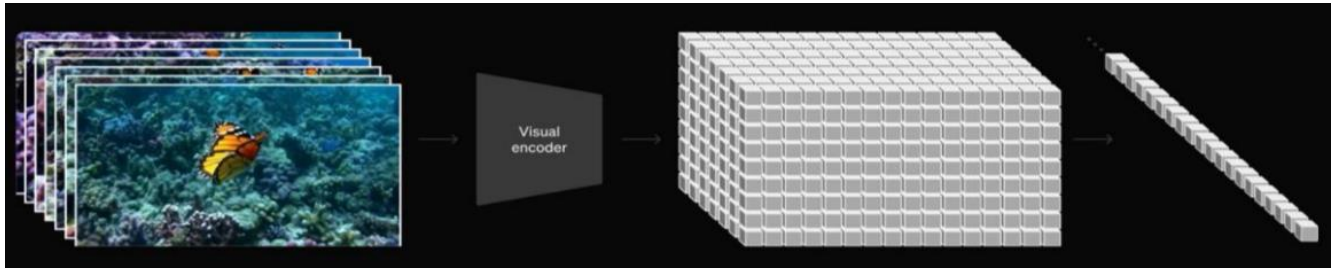
- 基于时空Patches生成
- 模型代表：Sora、清影、Pixverse V2、可灵
- 训练成本较高，但更容易保证生成视频的长度和一致性。。



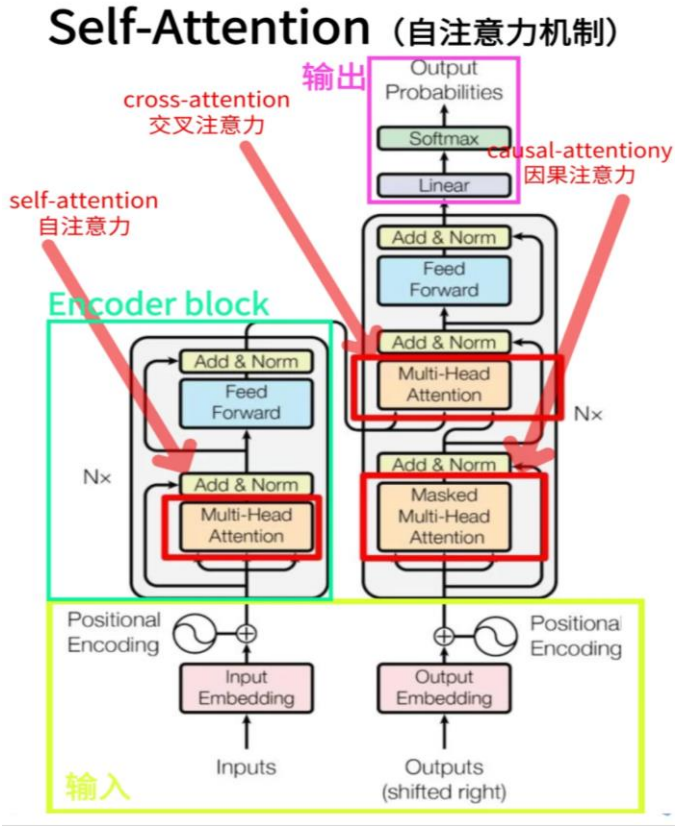
Transformer—生成技术的“游戏改变者”

Transformer不是一种生成范式，而是一种革命性的“序列处理”架构，其核心是自注意力机制，该机制赋予了AI视频模型前所未有的“全局视野”和“上下文理解能力”，是实现高质量、长时程、高一致性视频生成的关键所在。

- **时间一致性：** 通过全局视野，模型能记住一个物体在视频开头是什么样，确保它在结尾时不会“画风突变”，解决了GAN模型的“闪烁”难题。
- **理解复杂的动态关系：** 它能理解“A导致B”的逻辑。比如，一个人跑过水坑，模型知道接下来应该是“水花溅起”并且“地面变湿”，而不是毫无关联的画面。
- **可扩展性：** 对Transformer来说，长视频、短视频、宽屏、竖屏，都只是长短不一的“补丁(Patch)序列”。这使得Sora这类模型能灵活生成各种尺寸和时长的视频。



视频切成时空“补丁”，像单词一样排列，送给Transformer处理



Transformer架构

AI视频生成技术架构的演进路径

北京大学 PEKING UNIVERSITY

GAN的引入

以生成器与判别器相互博弈的对抗训练方式生成高质量图像，为AI生成领域（包括视频）提供了开创性的架构。

模型： GAN

2014

GAN的探索期

在AI视频生成的早期，研究者主要尝试将用于图像生成的GAN引入视频领域。GAN通过生成器和判别器的对抗来生成内容。

模型： VGAN、MOCOGAN、DVD-GAN

2016+

自回归Transformer的尝试

引入NLP领域的Transformer架构，利用其自注意力机制来理解和编码视频中的长距离时空依赖关系，显著提升了模型对复杂内容的理解能力。

模型： Phenaki、CogVideo、NUWA、VideoGPT

2020+

扩散模型的兴起（U-Net+扩散模型）

一种新的生成范式出现。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。在生成质量、多样性和稳定性上全面超越GAN。

模型： VDM、Make-A-Video, Runway Gen-2, Stable Video Diffusion、AnimateDiff

2022+

DiT路线（Transformer+扩散模型）

革命性地使用Transformer替代了传统扩散模型中的U-Net架构。通过将视频时空内容切分为Patches（令牌），直接在去噪过程中利用Transformer的全局建模能力，实现了卓越的可扩展性和长视频生成效果。

代表产品：Sora、KLing、Hailuo AI、PixVerse V2

2024+

GAN (生成能力) → Transformer (长时序理解) → 扩散模型 (更高质量生成) → Transformer + 扩散模型 (融合架构)

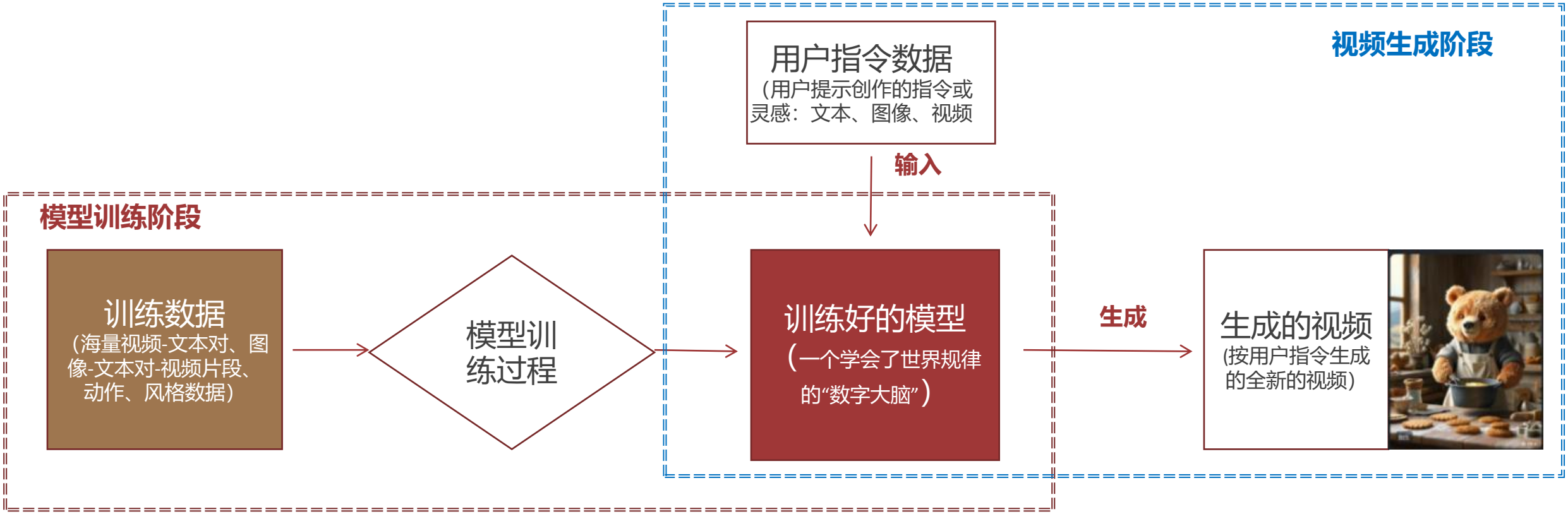
PART 02

AI视频技术原理与架构

1. 基础技术范式与模型架构
2. 从基础生成架构到商业模型
   —— 基本工作原理·模型结构详解·关键概念解析
3. AI视频生成的燃料：训练数据

理解视频生成基本工作原理

虽然各厂商的 AI 视频生成大体遵循相同的宏观流程——从海量数据准备，到大规模模型训练，再到最终的视频生成——但在**技术架构**、**数据质量**与**训练细节**等方面的差异，直接决定了模型在可生成视频时长、物理真实感和最终视觉效果上的优劣。



视频生成模型详解（推测的Sora为例）

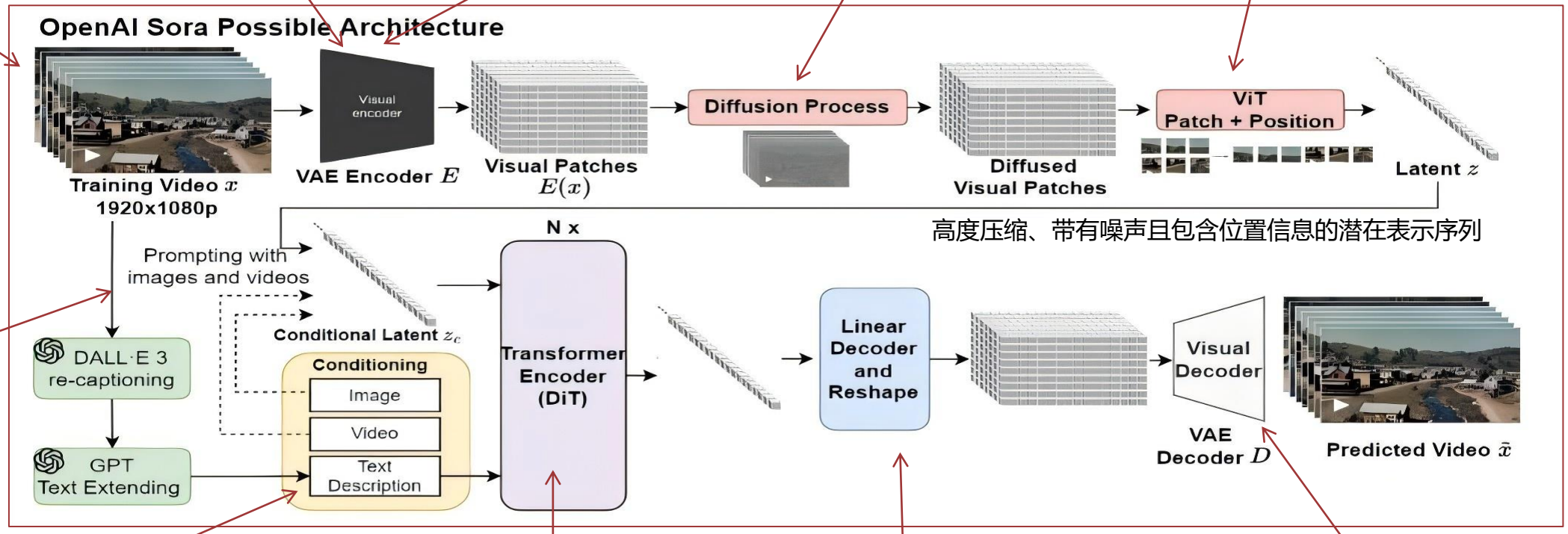
① Sora 的训练数据是原始高分辨率视频

② VAE编码器将视频压缩为低维、抽象的潜在表示

③ 压缩后的视频数据被切分为Patches，可看做是时空token

④ 清晰的潜在视觉补丁被转化为带有不同噪声水平的视觉补丁

⑤ 潜在补丁的时空关系被序列化，并添加位置编码，便于Transformer处理



⑥ 如果用户提示中包含图像或视频，首先由 DALL·E 3 将其转化为精细的文字描述，然后所有输入再交由 ChatGPT 扩展成更详细、更具描述性的提示信息。

⑦ 处理后的输入信息（图像、视频、扩展后的文本描述）都会被整合并编码成一个统一的条件潜在表示

⑧ DiT 从一个完全随机的噪声开始，经过多次迭代逐步去噪，在条件信息的引导下生成连贯、高质量的视频序列。

⑨ 线性解码器将扁平的序列恢复成 VAE 编码器输出的Patches 原始时空结构。

⑩ VAE 解码器将低维潜在空间解码回高分辨率的像素空间，生成最终的视频帧。

关键概念解析—像素空间与潜空间

通过将复杂的视频生成问题从难以驾驭的像素空间转移到更易于建模的潜空间，实现了计算效率和生成质量的巨大突破，使其能够更深刻地理解和模拟物理世界的动态规律。

■ 像素空间 (Pixel Space)

是我们直接感知世界的原始数据域。 它由图像或视频中每个像素的颜色值（如RGB）构成。

- **高维度、冗余:** 包含大量冗余信息，计算成本极高。
- **缺乏语义:** 像素级的微小变化通常不具备高级语义意义。例如，将两张人脸图像在像素空间中直接取平均，得到的是一张模糊的重影，而不是一张具有两者特征的新人脸。
- **描述的是“表象”:** 直接描述了视觉的表层样貌

■ 潜空间 (Latent Space)

是一个通过模型学习到的低维、抽象的表示空间。 它并非直接存储像素信息，而是通过一个编码器将原始数据压缩，捕捉其背后更高层次的语义特征和结构。

- **低维度、紧凑:** 维度远低于像素空间，信息密度高，计算效率也高。
- **富含语义:** 空间中的点或向量代表了抽象概念。在潜空间中进行数学操作（如插值、加减）是有意义的。例如，在潜空间中，从“男人”的向量平滑地移动到“女人”的向量，可以生成一系列自然过渡的面部图像。
- **描述的是“精髓”:** 捕捉了数据的本质结构和核心概念。

关键概念解析—时空压缩

将巨大且冗余的“像素空间”转换（编码）为小巧且高效的“潜空间”，这一从原始视频中提取深层时空信息并以低维形式表示的过程，被称为“时空压缩”。

视觉压缩网络

一个包含编码器（Encoder）和解码器（Decoder）的神经网络。（注：变分自编码器（VAE）是一种最常用的视觉压缩模型，但 Sora 所采用的具体模型名称并未公布。）

编码器：负责“压缩”。将高维度的、复杂的数据压缩成一个低维度的、包含核心信息的“代码”，这个代码被称为潜向量（Latent Vector）。

解码器：负责“解压”。它能接收这个潜向量，并将其还原成原始的、高维度的数据。

“时空压缩”在Sora中的应用过程

数据预处理：Sora在训练前，会用**视觉压缩网络**的编码器将所有视频的数据压缩成潜向量序列。这样做的好处是，后续模型处理的是这些信息密度极高、维度又低的潜向量，而不是原始像素，计算效率大大提升。

生成最终画面：当Sora的核心模型生成了一系列潜向量后，再由**视觉压缩网络**的解码器将这些潜向量“翻译”回我们能看到的像素画面，最终拼接成完整的视频。

关键概念解析—时空补丁

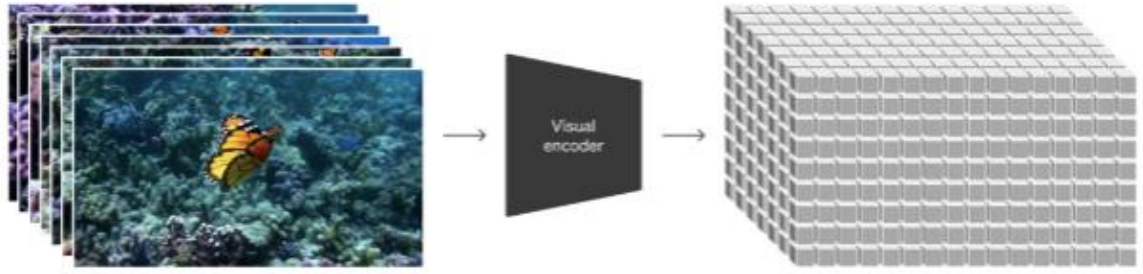
时空补丁（Spacetime Patch）

想象一下，把一张图片切成很多个小方块，就像拼图一样，每一个小方块就是一个 Patch。对于视频，Sora更进一步，它不仅在空间上（宽和高）切块，还在时间维度上切。所以它切出来的是一个“小方块视频”，即时空补丁。这就像从一叠胶片中，用一个立方体模具挖出一小块。

概念的重要性

统一数据格式：无论是长视频、短视频、宽屏视频、竖屏视频，还是单张图片（可以看作只有一帧的视频），都可以被统一处理成一连串的Patches。这极大地增强了模型的泛化能力和扩展性。

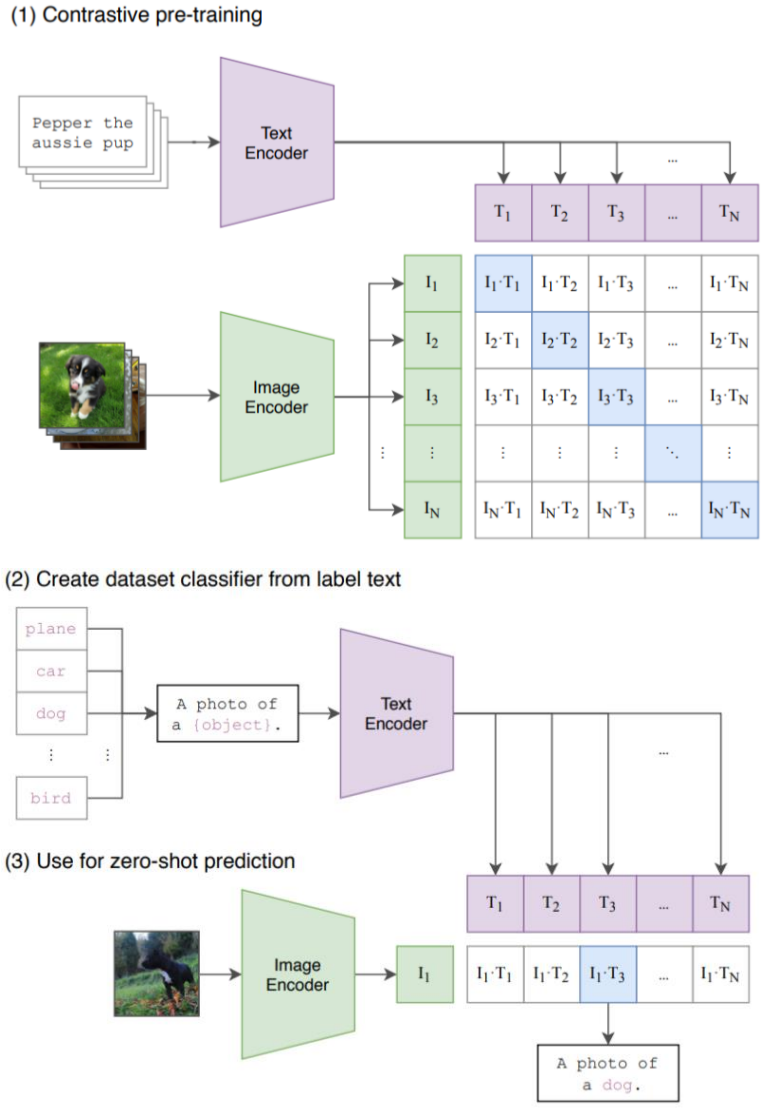
降低计算复杂性：直接处理原始像素数据非常耗费资源。将视频转化为一系列标准化的Patches，就可以发挥Transformer的优势进行高效地处理。这就像把一篇长文分解成单词，更容易分析语法和语义。



关键概念解析—CLIP

CLIP (Contrastive Language-Image Pre-training - 对比语言-图像预训练模型)

CLIP是OpenAI于2021年提出的一个能同时理解文本和图像的模型。利用海量“图片-文字描述”进行训练，学会了将内容相似的图片和文本映射到其内部“表示空间”中非常接近的位置。例如：“一只狗在玩球”的文本和一张狗玩球的图片，在CLIP的“脑海”里是挨在一起的。



CLIP在Sora中的作用

将用户文本提示转换为向量表示

当用户输入一段文字描述，比如“一只穿着宇航服的猫在火星上行走”，CLIP将这段文字转换成一个数学向量（文本编码）。

指导视频生成过程，确保生成内容与文本描述一致

文本编码作为条件被送入DiT后，在DiT的每一步去噪过程中，都不断指导DiT要生成“一行走的只猫，穿着宇航服，背景是火星”的视频，确保生成的视频内容与用户的文本描述高度一致。

PART 02

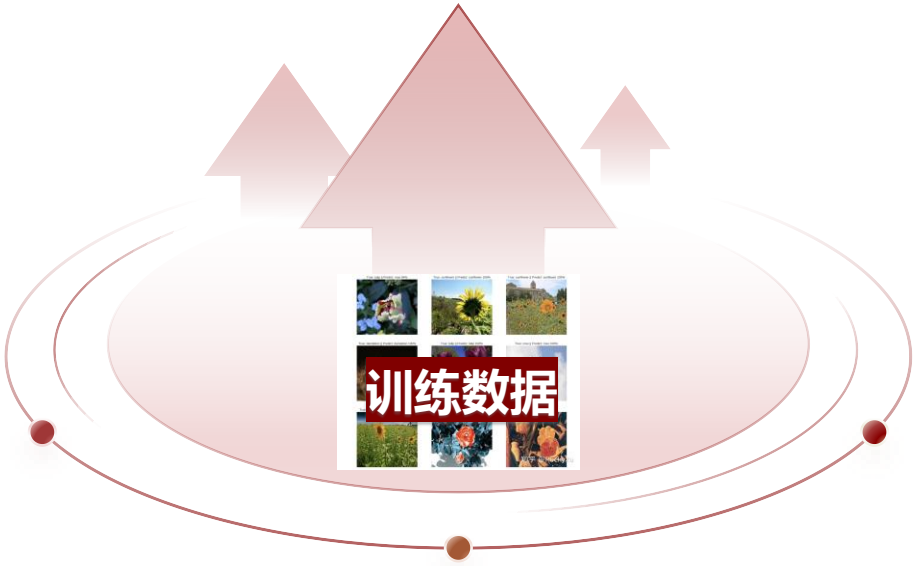
AI视频技术原理与架构

1. 基础技术范式与模型架构
2. 从基础生成架构到商业模型
3. AI视频生成的燃料：训练数据

—— 训练数据·数据重要性·主流数据集

AI视频生成的燃料：训练数据

视频生成模型的发展与高质量训练数据集密切相关。随着模型架构（如DiT）逐渐趋同，数据质量成为了决定模型生成效果上限的核心变量。



数据来源的多样性

不同场景/领域的数据具有不同的语言特征，对模型能力提升的侧重点也不同。

数据规模

影响泛化能力,影响生成质量

数据质量

- 高质量训练集能提高模型精度,减少训练时长
- 有重复、噪声、错误数据等低质量语料会损害模型性能。

训练数据从哪里来？—主流数据集

01

文本-视频对（Caption-level）数据集

- 公开的、专门用于文生视频的大规模数据集
- 视频 +文本描述,用于文本到视频生成
- 例:WebVid-10M, Panda-70M, HD-VILA-100M

02

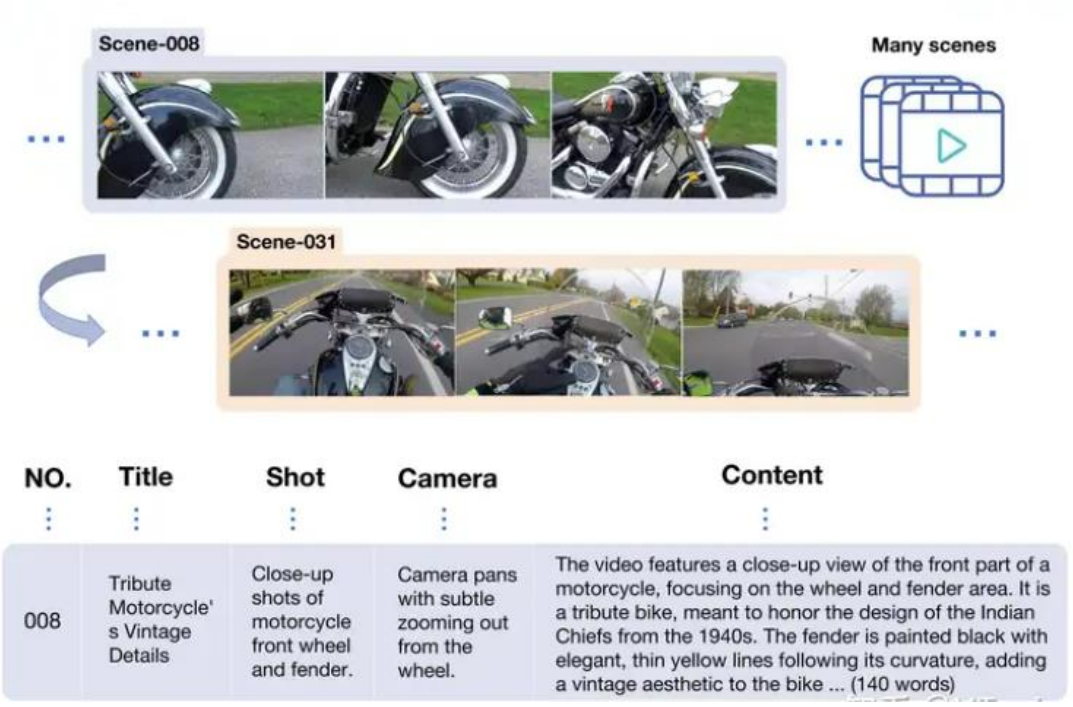
类别级（Category-level）数据集

- 经典的、用于视频理解/行为识别的基础数据集
- 视频按类别标注,用于无条件视频生成
- UCF-101, Kinetics, Something-Something

03

巨头公司使用的、不公开的私有数据集

- OpenAI (Sora) 的数据集
- Google (Imagen Video, Lumiere) 的数据集



以视频脚本方式注释的数据集示例

「AI视频技术原理与架构」要点总结

掌握AI视频的"诞生机制"与"完整流程"

- AI视频生成是一个数据驱动、模型学习的系统工程。
- 理解AI视频完整生命周期：数据收集→数据预处理→模型训练→视频生成四个核心环节。
- 明确"模型"本质——由算法和数据构成的学习系统，通过“输入到输出”的映射，实现从现实数据到视觉内容的智能转换。

洞悉AI视频的"范式变革"与"架构选择"

- 理解扩散模型成为主流的三重优势：训练稳定性、质量多样性、可控性，解决了GAN和自回归模型的核心缺陷。
- 掌握当前两大架构路线：U-Net架构（易训练但一致性较差）与DIT架构（成本高但长视频一致性强）。

理解AI视频的"关键组件"与"实现机制"

- 深入理解Transformer的核心价值：通过自注意力机制实现全局视野，是时间一致性、动态关系理解的关键。
- 掌握关键组件作用：VAE负责数据压缩与还原、时空补丁实现统一数据格式，降低计算复杂性、CLIP充当文本-视觉"翻译官"。
- 认识训练数据重要性：数据规模、质量、多样性直接决定模型泛化能力和生成质量。



PART 03

AI视频生成能力发展与突破

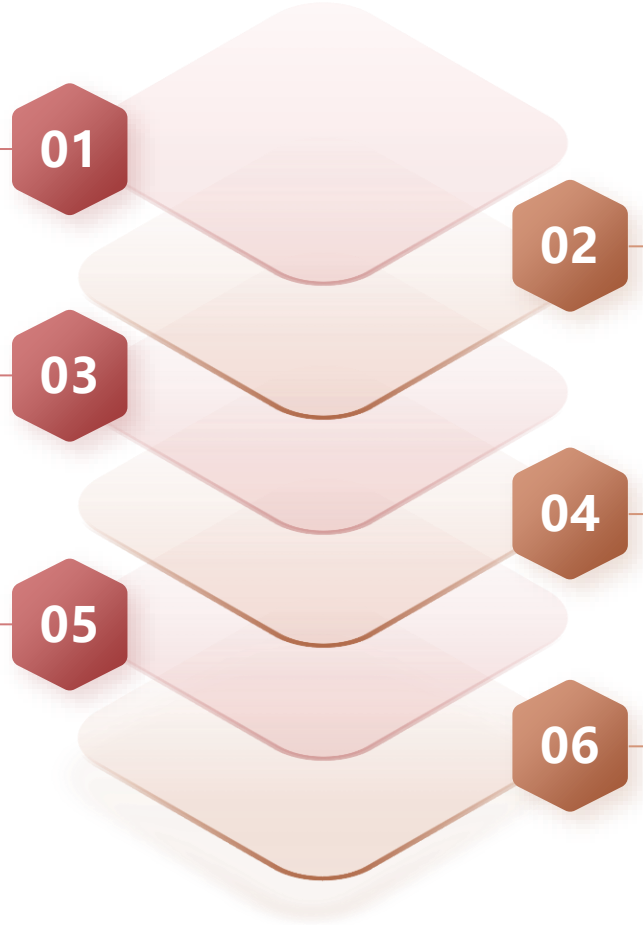
1. 厂商技术能力进展与突破

   —— 核心挑战·能力现状·示例

2. AI视频生成技术的能力评估

3. 推荐关注的AI视频厂商/模型

AI视频技术核心挑战—要解决的问题



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

时空一致性维持

保持整个视频的连贯性
确保人物特征在所有帧中保持一致
维持背景元素位置和外观的稳定性
随视频时长增加，一致性维持难度呈指数级增长

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续性

物理真实性模拟

确保生成内容符合现实世界物理规律
物体运动遵循惯性、重力等基本物理定律
复杂物理现象的真实模拟
光影效果随时间变化的自然过渡

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)
人物与物体之间的合理互动(拿起物品、推门等)
复杂动作序列的连贯表现

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

现状1—进展迅速，AI视频已具备专业级潜力

与早期模型相比，当前主流AI视频生成器在质量、时长、可控性和多模态能力上均实现了质的飞跃。

主要进展	详细说明
分辨率与时长的双重突破	当前，1080p分辨率已成为主流模型的标配，部分模型如Veo 2、可灵2.1已支持4K画质。在视频时长上，也从最初的几秒限制大幅延长。例如，快手的可灵（Kling）模型已支持生成长达2分钟的1080p视频，其他模型也在持续提升时长上限。
物理真实感与细节控制	新一代模型在理解和模拟物理世界上取得了显著进步。主流模型都能够理解重力、惯性等基本规律，生成如人物吃汉堡留下咬痕、液体流动等符合物理常识的场景。
运动与镜头语言的精准掌控	“可控性”是过去一年国内视频大模型最显著的突破。用户不再需要通过反复“抽卡”来获得满意的效果。模型能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。 同时，国产模型在中文语义理解上具备天然优势，能更准确地响应中文用户的复杂指令。
音画同步的革命性突破	2025年5月，谷歌在I/O大会上发布Veo 3模型，首次实现了原生音频生成，将V2A技术集成到视频生成过程中，真正实现了"有声时代"的突破。用户无需后期添加音频，Veo 3可以直接生成带有环境音、音效甚至人物对话的完整视频。
实时生成与交互技术	以字节跳动AAPT技术和初创公司Decart的MirageLSD模型为代表的实时视频生成技术正在兴起。 这些技术能实现每秒24帧甚至更高的生成效率，延迟极低，支持用户在生成过程中进行实时编辑和互动，彻底改变了传统视频“生成-等待-修改”的线性流程，为直播、游戏、实时通话等场景带来了巨大想象空间。

现状2—任重道远，仍有很长的路要走

尽管进步显著，但AI视频生成距离完美仍有很长的路要走，主要面临以下挑战：

主要挑战	详细说明
时序一致性难题	在较长视频中，人物、物体的外观和特征保持一致仍是当前AI视频生成的最大挑战之一。VBench等基准测试显示，随着视频时长增加，主体特征的飘移现象会逐渐加重。虽然具体的一致性数据因模型和场景而异，但业界普遍认为30秒以上视频的一致性维持仍需进一步突破。
物理与逻辑的鸿沟	虽然模型能模拟基本物理规律，但在处理复杂动态交互（如打碎物体、多人互动）和流体模拟时，效果依然不尽人意，甚至会出现异常失真画面。此外，AI对复杂的多对象空间关系和叙事逻辑的理解仍然有限。
细节与情感的缺失	在细节刻画上，AI生成的人物面部纹理、表情的细微变化与专业拍摄相比仍显生硬，缺乏真实感和情感感染力。复杂的内心活动和情感表达，依然是AI难以逾越的障碍。
算力与成本的制约	高质量视频的生成需要巨大的计算资源，这直接导致了高昂的训练和推理成本。清华大学等研究机构正在探索通过优化注意力机制等算法来降低算力消耗，以推动技术的普及。
数据质量重于数量	高质量、高相关性的数据比单纯增加数据量更能提升特定场景的生成效果。如"食物"类别数据量少于"人物"，但生成质量却普遍更高，说明数据策略优化很关键。

示例1：画面真实感

画面真实感：AI生成无限接近真实世界影像的能力

> 电影感特写镜头，照片级真实感： 一杯热咖啡放在窗边的深色木桌上。细腻的蒸汽缓缓升起。窗户玻璃上，雨滴正在滑落，窗外的城市霓虹灯光被模糊成柔美的光斑（焦外成像效果）。氛围感十足的灯光，4K画质，超高细节。

物体真实性：咖啡杯、咖啡、桌面、窗、蒸汽比较真实，雨滴真实感欠佳，混入杂物，窗户背影光斑真实感较差。

光影真实感： 咖啡杯上的高光、木桌的反射比较、雨滴的折射比较自然。

材质质感： 陶瓷的光滑、木纹的粗糙、液体的清澈感比较到位；

动态流畅性： 蒸汽的飘动和雨滴的滑落比较连贯，雨滴的滑落逻辑性欠佳，真实感欠佳

镜头感： 前景清晰，背景模糊，背景光斑体现明显的焦外成像效果



示例2：风格表现力

风格表现力：不同的艺术风格的有效呈现能力

吉卜力工作室风格，日式动画： 一杯热气腾腾的咖啡放在窗边的木桌上。窗外是宁静的雨天，远处是温馨的小镇街景。手绘动画风格，色彩温暖治愈，细节丰富。



- 风格还原度高，体现吉卜力动画的夸张或简化规律，呈现较好的出色彩饱和度。
- 氛围营造较好， 画面传达出“温暖治愈”的感觉。

赛博朋克，霓虹黑色电影风格： 一杯咖啡放在金属桌上，桌面倒映着全息广告。窗外是永恒的酸雨，背景是未来主义的摩天大楼和闪烁的霓虹招牌。高对比度，冷色调，充满科技感。



- 核心元素捕捉到位，准确生成了霓虹灯、雨夜、高科技感等赛博朋克要素。
- 光影表现较好，霓虹灯在桌子的反射效果炫酷且一致。
- 成功营造出高对比度、冷峻、压抑但又华丽的视觉氛围

示例3：场景和镜头运动能力

提示词：一台手持摄像机追踪着在沙漠中奔跑的机械公牛。它的动作搅动了跟随在机械生物身后的尘土。电影般的实景拍摄。





Runway Gen-4 场景运动

Direction	Value	Text Prompt	Portrait	Medium	Environment
Horizontal	-5.0	camera glides to the left			
Vertical	-4.0	camera glides down			
Pan	-2.0	camera pans			
Tilt	2.0	camera tilts upwards			
Zoom	-5.0	camera zooms out			
Roll	10	camera rotates to the right			

Runway Gen-4 精细化镜头控制

示例4：音画同步

> 一颗光滑的石头从画面上方掉入平静的池塘，激起水花和涟漪。整个过程用慢动作呈现。



可灵文生视频同时，根据画面生成了背景音，周围鸟鸣、石块落在水面激起水花的声音，非常贴切

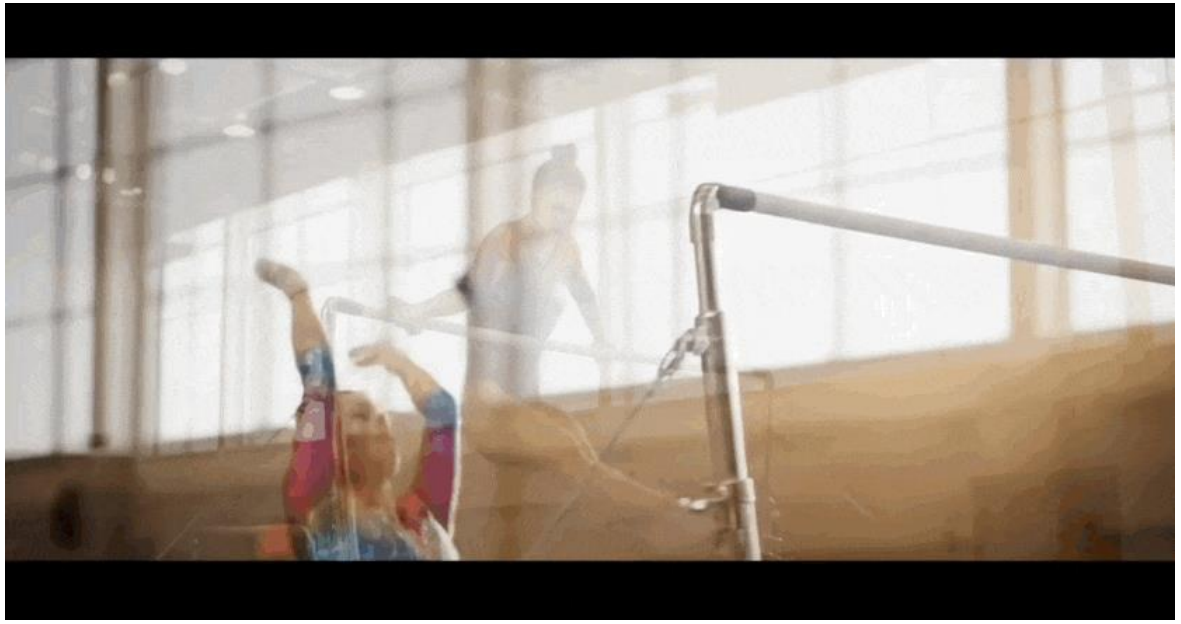


Veo 3 通过单个文本提示生成的视频和声音



Veo 3生成的具有惊人特效的创新音乐视频

示例5：运动场景动作与形态失真



Veo 3 在生成体操运动员表演时，出现了明显的身体扭曲和不合理的肢体动作：

- 旋转过程中胳膊呈现不自然的角度
- 身体从”正面”突然变成”背面”
- 手臂做出超出人体极限的360度旋转

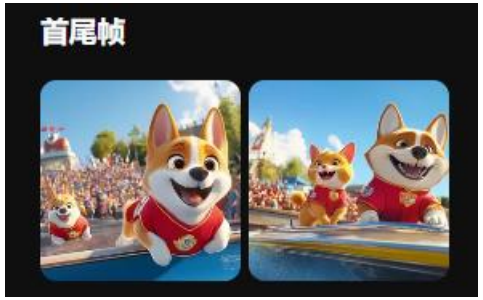


即梦3.0生成的运动员障碍跑的视频，人物动作处理不自然，缺乏真实感，尤其是踏步时，感觉处于“失去重力”的场景中。

示例6：多主体首尾帧生图效果不佳



一只柯基和一个橘猫，正在参加跑步比赛，它们在奋力奔跑，周围人群欢呼呐喊。





多主体，首帧生成单一场景，整体表达较好



多主体，首尾帧生成，主体一致性不佳，逻辑变化衔接不上，场景整体逻辑崩溃

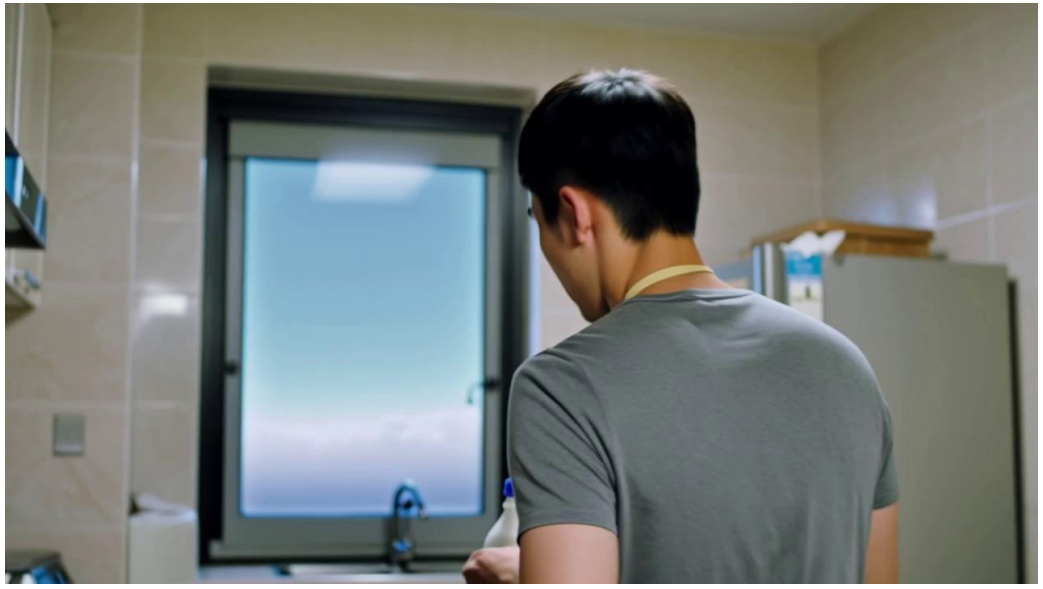
示例7：语义理解和提示词遵从

> 一个穿着红色T恤和蓝色牛仔裤的男人，从左向右走过屏幕，T恤上有一个白色的星星图案。



除了未按要求“从左向右走过屏幕”，其他都遵从提示词

> 一个男人走进厨房，从冰箱里拿出一个鸡蛋和一瓶牛奶。他把鸡蛋打在碗里，然后将牛奶倒进去。在此期间，他身上的围裙始终系着，厨房窗外的天色从未改变。



更复杂互动中，语义一致性很难保持

示例8：处理多对象和空间关系

> 男人坐在沙发上，拿起一个西红柿送至嘴边，轻轻咬下一口，镜头轻微拉近，聚焦在吃西红柿的动作上



即梦3.0：整体画面较好，但镜头拉近后，明显看出嘴部吃西红柿处理不佳，不符合真实物理规律

> 一只柯基犬在热带毛伊岛拍摄自己的Vlog



这个提示词的难点在于，如何表现柯基"拍摄自己”。Sora的演示虽然表现了这个细节，但自拍杆看起来像长在柯基身体上一样怪异。

「AI视频能力现状」小节

1. 与一年前相比，当前主流AI视频模型在语义理解和可控性上显著增强，用户需要反复“抽卡”来获得满意结果的情况已大幅改善。
2. 2024年的技术快速发展，在2025年得以延续，我们看到了厂商在核心技术指标上的突破性进展，例如快手的可灵模型支持生成长达2分钟的视频，谷歌的Veo 3则集成了V2A技术，标志着AI视频进入“有声时代”。
3. 当前市场涌现出众多模型，它们在不同维度上展现出各自的优势和短板。例如，有的模型在艺术风格渲染上表现惊艳，有的则精于精准的镜头运动控制。 这种“百花齐放”的局面表明，不同模型在技术路径和能力侧重上存在显著差异，用户的主观“体感”也因此千差万别。
4. 尽管进步显著，但AI视频生成距离完美仍有很长的路要走。时序一致性依然是当前最大的挑战，在较长视频中，主体特征“漂移”和逻辑不连贯的问题依然存在。 同时，在处理复杂的物理交互、精细的面部情感以及多主体的空间关系时，模型仍会暴露出“鬼畜”失真、情感缺失、逻辑错乱等问题。

模型能力评价维度多、不同模型又有各自优劣势，那么如何科学、客观地评价一个AI生成视频的“好”与“坏”？单纯依赖主观感受和个例展示，显然无法为技术发展提供清晰的指引。因此，建立一个统一、全面、可量化的测试基准和评估方法，变得至关重要。



PART 03

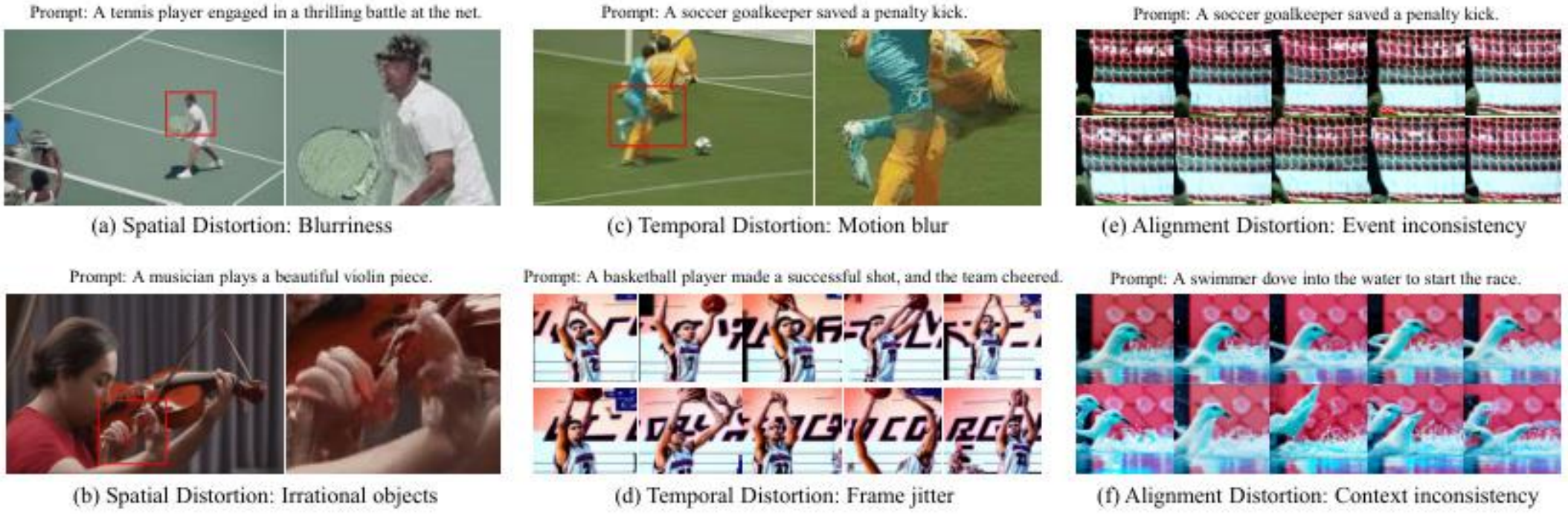
AI视频生成能力发展与突破

1. 厂商技术能力进展与突破
2. AI视频生成技术的能力评估
   —— 评估方法·评估工具·最新排名
3. 推荐关注的AI视频厂商/模型

如何定义和量化“好”的视频？—AI视频生成能力评估

AI视频生成的每一次突破性进展，都必须由更全面、更深刻的评估体系来验证其有效性并揭示新的挑战，二者共同构成了技术螺旋式上升的闭环。

- 形成标准化评估框架，消除了自我报告偏差，促进公平比较
- 建立多维度评价机制，防止单点优化，确保模型全面发展
- 动态演进的评估体系，通过提出新的、更难的“考题”，来牵引和指引整个领域未来的研究方向



AI视频生成能力评估方法

定量评估方法

通过数据量化质量，提供客观、可重复的评估指标

典型指标	说明
FVD	评估视频整体真实感的黄金标准，衡量生成视频在时空特征上与真实视频的分布距离。分数越低，代表视频越真实、动态越自然
Video IS	它衡量两个方面：生成样本的清晰度（quality）和多样性（diversity），IS分数越高越好。
CLIP Score	评估文本-视频一致性的核心指标，计算文本提示与视频帧在语义上的相似度，分数越高，代表视频内容与文本描述越匹配

特点：

- 客观、可重复、高效、适合大规模比较
- 可能与人类主观感受不完全一致，难以精准创意、美感

定性评估方法

依赖人类判断，捕捉定量指标无法衡量的细微差别。

主流方法	说明
A/B 测试	评估者在两个模型生成的视频中强制选择一个更优者，是模型迭代对比的最常用、最可靠方法
等级评分	从多个维度（如真实感、连贯性、美学）对单个视频进行打分。提供模型能力的绝对画像，诊断性强

特点

- 符合人类直观感受，能捕捉细微差别和抽象概念
- 主观性强、成本高、耗时长、难以规模化

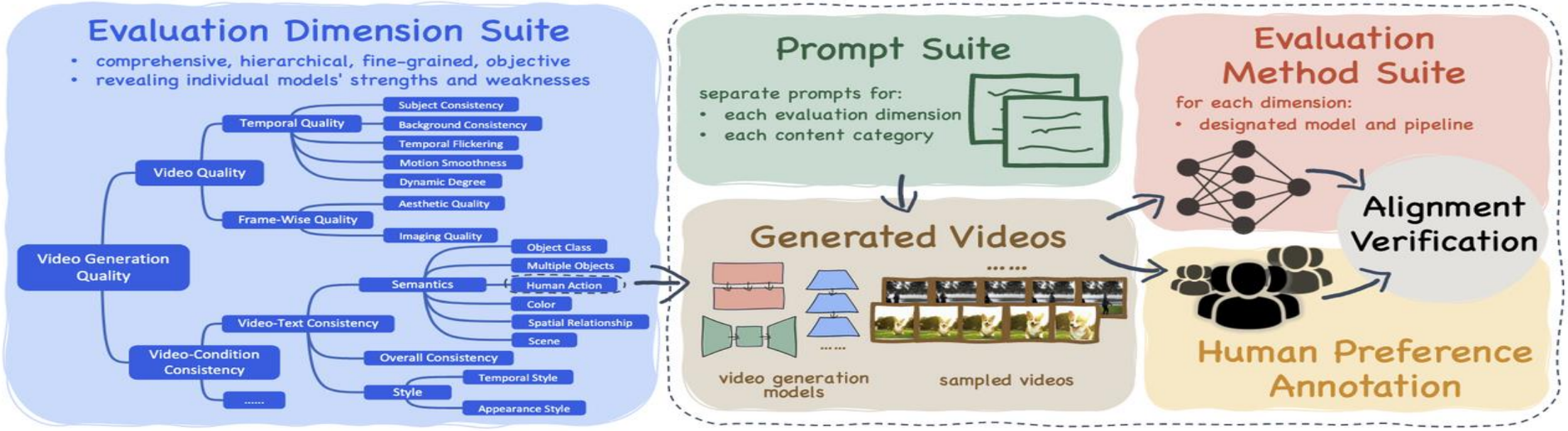
全面的评估系统

AIGCBench、VBench、FETV

VBench：视频生成多维度评估的诊断框架

VBench是一个系统性的、多维度的视频生成能力评估基准 (Benchmark)。它不是单一指标，而是一个整体评估框架，目标是提供一个标准化的“实验台”，全面诊断模型的能力。

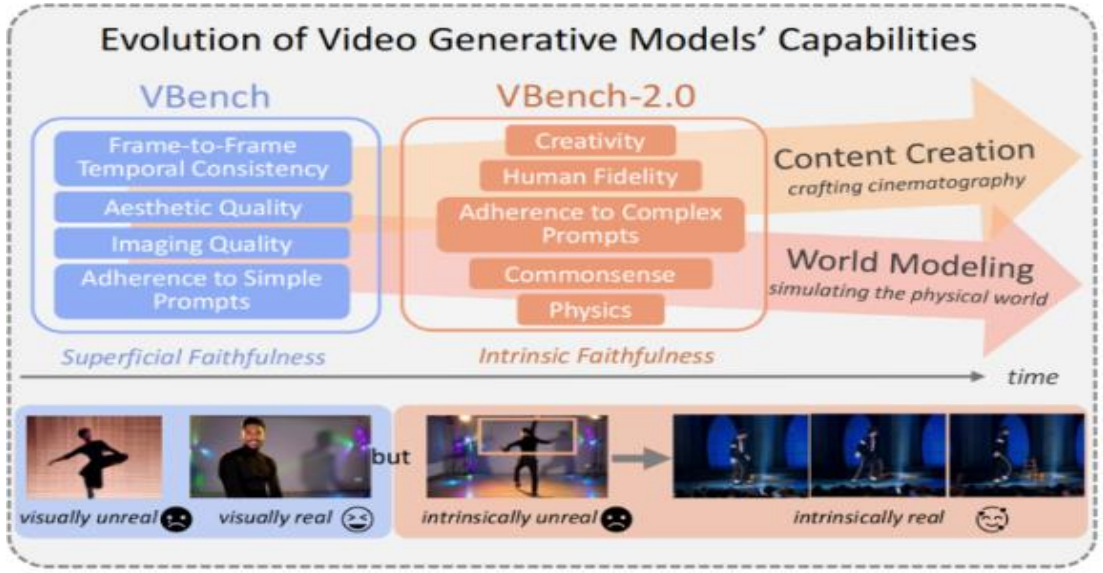
1. **全面的评估维度**：质量和一致性两个维度，共16个核心指标，涵盖视频质量、文本一致性、时间连贯性、特定能力（人类动作、美学质量、风格化）等，每个指标都设计对应100个左右的prompt。
2. **与人类感知的对齐**：通过收集人类偏好注释来验证VBench评估与人类感知的一致性。
3. **混合评估方法**：针对部分维度，开发了专门的AI自动化评估器（如使用CLIP评估一致性）。 对抽象维度（如想象力），定义了标准的A/B测试和等级评分流程。



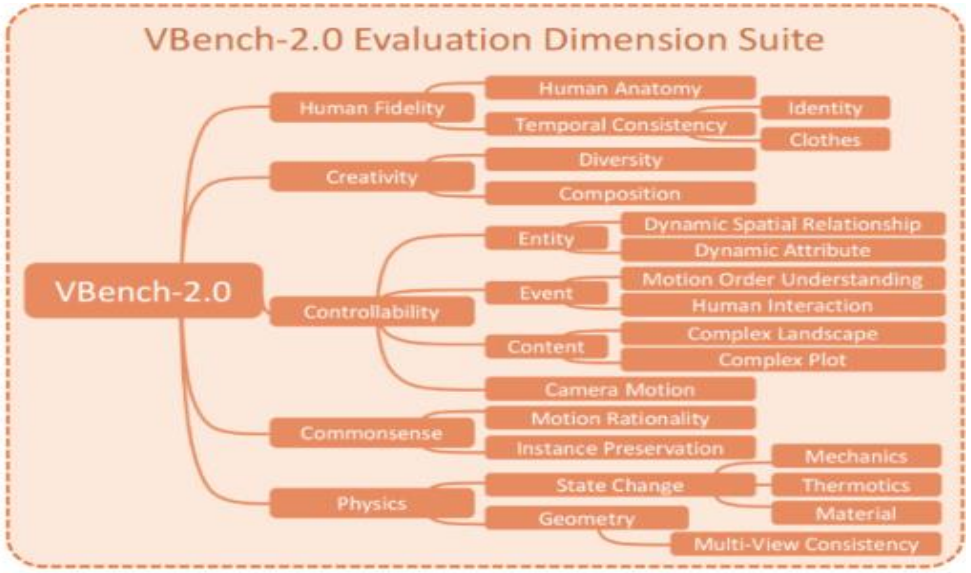
VBench2.0—从「表面真实性」到「内在真实性」

VBench-1.0侧重于视频生成视觉质量、文本匹配与整体流畅度等。VBench-2.0并非取代1.0，而是更聚焦视频的「内在真实性」。

1. **人体动作与结构**：关注做体操时动作是否连贯、角色动作是否合乎人体解剖常识等。
2. **可控性**：考察模型能否精确地执行用户给出的指令，例如相机运动、人物位置等微调效果。
3. **创造性**：观察模型在场景组合和故事情节拓展等方面的想象力。
4. **物理规律**：浮力、重力、碰撞效果是否合理？模型能否生成符合物理定律的动作或场景变化？
5. **常识推理**：在日常情景或逻辑推断中是否展现出合理性，例如「吃东西」时，食物是否真的进入了口中，角色行为是否合乎常理等。



(a) Scope of VBench-2.0



(b) VBench-2.0 Evaluation Dimension Suite

VBench1.0实时排名

北京大学 PEKING UNIVERSITY

截止到2025年11月25日的VBench1.0 排名（仅选取VBench团队采集和评估样本，按总分由高到低排序）

总分排名	模型名称	发布时间	总分	质量维度	语义维度	质量维度							视频条件一致性——视频文本一致性								
						时间质量					帧质量		语义理解						风格		整体一致性
						主题一致性	背景一致性	时间闪烁	运动平滑度	动态程度	美学质量	成像质量	对象类别	多个对象	人类动作	色彩	空间关系	场景	外观风格	时间风格	总体一致性
1	Sora	25-01-14	84.28%	85.51%	79.35%	96.23%	96.35%	98.87%	98.74%	79.91%	63.46%	68.28%	93.93%	70.85%	98.20%	80.11%	74.29%	56.95%	24.76%	25.01%	26.26%
2	Wan2.1-T2V-14B	25-07-25	83.69%	85.59%	76.11%	97.52%	98.09%	99.46%	98.30%	65.46%	66.07%	69.43%	86.28%	69.58%	95.40%	88.59%	75.39%	45.75%	22.64%	23.19%	25.91%
3	Luma	25-01-14	83.61%	83.47%	84.17%	97.33%	97.43%	98.64%	99.35%	44.26%	65.51%	66.55%	94.95%	82.63%	96.40%	92.33%	83.67%	58.98%	24.66%	26.29%	28.13%
4	Hunyuan Video	25-05-22	83.43%	85.07%	76.88%	97.22%	97.60%	99.39%	99.05%	71.94%	60.28%	67.24%	83.48%	66.71%	94.40%	89.79%	72.13%	54.46%	22.21%	24.52%	26.95%
5	MiniMax-Video-01	24-10-01	83.41%	84.85%	77.65%	97.51%	97.05%	99.10%	99.22%	64.91%	63.03%	67.17%	87.83%	76.04%	92.40%	90.36%	75.50%	50.68%	20.06%	25.63%	27.10%
6	Kling 1.6	25-05-08	83.40%	85.00%	76.99%	97.40%	96.84%	98.64%	99.13%	62.22%	64.81%	69.70%	93.34%	63.99%	96.20%	81.26%	79.08%	55.57%	20.75%	24.51%	26.04%
7	Wan2.1-T2V-1.3B	25-05-03	83.31%	85.23%	75.65%	97.56%	97.93%	99.55%	98.52%	65.19%	65.46%	67.01%	88.81%	74.83%	94.00%	89.20%	73.04%	41.96%	21.81%	23.13%	25.50%
8	HunyuanVideo (Open-Source Version)	24-12-16	83.24%	85.09%	75.82%	97.37%	97.76%	99.44%	98.99%	70.83%	60.36%	67.56%	86.10%	68.55%	94.40%	91.60%	68.68%	53.88%	19.80%	23.89%	26.44%
9	Gen-3	24-07-25	82.32%	84.11%	75.17%	97.10%	96.62%	98.61%	99.23%	60.14%	63.34%	66.82%	87.81%	53.64%	96.40%	80.90%	65.09%	54.57%	24.31%	24.71%	26.69%
10	Vchitect-2.0 (VEnhancer)	24-09-20	82.24%	83.54%	77.06%	96.83%	96.66%	98.57%	98.98%	63.89%	60.41%	65.35%	86.61%	68.84%	97.20%	87.04%	57.55%	56.57%	23.73%	25.01%	27.57%
11	CogVideoX1.5-5B	24-11-15	82.17%	82.78%	79.76%	96.87%	97.35%	98.88%	98.31%	50.93%	62.79%	65.02%	87.47%	69.65%	97.20%	87.55%	80.25%	52.91%	24.89%	25.19%	27.30%
12	CogVideoX1.5-5B	25-04-21	82.01%	82.72%	79.17%	96.56%	96.81%	98.53%	98.15%	56.16%	62.07%	65.34%	83.42%	67.28%	97.60%	88.40%	79.43%	53.28%	24.68%	25.42%	27.42%
13	Jimeng	24-11-15	81.97%	83.29%	76.69%	97.25%	98.39%	99.03%	98.09%	38.43%	68.80%	67.09%	89.62%	69.08%	90.10%	89.05%	77.45%	44.94%	22.27%	24.70%	27.10%
14	RepVideo	25-01-16	81.94%	82.70%	78.91%	96.25%	96.56%	99.16%	98.13%	57.78%	62.40%	63.16%	87.83%	71.18%	98.00%	82.51%	74.74%	52.96%	25.12%	25.31%	26.96%
15	CogVideoX-5B (Diffusers)	25-03-04	81.91%	83.05%	77.33%	96.45%	96.71%	98.97%	97.20%	69.51%	61.88%	63.33%	85.07%	63.94%	98.60%	83.03%	68.91%	51.96%	24.98%	25.42%	27.65%
16	Vidu	24-11-15	81.89%	83.85%	74.04%	94.63%	96.55%	99.08%	97.71%	82.64%	60.87%	63.32%	88.43%	61.68%	97.40%	83.24%	66.18%	46.07%	21.54%	23.79%	26.47%
17	Kling (2024-07 high-performance mode)	24-08-01	81.85%	83.39%	75.68%	98.33%	97.60%	99.30%	99.40%	46.94%	61.21%	65.62%	87.24%	68.05%	93.40%	89.90%	73.03%	50.86%	19.62%	24.17%	26.42%
18	Step-Video-T2V	25-03-13	81.83%	84.46%	71.28%	98.05%	97.67%	99.40%	99.08%	53.06%	61.23%	70.63%	80.56%	50.55%	94.00%	88.25%	71.47%	24.38%	23.17%	26.01%	27.12%
19	MAGI-T2V-24B-distill	25-08-06	81.83%	84.69%	70.37%	97.26%	97.67%	99.68%	98.84%	72.50%	59.33%	65.27%	84.10%	50.61%	88.20%	87.54%	72.95%	28.91%	21.69%	24.24%	26.31%
20	CogVideoX-5B (SAT prompt-optimized)	24-09-02	81.61%	82.75%	77.04%	96.23%	96.52%	98.66%	96.92%	70.97%	61.98%	62.90%	85.23%	62.11%	99.40%	82.81%	66.35%	53.20%	24.91%	25.38%	27.59%

VBench2.0实时排名

截止到2025年11月25日的VBench2.0 排名（仅选取VBench团队采集和评估样本，按总分由高到低排序）

序号	模型名称	采样	评估	开源闭源	采样时间	总分	创造性总分	常识推理总分	可控性总分	人体动作与结构总分	物理规律总分
1	Veo 3	VBench Team	VBench Team	API	2025-9-4	66.72%	60.85%	51.23%	61.06%	86.88%	53.94%
2	Wan2.1	VBench Team	VBench Team	Open Source	2025-3-28	60.20%	55.26%	48.22%	51.93%	81.60%	45.16%
3	Kling 1.6	VBench Team	VBench Team	API	2025-3-28	59.00%	48.58%	44.67%	46.69%	83.56%	50.87%
4	Sora-480p	VBench Team	VBench Team	API	2025-3-28	58.38%	60.57%	36.67%	38.82%	87.72%	41.73%
5	StepVideo	VBench Team	VBench Team	Open Source	2025-3-28	55.78%	51.27%	30.89%	39.37%	88.18%	37.62%
6	HunyuanVideo	VBench Team	VBench Team	Open Source	2025-3-28	55.30%	41.85%	42.78%	44.53%	82.41%	42.65%
7	CogVideoX-1.5	VBench Team	VBench Team	Open Source	2025-3-28	53.35%	43.66%	46.17%	46.14%	72.14%	40.23%

模型名称	人体动作与结构			创造性		可控性							常识推理		物理规律			
	人体解剖学	时间一致性		构图	多样性	实体		事件		内容		相机运动	动作合理性	实例保存	状态变化			几何结构
Veo 3	90.26%	99.53%	70.84%	68.57%	53.13%	81.82%	84.11%	75.51%	35.95%	45.89%	63.74%	40.40%	80.67%	21.78%	21.87%	54.94%	45.98%	92.98%
Wan2.1	82.59%	96.48%	65.73%	54.51%	56.00%	71.74%	75.68%	67.16%	36.79%	31.40%	45.42%	35.35%	80.00%	16.44%	13.14%	39.51%	39.08%	88.89%
Kling 1.6	86.99%	91.75%	71.95%	43.89%	53.26%	65.55%	68.00%	59.46%	64.38%	20.77%	19.41%	29.29%	72.00%	17.33%	10.83%	61.73%	38.51%	92.40%
Sora-480p	86.45%	98.15%	78.57%	53.65%	67.48%	62.22%	64.94%	43.36%	58.22%	19.81%	8.06%	15.15%	58.00%	15.33%	11.11%	27.16%	34.48%	94.15%
StepVideo	85.62%	96.04%	82.87%	35.48%	67.05%	60.48%	57.89%	55.38%	71.98%	14.98%	8.06%	6.83%	54.67%	7.11%	7.29%	21.60%	37.36%	84.21%
HunyuanVideo	88.58%	82.97%	75.67%	43.96%	39.73%	76.09%	64.37%	56.52%	43.80%	21.26%	22.71%	26.94%	66.67%	18.89%	9.78%	33.95%	34.48%	92.40%
CogVideoX-1.5	59.72%	87.18%	69.51%	44.70%	42.61%	80.80%	83.19%	67.13%	21.79%	19.32%	24.18%	26.60%	72.33%	20.00%	11.21%	33.33%	33.91%	82.46%

VBench通过多模型对比给出的有价值结论

“时序连贯性”及“视频的动态程度”应同时提升

时序连贯性与视频中运动的幅度之间有一定的权衡关系。部分模型在背景一致性和动作流畅度方面表现较好，但在动态程度方面得分较低，而另一部分模型则相反。

分场景内容进行评测，发掘不同模型潜力

部分模型在不同场景上表现出的性能存在较大差异，需深度挖掘模型在某个能力维度的上限，进而针对性地提升短板。

复杂运动类别中的时空表现均不佳

在空间上复杂度高的类别，模型美学质量维度得分往往低。这表明当前模型在处理时序建模方面仍然存在一定的不足，时序上的建模局限可能会导致空间上的模糊与扭曲，从而导致视频在时间和空间上的质量都不理想。

对于难生成的类别，提升数据量收益不大

研究人员对视频数据集 WebVid-10M 进行了统计，发现其中约有26%的数据与 “Human” 有关，占比最高。然而，在评估结果中，“Human” 类别却是模型表现最差的场景。表面在复杂的生成类别中，仅仅增加数据量可能不会对性能带来显著的改善。

提升数据质量应优先于数据量

“Food” 类别在WebVid-10M中仅占据11%，但在评测中几乎总是拥有最高的美学质量分数。这意味着，在百万量级数据的基础上，筛选或提升数据质量，比增加数据量更会对模型能力带来帮助。

待提升能力：准确生成生成多物体，表现物体间的关系

多数视频生成模型在多对象生成和空间关系方面不及图片生成模型，提升组合能力在未来研究中具备重要性。

SuperCLUE-Video：首个中文视频大模型权威评测基准

北京大学 PEKING UNIVERSITY

SuperCLUE 是目前中国国内领先且最具影响力的大模型评测基准之一。

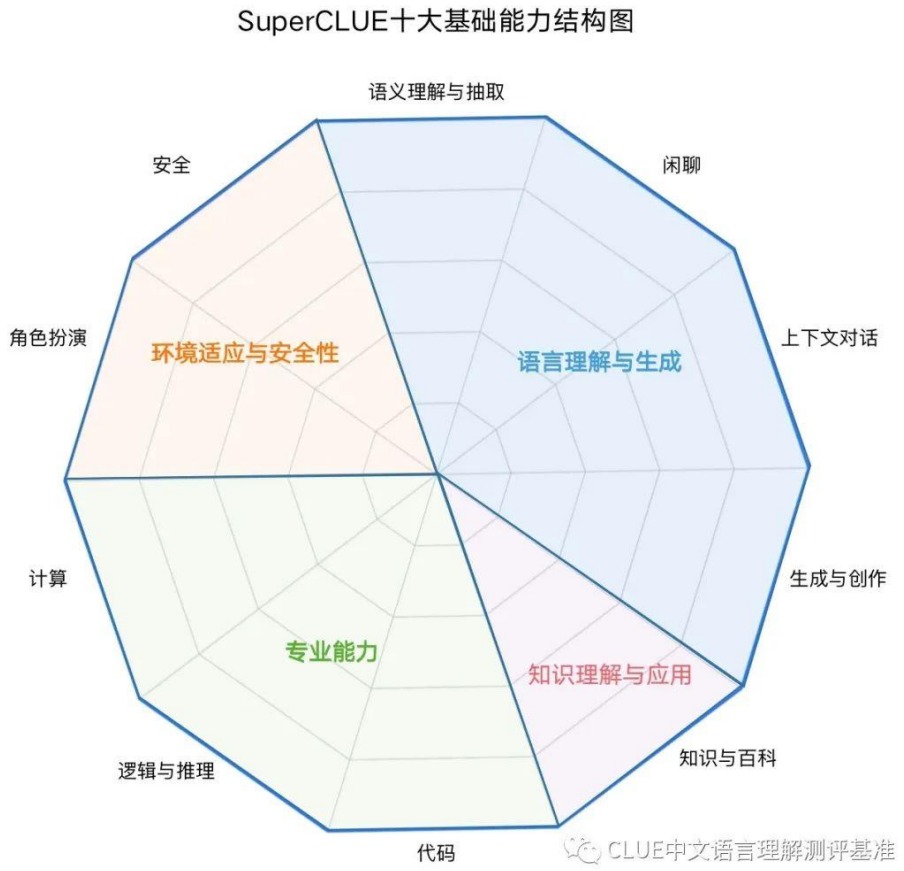
1. 通用大模型总榜：

最核心、最知名的榜单，每月更新，展示了文心一言、通义千问、Kimi、豆包、混元、GPT系列等国内外主流模型的综合排名。

2. 专项榜单：

针对特定能力或模态进行深度评测。

- **SuperCLUE-Code：** 专测代码能力
- **SuperCLUE-Agent：** 专测智能体的任务规划和工具使用能力
- **SuperCLUE-Video：专测视频理解和生成能力**
  - ① **视频理解能力：**视频问答、视频摘要、安全伦理
  - ② **视频生成能力：**
    - **SuperCLUE-T2V：文生视频专项**
    - **SuperCLUE-I2V ：图生视频专项**



SuperCLUE-Video：文生视频评测

北京大学 PEKING UNIVERSITY

SuperCLUE-T2V
文生视频评价标准

大类	一级维度	二级任务	评价标准
通用能力	视频画质渲染	主体外观画质	清晰度 流畅度 真实感
		背景画面画质	
	要素搭建能力	主体外观准确性	主体外观准确性
		背景画面准确性	背景画面准确性
		数量精准性	主体数量精准性 背景要素数量精准性
		空间关系	主体空间精准性 背景空间精准性
		运镜准确性	始末点位准确性 运镜轨迹准确性 运镜模糊度
	动态塑造能力	主体动态准确性	动作准确性 动作自然性
		主体交互准确性	交互准确性 交互自然性
		动态排列准确性	主体动作排序精准性 要素运动连贯性
场景应用能力	世界模拟	人物	要素特征准确性 场景应用贴切性
		生物	
		地理	
		人文	
	影视素材	复古风格	要素特征准确性 场景应用贴切性 视频可应用性
		科幻创作	
		动漫再现	
		极限镜头	
	应用设计	广告设计	要素特征准确性 场景应用贴切性 视频可应用性
		创意视频	
		产品宣发	
		直播模拟	
	中国文化	传统节日	要素特征准确性 场景应用贴切性
		中华饮食	
		历史文化	
		非遗特色	

测评集构建：针对评估维度撰写生成视频的提示词。

评分标准：Gemini-2.5-Pro根据预先定义的评测流程进行评估。

自动化评分：Gemini-2.5-Pro结合评价标准进行自动评分（同一个模型生成的全部视频采取相同的评价流程进行前后三次测评，取平均值）。

综合评价与反馈：给出每个任务的单独评价和总评分。

示 例

【考察大类】：通用能力

【测评维度】：视频画质渲染——主体外观画质

【提示词】：一个小女孩抬头向上看，这时一道光照在她的脸上，女孩缓缓闭上了眼。

【测评模型】：Doubao-Seedance-1.0-pro

【生成视频】：



----- 评价开始 -----
初步质量定性 (0-10 分档)：优秀 (9-10 分档)
我的思考与总体评价（重点列举缺陷）：
这个视频整体表现极为出色，在画质、动态表现和真实感等多方面都达到了很高的水准。视频成功地将提示词的要求生动地呈现出来，给人留下深刻印象。不过，以近乎苛刻的标准审视，仍存在一些极为细微的可提升之处。
优点简述： 视频的清晰度极高，人物细节和背景元素都清晰可辨。流畅度方面表现卓越，动作过渡自然流畅。真实感强烈，女孩形象逼真，光影互动自然，情感传达也较为到位。
必须详细列出的缺陷、不足、与提示词偏差之处： 尽管整体表现卓越，但仍有一些极细微的地方可进一步完善：

- 缺陷 1：皮肤纹理细节。在特写镜头下，女孩的皮肤虽然质感逼真，但相较于真实皮肤，细微的纹理和毛孔细节略有不足，在超高分辨率下能察觉到一些数字生成的痕迹。
- 缺陷 2：光线反射的细微瑕疵。在强光照射下，人物面部某些区域的光线反射效果，与现实中光线在皮肤上复杂的反射效果相比，略显简单，缺少一些更细腻的变化。

逐项打分与评价（0-10 分，严格区分）：

- 基础要求检查：视频完全满足基础要求，与提示词“女孩缓缓闭上了眼”高度吻合，内容健康，无任何违规。
- 标准 1 < 清晰度 > 的评价： 视频的清晰度令人赞叹。分辨率极高，女孩的面部特征如睫毛、眉毛、嘴唇轮廓等都锐利清晰，发丝根根分明，衣领上蕾丝花边的细微纹理也展现得淋漓尽致。背景的散景效果自然，没有出现模糊、色块或噪点等问题。整体画面干净，细节丰富。即便以最严格的标准衡量，与顶级设备拍摄的素材相比，也只是在最细微的纹理上有极微弱的差别，几乎可以忽略不计。 基于标准 1 的得分：【9】
- 标准 2 < 流畅度 > 的评价： 视频的流畅度达到了近乎完美的程度。从女孩开始闭眼到完全闭上的过程，动作过渡极其自然，帧率稳定，帧间变化细腻且符合物理世界的运动规律，毫无卡顿、跳帧或不自然停顿的现象。无论是整体的动作连贯性还是细节的流畅性，都表现得非常出色。 基于标准 2 的得分：【9】
- 标准 3 < 真实感 > 的评价： 视频的真实感极为突出。女孩的形象摆脱了 AI 常见的不真实感，面部结构、皮肤光泽、眼神以及对光线的反应都极度逼真。闭眼动作自然，面部肌肉的联动也符合生物规律，传达出了因强光刺激而产生的不适感。光线与人物的互动效果真实，面部光照区域和阴影的过渡自然，很好地模拟了真实光线的次表面散射和高光变化。尽管存在上述提到的极细微缺陷，但整体真实感表现非常优秀。 基于标准 3 的得分：【9】

综合得分计算： 计算过程：(9 + 9 + 9) / 3 = 9.00 综合得分：【9.00】
评价复核确认： 我已再次检查评估，确保：评分 极度严格，充分利用了 0-10 分 范围；重点关注并列出了所有缺陷；分数能 有效区分 不同质量水平；理由 具体充分；格式 完全符合要求。
----- 评价结束 -----

SuperCLUE-Video：图生视频评测



#示例1 运动流畅性

【Prompt】：「人物在墙上做着涂鸦，镜头跟随着人物的手由左往右缓缓移动。」



SuperCLUE 最新榜单

北京大学 PEKING UNIVERSITY

注：

- 最新文生视频排行榜发布时间：2025年7月
- 最新图生视频排行榜发布时间：2025年6月

SuperCLUE-T2V
文生视频基准测评总榜

排名	模型名称	所属机构	总分	通用能力	场景应用	使用方式
-	veo-3.0-generate-preview	谷歌	**55.40**	**50.79**	**60.01**	**API**
1	Hailuo-02	MiniMax	**51.67**	47.09	**56.25**	**API**
2	Doubao-Seedance-1.0-pro	字节跳动	**49.07**	**53.93**	**44.21**	**API**
3	kling 2.1 大师版	快手科技	44.84	**48.59**	41.09	**API**
-	pika 2.2	Pika	40.02	43.40	36.65	网页获取
4	Pangu-T2V	华为	39.11	40.29	37.92	**API**
-	Sora	OpenAI	37.99	38.33	37.65	网页获取
5	Vidu Q1	生数科技	36.66	35.65	37.67	**API**
6	Wan2.1-T2V-14B	阿里巴巴	33.03	37.21	28.84	本地部署
6	pixverse-4.5-video	爱诗科技	32.73	35.13	30.33	**API**
6	wanx2.1-t2v-plus	阿里巴巴	32.52	32.53	32.51	API
-	Mochi	Genmo	29.02	32.31	25.73	本地部署
7	CogVideoX-5b	智谱AI	25.67	25.36	25.98	本地部署
-	Open-Sora-v2	潞晨科技	25.50	27.21	23.78	本地部署
7	Step-Video-T2V	阶跃星辰	25.31	28.62	22.00	本地部署
-	Cosmos-1.0-Diffusion-14B-Text2World	英伟达	24.43	31.56	17.29	本地部署
8	Hunyuan Video	腾讯	22.32	25.54	19.11	本地部署

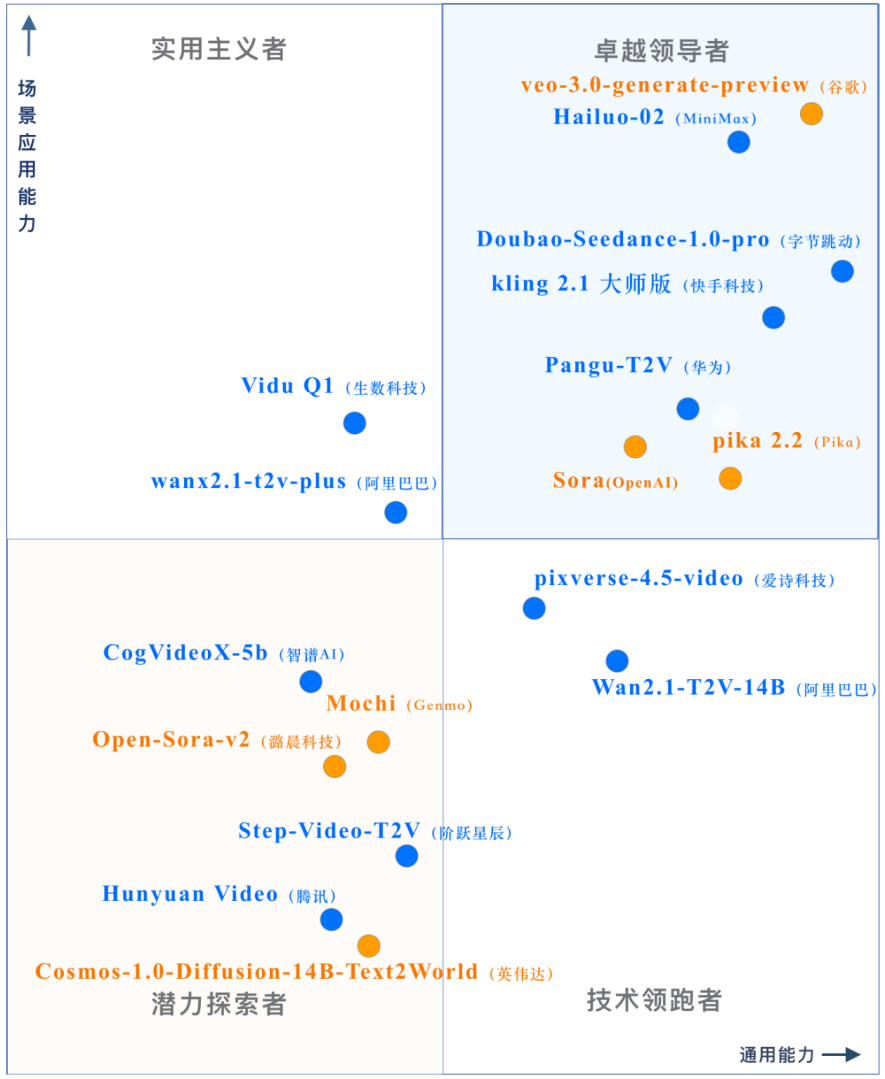
SuperCLUE-I2V
「图生视频」基准测评总榜

排名	模型名称	机构	总分	基础能力			应用能力			使用方式
				运动流畅性	内容一致性	物理真实性	动漫风格	写实风格	奇幻风格	
1	即梦-视频3.0Pro	字节跳动	**72.22**	**81.85**	**76.67**	**76.30**	**69.26**	**75.19**	54.07	官网
2	可灵 2.1	快手科技	**69.69**	**78.52**	**68.89**	**67.78**	**57.41**	**87.41**	**58.15**	官网
-	Pika 2.2	Pika	**63.46**	65.19	58.15	65.56	55.93	74.81	61.11	官网
3	Pixverse V4.5	爱诗科技	**57.04**	**65.56**	53.70	**61.85**	37.04	70.74	53.33	API
4	清影-AI生视频	智谱AI	**54.26**	53.33	56.30	54.44	33.70	**72.59**	**55.19**	官网
4	Vidu Q1	生数科技	**53.64**	58.52	47.41	54.81	**46.67**	66.30	48.15	API
4	阶跃视频	阶跃星辰	**53.51**	54.76	**60.00**	51.48	41.48	62.22	51.11	官网
5	wanx2.1-i2v-plus	阿里巴巴	**51.67**	51.85	42.59	52.22	36.30	70.37	**56.67**	API
6	I2V-01-Director	MiniMax	**49.20**	53.70	54.81	43.70	39.63	61.48	41.85	官网
7	WHEE	美图	**36.30**	46.30	40.74	47.78	20.37	41.48	21.11	官网
-	Sora	OpenAI	**30.31**	35.56	38.15	32.59	14.44	40.74	20.37	官网

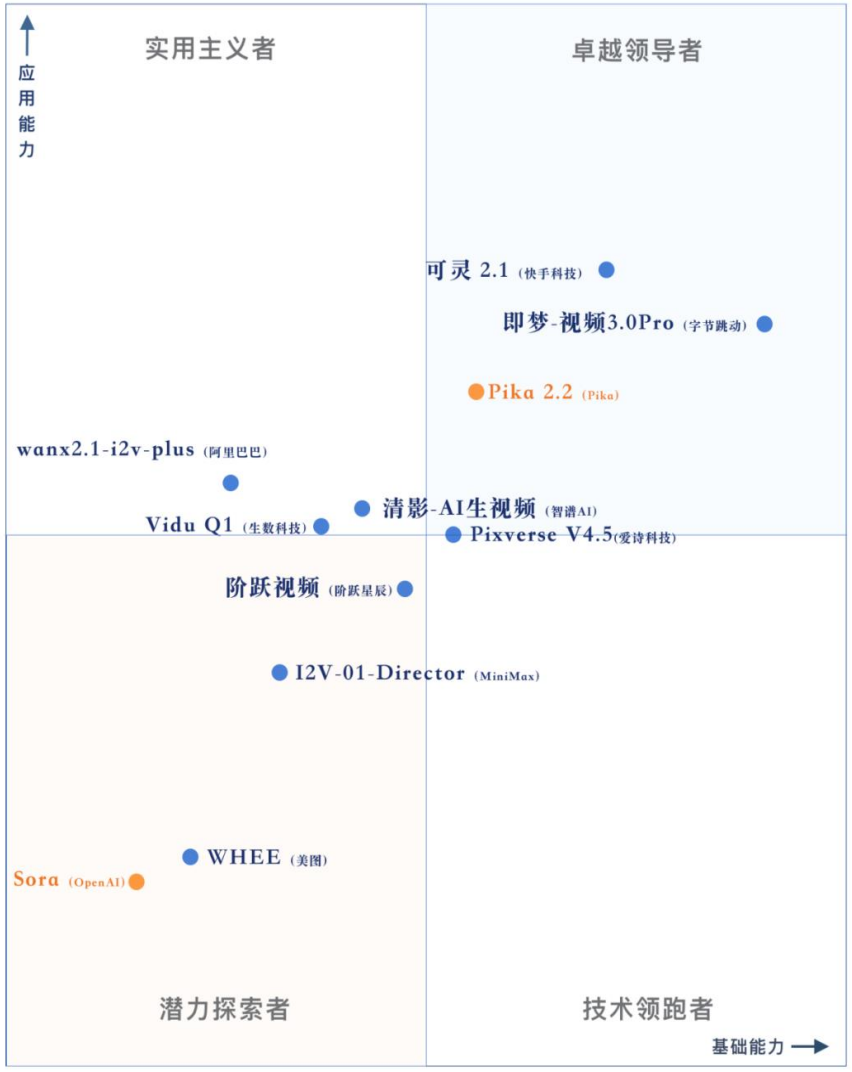
SuperCLUE视频生成模型象限图

北京大学 PEKING UNIVERSITY

SuperCLUE- T2V文生视频模型象限图



SuperCLUE-I2V 图生视频模型象限图



SuperCLUE 新增竞技场排名

SuperCLUE借鉴Artificial Analysis，于2025年10月推出全新的交互式测评功能——竞技场，通过让用户直观对比两个适配模型并投票，最终根据用户投票结果汇总产生排名（第一期文生视频排名已经出炉）。





排名	模型名称	机构	排位分	95%CI	投票数	发布时间
1	veo-3.0-generate-preview	Google	1330.5	+32.3/-30.8	1777	2025.05
2	Hailuo-02	MiniMax	1265.1	+30.5/-28.8	1720	2025.06
3	kling 2.1 大师版	快手	1212.6	+29.3/-29.6	1603	2025.05
4	Wan2.1-T2V-14B	阿里巴巴	945.9	+30.6/-27.0	1189	2025.02
5	pika 2.2	Pika Labs	910.2	+27.5/-27.7	1119	2025.02
6	Hunyuan Video	腾讯	741.2	+29.9/-30.6	836	2025.01
7	Open-Sora-v2	潞晨科技	594.3	+30.9/-33.5	614	2025.03

我们使用基于 Bradley-Terry 模型的算法得出 排位分 与 95%CI。
95%CI 表示 95% 置信区间，即有 95% 的把握认为模型的真实水平位于该区间内。表格内的"发布时间"指的是模型对外发布时间。

Artificial Analysis基准测试排名

北京大学 PEKING UNIVERSITY

Artificial Analysis是一家专注于AI基准测试和分析的独立平台，其采用的核心策略是竞技场模式，根据用户评测提供实时排名（数据取自2025年11月25日）

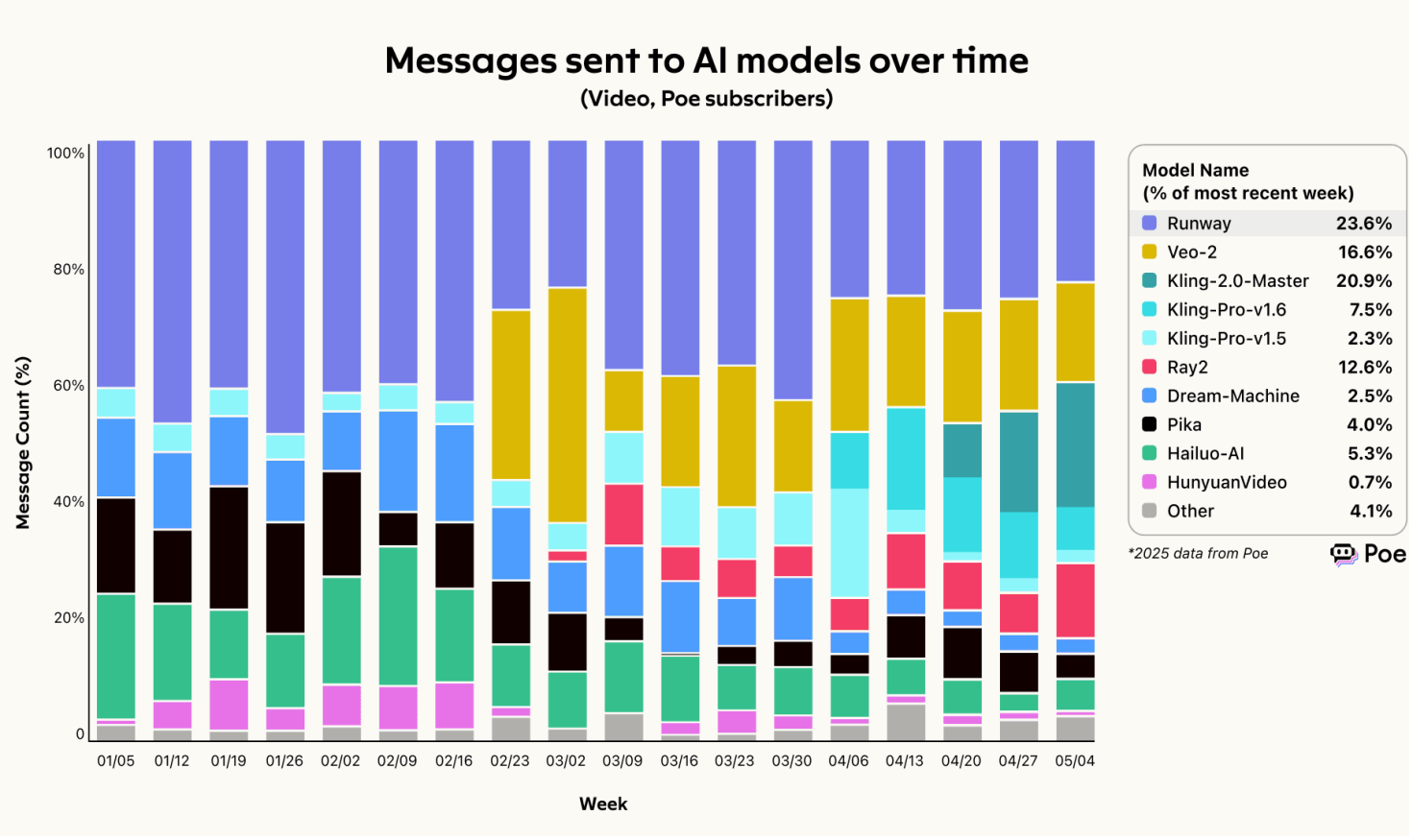
文生视频排名	公司	模型名称	发布时间
1	Google	Veo 3 (No Audio)	25-07
2	Kuaishou KlingAI	Kling 2.5 Turbo 1080p	25-09
3	Google	Veo 3.1 Preview (No Audio)	25-10
4	Google	Veo 3.1 Fast Preview (No Audio)	25-10
5	Luma Labs	Ray 3	25-09
6	OpenAI	Sora 2 Pro (No Audio)	25-09
7	MiniMax	Hailuo 02 Standard	25-06
8	Lightricks	LTX-2 Pro (No Audio)	25-10
9	MiniMax	Hailuo 2.3	25-10
10	Bytedance	Waver 1.0	25-08
11	Google	Veo 3 Fast Preview (No Audio)	25-06
12	Lightricks	LTX-2 Fast (No Audio)	25-10
13	Vidu	Vidu Q2	25-10
14	MiniMax	Hailuo 02 Pro	25-06
15	Alibaba	Wan 2.5 Preview	25-09
16	PixVerse	PixVerse V5	25-08
17	OpenAI	Sora 2 (No Audio)	25-09
18	ByteDance Seed	Seedance 1.0	25-06
19	Kuaishou KlingAI	Kling 2.1 Master	25-05
20	Alibaba	Wan 2.2 A14B	25-07

图生视频排名	公司	模型名称	发布时间
1	Kuaishou KlingAI	Kling 2.5 Turbo 1080p	25-09
2	Google	Veo 3.1 Fast Preview (No Audio)	25-10
3	PixVerse	PixVerse V5	25-08
4	Google	Veo 3.1 Preview (No Audio)	25-10
5	Baidu	GenFlare	25-10
6	MiniMax	Hailuo 02 Pro	25-06
7	Bytedance	Waver 1.0	25-08
8	Lightricks	LTX-2 Pro (No Audio)	25-10
9	Vidu	Vidu Q2 Turbo	25-09
10	MiniMax	Hailuo 2.3	25-10
11	Vidu	Vidu Q2 Pro	25-09
12	ByteDance Seed	Seedance 1.0	25-06
13	Lightricks	LTX-2 Fast (No Audio)	25-10
14	Alibaba	Wan 2.5 Preview	25-09
15	MiniMax	Hailuo 02 Standard	25-06
16	MiniMax	Hailuo 2.3 Fast	25-10
17	Luma Labs	Ray 3	25-09
18	Google	Veo 3 (No Audio)	25-07
19	Kuaishou KlingAI	Kling 2.1 Master	25-05
20	MiniMax	Hailuo 02 Fast	25-06

代表用户使用倾向的AI视频网站访问量排名

右图： AI产品榜发布的2025年10月AI视频生成网站访问量排名

下图： 摘自POE《2025年春季AI模型使用趋势报告》，截止2025年5月4日前一周AI视频生成使用份额占比



全球排名	产品名 AI产品榜	网站(web)分类 aicpb.com	10月上榜网站 Web访问量	10月上榜网站 变化
1	Sora | ChatGPT	AI Video Generators	68.58M	93.90%
2	Loom	AI Video Generators	18.41M	1.74%
3	klingai | 快手	AI Video Generators	11.03M	-9.42%
4	HeyGen	AI Video Generators	8.8M	22.83%
5	Runwayml	AI Video Generators	8.43M	31.37%
6	PixVerse	AI Video Generators	7.47M	-19.46%
7	Pollo	AI Video Generators	6.94M	8.47%
8	Hailuo AI | MiniMax	AI Video Generators	6.86M	11.67%
9	Wan AI - 阿里	AI Video Generators	5.62M	38.42%
10	Vidnoz AI	AI Video Generators	5.34M	-0.03%
11	Viggle	AI Video Generators	2.84M	39.48%
12	Clipfly	AI Video Generators	2.61M	28.13%
13	Vidu 海外 | 生数科技	AI Video Generators	2.24M	-0.86%
14	Synthesia	AI Video Generators	2.05M	7.11%
15	Pika Art	AI Video Generators	2.04M	12.86%
16	Luma AI	AI Video Generators	1.89M	-1.53%
17	Creati	AI Video Generators	1.68M	1475.58%
18	Animaker	AI Video Generators	1.65M	3.91%
19	Fliki	AI Video Generators	1.17M	3.71%
20	Hedra	AI Video Generators	1.14M	-15.98%



PART 03

AI视频生成能力发展与突破

1. 厂商技术能力进展与突破
2. AI视频生成技术的能力评估
3. 推荐关注的AI视频厂商/模型

推荐关注的AI视频生成厂商/模型

北京大学 PEKING UNIVERSITY

工具名称	最新模型	所属公司	国内国外	推 荐 理 由
Veo	Veo 3.1	Google	国外	Google的王牌，对标Sora。最近的SuperClue评测中，文生视频排名第一，图生视频排名第4。
可灵AI	O1	快手	国内	国内热度极高，发布一年多就迭代了20多个版本。能生成长达2分钟的1080p视频，DiT架构，Sora最强力的挑战者之一。
Sora	Sora 2	OpenAI	国外	AI视频的“引爆点”，让全世界看到AI视频的潜力，其目标是世界模型。在AI视频生成工具网站中，访问量仍居榜首。
海螺AI	Hailuo 2.3	MiniMax	国内	国内AI创业公司MiniMax的力作，海外用户多，性价比高。在最近的多个技术评测中，文生视频和图生视频排名靠前。
即梦AI	Seedance 1.0	字节跳动	国内	背靠剪映巨大的用户量，功能迭代快，尤其在人物和风格化视频方面有优势，是普通用户最容易接触到的工具之一。
Runway	Gen-4.5	Runway	国外	AI视频创作的先驱，教育了早期市场。强大的视频创作套件，提供运动笔刷、导演模式等专业功能。
PixVerse	PixVerse V5.5	爱诗科技	国内	海外认知度高，在网页和Discord上都能用，免费额度高，很多新手入门的首选，网站单月访问量达8000W。
Vidu	Vidu Q2	生数科技	国内	被称为“中国版Sora”。基于U-ViT架构，演示效果惊艳，在生成具有中国元素的内容上表现突出，是国家队的重要代表。
Luma AI	Ray 3	Luma AI	国外	从3D建模领域杀入视频赛道的黑马，收尾帧功能的鼻祖，新推出的Ray 3模型在文生视频中排名靠前，值得关注。
通义万相	WAN 2.5	阿里巴巴	国内	开源中的佼佼者，能力全面，一直在稳步迭代。在动漫、插画等风格化视频的生成上比较有特色。

「AI视频生成技术发展现状」要点总结

AI视频已具备专业级潜力，但距离“长剧情、稳定角色、复杂交互”的全面生产级替代仍需在时序一致性与“内在真实性”上持续攻坚，并以统一、演进的评测体系牵引技术与应用的螺旋式进步。

技术能力进展神速，但瓶颈依旧突出。

- 在过去一年中，AI视频在分辨率、时长、可控性和多模态等核心指标上取得了里程碑式突破，已具备专业级潜力。
- 时序一致性、物理与逻辑真实性、以及细腻情感表达依然是当前模型难以逾越的鸿沟，导致生成内容时常出现“鬼畜”失真和逻辑错乱。

头部厂商引领，模型各有专攻。

- 国内外科技巨头与初创公司纷纷入局，形成了以Sora、Kling、Veo、Vidu、即梦等为代表的激烈竞争格局。
- 不同模型在技术路径和能力上展现出差异化，有的擅长艺术风格渲染，有的精于物理世界模拟，有的则在镜头语言控制上表现出色，用户的主观“体感”因此千差万别。

从“主观感受”迈向“客观评估”：标准化基准成关键。

- 随着技术成熟，单纯依赖个例展示和主观评价已无法满足行业需求。建立统一、全面、可量化的评测基准变得至关重要。
- 以 VBench 和 SuperCLUE 为代表的评测体系，正推动行业从评估“表面真实性”向“内在真实性”深化，为技术迭代提供了清晰的指引和公平的比较平台。

未来，数据质量优先，用户真实需求为王。

- 评测数据显示，高质量、高相关性的数据对模型性能的提升效果远超单纯增加数据量。数据策略的优化是未来突破的关键。
- 技术指标的领先并不完全等同于市场成功。最终，用户访问量和使用倾向反映了产品的易用性、成本和生态。技术发展需更加关注解决用户的实际痛点，才能在激烈的市场竞争中胜出。

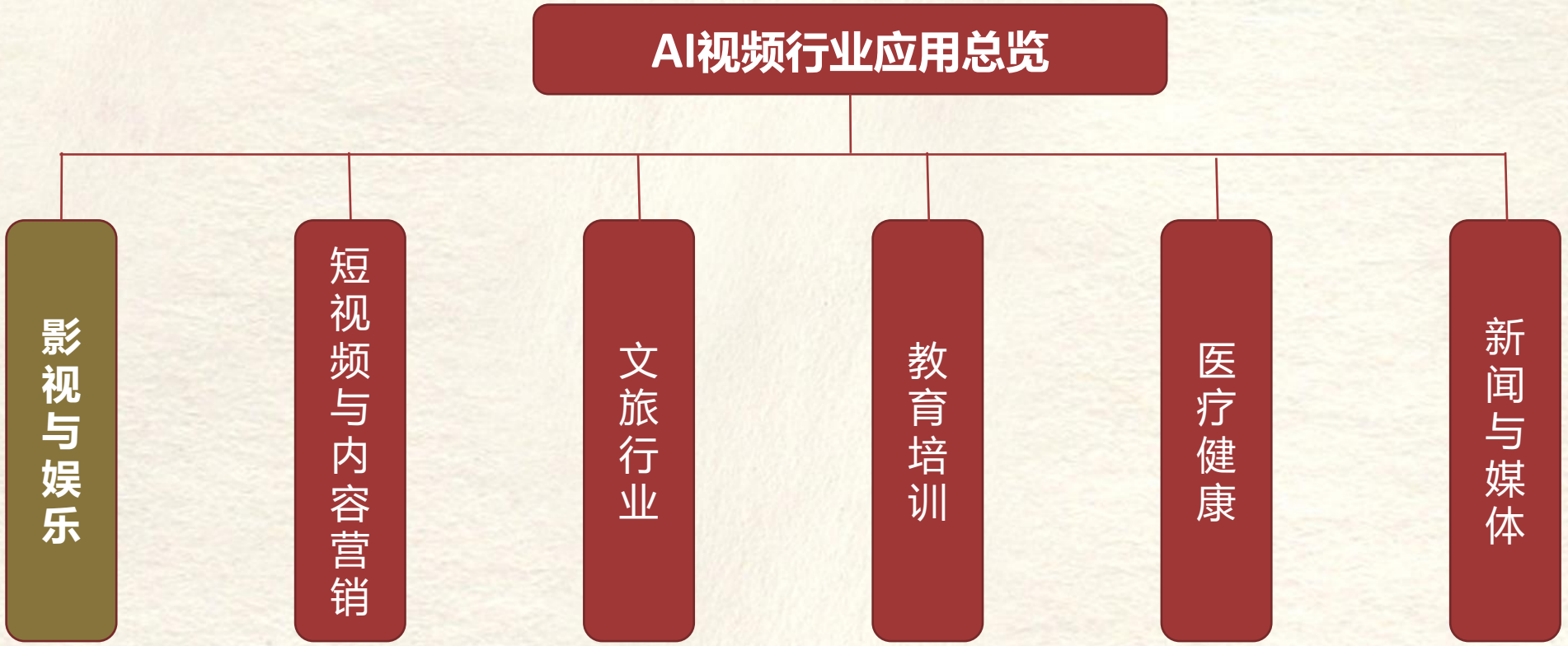


PART 04

AI视频生成行业应用现状

- 影视与娱乐
- 短视频与内容营销
- 文旅行业
- 教育培训
- 医疗健康
- 新闻与媒体

AI视频行业应用：从创意辅助到全流程重构



AI视频行业应用—影视娱乐行业迈进商业化阶段



















2016年《Sunspring》

AI编剧的首次尝试使用LSTM递归神经网络编写剧本，训练数据来自1980-90年代科幻电影剧本，在伦敦科幻电影节48小时挑战赛中入围前十，YouTube观看量超过100万次，被CNet评为"美丽而奇异的科幻新奇作品"

2018年《Zone Out》

《Sunspring》同一团队制作，采用神经网络和换脸技术，整合大量电影素材和演员语音生成新场景和对话，首部完全由AI独立制作的科幻电影，时长6分钟.

2023年1月《犬与少年》

首部AI生成背景的商业动画，Netflix联合小冰公司、WIT STUDIO制作，讲述少年与机器狗重逢的故事，背景绘制由AI技术完成，简化传统动画制作流程

2024年3月《中国神话》

国内首部AI全流程微短剧，共6集，美术、分镜、视频、配音、配乐全部由AI完成，实现全流程AI制作是中国首部完整的AI生成影视作品，标志着AIGC在国内的正式商业化应用

2024年3月《Our T2 Remake》

全球首部AI长片电影首映时间，90分钟完整长片，由50位AI艺术家协作完成，技术上基于《终结者2：审判日》框架，完全使用AI技术重制，是首部达到院线标准的AI长片电影

2024年4月《海上女王郑一嫂》

FizzDragon公司制作，时长70分钟，解决了长对白、长镜头处理和角色一致性等技术难题，是全球首部政府批准进入院线公映的华语AIGC大电影

2024年6月《三星堆：未来启示录》

中国首部AIGC连续性叙事科幻短剧集，博纳影业制作，抖音联合出品，即梦AI为该剧提供生成式人工智能技术支持。是传统影视公司与AIGC视频产品首次联合，该剧第一季共12集，采用全流程AIGC技术进行制作。

2025年1月《美猴王》

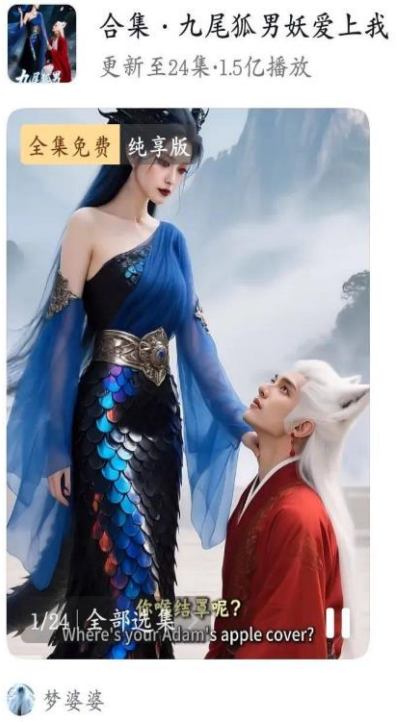
连云港文旅联合央广推出的AI贺岁微短剧，每集约五分钟，共十集，以《西游记》前七回为蓝本，讲述孙悟空成长故事。全剧采用AIGC全流程制作，是AI技术与传统文化IP的成功结合典范

2025年2月《兴安岭诡事》

国内首部实现盈利的AIGC短剧，90%情节AI创作，降低90%外景成本。抖音首次上线便引发关注。其前五集免费内容就吸引了高达4159万的播放量，展示了AI在制造悬疑、惊悚氛围方面的独特优势，成功吸引了大量观众

AI仙侠、玄幻短剧逐渐走红

2025年1季度，传统短剧市场陷入低迷期，内容同质化严重且制作成本高企，恰逢AI技术快速发展，AI短剧凭借低成本、高效率和个性化定制的优势，填补了市场空缺并开创了内容创作新模式，成为行业转型的契机。



抖音账号“梦婆婆”连载的《九尾狐男妖爱上我》目前累计播放量已超1.1亿，冲上抖音+快手短剧综合热度榜TOP 20



《玄幻：从拉二胡开始》幕后是只是一个7人小团队



《白骨精前传》抽卡素材生成成本仅6000元。



《我在阴间送外卖》总体制作成本5000元左右，在快手获得超20万点赞。确定分镜头后，一集内容最快只用两天就能完成。



快手“星芒剧场”上线的《山海奇镜之劈波斩浪》则创下3集播放量破千万的成绩，主角颜值被夸“艳压”内娱

“大片”中的AI视频生成技术加持

尽管AI视频生成技术还有很多局限，无法作为独立创作媒介，但越来越多影视作品已经将AI技术作为重要的辅助工具，如《异人之下》应用真人摹片动画结合人工智能技术；《传说》使用AI 技术还原 27 岁的成龙形象等



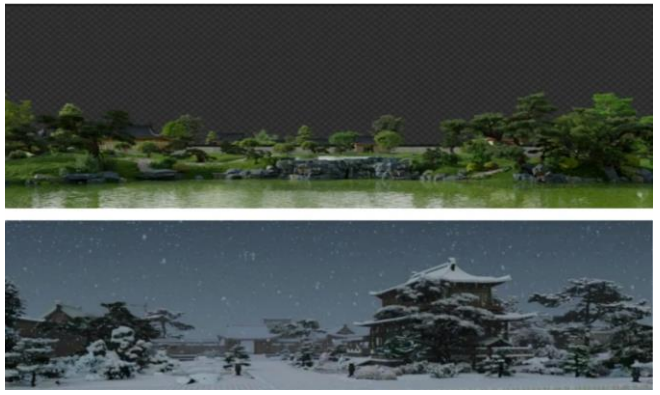
《复仇者联盟3》使用机器学习算法进行面部表情捕捉，AI开始在色彩校正、声音合成、视觉效果生成等后期制作环节发挥作用



《流浪地球2》借助深度学习、神经网络和生成对抗网络等技术“复活”吴孟达、让年过半百的吴京“重返二十岁”



《珠帘玉幕》的片头采用了由AI生成的国风元素



《只此青绿》翰林图画院的植被（上）和雪景（下）制作



《异人之下》真人摹片动画+AI技术的方式呈现

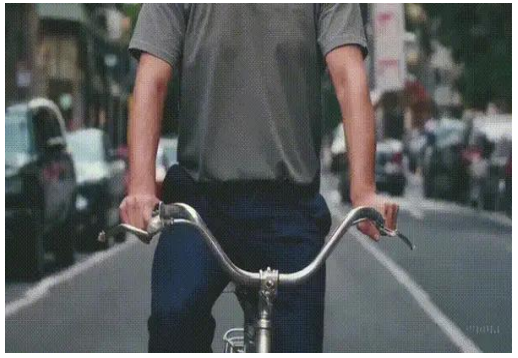


《传说》使用AI 技术还原 27 岁的成龙形象

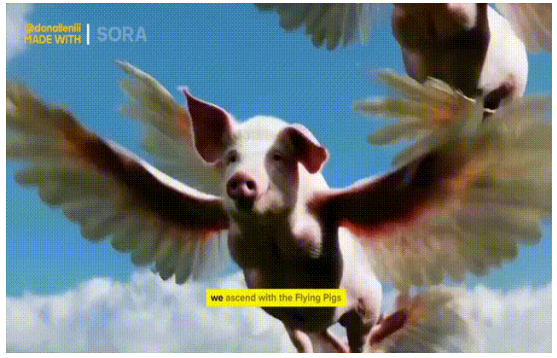
传统影视行业知名人士积极跟进AIGC

2024年6月：OpenAI Sora与好莱坞合作

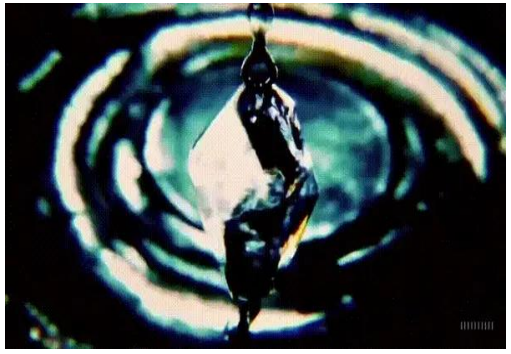
OpenAI邀请7个不同工作室和艺术家使用Sora创作，发布多部由专业创作者使用Sora制作的作品，标志着顶级AI技术公司与传统影视行业的深度融合



气球人



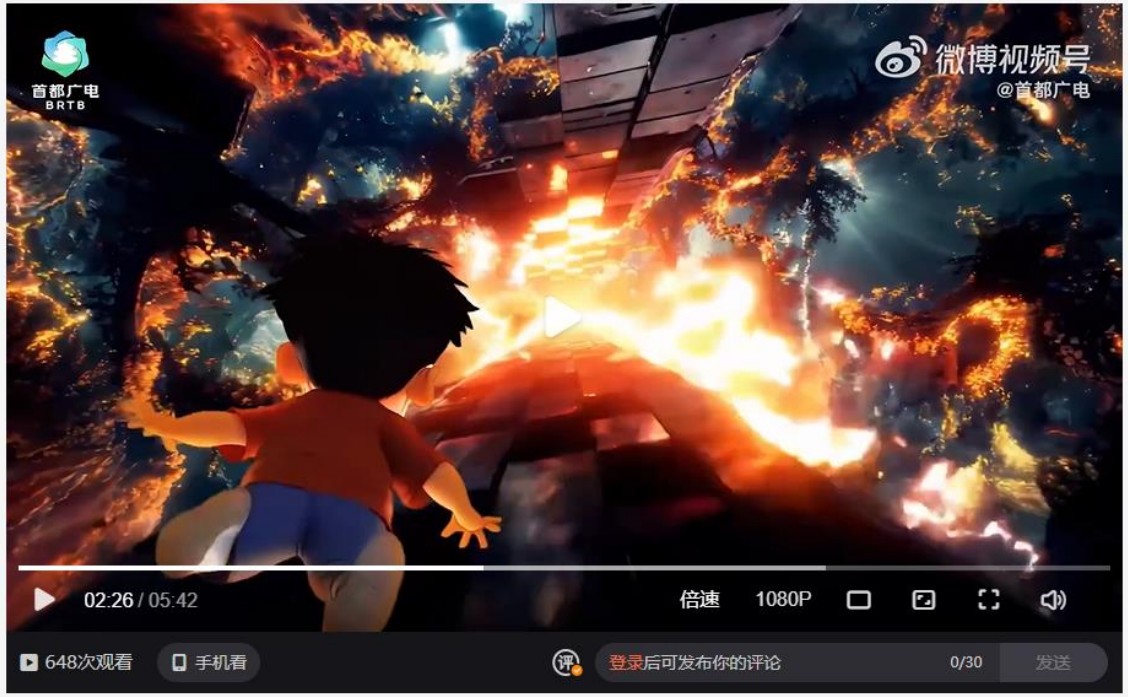
飞猪来了



人类寄给外星人的地球名片

2024年12月：可灵AI导演共创计划

国内李少红、贾樟柯、叶锦添、薛晓路、俞白眉、董润年、张吃鱼、王子川、王卯卯等9位知名导演，集体参与AI电影创作实验，发布9部AIGC电影短片，题材涵盖奇幻、志怪、亲情、动画等多元类型。



“可灵AI导演共创计划”导演王卯卯作品《再见兔子》

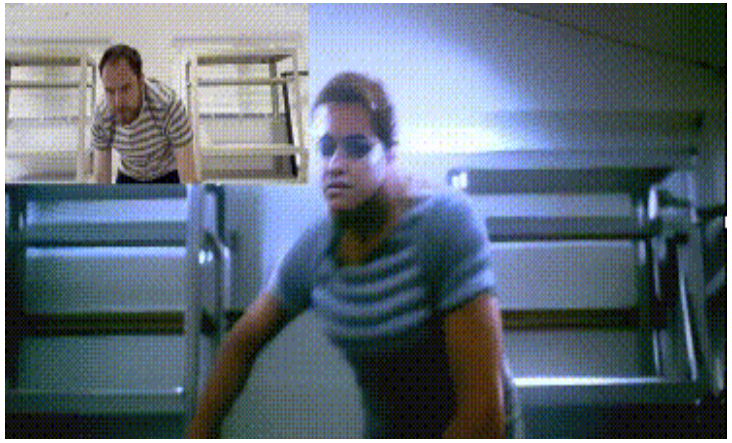
示例1：《Our T2 Remake》

全球首部完全由AI生成的长篇电影《Our T2 Remake》，长度90分钟，由50位AI领域的艺术家共同创作，历时三个月完成。影片通过使用多种先进的生成式人工智能工具，实现了从剧本创作到角色建模、场景生成等一系列制作环节的自动化。整个过程并未使用任何原版电影中的镜头、对话或音乐，所有内容均为AI原创。

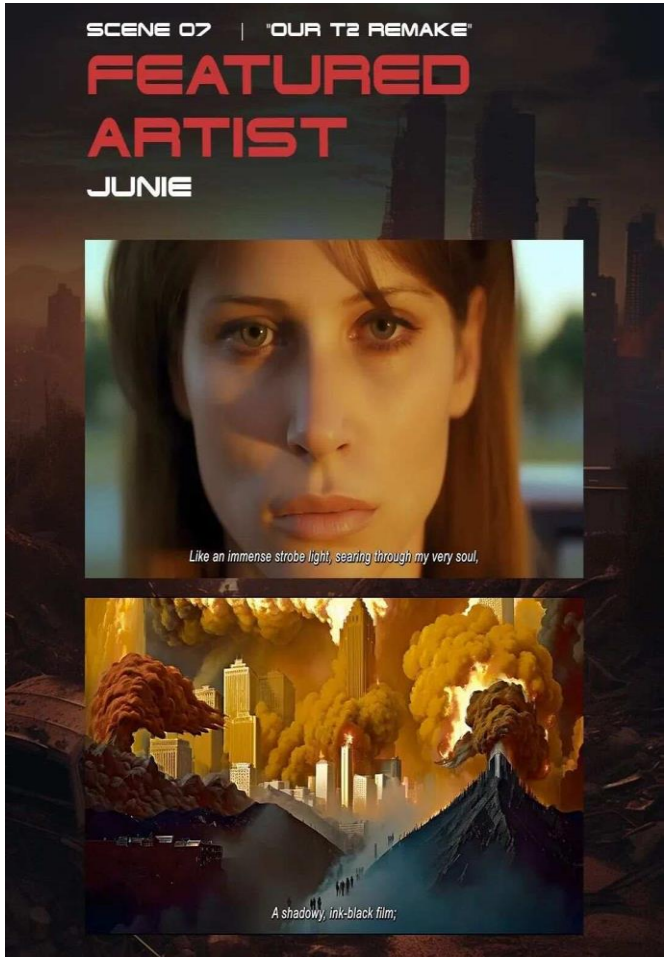
- **AI图像生成**：SD/MJ/DALL·E
- **AI 动画**：Pika/Runway/Deforum
- **口型同步**：D-ID AI
- **旁白**：ElevenLabs
- **剪辑**：Premiere
- **文字翻译**：ChatGPT



真人绿幕实拍+Kaiber的工作流，人分饰两角完成的武打场景



实拍绿幕+Runway



AI 在处理复杂的动作序列，尤其是在模拟人类动作的真实性和流畅性方面，仍然存在局限性。团队采用的工作流是实拍/CG/AE 等传统影视制作方法+AIGC 工具的这样一个结合

示例2：《新世界加载中》

快手可灵AI与星芒短剧联合出品、异类Outliers团队制作的全球首部AI单元故事集《新世界加载中》，是近期影视市场最有存在感的创新内容之一



呈现出电影级的镜头语言与画面质感，跨多类型题材的内容创中，大胆挑战了多元世界观建设、多样审美风格呈现、多类人物群像构建等大题





《新世界加载中》的难度挑战在于：历史、奇幻、都市、喜剧、科幻，多种题材的同时存在。在可灵AI的技术能力支撑之下，年轻创作者与AI成功共创，达成了阶段性的最好效果，Z时代的想象力自由落地，导演们的“脑中创意”最后找到了精准表达。





不仅仅是视频生成—AI驱动的影视行业全链革新

1. 前期制作：从“依赖经验”到“数据驱动”

剧本创作与分析： AI模型（如GPT-4、Claude）可以辅助进行剧情构思、角色设定、对白撰写，甚至进行剧本的市场潜力分析（预测受众、票房等）。

概念设计与故事板： 文生图工具（Midjourney, Stable Diffusion）和文生视频工具（Sora, Pika, Runway）成为核心生产力。导演和美术指导只需输入文本描述，就能在几分钟内生成高质量的概念图、角色造型和动态故事板。极大地压缩了视觉探索的时间和成本，让“所想即所得”成为可能，导演可以在开拍前就“看”到完整的影片雏形。

虚拟勘景： AI可以根据剧本描述生成或改造虚拟场景，团队无需长途跋涉即可完成勘景，并进行光照、镜头角度的预演。

2. 中期拍摄 ： 虚实结合，降本增效

虚拟制片： AI与LED虚拟影棚（如《曼达洛人》使用的StageCraft技术）的结合是中期拍摄的核心革命。AI可以实时渲染和调整LED屏幕上的背景，使其与摄像机的移动完美同步，演员仿佛置身于真实场景中。 将大量后期特效工作前置到拍摄环节，减少了后期合成的复杂度和不确定性。

AI“参演”与数字人： 可以更低成本、更逼真地实现演员的年轻化或年老化。

数字替身： 为演员创建高精度数字替身，用于完成危险动作或在演员无法到场时进行拍摄。解决了演员档期、年龄限制、高危动作等难题，并催生了全新的“虚拟偶像”经济。

3. 后期制作：从“劳动密集”到“智能高效” （目前AI应用最成熟）

智能剪辑与调色： Adobe Premiere Pro等主流剪辑软件已集成AI功能，如自动识别语音生成字幕、智能色彩匹配、自动重构图（适应不同社交媒体尺寸）。AI自动字幕功能的准确率已超过95%，将原本数小时的工作缩短到几分钟。

视觉特效： AI在Roto（影像抠图）、物体移除、场景修复等重复性劳动中展现出巨大威力。工具如Runway ML可以一键移除视频中的移动物体。

AI配音与本地化： AI语音合成技术（如ElevenLabs）可以克隆演员的声音，并用其声音生成不同语言的配音，保留了原声的情感和音色。极大地降低了多语言发行的门槛和成本，提升了海外发行的效率和质量。

示例1：《哪吒之魔童降世》

2025年开创性的动画电影《哪吒之魔童降世》的制作团队通过构建多模态生成技术栈**重构了60%的工业流程**，实现了**生产周期减少30%**，**人均生产力提高470%**，**成本降低20%**，**画面质量提升40%**的效果。该电影的技术应用涵盖了从艺术设计、动作捕捉到多语言推广的整个链条，成为AI驱动的艺术工业化的一个里程碑案例。

制作前期

剧本创作：

情节生成：AI分析“封神”文本，生成符合世界观的情节分支。

创意筛选：“元宇宙天庭”的赛博朋克概念，从AI生成的 300+ 种方案中选出。

美术设计

核心工具：ControlNet

效率飞跃：概念图创作周期从 数周 → 3天。

海量迭代：迭代 1200+ 张概念图，生成 800+ 版哪吒形象，最终版获 92.3% 观众满意度。



制作中期

场景搭建

AI生成：敖丙的“冰晶王座”由AI生成 37 个版本供美术挑选组合。

动画制作

关键帧：利用Stable Diffusion (图生图)，艺术家勾勒草图，AI补全光影色彩。哪吒“风火轮1080度旋转”镜头经AI 200+ 次迭代优化。

面部动画：建立含 10万+ 微表情数据的AI库；通过GANs技术将真人演员 52 种表情精准映射至角色，情感传递准确度提升70%。

特效制作

AI接管：30% 的特效镜头（全片1948个）由AI处理重复性工作。

降本增效：AI处理部分使人工时间 减少60%，成本降低45%。

渲染革命：“万龙甲”场景渲染效率 提升300% (单帧72h→8h)，成本降低40%。单帧渲染成本从 300元 降至 18元。

制作后期

声音处理

AI配音：AI语音合成技术生成 7个方言版本，助推区域票房提升15%-20%。

智能调色

核心系统：“CV-Scene”系统

数据驱动：基于 300万 组光照数据训练。

极限提速：将传统 3天 的人工调色流程压缩至 0.8秒。



by《NOILDAGOHd NOLIVWINY NI ASOTONHDHL, IALLVHHNH9 TVGOWII,IOW HO INIWdOTHAHG GINY NOLILVDITIdV HH., NO HHAVd GLIHM》

示例2：腾讯视频影片宣发视频制作

腾讯视频拥有丰富的版权视频资源，在内部和外部平台(如抖音视频号)进行宣发，需要大量高质量的素材支持，利用自研的基于意图感知的多模态视频素材Agent解决方案，实现任务自动匹配与宣发视频素材交付。



腾讯视频智能创作平台部算法专家 陈 曦《腾讯视频在剧本分析和素材宣发方向AI的应用和实践》

“AI+”逐渐从影视制作的创新工具升级为核心生产力

北京大学 PEKING UNIVERSITY

AI技术已经从辅助工具升级为影视工业的核心生产力之一。这不仅仅是技术上的“炫技”，而是深刻影响了从创意源头到最终成片的每一个环节，AI的全面介入，正在从根本上改变影视行业的成本构成。

传统成本构成	AI重构后的成本模型
高昂的实景成本（场地租赁、美术搭建、差旅）	**降低的虚拟制片成本**（LED影棚租金、数字资产购买/制作）
庞大的人力成本（大量美术、后期技师、群演）	**优化的人力结构**（需要更少的初级技工，需要更多AI艺术家、技术指导）
不可控的时间成本（天气、演员档期、后期返工）	**可控的制作周期**（虚拟制片不受天气影响，AI工具加速迭代）
高风险的创意成本（剧本不受市场欢迎，项目失败）	**降低的试错成本**（AI快速生成概念原型，用于早期市场测试）



博采AI虚拟影视基地，单体直径达50米的LED虚拟影棚，设计为270度环幕，由26600块LED屏幕组成，总像素达6亿，单体影棚总面积达5000平方米



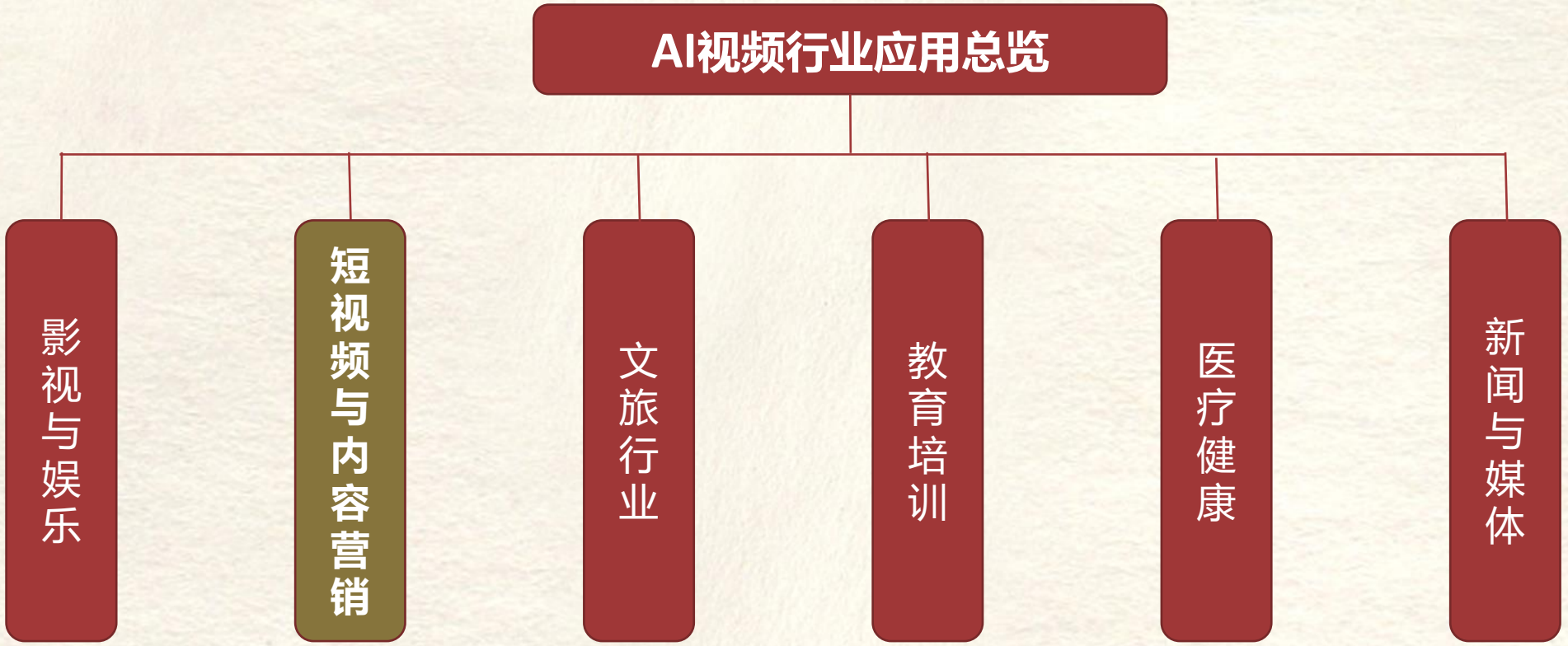
基于专业制作思维开发的AI智能体—墨客，用非线性逻设计逻辑破解专业创作的“抽卡困局”，实现从文本创作一直到终版视频的全流程AI输出，帮助专业的影视工作者更快捷更精准更有效地完成创作





这场悬崖戏，现场仅仅依靠一块岩石，就拍出万丈深渊的场景。

AI视频行业应用：从创意辅助到全流程重构



5亿人次被10只假兔子骗了

2025年夏天，一段8秒的监控视频席卷全球：后院蹦床上，10只兔子蹦跳玩耍，画面毫无特效痕迹，却因“真假难辨”引发全民热议。这场“全民找茬”的狂欢，最终助推播放量突破5亿大关。网友直呼“被AI骗麻了”🤯！







传统AI视频常以“奇观化”吸引眼球，例如”葡萄变猫、唐僧吃汉堡”等猎奇内容。然而这条兔子视频反其道而行，用“伪装真实”制造反差：

极致仿真：AI模拟监控画质的噪点、低帧率，甚至还原后院植被的随机晃动；

心理博弈：利用“监控偷拍”的叙事惯性，让观众默认画面真实性，再通过讨论发酵引爆悬念；

开放结局：不主动声明AI生成，留白式设计激发用户二次创作，衍生出“浣熊开会” “猫咪蹦迪”等病毒式变体。

AI视频自带“吸睛”光环

AI视频激发了人们对新技术的好奇心，又呈现出一种超现实的独特美学风格，同时其低成本高效率的生产方式颠覆了传统内容创作的认知，引发观众对技术与创意未来的深入思考。这些特质恰好满足了短视频平台对新奇内容的渴求和内容营销者快速吸引注意力、提高参与度的需求。







2023年，一段由Runway Gen-1生成的“威尔·史密斯疯狂吃意大利面”的鬼畜视频迅速燃爆各大社交媒体、数字媒体和广播公司，仅在 Twitter 上，这一视频的观看次数就超过了 800 万。这段视频怪异、扭曲但又极具魔性，完美符合Meme的传播特质。它不是一个商业广告，但其传播广度远超多数广告。

YouTube"AICat7"账号通过AI生成猫咪拟人化故事视频，最高视频播放量高达近9000万，收益丰厚。

这条30秒的NBA 广告完全由Veo 3制作，成本2000美金，降低 95%，单人次独立完成。决赛期间曝光量超 1 亿次，品牌认知度提升 25%。

短视频和内容营销是目前AI视频应用热度最高的行业

基于降本增效+海量需求，短视频和内容营销是当前AI视频生成最火热、应用最广泛、商业模式最清晰的领域。

（一）AI作为“生产工具”：如用AI生成电商产品广告、社交媒体信息流广告、品牌概念宣传片、数字人播报等。

（二）AI本身作为“营销叙事”：将“用AI创作”这件事本身作为营销事件的核心，追求话题性、品牌创新形象和病毒式传播。

成本	传统视频广告或内容制作成本高昂（拍摄、剪辑、演员、场地等），AI可以用极低的成本快速生成大量视频素材。
效率	一个营销团队可以在一天内生成并测试数十个不同风格、不同文案、不同画面的广告视频，这是传统流程无法想象的，极大地提高了A/B测试和内容迭代的速度。
创意	AI可以突破真人实拍的限制，生成充满想象力的、超现实的视觉效果，为品牌提供全新的创意表达方式。
节奏	短视频平台需要高频、海量、快节奏的内容。AI的生成能力恰好满足了这一需求。无论是商品展示、品牌故事还是信息流广告，AI视频都已经有了大量落地案例。

场景1：AI视频生成品牌创意广告

随着AI视频生成细节更真实、可控性更强，AI在广告领域的应用会更加深入和普及。但最终决定一个广告成功与否的，依然是其背后的洞察、策略和能否触动人心的情感连接。能够巧妙利用AI放大人类创意，而不是简单替代人类，方能胜出。



可灵AI做的小米AI眼镜广告片



元气森林广告



全球第一部由用户自己选择剧情、自己出演的AIGC广告片

- 只需选择性别、脸型、旅行起止地点、上传照片简单几个步骤，三分钟左右的时间，便可获得一部由AI生成，自己主演的个性化驾驶大片；
- 该小程序在广州车展期间一经发布，便有数万名全球网友在第一时间加入到"主演"行列



宁德时代推出“AI大片,邀您主演”的广告



Sora首个反斗城品牌商业广告，褒贬不一



失败案例：可口可乐AI广告《魔法假日来了》

场景2：社交媒体与UGC创作

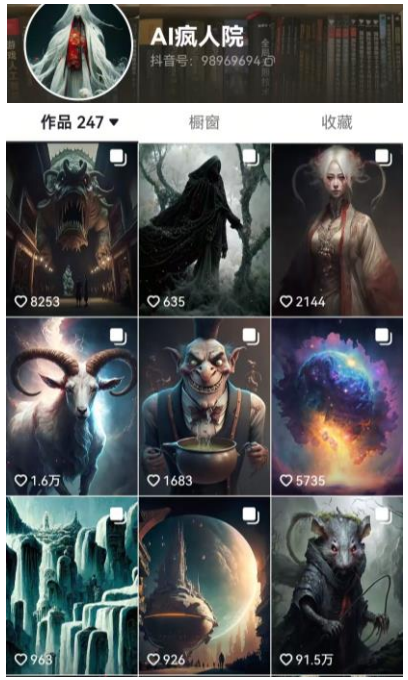
C端普通用户利用AI视频天然的话题性，根据自身风格和受众偏好定制独特的短视频素材，快速生产适合不同平台传播规则的内容，追求话题性、创新形象和病毒式传播，拓展影响力和合作机会，提升创作效率，吸引流量。



唐朝胖美人：通过 AI 生成的唐代丰腴美人形象，与现代 "白瘦幼" 审美形成强烈反差。37 个作品，在短短几个月内涨粉 20.7 万，单条视频最高获赞超 8 万，登上抖音热榜第 17 位。



周杰伦：抖音账号首条作品是他从小到大以及不同时期专辑封面照片的AI转场，发布1个多小时后点赞量就突破300万



AI疯人院：完全依靠AI工具，制作了一系列暗黑、怪诞、充满哲学思辨的短视频故事，吸引了数百万粉丝，成功开创了一个全新的内容赛道，证明了纯AI内容也能建立强大的品牌IP。



Act-two：利用工具新功能创造热度也是很多自媒体作者的玩法，最近Runway上线act-two功能，就被众多用户纷纷尝试，非常吸睛。



"LABUBU"AI 二创的病毒式传播，将 LABUBU 与各种热门 IP 进行创意结合，如《藏海传》《复联系列》等，扩大影响力和话题性

场景3：虚拟主播/真人对口型营销视频

在AI视频领域，真人形象对口型旨在高效复制和放大真人IP的影响力以实现内容量产，而虚拟IP形象则专注于打造新颖、可控的品牌数字资产，以在直播和推广中建立独特的品牌认知



义乌商贸城45岁的袜子摊主傅江燕对着手机镜头喊出"12345"，AI自动生成英语、阿拉伯语、斯瓦希里语等36种版本的营销视频。这位初中文化的老板娘，去年单凭AI制作的英语视频"卖空"了三个集装箱的货品。

图中这位老板娘能够自如地“说出”英语、法语、西班牙语、阿拉伯语、中文来介绍其产品，视频中她一口流利的外语让海外客户们纷纷震惊，殊不知背后其实有AI加持





在小红书平台上，我们会看到一些用动漫人物来讲解冷知识的视频，关注度很高。严格意义上来讲，这种平面动漫人物生成的AI视频不算数字人，但可以实现和数字人类似的效果。

今年6月15日，真人形象罗永浩数字人亮相百度优选的直播间，7个小时，观看人次超过130万人，GMV突破5500万，其中，1/3的产品销量超过罗永浩5月底的真人直播场次。

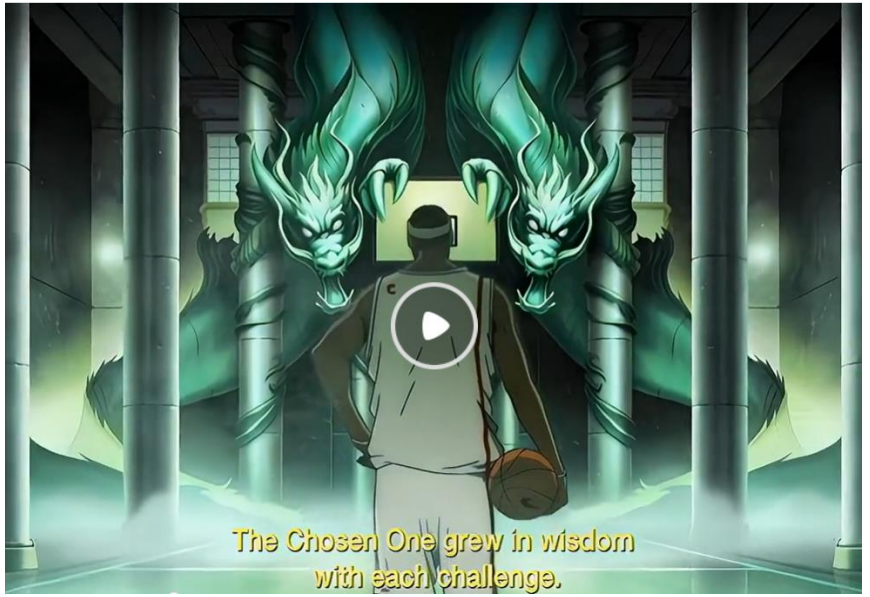


场景4：AI视频广告混剪

AI视频广告混剪将零散素材（图片、视频、文案）智能地批量重组，一键生成海量且适配多平台的营销短视频，实现低成本、高效率的广告内容生产与投放。广告混剪在“效率提升”和“规模化生产”方面已相当成熟，但在“顶级创意”和“情感表达”方面仍处于初级阶段。

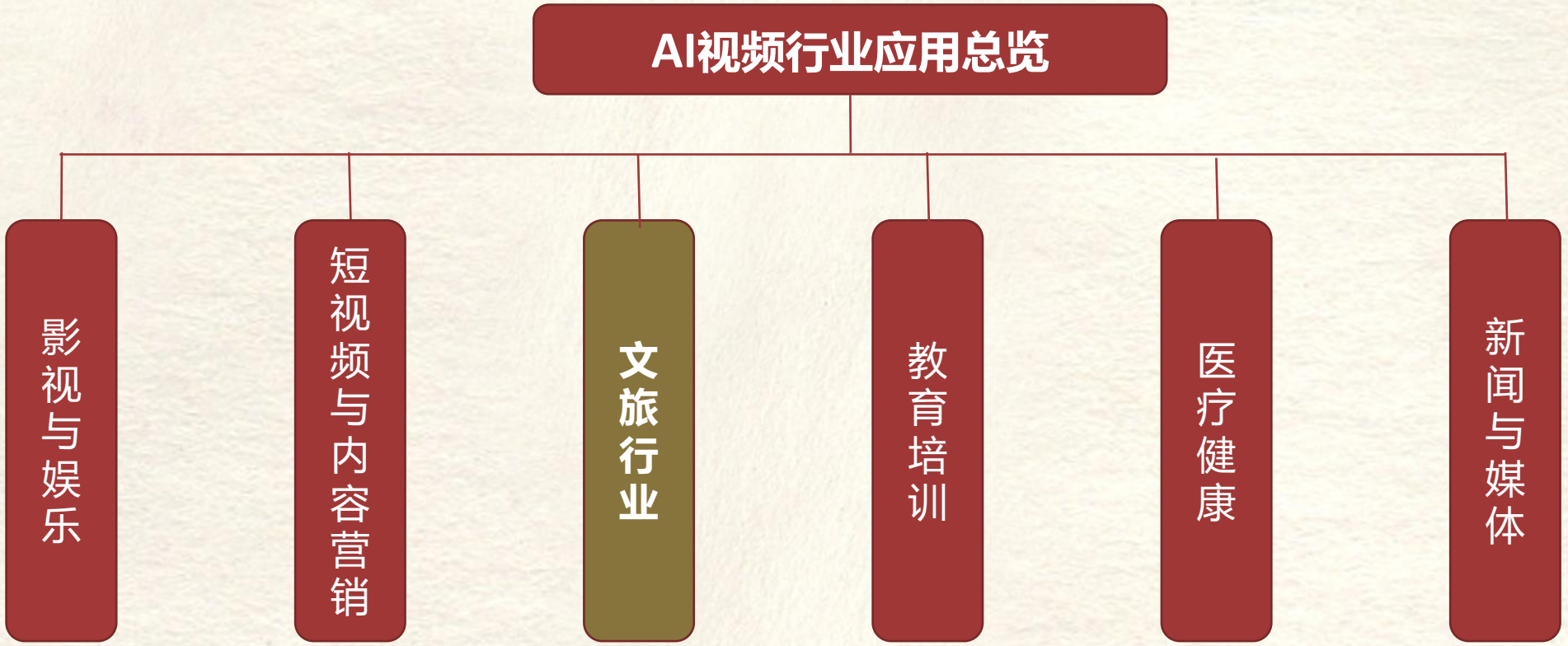


来自微播易《2024年AI营销应用解析报告》



耐克利用AI工具，将其精心制作的长篇品牌故事片（如纪录片、运动员专访）自动“混剪”成数以百计的、适合不同社交媒体平台（如TikTok, Instagram Reels）的短视频、动图和AR体验素材，

AI视频行业应用：从创意辅助到全流程重构



AI视频行业应用-文旅行业

AI视频生成技术在文化和旅游（文旅）行业中的应用主要集中在城市宣传、历史教育、节日活动推广、旅游营销和内容创作等领域。这些技术通过生成高质量、创意的视频内容，提升了文旅内容的吸引力和传播效率



文旅宣传片制作 01

智能化内容生成：

AI可以自动生成景区介绍、旅游攻略、文案等文本内容，并将其转化为视频

多模态创作：

结合文本、图像、音频等多种元素生成综合性宣传内容



沉浸式体验场景 02

VR全景体验：

通过AI技术构建虚拟旅游场景，让用户身临其境

AR实景导航：

结合现实场景提供智能导览服务

数字孪生技术：

创建景区的数字化复制品，提供虚拟游览体验



个性化营销推广 03

精准内容推送：

基于用户画像和偏好生成个性化推荐视频

社交媒体营销：

自动生成适合不同平台的短视频内容

品牌IP打造：

通过AI虚拟人等技术创建景区专属IP形象

AI文旅宣传片

AI制作文旅宣传片，无需大型拍摄团队和设备，成本低，周期短，不受天气季节限制。AI视频能呈现更加个性化的场景，可针对不同受众制作多个版本。此外，AI宣传片易于实现多语言本地化，还能创造传统摄影难以达到的视觉效果。









2024年3月，央视网用AIGC生成了一条全国文旅宣传片，全流程使用AI制作，利用AI文生视频和AI配音技术，让34个省级行政区的名字和当地特色巧妙结合了起来，展现大美中华



江苏传媒2024年"五一"假期推出的城市宣传片《AI你·南京》，AI技术全流程制作，融合玄武区湖光山色、秦淮区桨声灯影等区域特色，展现南京历史底蕴与现代科技的交融。上线一周左右就在在抖音平台获13.9万人次观看，视频号获11.8万人次观看并登上抖音南京热榜TOP4



快手可灵AI携手淮安打造首支城市定制主题大片《脉承淮水》联动11位超级创作者演绎「鲜活」淮安，话题总曝光17亿+，视频总播放2亿+，总互动量3000万+





在北京中轴线成功申遗一周年之际，生数科技与北京城市学院联合打造了一部以“中轴线”为主题的《中轴线上的北京文化》AIGC系列宣传片。利用Vidu 首尾帧功能，轻松实现从钟鼓楼、万宁桥、景山、故宫、端门、天安门穿越到天坛、先农坛、永定门的‘一镜到底’镜头。

AI+IP：AI数字人/AI文旅推荐官/AI伴（导）游

AI+IP的价值在于通过将人工智能与文化IP融合，创造全新的文旅体验和商业模式。这种融合能够让抽象文化变得具体可感，帮助文旅产业进入深度体验阶段。同时可挖掘微短剧、"灵境"体验、数智人和AI衍生品等更多商机。



2025大阪世博会上，科大讯飞的“AI孙悟空”腾云驾雾现身于水墨云海中，用中、日、英三语，实时与观众交互。用“科技+文化”的力量和底蕴，让中国文化“活”了起来。



广西"三姐数字人助游"项目入选“2024年文化和旅游数字化创新示范优秀案例”，项目基于"刘三姐数字人"形象，依托广西首个文旅大模型和广旅集团知识库，结合AR、AI等技术，整合广西文旅资源，让数字人能与游客多维交互，提供导游、客服、导览和AI对话等智慧旅游服务。



云南文旅与腾云公司、魔珐科技达合作共建的全国首个AI虚拟人文旅示范级项目。该项目利用虚拟人视频和直播技术，赋能全省景区的导游培训、导游服务、景区点位导览服务及乡村助农等领域，传播云南省丰富的历史文化、民俗风情、特产商品、非遗技艺技法。

AI融入VR、AR，打造更精准的沉浸式体验

近年来，“AI+VR”、“AI+AR”等技术打造的沉浸式数字文旅场景，正在成为文旅新秀，不断催生新产品、新场景、新体验。AI越来越多融入VR、AR，增强场景识别与交互能力。例如：通过物体标注，自动识别并解释场景中的建筑、艺术品等目标；实时跟踪人物与背景的遮挡关系，增强画面真实性；捕获并重现游客动态，提升沉浸感等。



青岛海上皇宫U·SEA奇幻海世界综合运用了AI、XR、AR、MR 等技术“还原”了远古海洋，让游客身临其境感受海洋魅力



全景故宫



数字敦煌

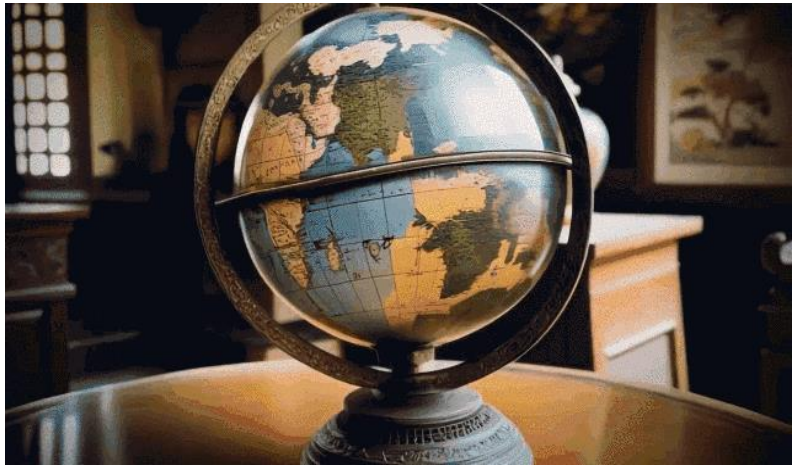
日常主题宣发小视频

在文旅行业的日常数字化营销中，可利用AI技术生成丰富多样的主题小视频，内容包括传统节日庆典、四季风光变化、二十四节气特色、实用旅游攻略以及当地特色美食展示等。这些AI生成的视频素材能有效丰富营销内容，吸引不同兴趣点的潜在游客。





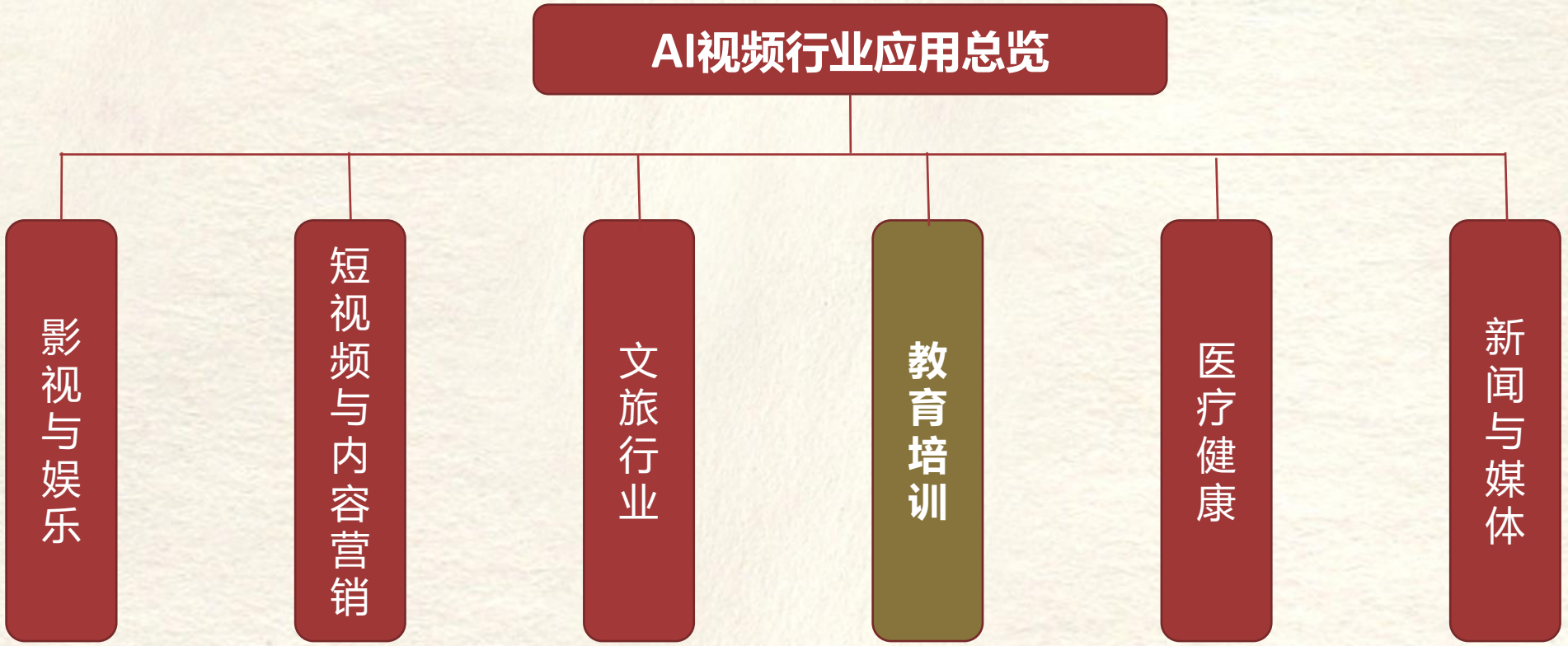








AI视频行业应用：从创意辅助到全流程重构



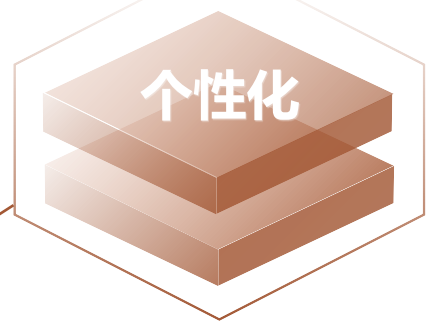
从“辅助工具”到“核心引擎”

AI视频生成技术正逐渐渗透到教育培训行业的各个角落，从个性化学习、课程开发到虚拟教学，展现出巨大的潜力和价值。已不再是零星的尝试，而是呈现出规模化、体系化的发展趋势。



降本增效，普惠教育资源

传统教学视频的制作成本高且周期漫长。AI视频生成技术能够将文本、图片甚至简单的指令快速转化为内容丰富、形式多样的教学视频，极大降低制作门槛和成本，让优质教育资源的生产和传播更加高效、广泛。



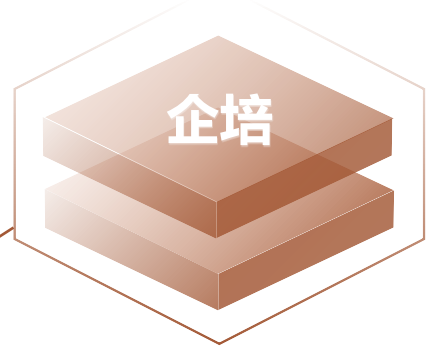
个性化学习成为可能

AI技术能够根据学习者的知识水平、学习进度和兴趣偏好，动态生成或推荐个性化的学习视频，实现“千人千面”的精准教学，有效提升学习效率和参与度。



交互性与沉浸感增强

结合AI虚拟数字人、语音交互等技术，AI视频能够创造出更具交互性和沉浸感的学习场景，例如与虚拟教师进行实时问答、在模拟环境中进行实操演练等，从而激发学习兴趣，加深知识理解。



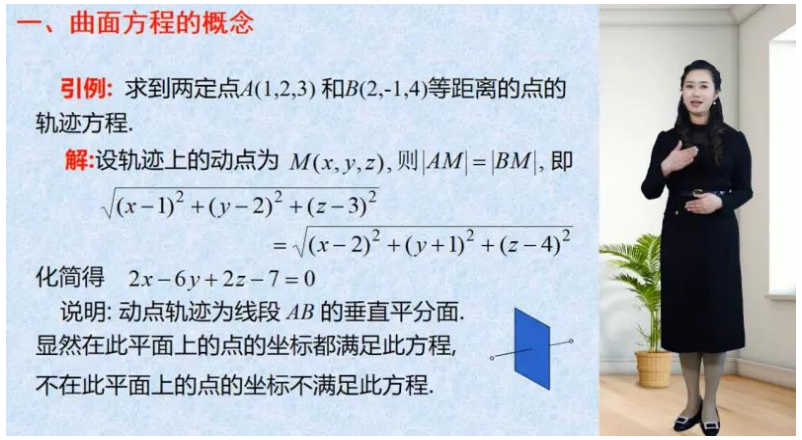
企业培训的创新实践

在企培领域，AI视频生成技术可用于新员工入职培训、产品知识讲解、操作技能演示等方面。标准化的内容和生动的展现形式，确保培训质量的统一性和高效性。

场景1：批量课程/微课生产与更新

采用文本转视频+数字人（如Synthesia/HeyGen），将教学资产（如教材、PPT、脚本等）转化为标准化视频，结合模板化与术语库以保障一致性与可维护性







英国博尔顿学院是一所继续教育学院，信息与学习技术和学习资源中心只有 7 个人，但服务的在线学习的学生却超过 10,000 名。采用Synthesia快速制作AI视频，将 10 分钟视频制作时间从 3 天缩短到 30 分钟，节省了80%的时间，仅一年时间，就向博尔顿学院的在线课件库添加了 400 多个视频 。

科大讯飞“智作”平台提供从文本到语音再到视频的一站式服务，能够快速将教学文稿转化为带有配音和动画效果的教学视频，常用于制作微课、知识短视频等。上图为青海大学一名老师使用该平台制作的高等数学课程视频。

拉美航空需要为整个组织 37,000 多名员工进行技术或监管内容培训，航空业培训存在跨语言、跨地区团队沟通的压力，传统外包视频方式无法跟上频繁的运营变化。利用AI工具，培训部门用较短时间制作了 300 多个视频在公司内传播，覆盖超过 16,000 名学习者。除了效率提升， Synthesia 的一键翻译功能，可以轻松创建西班牙语、葡萄牙语和英语的多语言培训视频，不仅缩短时间，视频质量也大大提升。

场景2：K-12 教育领域的应用

利用AI视频生成工具，将抽象的知识点（如古诗词意境、科学实验原理）转化为直观的视频画面，辅助教学。



AI“复活”历史名人：济南市莱芜区花园学校文昌校区语文教师王聪教一年级学生背诵《静夜思》时，用AI生成李白的虚拟形象，让他与学生对话，增强课堂的情境代入感这种寓教于乐的形式深受低年级学生欢迎，拉近了他们与历史人物的距离。



制作绘本故事：知乎作者清风路晓，使用DeepSeek+即梦为孩子制作儿童绘本故事的视频。

【工具】

DeepSeek、剪映

【学科知识点】

小学语文三年级古诗《山行》。古诗语言凝练、意境深远，引入视频可以将古诗中的抽象意象具体化、形象化，帮助学生理解内容，感受诗歌的美。

【步骤与方法】

（1）生成视频脚本

打开 DeepSeek 官网，勾选【深度思考（R1）】与【联网搜索】，输入提示词：

我需要为古诗《山行》制作一个课堂朗诵引入视频，具体要求如下：

(1)视频功能：通过逐句朗诵+画面呈现，快速带进诗意氛围，用于课堂开头吸引学生注意力，无需教学解析或互动环节；

(2)内容结构：

- 仅保留“逐句朗诵与对应画面”部分，每句诗独立成段；

- 每段包含：诗句朗诵语调指导（如重音、节奏）、匹配的画面描述、字幕设计建议；

(3)风格要求：

- 画面：国风动画/实拍+水墨特效，突出秋日清冷与枫叶灿烈对比；

- 声音：背景音乐用古琴/箫，朗诵者用沉稳男声，句尾留白 2 秒供学生感受意境。

（2）用剪映生成视频

打开剪映，软件首页点击【图文成片】，进入视频生成界面。



图 1-2-13 剪映图文成片使用流程

在自由编辑文案页面，将 DeepSeek 生成的脚本复制粘贴到输入区域，然后在右下角选择视频解说的音色，最后点击右下角【生成视频】按钮，在弹出框中选择【智能匹配素材】，随后剪映开始生成视频。



图 1-2-14 自由编辑文案操作过程

（3）编辑视频

在剪映中调整视频的背景音乐、文本配音、画面内容和字幕文本。

（4）导出视频

点击左上角【菜单】按钮，展开【文件】选项，点击【导出】，进入导出选项编辑页面。在导出选项编辑页面中，可以设置视频的标题和保存路径，视频导出的分辨率等建议使用默认推荐配置。

【最终效果】

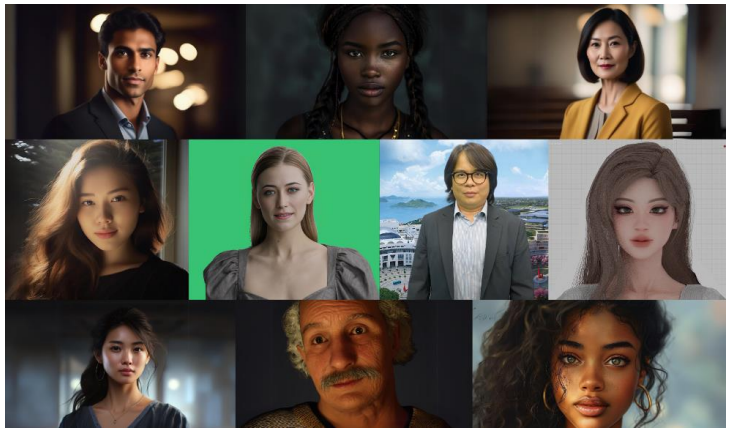


制作古诗词视频：上海市浦东教育发展研究院为中小学教师示范如何利用剪映AI功能制作古诗词视频。

场景3：AI虚拟教师与数字人讲师

AI虚拟教师具有不知疲倦、形象多变、7x24小时在线的优势。通过AI驱动的虚拟数字人进行授课、辅导和答疑，不但为学生提供了全新的互动体验，而且这些老师本身也是IP，受到学生的喜爱和追捧，甚至会形成独特的“粉丝文化”。







仁川国立大学与DeepBrain AI合作，将教育创新研究所所长、韩语教育系教授金炳元以数字化形式呈现。标志着教育领域的一个里程碑，打造出韩国首位人工智能人类教授。

香港科技大学的“AI讲师”：2024年3月，香港科技大学推出了亚洲首批十位AI讲师，这些虚拟教师来自不同国家，拥有不同的背景和专业，能够用多种语言进行教学，极大激发了学生的学习兴趣和课堂参与度。2024年12月，科大举行“最受学生欢迎”评选活动，并聘任选票最高的Fiona出任科大首位AI主播，肩负报导大学发展、科研成就及重大活动的重任。

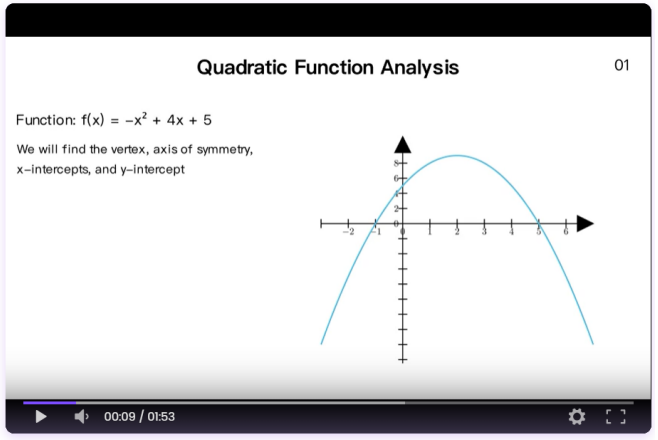


场景4：生成个性化学习视频

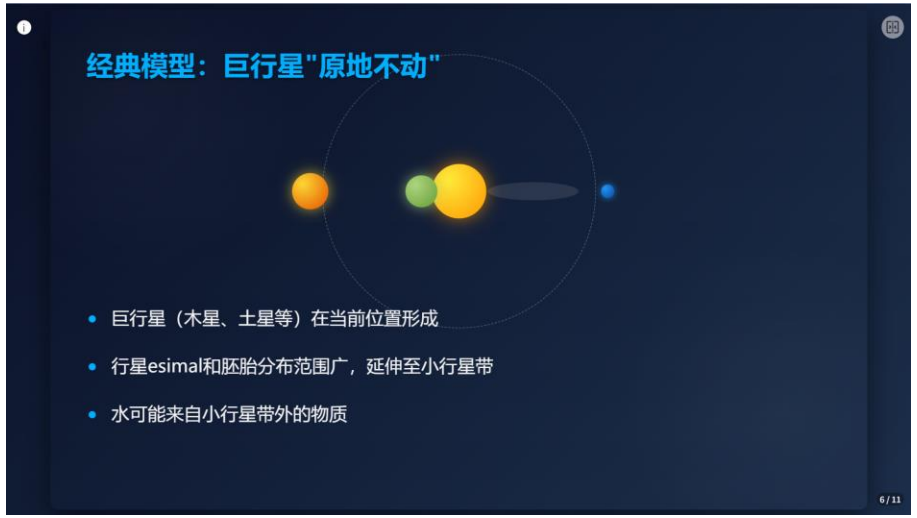
AI 视频生成技术还可以个性化的学习路径和提供学术支持，通过收集有关学业成绩、学习风格以及学生优势和劣势的信息，制定个人需求的教学策略，创建针对特定学科和主题的个性化讲解视频，帮助学生更好地理解复杂的学术内容。



芬达为什么冰的才好喝



f(x) = -x² + 4x + 5. Show the vertex, axis of symmetry, x-intercepts, and y- intercept

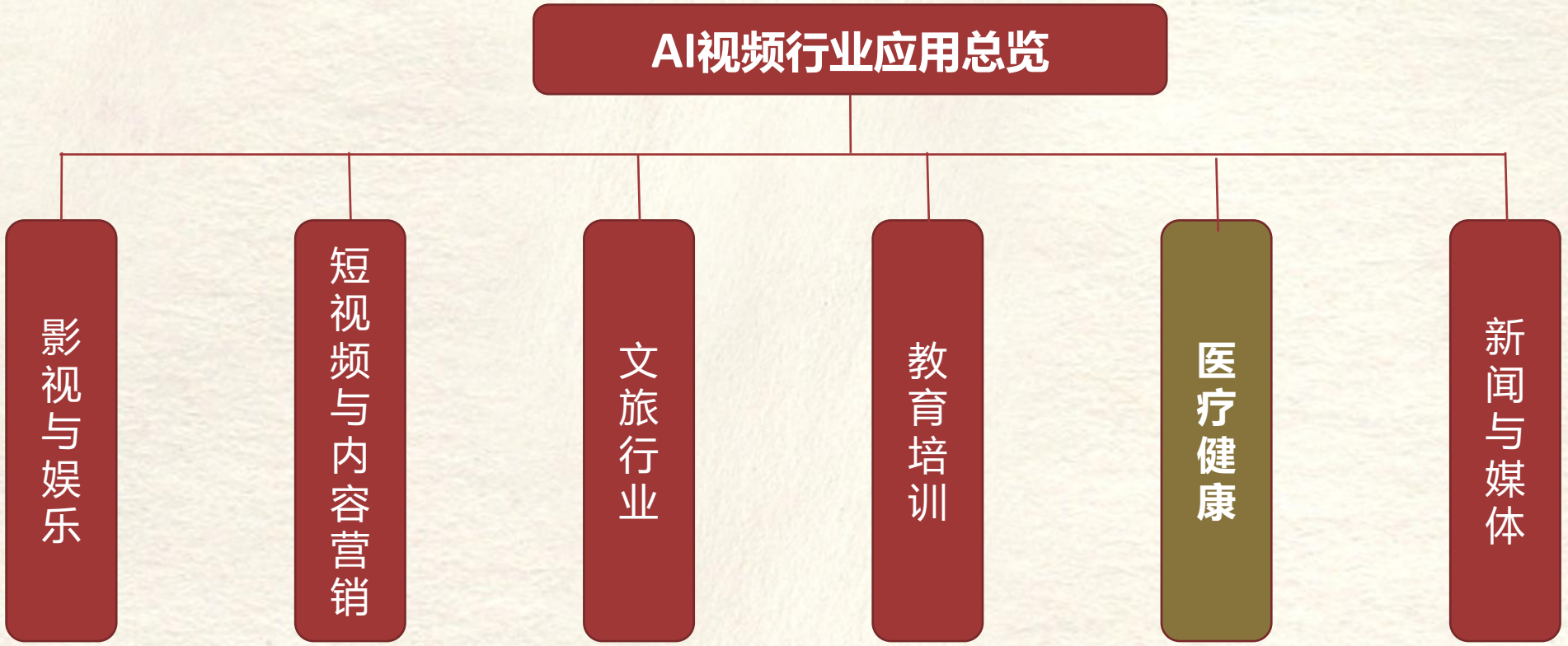


斯坦福大学的VideoTutor学习平台，能够根据学生提出的任何问题，迅速制作出包含结构化脚本和专业配音的视频讲解。这些视频不仅内容丰富，更通过生动的动态视觉效果，如图表、动画、公式演示等，将知识以直观易懂的方式呈现给学生。

秘塔AI “今天学点啥”可将网页、学术文献等复杂内容转化为个性化讲解内容，用户通过输入指定链接或文本，系统将自动生成适配用户知识水平的讲解幻灯片及音频。通识支持多种讲解风格，包括课堂、小说、故事等个性化定制模式。该工具旨在通过AI技术推动教育平权，帮助用户以个性化方式探索各领域知识。

注：严格来说，此案例并不属于本报告讨论的AI视频生成技术，但考虑到从用户使用角度看，属于”生成个性化学习视频”的范畴，因此给出此应用案例供参考。

AI视频行业应用：从创意辅助到全流程重构



AI视频行业应用—医疗和健康行业

目前，AI已经深入医疗健康的各个层面，AI视频技术也在各个场景得到有效应用。目前的AI视频技术中，用于辅助诊断和治疗优化主要是视频分析技术，而视频生成技术主要用于患者教育、医学培训和指导。



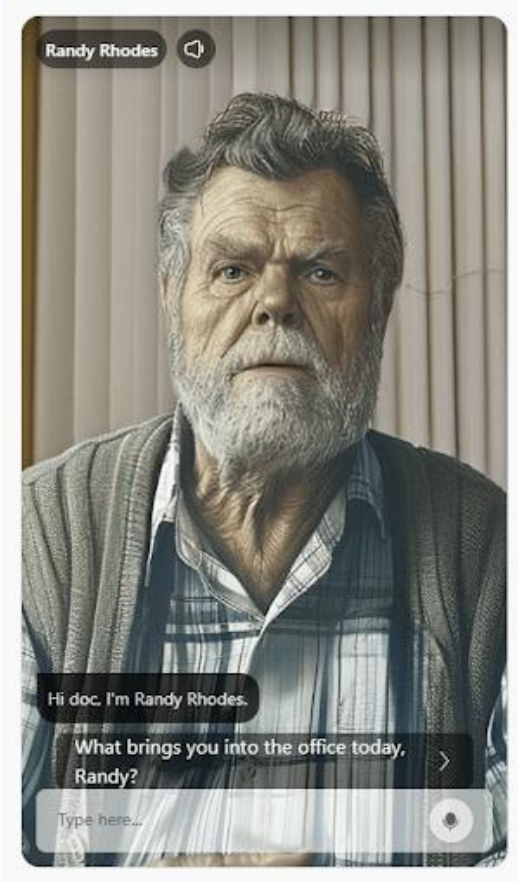
注：① 上图来自《人工智能+"医疗健康行业应用白皮书》
② 图中“√”标记为AI视频可以辅助应用的场景

AI视频生成在医疗健康行业主要应用场景

AI视频生成技术主要用于患者教育（虚拟药师）、医学培训（动态教学视频）、康复指导（定制化演示）等，其核心价值是通过合成新视频内容降低专业知识的传播门槛，节本增效。

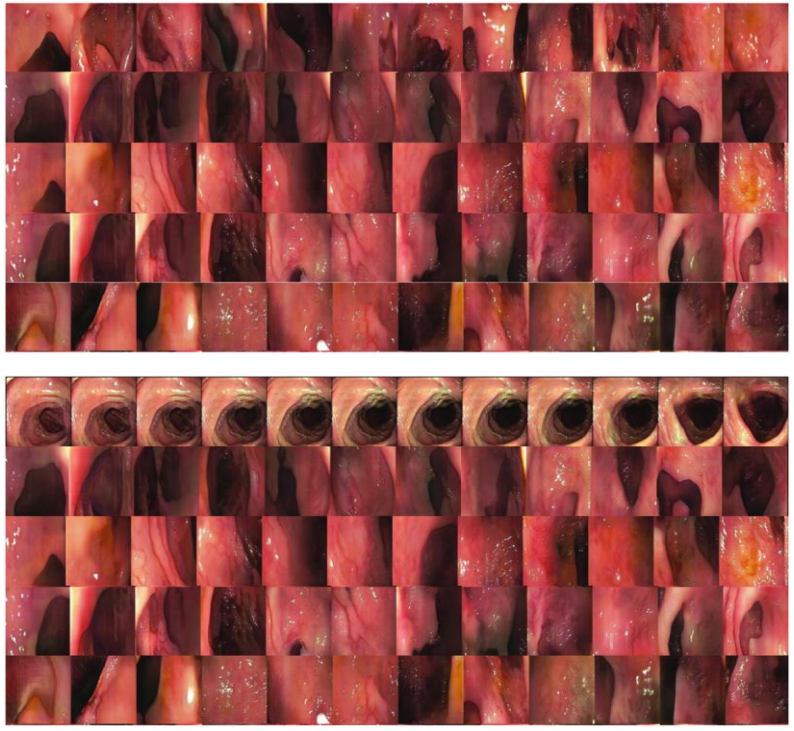
主要场景	示例	关键价值
医学教育与培训	虚拟患者模拟、内镜/结肠镜视频模拟、护理技能培训、HCP(医疗专业人员)教育、政策与流程SOP演示	提供安全的临床技能练习环境，解决稀有病例数据匮乏问题，缩短培训内容制作周期，确保政策流程标准一致性，加速医学生和医护人员胜任力培养
患者服务与教育	多语种患者教育视频、手术知情同意AI视频、智能问询与导诊系统、慢病管理与照护者支持、院内导航	克服语言和文化障碍，提升患者理解度，提供全天候服务，减少等待时间，优化就医体验，为患者和照护者提供持续支持和知识赋能
手术与远程医疗协作	术中视频增强处理、术前三维与术中视频融合、远程"虚拟洗手"协作	实现低延迟术中视频增强，保障数据隐私，支持专家远程协作，突破地域限制提升复杂手术资源可及性
医学传播与内容创作	会场互动AI视频、多语言健康知识库、患者案例分享、内部沟通材料	提升内容创建效率和多语言本地化能力，确保沟通一致性，创建交互式内容，加速医学知识传播，提高医疗教育可及性
医疗数据与隐私保护	合成医疗影像/视频生成、隐私过滤与去标识化、合规框架建设	解决医疗数据稀缺与隐私保护矛盾，生成高质量合成数据，为AI任务提供充足训练素材，同时符合监管要求，促进医疗AI合规发展

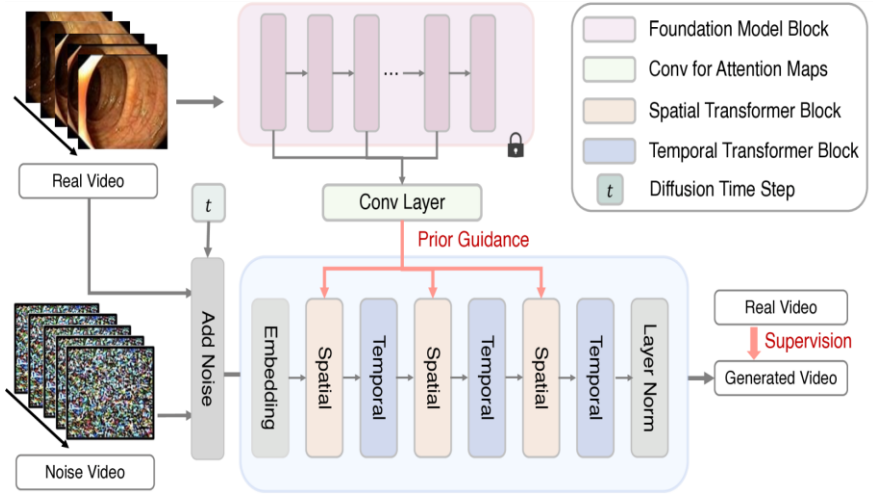
场景1：医学教育培训与训练

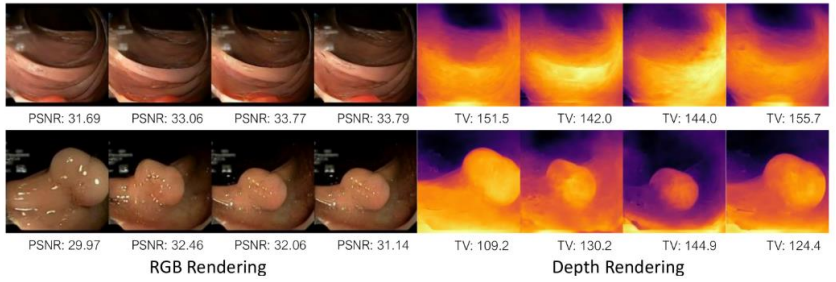


南伊利诺伊大学医学院虚拟病人“兰迪”，帮助学生练习、犯错、建立自信，以及最重要的，学会赢得信任。

视频生成对抗模型—MeVGAN，专门用于制作结肠镜检查视频，可应用于结直肠癌的筛查和治疗。MeVGAN能够生成高质量的合成结肠镜检查视频。下图为MeVGAN 能够生成结肠镜检查过程的许多不同阶段的示例





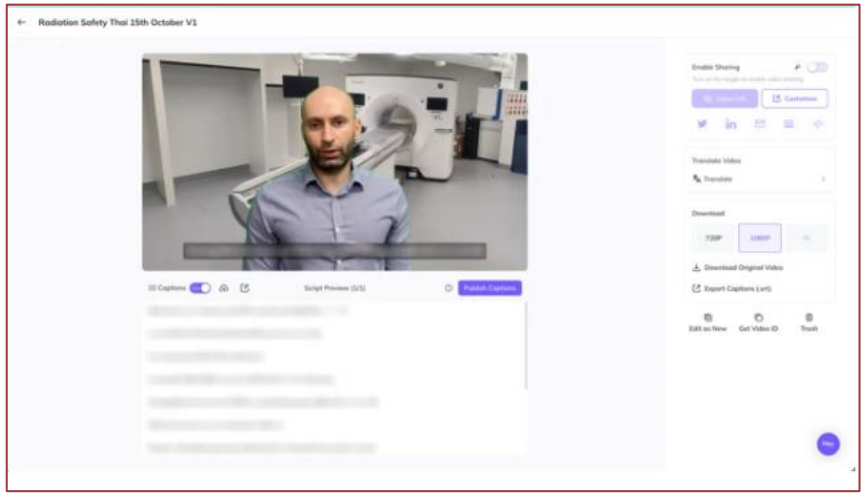


Endora，这是一个开创性的医学视频生成框架，能够创建高质量、动态且逼真的内窥镜模拟。Endora 甚至能够通过现成的内窥镜重建方法实现 3D 手术场景模拟。

场景2：患者教育与知情同意

案例1：AI生成诊疗沟通视频

朱拉隆功大学医学院（泰国曼谷）放射科核医学部门采用HenGen技术生成数字人视频，用于增强影像和放射治疗患者体验。该系统能根据患者的语言、文化水平、年龄段和文化背景提供个性化沟通，显著提升患者参与度，加深医疗程序理解，并提高整体护理满意度。



HenGen生成的诊疗沟通数字人，用于和增强影像和放射治疗患者实事沟通

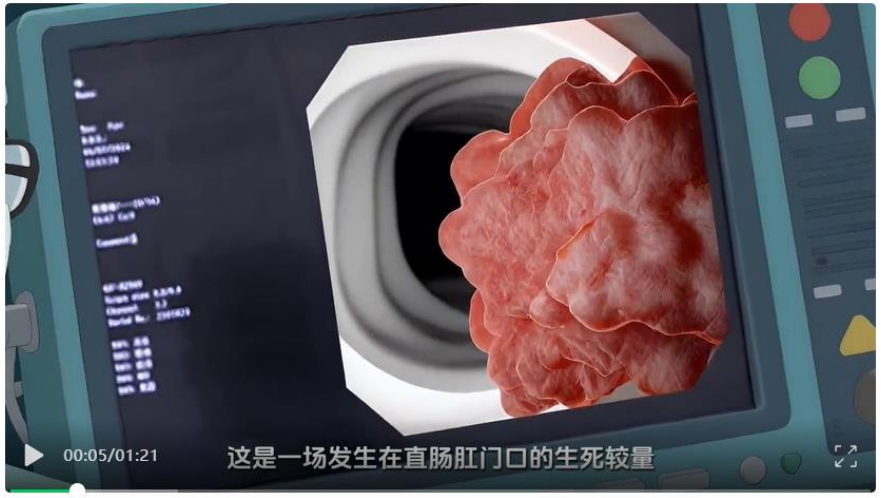
案例2：AI生成手术知情同意视频

英国Moorfields眼科医院应用人工智能技术制作白内障手术知情同意视频，成功将AI融入患者护理和手术知情同意流程。这一创新应用显著提升了患者对医疗信息的理解水平，有效缓解患者紧张情绪，并加速了医疗知识的普及传播。

案例3：AI生成患者教育视频

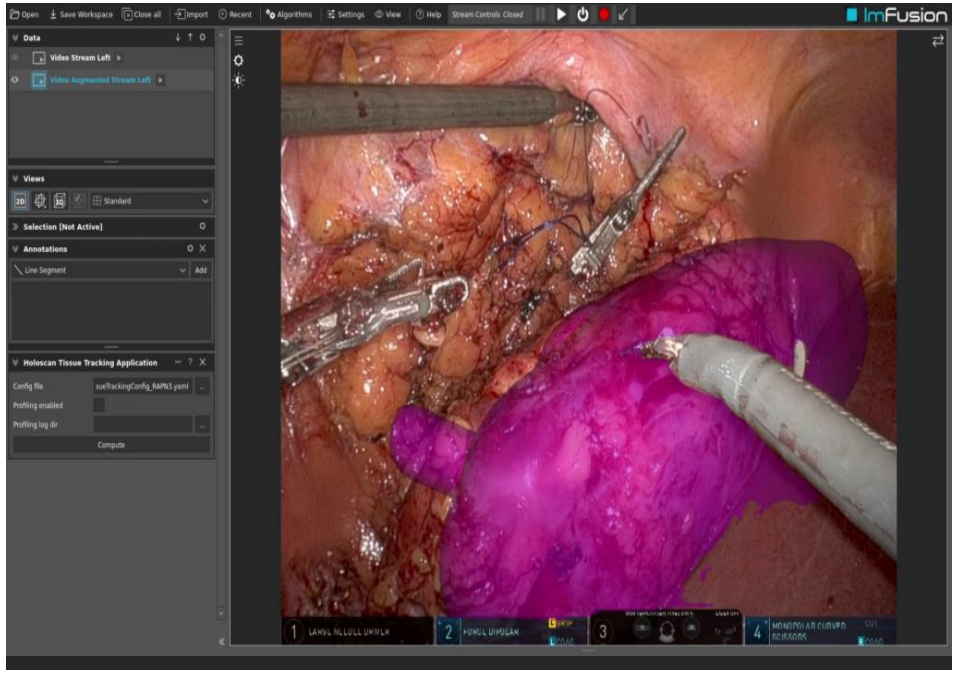
英国赫尔大学教学医院运用人工智能技术为乳腺癌患者开发了系列教育视频，共计七个，精准匹配患者健康历程各阶段的信息需求。其中包括"乳腺癌随访计划"，采用虚拟人物旁白形式指导患者进行乳房术后伤口护理。这些AI生成的教育资源获得了患者的广泛好评和欢迎。

Vidu赋能301医院，制作包括医疗手术宣传视频、疾病科普视频、超级微创诊疗书籍IP到AI医疗科普等一系列科普视频

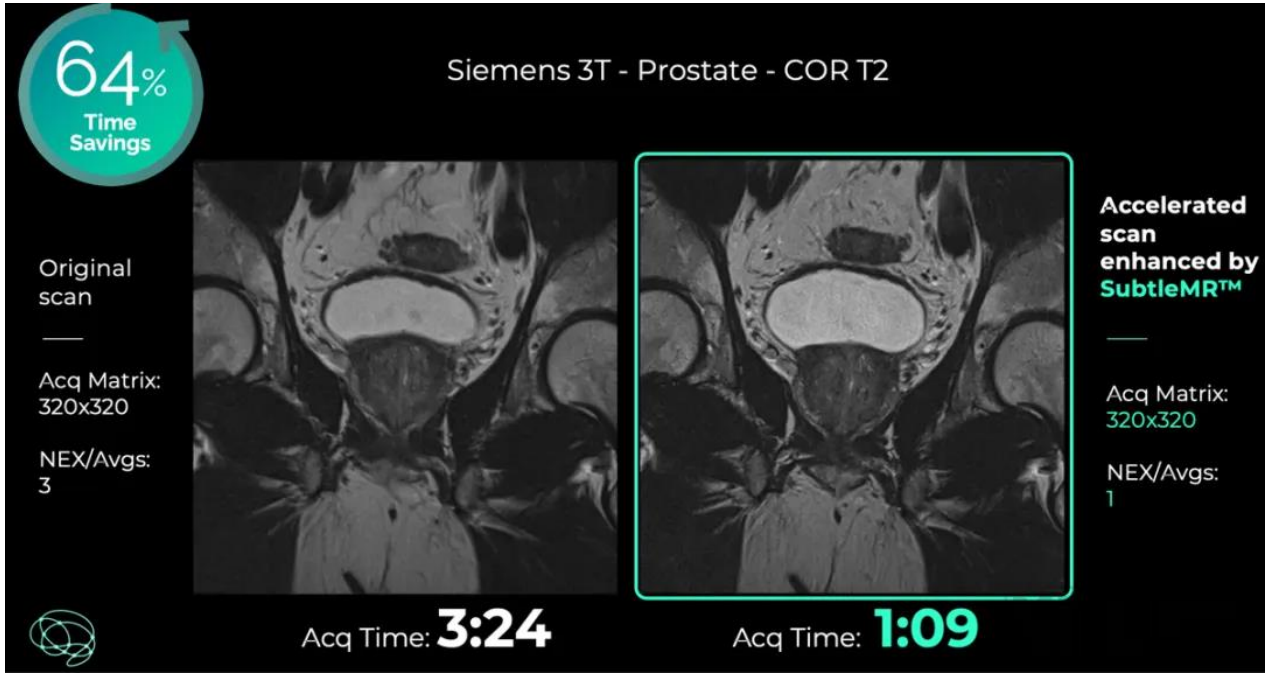


由Vidu AI生成的医疗科普视频：《超级微创破局低位直肠癌！保肝护命不再是奢望》

场景3：术中视频引导与影像增强



总部位于慕尼黑的 ImFusion 公司和 NVIDIA Connect 计划成员使用 NVIDIA Holoscan 创建了一个可以将术前数据与实时术中成像相结合的系统。通过将多模态成像与 NVIDIA Holoscan 融合来实现实时手术指导



SubtleMR能够从零剂量或低剂量成像中还原出接近全剂量扫描的效果，从而将钆剂剂量减少90%，同时不损害诊断质量。此外，SubtleMR还能加速MRI成像过程，可在4到10倍速的时间内完成成像，确保诊断级别的精确度。

场景4：医院导诊与客服数字人

AI导诊数字人可以进行健康宣教、预约挂号、问诊、导诊，还有多轮对话，比如推荐科室、医生。通过多轮回答可以更深入地了解患者的情况，给患者提供更准确的推荐。

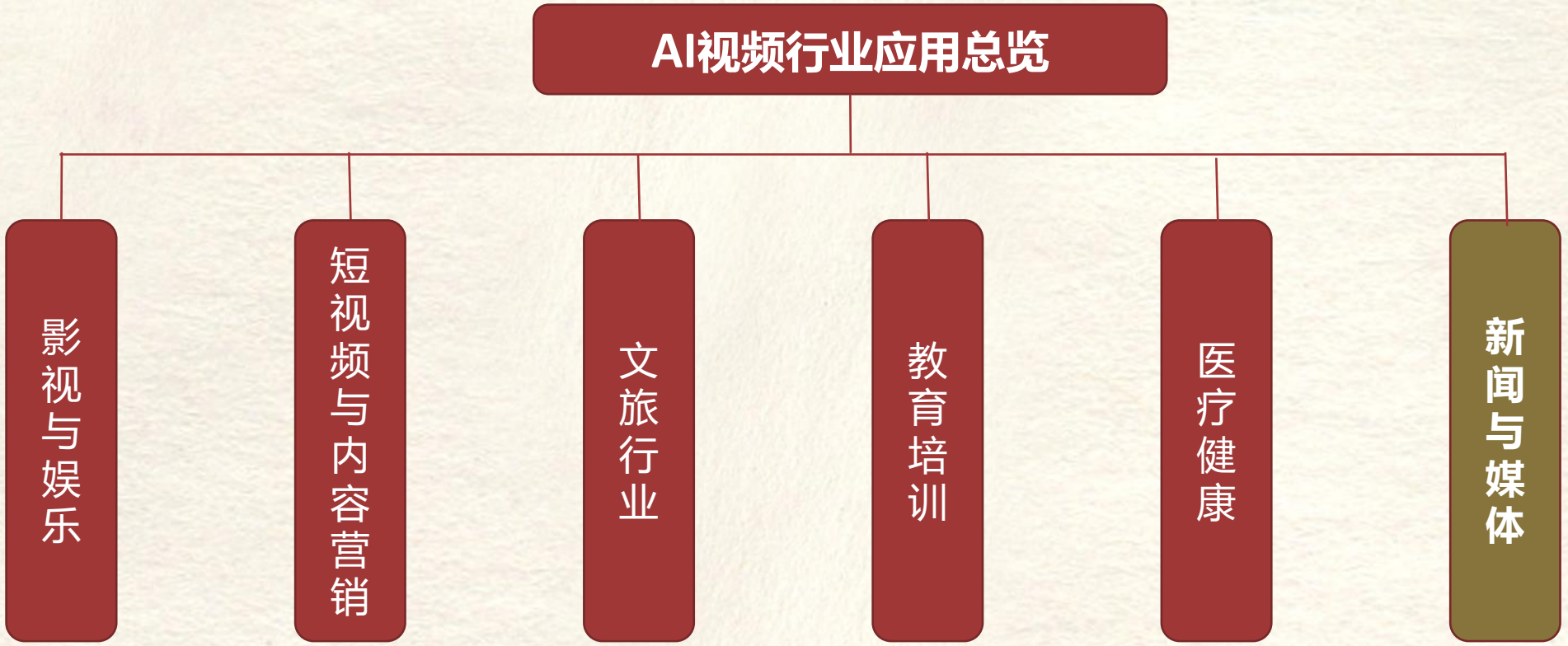


中山市人民医院的新应用服务——AI导诊数字人，7*24小时服务，节省了医院的导诊人力，也受到市民的欢迎。



南阳市中心医院日均门诊量超4000人次，导诊台前常常排起长龙。如今，一位神秘的"数字医生"悄然改变了这一切——通过世优波塔AI数字人" 诊前导诊 - 诊中陪诊 - 诊后管理 "全周期服务，实现了“永不离线的智能守护”。

AI视频行业应用：从创意辅助到全流程重构



新闻和媒体行业的两个重要趋势

1. 视频成为新闻消费的主导力量

- **新闻消费正加速从传统媒体（电视、纸媒、新闻网站）转向社交媒体和视频平台。**全球通过社交媒体观看新闻视频的比例从2020年的52%增长到2025年的预计的65%。观看任何形式新闻视频的比例从67%增长到75%。推动了新闻向视频化和个性化创作者的转变。
- **“视频播客”迅速兴起，**播客不再局限于纯音频。许多热门的新闻类播客现在都会进行视频录制，并通过YouTube和抖音、视频号等视频平台分发。
- **以视频作为传播媒介的影响力人物/创作者生态系统的崛起：**以YouTube博主、TikTok创作者和播客主持人为代表的“影响力人物”和“个性化创作者”正在成为重要的新闻来源，他们主要通过视频来传播观点和信息，对传统新闻机构构成巨大挑战。

2. AI正在深刻地重塑视频新闻的生产和分发方式

- **AI虚拟主播/生成式记者：**这是AI在视频新闻领域最引人注目的应用。一些亚洲国家的广播公司已经开始使用AI生成的虚拟形象来播报新闻。
- **自动化内容转换（文生视频）：**AI技术使得将传统新闻文本自动转换为视频格式成为可能，这是出版商正在探索以提高效率和吸引年轻受众的方式。
- **AI辅助视频制作与分发：**AI不仅用于前端呈现，也越来越多地被用于视频制作的幕后流程。如自动化脚本与摘要、个性化推荐等

以上来自《路透社数字新闻报告2025》

AI视频生成技术在新闻媒体中应用场景和价值

北京大学 PEKING UNIVERSITY

效率提升与成本优化	**自动化新闻快讯**：快速生成突发新闻短视频，抢占社交媒体先机 **图文内容视频化**：自动将文章转为视频格式，提取核心观点，配合影像和AI配音 **社交媒体内容量产**：为不同平台快速生成多版本视频内容
内容创新与格式拓展	**AI数字人/虚拟主播**：24/7不间断新闻播报、多语言新闻服务或特定栏目 **可视化解释性新闻**：将复杂概念（如原理）通过视觉方式具象化，更易懂 **历史影像再现**：制作历史纪录片或回顾性报道时，将残缺历史资料生成动态场景视频
个性化与可及性	**个性化视频新闻简报**：用户可以像定制文字简报一样，定制自己的“每日视频新闻”。 **增强新闻可及性**：如为特殊群体提供AI手语翻译和增强字幕服务

新闻媒体行业的垂直大模型

新闻媒体行业纷纷打造独属的垂直大模型，其中自动生成视频新闻稿件是各个模型的核心功能之一。



央视听媒体大模型： 具备了强大的视频理解能力和视听媒体问答能力和通用的生成能力，可通过文本直接生成视频。同时具备快速生成“数字人主播”的能力，使用较短的真人采集视频即可生成对应的数字人，不仅能根据既定文案和背景场景快速生成播报视频，还可自动学习真人的语言及动作习惯，做到形象更逼真，表情更自然。

芒果大模型：广电行业首批通过中央网信办备案的大模型，以其为基础的“芒果智能体平台”孵化70余款应用，技术深度赋能内容创制全流程。内容生产层面，专注于AI视频生成的“山海研究院”使虚拟拍摄技术成本下降40%；





第一财经星翼大模型的“星翼·财经智能生产套件”实现财经内容生产的智能化，可自主识别、提炼输入内容的核心观点，即时生成金句、海报，适配多渠道传播特点的标题，同步完成文稿和短视频的内容创作并发布。

河南广播电视台“大象元”AI创作和应用平台
致力于探索AI技术在内容创作和资源生成方面的应用。目前已经存储了数千个高精度的、 以“东方元素+国潮国风”为风格的三维模型，还兼容了VR、AR、XR等新兴媒介的三维素材需求， 形成了一个内容丰富、易于检索的数字资产库。



场景1：新闻虚拟主播

从文本、语音到视频，人工智能正推动新闻叙事形式多元演进。以AI合成主播为代表的虚拟播报是近年的热点应用。通过语音合成、面部表情识别、动作捕捉等技术，AI主播可模拟专业播音员的形象与声音，实现全天候滚动播报



2022年“AI王冠”在央视5G新媒体平台亮相时，观众们对这位超写实AI主播与人类原型——首席时政记者王冠主播惊人的相似度感到震惊与敬畏。从“AI王冠”自然逼真的声音，到栩栩如生的动作和外貌，这位全球首个超写实视频AI主播堪称媒体创新领域的一个重要里程碑。



India Today AI 主播 Sana，获 INMA 2024“AI 驱动新闻间转型”大奖。已经进入晚间 9 点黄金时段，具多语言能力，常态化承担天气、股市、星座等“快周转、强结构”的板块，而人类主播聚焦深度与现场报道；



2025年春节期间，《 杭州新闻联播 》于2025年春节期间首次采用六位 数字人主播进行新闻播报，六位数字主播具备高度逼真的形象和表情动作，以双播模式完成新闻播报，实现了零失误播报。

场景2：AI生成沉浸新闻叙事

AI视频技术通过智能生成和重构无法实拍的场景，打破了传统影像的局限，让沉浸式新闻能以更强的视觉冲击力和情感共鸣将观众“置于”事件核心。。





《纽约时报》的 VR 新闻《The Displaced》，通过 AI 优化场景渲染，让用户 “走进” 难民儿童的日常生活空间，发布后获得超 300 万次体验，是早期 “沉浸式新闻的里程碑”；



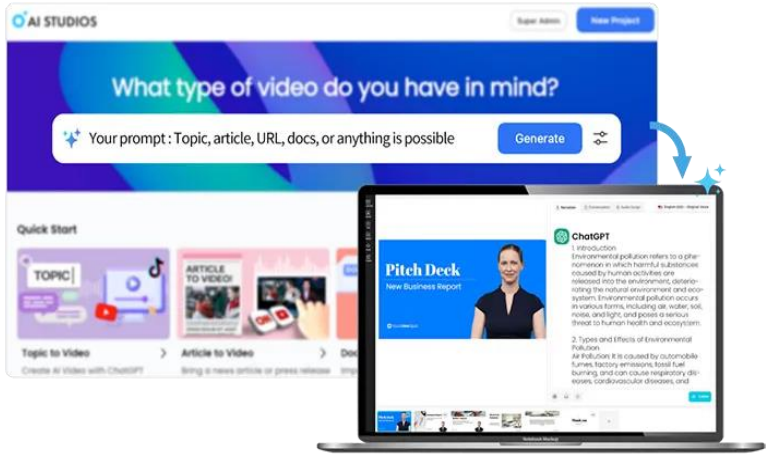
短剧《AIGC｜追光之旅！沉浸式对话中轴线上的“神秘嘉宾”》用AIGC技术对中轴线进行全景展示，让观者从这条“壮美秩序”中对北京城市构造尽收眼底，了解到城市规划中蕴含的价值取向和设计、建设的中的匠心独运。



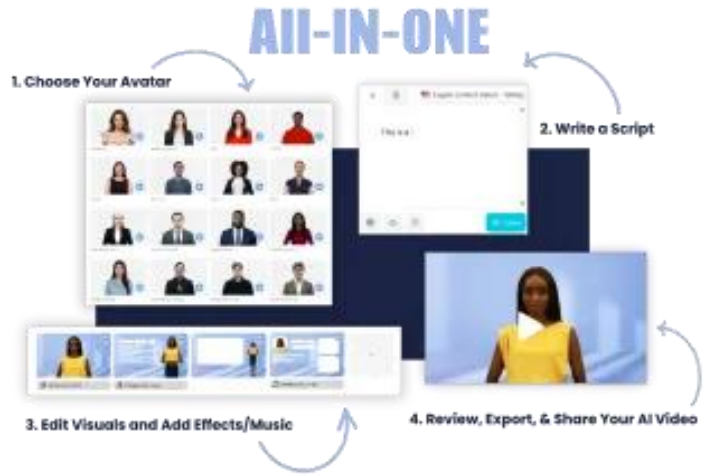
清华大学新闻与传播学院新媒体研究中心与新华社AIGC应用创新工作室合作《AIGC说真相｜“打劫”叙亚，美国盗抢成瘾》。使用AIGC技术生成叙利亚新闻动态视频，创新新闻生产方式，提高新闻报道效率，成为央媒报道首创。

场景3：使用DeepBrain AI制作人工智能新闻主播视频

DeepBrain AI是AI新闻视频生成领域较为知名的供应商，央视“AI王冠”和北京卫视“春妮”数字人均由其合作完成。其AI Studios产品的AI新闻记者生成器，可以快速创建突发新闻片段、最新消息和报道。







01 选择来源

粘贴一个网址或复制粘贴一篇新闻文章，几分钟内即可即时创建一个可用于广播的视频，包括脚本、镜头和人工智能虚拟人旁白。

02 选择模版

AI Studios 拥有多种新闻模板，非常便捷，直接使用专业设计的新闻媒体模板，可快速适用于多个平台，如电视、YouTube、TikTok等。

03 定制主播

AI Studios提供2000多个AI虚拟形象，涵盖不同种族、年龄和风格。逼真的人工智能模型，能够直接根据文本说话。用户还可以创建自定义虚拟形象。

04 定义参数和语言

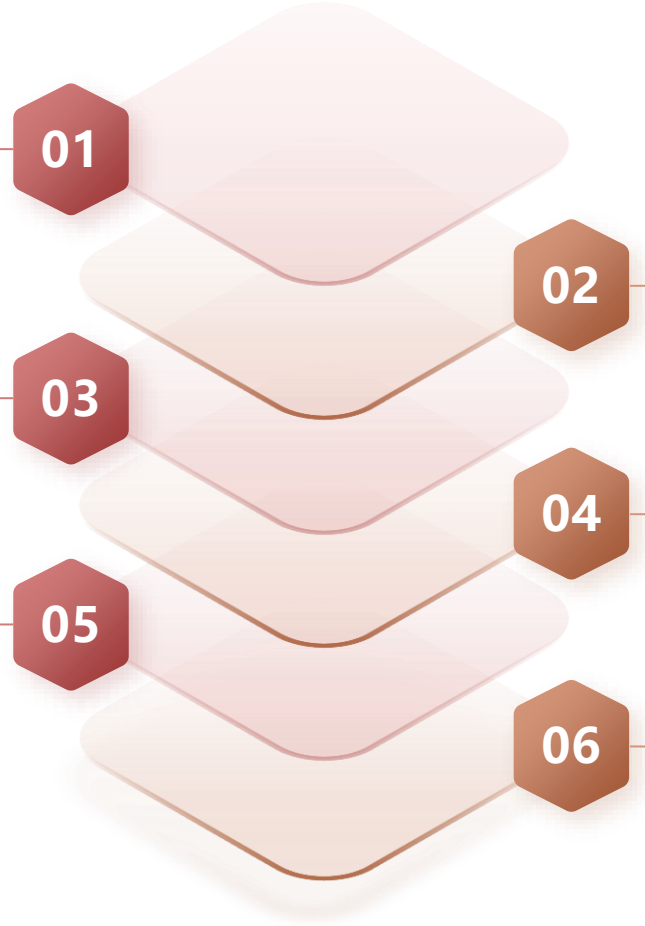
AI Studios支持定义视频时长、图片和镜头的来源、语言。会自动在来源识别关键词，如体育新闻中的“巴萨”、“曼联”、“2009年”、等，并自动搜索匹配图片。

05 预览和编辑、下载

点击“生成”，可以预览视频实时生成，一般生成时间在几分钟内，可以同时进行编辑。

媒体+AI：发展热潮下的冷思考

AI视频在新闻业落地尚处爬坡期，面临质量可控与可信度、合规与版权、算力成本、流程系统对接及人才机制等多维挑战，需技术打磨、治理规范与组织变革协同推进方可规模化应用。



01

生成质量与可控性不足

穿模、跳帧、口型不同步，复杂镜头难精确控制

02

真实性与信任风险

深度伪造、幻觉与偏见，溯源/水印鲁棒性不足，易误导受众

03

规范与合规缺口

AIGC标识、使用边界、审核流程与取证标准尚未统一

04

版权与权利链不清

训练数据、风格挪用、肖像/声音人格权与生成物权属存在不确定性

05

成本与算力压力

高分辨率/高帧率生成、存储分发与实时性要求高，供应商锁定与技术迭代不确定

06

流程与系统对接难

与采编/剪辑/播控对接不顺，元数据、时间码、质检与归档缺标准化

「AI视频行业应用现状」要点总结

北京大学 PEKING UNIVERSITY

从“工具”到“引擎”

AI视频已超越了单纯的“效率工具”范畴，正演变为驱动内容生产全流程革新的核心引擎。其根本价值在于“降本增效”与“创意赋能”，正系统性地改变着行业的成本结构与创新模式。

行业应用重点

在影视、营销、媒体等内容密集型行业，AI的应用已进入“全链条改造”阶段，催生了AI短剧、AIGC广告等新业态。在教育、文旅、医疗等垂直领域，AI则通过生成个性化、可视化、沉浸式的视频内容，极大地提升了知识传播的效率和用户体验的深度。

落地成熟度

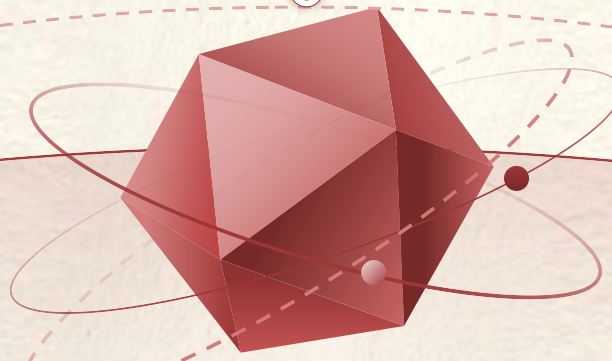
后期制作（如智能剪辑、VFX、AI配音本地化）及素材混剪与多平台适配已实现规模化应用；数字人与自动化剪辑技术也已广泛落地。相比之下，端到端的全流程AI生成虽有突破性案例，但尚未普及，仍处于探索阶段。

瓶颈与风险

在技术飞速发展的热潮下，我们也必须清醒地认识到其面临的挑战。生成质量的可控性、内容的真实性与伦理风险、版权归属的明确性、以及产业集成的现实障碍，是决定其能否健康、可持续发展的关键。

未来展望

未来，AI视频的规模化应用需要技术打磨、治理规范与组织变革的协同推进。它将不再是孤立的技术，而是会深度融入产业生态，成为数字内容时代不可或缺的基础设施，最终实现与人类创作者的高效协同与价值共生。



影视娱乐
影视制作+批量剪辑+虚拟制片

营销短视频
品牌创意+混剪批量化+A/B测试闭环

新闻/资讯
虚拟主播+文图转视频+多平台分发

教育培训
微课/课件批量化+多语本地化+数字人讲师

文化旅游
城市/景区宣传片+AI数字人IP+沉浸式XR联动

医疗健康
患者教育/知情同意+导诊客服数字人+教学仿真



PART 05

平台/工具选用推荐

1. 主流AI视频工具/平台介绍
2. 平台/工具选用建议

主流AI视频平台/工具介绍

结合前面的模型能力和评估排名、用户访问情况以及工具本身的生态和定位，日常可关注以下AI视频平台：

国内/国际	工具名称	推荐理由
国内	可灵 AI	全球第一梯队。具备极强的物理规律模拟能力，擅长写实风格与长视频续写。
	即梦 AI	无缝衔接短视频生态。拥有海量特效模板，生成速度快，性价比高。
	海螺 AI	图生视频能力卓越。语义理解强，能将2D插画完美转化为电影级动态视频，动漫感极佳。
	Vidu	独创多主体一致性技术。对中国元素理解深刻，适合处理复杂场景的叙事。
	Pixverse	创意表现力强。内置海量动态特效，支持多风格切换及音效生成，海外表现亮眼。
国际	Sora	现象级平台，行业技术标杆，拥有顶尖的物理世界模拟能力与长视频连贯性。
	Runway	专业影视后期首选。提供精细的“导演级”运镜控制与强大的视频编辑工具链。
	Pika	社交媒体“爆款机”。以独家的病毒式特效（如融化、爆炸）和完美循环视频见长。
	Veo	生态集成。在生成质量、速度与成本间取得最佳平衡，支持超长叙事与4K画质。
	Luma AI	3D与特效专家。擅长生成大幅度运镜、产品展示及具有电影级光影效果的视频。

可灵 AI

厂商背景： 快手自研

核心优势： 全球视频生成第一梯队，本土适配率高。物理规律模拟真实，功能多，支持1080P直出，最长可达2分钟（续写）。

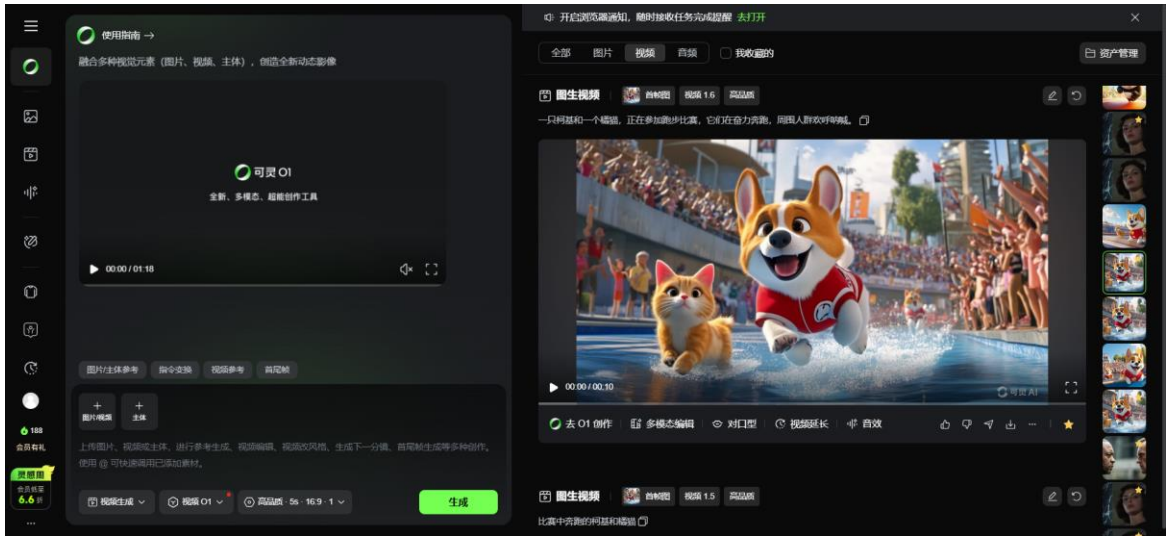
特点：

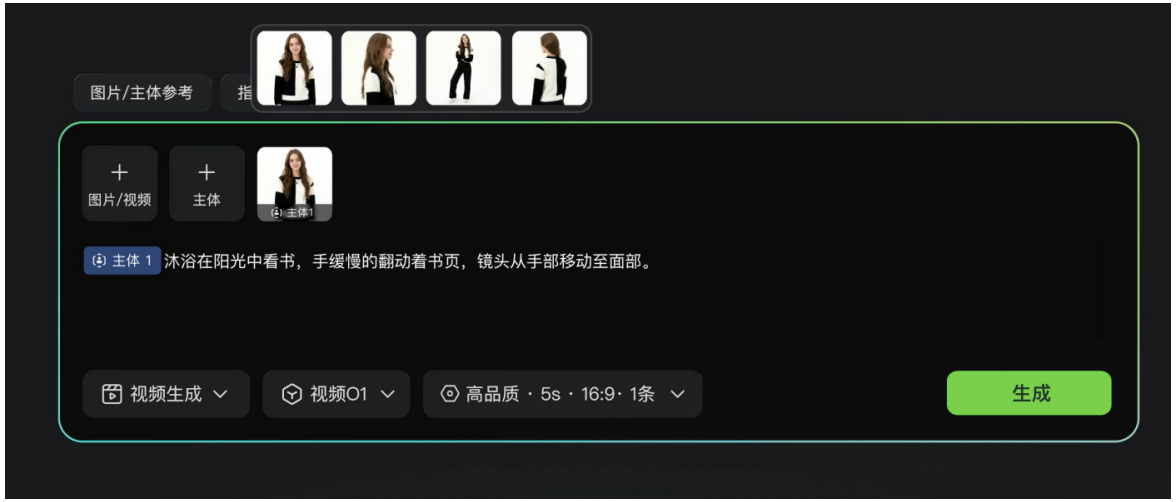
- 文生视频、图生视频、多图参考。支持首尾帧控制、运镜控制（平移/倾斜/缩放）
- O1模型实现在单一输入框内无缝融合文字、图像、视频等多模态输入。
- 支持动态画笔，可精确控制局部运动。
- 电影级画质，写实风格极强，适合剧情片和真实世界模拟。
- 支持音效生成、支持口型同步
- DeepSeek反推提示词、提示词优化

价格：

- 免费额度：登录每月送166积分（约11个5s标准视频）。
- 会员特权：并发和速度、去水印、高品质视频、视频延长、新O1模型。

适用场景： 剧情短片、广告二创、电商带货预热、专业影视制作。





即梦 AI

厂商： 字节跳动

核心优势： 与抖音生态无缝衔接，速度快，模板特效极其丰富，价格亲民，适合预算有限的创作者。

特点：

- 核心功能：文生视频、图生视频、支持首尾帧控制、主体参考、智能多帧，高级运镜模板、支持1080P。
- Agent模式支持自然语言交互、需求拆解、任务执行。
- 支持数字人功能、支持动作模仿、音效生成、口型同步
- 支持智能画布功能辅助生成视频分镜、故事创作模式
- 画面美学高，色彩饱满，结构稳定。且风格多样，涵盖写实、动漫、3D等适合短视频平台的主流风格。

价格：

- 免费额度：每日60积分（约10个5s标准视频）。
- 会员特权：速度、去水印、高品质视频、对口型、视频补帧。
- 提供1元7天会员试用。

适用场景： 抖音短视频、信息流广告、节日营销、电商产品展示。





海螺 AI

厂商： MiniMax

核心优势： 国内图生视频能力极强，能将2D插画转化为动画电影级效果。语义理解能力优秀。

特点：

- 核心功能：文生视频、图生视频、支持首尾帧控制、主体参考、自然语言控制运镜。
- Agent自动规划剧本，完成剪辑和配乐等工作，生成一个完整视频
- 导演级控制：可在特定时间点进行平移、倾斜、缩放。
- 动画电影级效果，人物表情和肢体动作非常自然。在二次元、插画转动态视频方面表现卓越。
- 分辨率上限：目前多为720p/1080p，视频时长6-10秒。

价格：

- 免费额度：无，可单独购买积分，价格高，6S768P视频约1.25元/个，且权益受限
- 会员特权：生成速度、多任务、去水印、高品质视频，单价降低（右图）。

适用场景： 动漫/插画动态化、表情包制作、人物口播视频。





Vidu

厂商： 生数科技 & 清华大学

核心优势： 独创多主体一致性功能。主打长视频与中国元素，技术硬核。

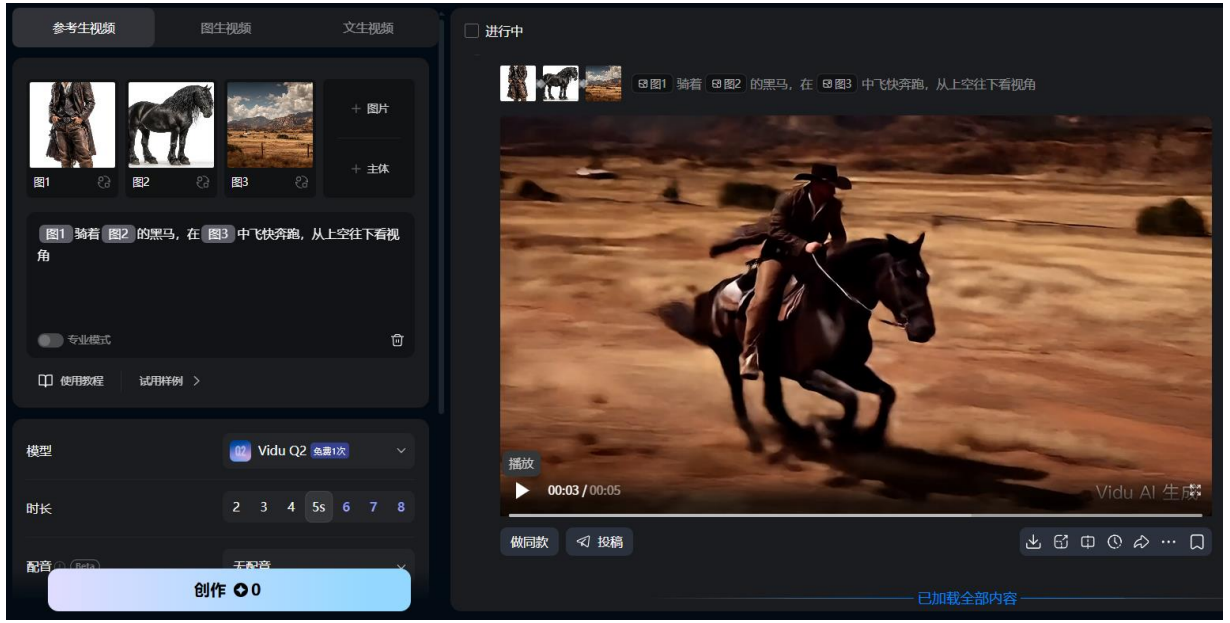
特点：

- 核心功能：文生视频、图生视频、多帧生图控制、主体参考、推荐运镜。
- 1080P高清，动态范围大，光影物理模拟优秀。
- 多主体一致性表现优秀，适合复杂场景
- 风格综合，对龙、熊猫等中国文化元素理解深刻。
- 相对其他平台，生成速度较快
- 新手友好：新手引导、提示词示例、平台手册

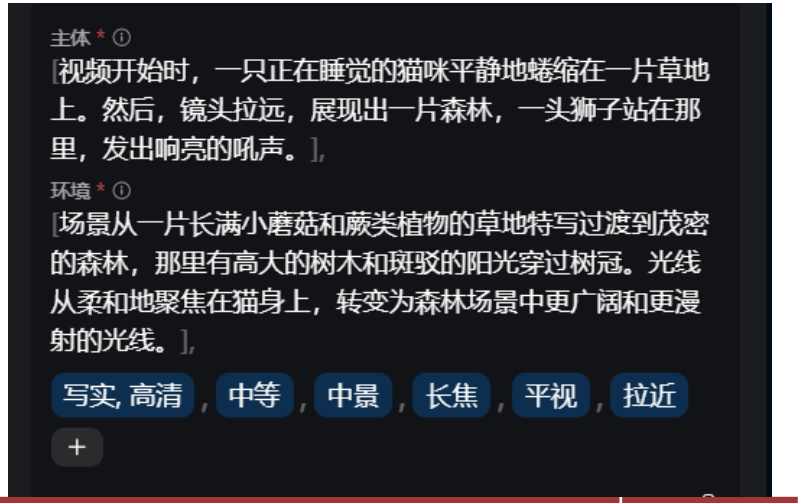
价格：

- 免费额度：有限免次数，登录、做任务可得积分
- 会员特权：长视频，高清画质、快速通道、可商用、多任务、单价低。
- 基础版会员：48元/月，1.2元/5S（1080P）

适用场景： 教育课程、企业宣传、包含复杂多角色的视频。







Pixverse

厂商：爱诗科技

核心优势：功能丰富、多参数灵活配置，细节细腻，海外市场表现亮眼，创意表现力强，自带音效与配音生成。

特点：

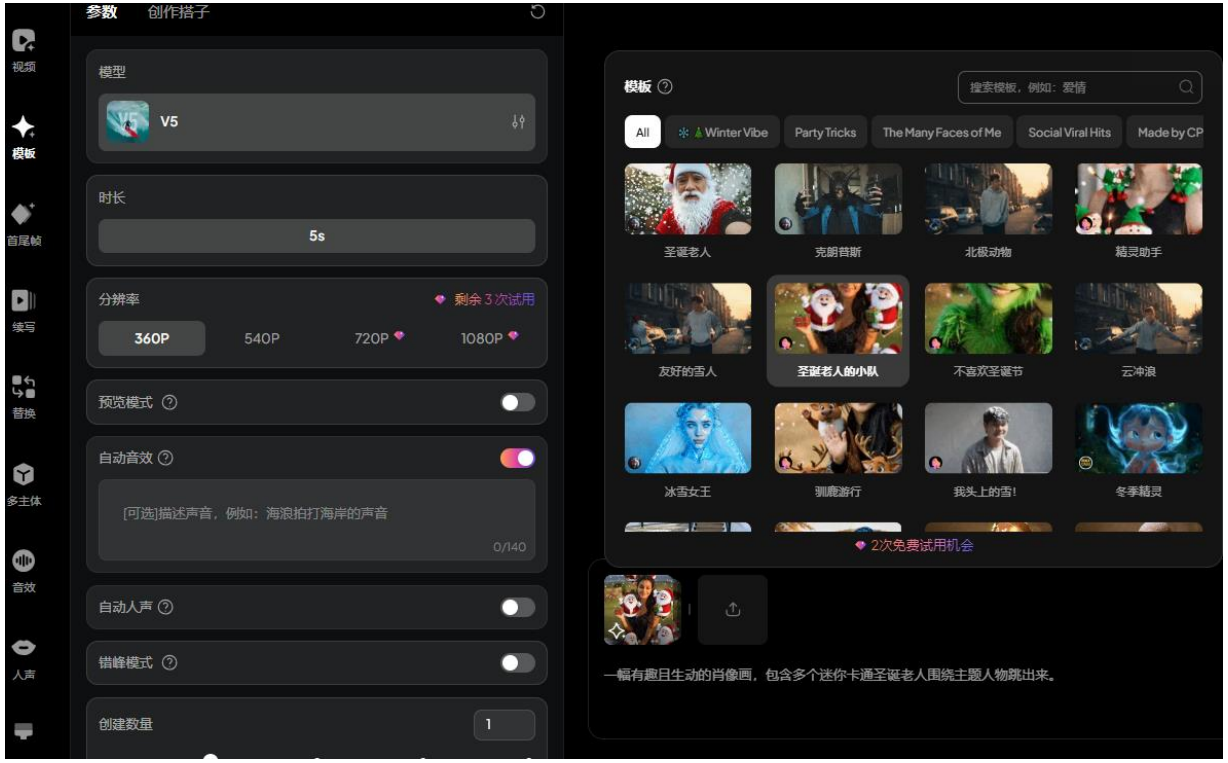
- 核心功能：文生视频、图生视频、首尾帧、多主体生成、续写、局部替换、多镜头控制、运动笔刷、多模态输入、Agent 创作助手。
- 创意表现有独特优势，内置了300 + 动态特效模板
- 自带音效还原与人声配音，口型同步免后期
- 多比例、多风格（写实/卡通/动漫）一键切换。
- 支持预览模式，低分辨率抽卡，然后再进行4K超清处理
- 支持视频延长、视频拼接等高级编辑功能

价格：

- 免费额度：免费次数、登录送积分，做任务领积分
- 会员特权：高清画质、快速通道、独家特效、多任务、单价低。
- 基础版会员：$7.9元/月，约60个视频

适用场景：广告分镜、创意视频、MV制作、动漫创作、产品发布视觉稿、多风格内容批量生成。





Sora

厂商： OpenAI

核心优势： 首个真正理解物理、运动和电影构图的模型，重力、碰撞、流体动力学模拟行业顶尖。

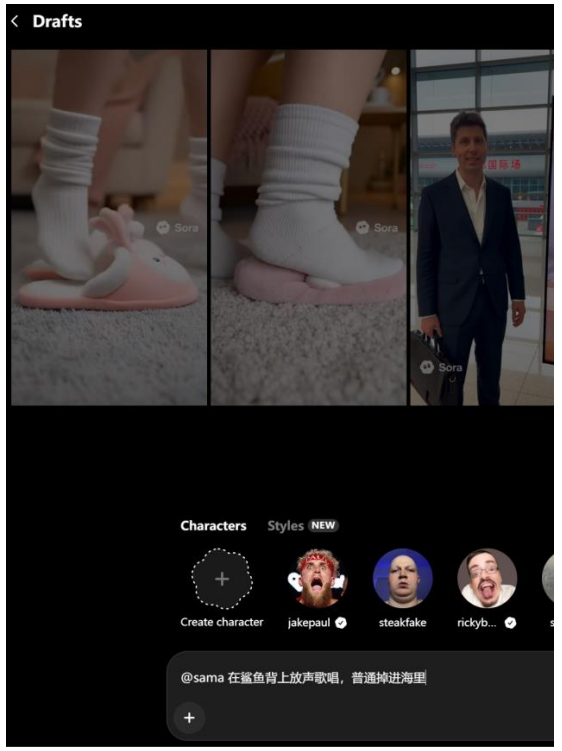
特点：

- 优秀的物理模拟能力，如重力与运动规律、复杂物理场景、失败场景的合理建模等。
- 支持跨镜头复杂指令，保持"世界状态"一致性，适配长视频生成
- 多种视觉风格上都表现出色，包括现实主义、电影质感和动漫风格。
- Cameo功能：允许用户嵌入真人影像，实现虚拟与现实无缝融合
- 同时生成画面、环境音、物体交互声及对话，真实感远超传统AI视频

价格：

采用与ChatGPT统一的账号系统，每天30个免费视频生成额度，之后的价格较贵，见右图。

适用场景： 顶尖创意工作室、高预算影视项目、概念片。







Runway

厂商： Runway ML

核心优势：老牌AI视频平台，综合创作能力和可控性获全球用户认可。产品精雕细琢，专业影视后期首选。

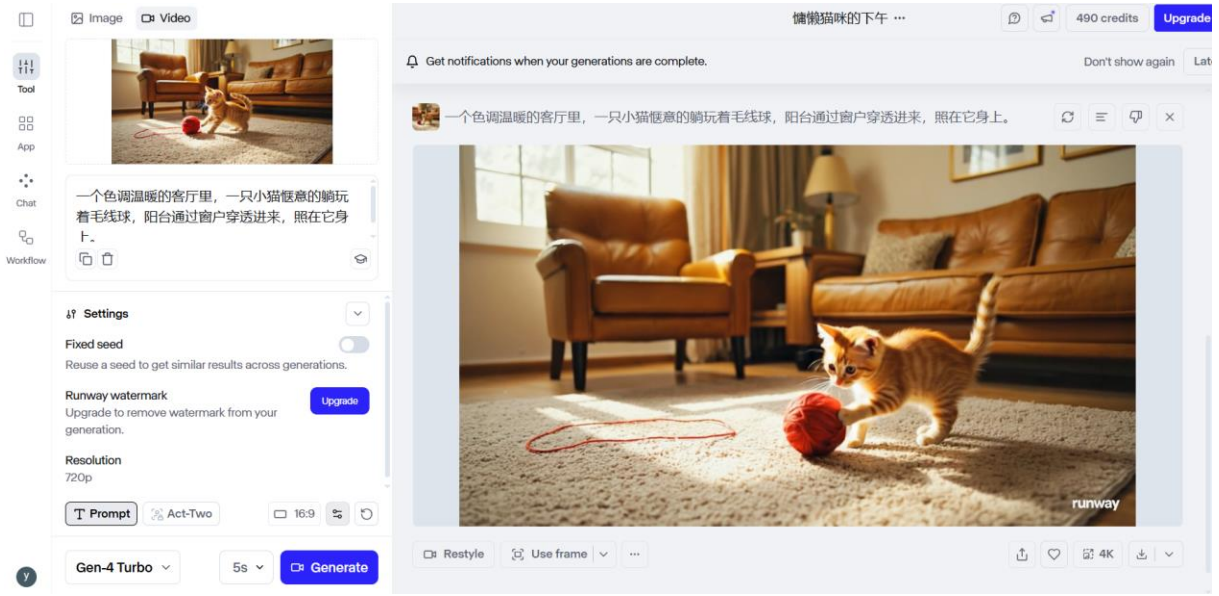
特点：

- 功能强大：文本生成视频、图像生成视频、视频编辑、风格转换、运动笔刷、一键抠像、绿幕去除、图像修复等
- 内置时间线编辑器，含动画、转场效果及视觉特效，实现一体化工作流程。
- 导演模式：精准控制镜头推拉摇移。
- Gen-4.5动态质量、物理一致性、角色外观稳定性表现俱佳，尤其擅长 电影级写实风格；
- 具备跨场景角色一致性、4K放大功能
- 集成了Veo 3.1模型，支持带音效视频生成。

价格：

- 按秒计费，注册送125积分，可生成25秒Gen-4 Turbo视频。会员最低$15/月，含625积分。不同模型消耗积分不同（右图）。

适用场景：专业广告片、电影后期特效、需要精准控制画面的项目、专业视频编辑。



Video Generation	*per second*
Gen-4 Aleph	15
Gen-4 Video	12
Gen-4 Video Turbo	5
Act-Two (3 second minimum)	5
Veo 3 or 3.1 Video with Audio	40
Gen-3 Alpha Video	10
Gen-3 Alpha Video Turbo	5

Pika—社交媒体的宠儿

厂商： Pika Labs

核心优势： Pikaffects特效火爆社交媒体，超逼真生成效果和增强物理表现，在艺术创作者中口碑良好

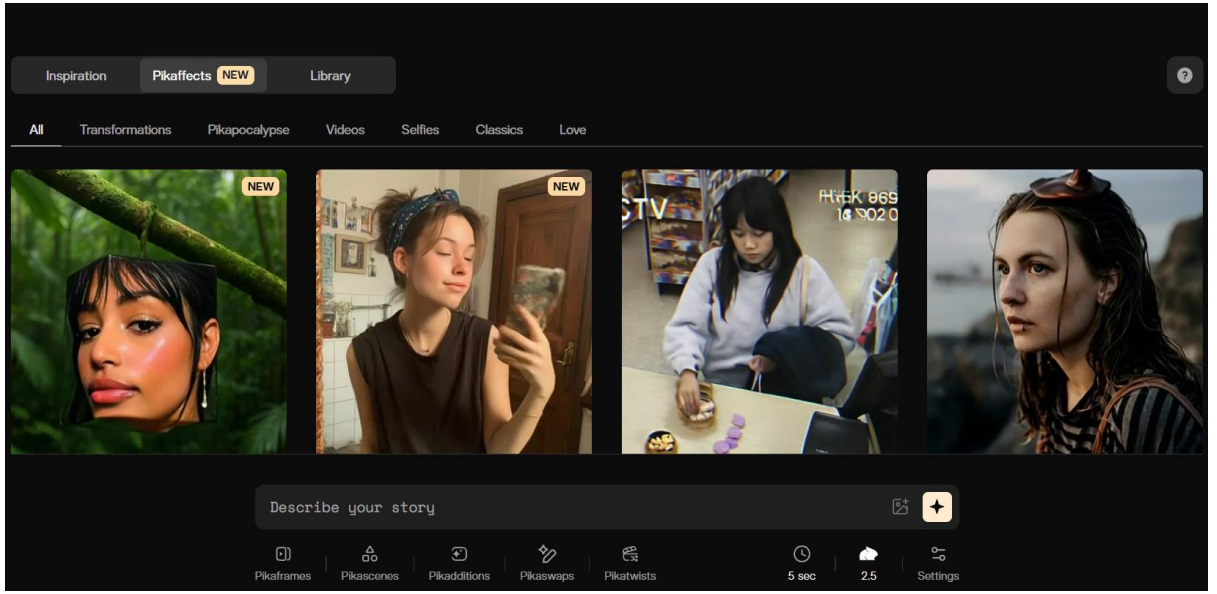
特点：

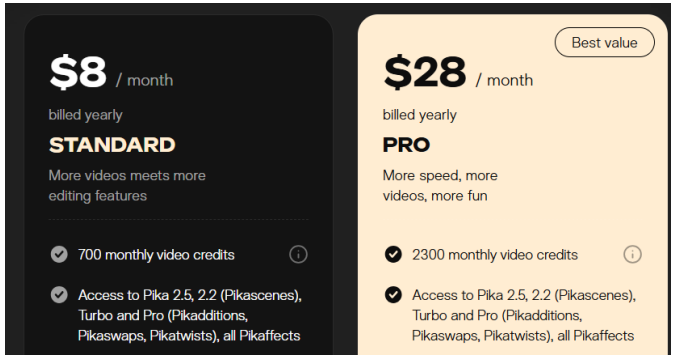
- Pikaffects（特效）： 独家的“融化”、“爆炸”、“蛋糕化”特效，自带病毒传播属性。
- 非常擅长制作3-4秒的完美循环视频，适合做背景或表情包。
- 支持将现有视频转换为完全不同的风格，如黑白风、动漫风、自然风等；
- 支持AI视频编辑 和视频实时重绘。
- 与X、TikTok、Discord、Youtube等无缝对接。

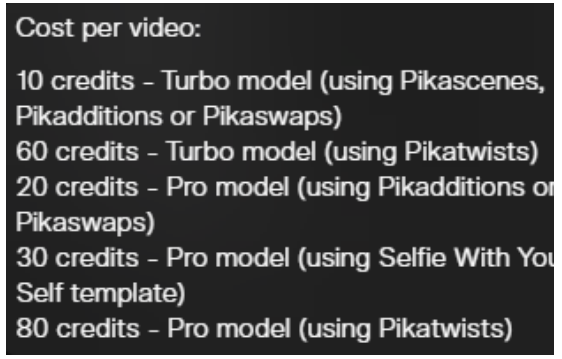
价格：

- 每月赠送80积分，免费会员功能受限
- 最低会员￥10/月，含700积分，不同模型消耗积分不同（右图）。

适用场景： 创意短循环视频、社交媒体特效视频、风格化动漫、概念设计。







Veo—质量、速度和价格的最佳平衡

厂商： Google DeepMind

核心优势： 人物一致性好，电影感稍逊于Sora，在速度、成本和稳定性上达到了商业最佳平衡，尤其在生成有人物的视频时大放异彩。

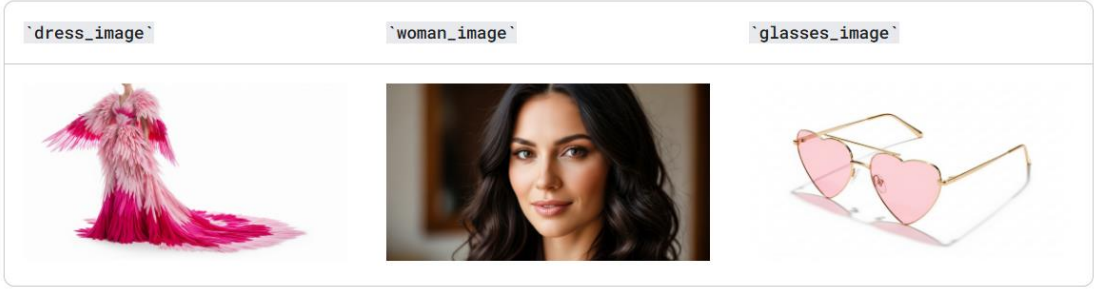
特点：

- 集成在Google生态（Gemini、flow）中使用
- 在音频生成、画面编辑和真实感方面具有优势
- 沉浸式叙事新体验：仅需文字提示即可自动生成带对白、唇形同步和音效的1080P高保真视频
- 理解长提示词、视频理解像素级、渲染质量高
- 可延续现有视频的叙事（长达141秒），并保持视觉风格与背景音频的连贯性。

价格（Gemini API计费）：

- 免费版：每月10个视频（480p，带水印）； 标准版：每月30美元（100个视频，1080p）； 高级版：每月90美元（500个视频，4K画质）。

适用场景： 企业营销视频、专业影视制作、高端人物场景广告、需要大量迭代和精細调整场景。



以一个中景镜头开场，镜头平视一位拥有深色秀发和温暖棕色眼眸的美丽女子。她身着一件华丽的高级时装火烈鸟礼服，裙摆由层层叠叠的粉色和紫红色羽毛组成，搭配一副俏皮的粉色心形太阳镜。她神态自若地漫步在阳光普照的泻湖中，湖水清澈碧绿，水面波光粼粼。



Luma AI—3D空间与视觉特效专家

厂商： Luma AI

核心优势： 较早进入AI视频领域，最早实现首尾帧功能，已服务超过 3000 万用户，具有广泛的市场基础，新推出的Ray 3模型表现优异。

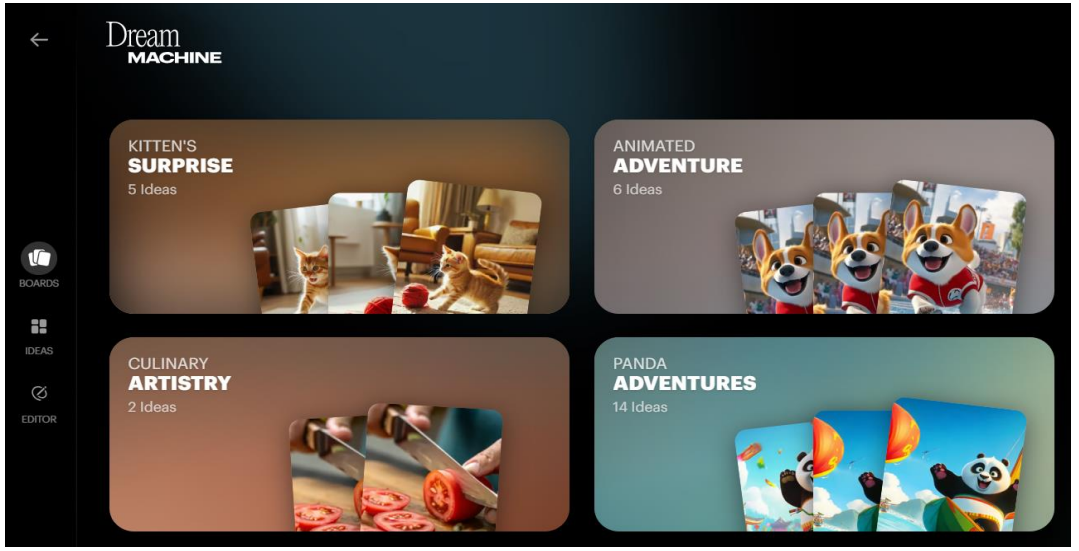
特点：

- 在3D创作者社区有影响力，擅长生成产品旋转、空间镜头、视差动画。
- 新发布的Ray模型是全球首个推出推理视频模型的 AI 公司，具备理解意图、评估输出、即时优化的推理能力，在多项基准测试中表现优秀
- 原生HDR生成：首个支持10-16 位 HDR 视频生成的 AI 模型，原生 1080p分辨率，支持4K超分辨率，达到电影级视觉效果
- 用于迭代的草稿模式，可以快速生成草稿，然后选择喜欢的进行“精修”，得到高分辨率、全质量的输出。

价格：

- 免费额度：每月8个草稿模式视频，等待时间长
- 最低会员$9.99/月，可生成50个草稿视频。

适用场景： 3D内容创作、产品展示、建筑可视化，Ray3的电影级视觉效果和推理能力适合影视制作、广告创意。



Free	Lite	Plus	Unlimited
$0 /month	$9.99 /month	$29.99 /month	$94.99 /month
8 Videos (draft mode)	50 Videos (draft mode)	160 Videos (draft mode)	
Select	Select	Select	Select
Limited usage	3,200 monthly credits	10,000 monthly credits	10,000 monthly credits
Draft resolution	Full Ray3 access	Full Ray3 access	Unlimited use in Relaxed Mode
Lower priority processing	4k with up-res	4k with up-res and HDR	Full Ray3 access
Non-commercial use only	High priority processing	High priority processing	4k with up-res and HDR
Watermarks	Non-commercial use only	Commercial use allowed	High priority processing
	Watermarks	No watermark	Commercial use allowed
			No watermark



PART 05

平台/工具选用推荐

1. 主流AI视频平台/工具介绍
2. 平台/工具选用建议

工具选用建议—按使用场景

应用场景/需求	首选推荐工具	推荐理由
专业影视/剧情短片	Runway、可灵 AI	需要极高的画质、写实度以及对镜头（推拉摇移）的精准控制；Runway适合精细后期，可灵擅长写实生成。
短视频/社交媒体运营	即梦 AI、Pika	需要快速出片、丰富的模板特效和病毒传播属性；即梦性价比高适合高频更新，Pika特效吸睛。
动漫/插画动态化	海螺 AI、Pixverse	对二次元、插画风格的还原度要求高；海螺AI在图生视频上的动态演绎非常自然，Pixverse风格切换灵活。
电商/产品展示	Luma AI、即梦 AI	需要展示产品的3D空间感或快速生成带货视频；Luma擅长物体旋转与空间运镜。
中国风/传统文化内容	Vidu、国内工具	对龙、熊猫等中国元素理解深刻，且多主体一致性好，适合制作包含复杂角色的国风故事。
企业宣传/长视频叙事	Veo、Sora 、可灵AI	需要生成较长时长的连贯内容；Veo支持长达141秒叙事且音画同步，Sora物理模拟最强但成本高。
MV制作/创意视频	Pixverse、Runway	重视创意特效与音乐的结合；Pixverse自带音效且创意模板多，Runway支持复杂的风格化转换。

工具选用建议—按易用性与访问门槛

级别分类	包含工具	核心特点与门槛	适用人群
第一梯队 (国内直连)	**即梦 AI 海螺 AI 可灵 AI Vidu**	【零门槛】 1. 网络： 国内直连，无需特殊网络环境 2. 语言： 全中文界面与提示词支持。 3. 支付： 支持微信/支付宝。 4. 账号： 手机号/微信直接注册。	所有国内用户 (小白、运营、企业团队)
第二梯队 (海外独立站)	**Runway Pika Luma AI Pixverse**	【进阶门槛】 1. 网络： 需科学上网环境。 2. 语言： 英文界面（需配合翻译工具）。 3. 支付： 通常需外币信用卡（Visa/Mastercard）。 4. 功能： 不排斥复杂一些的参数设定和操作。	专业创作者 (拥有海外网络条件与支付能力的设计师/后期)
第三梯队 (生态受限)	**Sora Veo**	【较高门槛】 1. 形态： 非独立应用，深度绑定巨头生态或仅提供API。 2. 风控： 对IP和账号区域有极其严格的封锁。	技术极客 / 开发者 (仅关注技术动态，无需尝试落地)

「AI视频工具选取」要点总结

拒绝“工具唯粉”，构建多模型协作的工作流

没有任何一个平台能完美覆盖所有需求，真正的高手懂得利用多个工具来取长补短，以组合拳代替单点生成。

场景定义工具，而非技术定义创作

抛弃对技术的盲目追求，“适合”才是最高标准。比如：做二次元选海螺，做写实选可灵，做特效选即梦，精准匹配不仅为了省钱，更是为了极致的效率。

不盲目崇拜国外的工具/模型

国产第一梯队（可灵/即梦）的能力已经处于领先水平，且带来的网络直连、语义理解优势，比Sora的理论高画质更能快速创造商业价值。

驾驭工具的核心在于“导演思维”

软件迭代极快，今天的主流明天可能被淘汰。不要沉迷于按钮操作，掌握镜头语言、分镜构思与提示词逻辑，才是跨越工具周期、不被AI取代的核心竞争力。

保持敏锐，多跟进多尝试

AI视频领域技术更迭以“周”为单位，绝大多数平台都提供免费额度。建议创作者多跟进、多尝试、多对比，第一时间捕捉新模型带来的能力，有助于自己的选择和判断。

结束语

北京大学 PEKING UNIVERSITY

“

技术没有终点，只有不断迭代的里程碑。AI视频技术的发展，最终将指向“人机共创”的新范式。**它不会简单地替代人类创作者，而是将人类从繁琐的重复劳动中解放出来，去专注于更具价值的创意构思与情感表达**。

现在的AI，是AIGC和Agent的阶段，下一步将会是端侧智能、硬件智能、具身智能。

现在，还有使用AI的组织、使用AI的个体；将来所有的组织和个人都将基于AI、与AI融合，AI将成为水电一样基础设施，正如今天的互联网技术一样。

”