信达证券 CINDA SECURITIES

Research and Development Center

生成式 AI+视频行业深度报告：

AI+视频的星辰大海远不止于创意视频的生成

证券研究报告

行业研究

行业专题研究（深度）

AI 行业

投资评级 看好

上次评级 看好

冯翠婷 传媒互联网及海外 首席分析师
执业编号：S1500522010001
联系电话：17317141123
邮 箱：fengcuiting@cindasc.com

信达证券股份有限公司
CINDA SECURITIES CO.,LTD
北京市西城区宣武门西大街甲127号金隅大厦B座
邮编：100053

AI+视频的星辰大海远不止于创意视频的生成

2024 年 07 月 28 日

本期内容提要：

- **站在当下，本报告研究 AI+视频的意义：技术和产品迭代升级较快导致目前市场大多数报告无时效性，且往往缺少对产品的实际测试以及对相同提示词的比较分析，而 AI 视频生成正成为当前 AI 产业发展的关键节点。**视频杂糅了文本、语音、图像等多维度内容，其训练的难点也往往在于视频数据对数量和质量的不足、算法架构需要优化、物理规律性较差等等，但随着 AI+视频的技术和产品升级迭代，众多行业有望受益，诸如电影、广告、视频剪辑、视频流媒体平台、UGC 创作平台、短视频综合平台等，而目前正处在 AI+视频发展的关键性时刻。

- **市场主流的 AI 视频生成技术迭代路径经历了早期的 GAN+VAE、Transformer、Diffusion Model 以及 Sora 采用的 DiT 架构（Transformer+Diffusion），技术迭代升级带来视频处理质量上的飞跃性提升。**VAE 引入了隐变量推断，GAN 生成的图像真实清晰，VAE+GAN 的串联融合可以实现数据的自动生成+高质量图像生成；Transformer 在并处处理、长时间序列数据处理、多注意力处理上有着强大的优势，通过预训练和微调可提高模型性能；扩散模型可解释性强，可生成高质量图像和视频；李飞飞联合谷歌研发的 WALT 视频大模型将图像和视频编码到共享潜在空间中。Sora 采用的 DiT 架构有效进行结合，利用 Transformer 处理潜在空间中的图像数据块，模拟数据的扩散过程以生成时长更长、质量更高的图像和视频。

- **我们认为，国内 AI+视频产品单条价格低于海外产品，其中 Runway Gen-3 Alpha 和快手可灵为目前 AI 视频生成的全球第一梯队，在视频分辨率、生成速度、物体符合物理规律、提示词理解、视频时长等诸多维度上表现均较为优秀。**核心梳理国内和海外市场 AI 视频生成的核心参与者，如海外 Luma AI（Dream Machine）、Runway （Gen 1-2 & Gen-3 Alpha）、Pika、Sora，国内快手可灵、美图、PixVerse、剪映即梦、清华 Vidu、七火山 Etna 等，集中梳理了众多产品的融资历程、产品迭代、核心功能、实测效果比较等多方面。经过我们测算，目前 AI+视频主流产品的单条视频生成价格分别为：Luma AI 0.16 美元（1.17rmb）、Pika0.05 美元（0.364rmb）、Runway 0.48 美元（3.49rmb）、快手可灵 0.5rmb、字节剪映即梦 0.04rmb、爱诗科技 Pixverse V2 为 0.02 美元（0.174rmb）、美图 WHEE 为 0.32rmb，国内 AI+视频产品单条价格较低，质量不差。

- **不止于视频生成，从 AI 生成到 AI 工作流，一站式 AI 视频生成+剪辑+故事创作有望成为产业核心发展方向。目前，AI+视频大多数用于创意内容生成，直接用于 ToB 商业化较少。追溯原因，首先生成视频的人物一致性、所需时长、画面质量尚且不满足立即商业化水准。其次，我们发现目前主流 AI 视频工具还处在视频生成竞争的阶段，且大多数为单一功能产品。在视频生成之后，诸如准确的提示词生成、修改视**

频片段、添加字幕、脚本生成、转场衔接、背景音乐添加等众多细节功能暂未集成，因此现今阶段还需要多种不同的视频创作工具串联使用才能达到直接输出可商业化视频的效果，环节繁琐、多工具之间的格式也可能存在不兼容的可能性，给用户带来使用上的不便。因此我们认为，后续需要持续关注能够一站式提供视频生成+编辑等功能的企业，了解用户痛点，打磨产品细节，才能真正将技术用于生产工作、娱乐等众多环节，带来商业化变现的潜在空间。一站式 AI 视频生成&剪辑&UGC 创作有望解决市场一直在质疑的“AI+视频没有实质作用问题”。

- **AI+视频时代来临，思考哪类公司存在商业化变现的可能性？我们认为，1）一站式平台型公司，如 Adobe、美图公司；2）AI+视频技术头部服务商转型产品类公司，如 Runway、商汤科技；3）视频剪辑类公司，如快手；4）广告营销类公司，如易点天下、蓝色光标、因赛集团、利欧股份；5）UGC 社区类公司，如 Bilibili；6）视频数据类公司，如捷成股份、华策影视、视觉中国、中广天择；7）IP 类公司，如上海电影、阅文集团、汤姆猫、中文在线、果麦文化；8）探索 AI 视频工作流及其他创作方向类公司，如博纳影业、超讯通信、柠萌影视。9）其他建议关注猫眼娱乐、光线传媒、芒果超媒、万达电影等。**

- **风险因素：**AI 底层大模型发展不及预期、AI 视频技术迭代不及预期、AI 视频产品付费渗透率提升不及预期。

目 录

一、生成式 AI 发展进程，文生视频正成为当前 AI 行业关键发展节点 6
二、目前市场主流的海外生成式视频参与者 15
三、目前市场主流的国内生成式视频参与者 28
四、从 AI 生成到 AI 剪辑，一站式 AI 视频生成+编辑有望成为另一核心方向 34
五、 AI+视频发展方向展望 37
六、风险因素 39

表 目 录

表 1：Transformer、Diffusion、DiT 模型的产品梳理 12
表 2：Runway 历年融资轮次、融资金额及对应估值 18
表 3：Luma AI、Pika、Runway Gen-3 Alpha、Sora 相同提示词生成视频的效果多维度比较 27
表 4：海内外视频生成产品单视频所需成本比较（1 美元=7.28 人民币） 28
表 5：Adobe 数字媒体业务和数字体验业务预估市占率 42
表 6：快影和剪映产品相关数据 46
表 7：相关公司提供视频数据用于训练多模态大模型 48
表 8：IP 类公司可基于 AI+视频开发更多 IP 衍生品 49
表 9：部分公司对 AI 短剧/AI 短片方向上的探索 49
表 10：相关上市公司估值表（截至 2024.07.24） 50

图 目 录

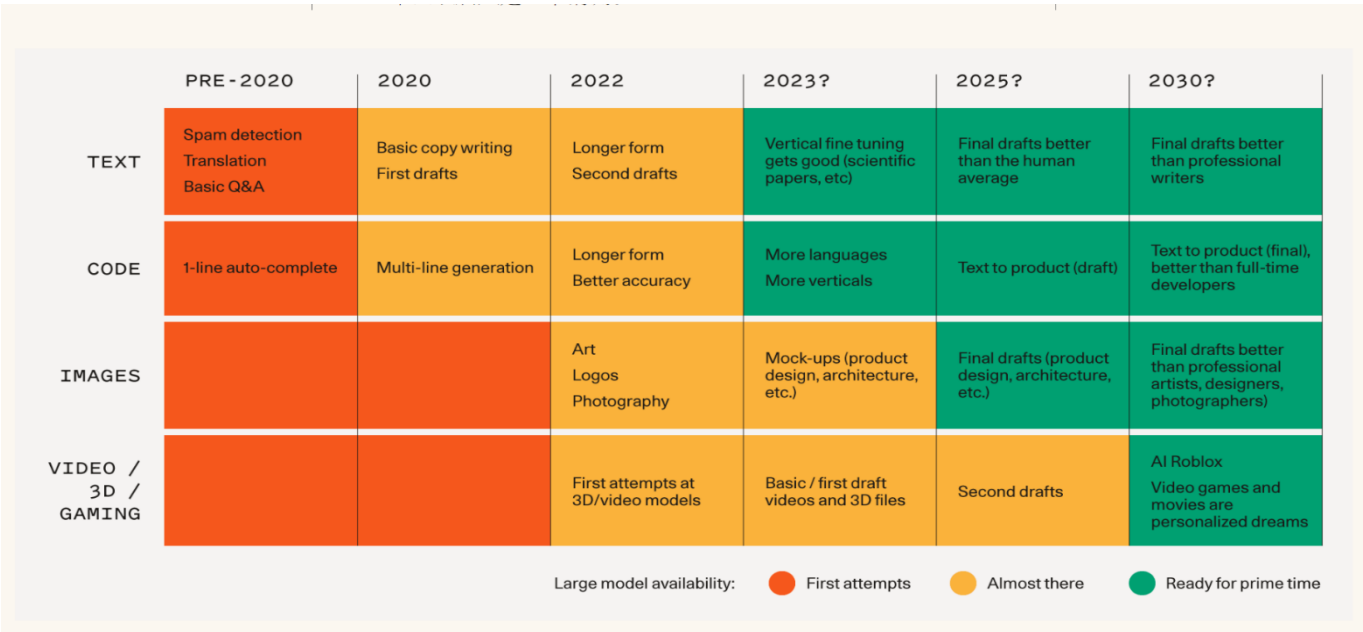
图 1：生成式 AI 发展进程 6
图 2：AI 应用地图梳理 6
图 3：主流文生视频技术的演进路径 7
图 4：AIGC 视频生成的技术演进路径 8
图 5：GAN 生成对抗网络运作原理 8
图 6：Diffusion 扩散模型运作原理 10
图 7：WALT 视频生成模型搭建原理示意图 11
图 8：Sora 基于 DiT 模型生成图像视频 12
图 9：2023 年生成式 AI+视频时间表 15
图 10：Luma AI Dream Machine 官网宣传文生视频功能 16
图 11：Luma AI Dream Machine 官网宣传前后帧输入图片生成连贯视频功能 17
图 12：Luma AI Dream Machine 实测演示中会遇到不符合物理规律、物体对象缺失等问题 17
图 13：Runway 产品定价模式 18
图 14：Runway Gen-1 视频生视频 19
图 15：Runway Gen-1 视频生视频演示 20
图 16：Runway Gen-2 文生视频效果表现较好 20
图 17：Runway Gen-2 图生视频效果及笔刷功能表现较好 21
图 18：Runway Gen-3 Alpha 通过运动画笔、高级相机控制、导演模式可以更精细控制运动 21
图 19：Runway Gen-3 Alpha 两端提示词测试，效果较强 22
图 20：Adobe 产品中引入第三方视频模型 Pika 优化用户体验 23
图 21：Pika 文生视频界面及视频编辑核心功能 23
图 22：Sora 合成的 60 秒视频 24

图 23：OpenAI 扩散模型过程 24
图 24：Sora 可进行多个视频的组合 25
图 25：Luma AI Dream Machine 生成效果（电影质感，略微不符合物理规律 25
图 26：Pika 生成效果（提示词理解、画面质感等方面有差距） 26
图 27：Runway Gen-2 生成效果（主角没有跟随镜头移动） 26
图 28：Runway Gen-3 Alpha 生成效果（各方面表现均优秀） 27
图 29：快手大模型产品矩阵及可灵 AI 产品功能升级 30
图 39：后续 Firefly 关于多模态音频、视频方向上的功能展望 36
图 40：Adobe Firefly 集成第三方大模型如 Runway、OpenAI Sora 用于视频剪辑 36
图 41：Captions AI Shorts 功能 37
图 42：Captions AI AD Creator 功能 37
图 43：阿里达摩院“寻光”一站式视频创作平台视频编辑功能 38
图 44：阿里达摩院“寻光”视频素材创作功能 38
图 45：美图 MOKI AI 短片产品 39
图 46：商汤 Vimi 人物视频生成 40
图 47：智象大模型升级 2.0 版本 40
图 48：智向未来即将上线一站式分镜头故事创作视频生成功能 41
图 49：Adobe Creative Cloud TAM 市场规模预测 42
图 50：Adobe Express 在 24 年 4 月迭代 AI 功能后，日活数骤然抬升并稳定提高 43
图 51:Adobe Premiere Pro 引入第三方模型如 Pika、OpenAI、Runway 生成视频 43
图 52：美图公司底层、生态层、应用层架构 44
图 53：Vimi 在人物一致性功能支持下打造的数字分身打造 AI 视频功能、AI 表情包功能 45

一、为什么要研究 AI+视频——AI 视频生成正成为当前行业发展关键节点

2023 年红杉资本在关于生成式 AI 发展进程的预测报告中表明，在历经文生文、文生图的升级迭代后，我们目前正处在 AI+生产力办公&设计、AI+视频和 AI+3d 渗透的历史节点上。在底层大模型技术迭代逐渐加速的今天，AI 文本对话、AI 文生图、AI 陪伴等方向已经逐渐成为竞争激烈的主要方向，展望未来我们需要对更多 AI+做深入的研究，而视频方向一直是业内关注的重点方向之一。视频杂糅了文本、语音、图像等多维度内容，其训练的难点也往往在于视频数据对数量和质量的不足、算法架构需要优化、物理规律性较差等等，但我们相信，随着 AI+视频的技术和产品升级迭代，众多行业有望受益，诸如电影、广告、视频剪辑、视频流媒体平台、UGC 创作平台、短视频综合平台等，而目前正处在 AI+视频发展的关键性时刻，正从 AI+视频创意生成逐渐过渡到一站式视频生成+剪辑+UGC 的后续阶段。

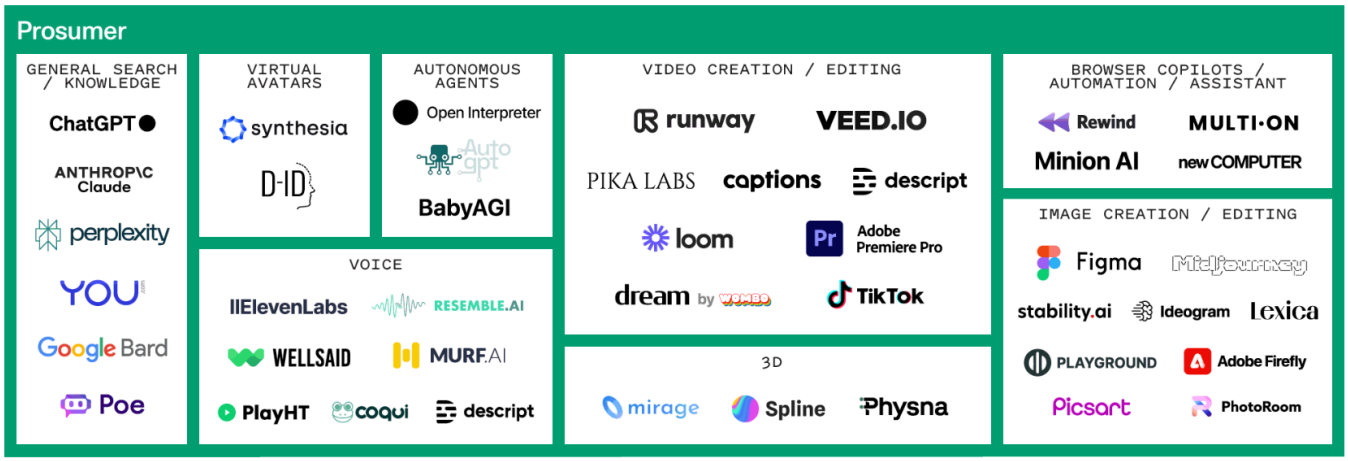
图 1：生成式 AI 发展进程



资料来源：红杉资本官网，信达证券研发中心

在红杉资本 2024 年关于 AI 应用的地图梳理中反映了市场中的两个重要趋势：生成式人工智能从技术趋势演变为实际应用和价值，以及生成式人工智能应用日益呈现多模态的特性。可以看到，AI 视频生成及编辑的版图占比较多，重要性和产品推进速度目前较快。

图 2：AI 应用地图梳理



资料来源：红杉资本官网，信达证券研发中心

文/图生视频面临着众多方面的挑战，例如：

1）计算成本：确保帧间空间和时间一致性会产生长期依赖性，从而带来高计算成本；

2）缺乏高质量的数据集：用于文生视频的多模态数据集很少，而且通常数据集的标注很少，这使得学习复杂的运动语义很困难。文生视频模型需要依赖于大量数据来掌握如何将文本描述转化为具有写实感的连续帧，并捕捉时间上的动态变化；

3）视频生成质量：时空一致性难以保持，在不同镜头、场景或时间段内较难确保角色、物体和背景的一致性。可控性和确定性还未充分实现，确保所描述的运动、表现和场景元素能够精确控制和编辑。视频时长的限制，长视频制作仍面临时间一致性和完整性的挑战，这直接影响到实际应用的可行性；

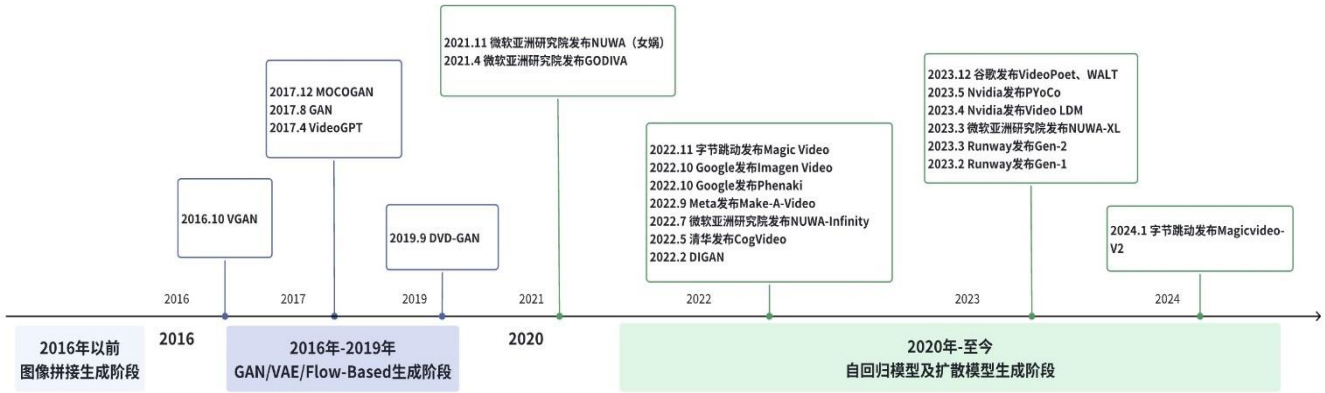
4）语义对齐：由于自然语言具有复杂性和多义性，文本语义理解、文本与视频元素的映射关系仍是挑战；

5）产品易用性：对于文生视频，产品的易用性和体验仍需改进。个人用户希望制作流程易上手、符合习惯，并支持快速素材搜索、多样模板、多端同步和一键分享；小 B 端用户关注成本可控下的快速营销视频制作和品牌传播效果；行业用户则需要内容与交互性的融合，包括商用素材适配性、快速审核和批量制作分发能力；

6）合规应用：文生视频的应用面临素材版权、隐私安全和伦理道德等风险。

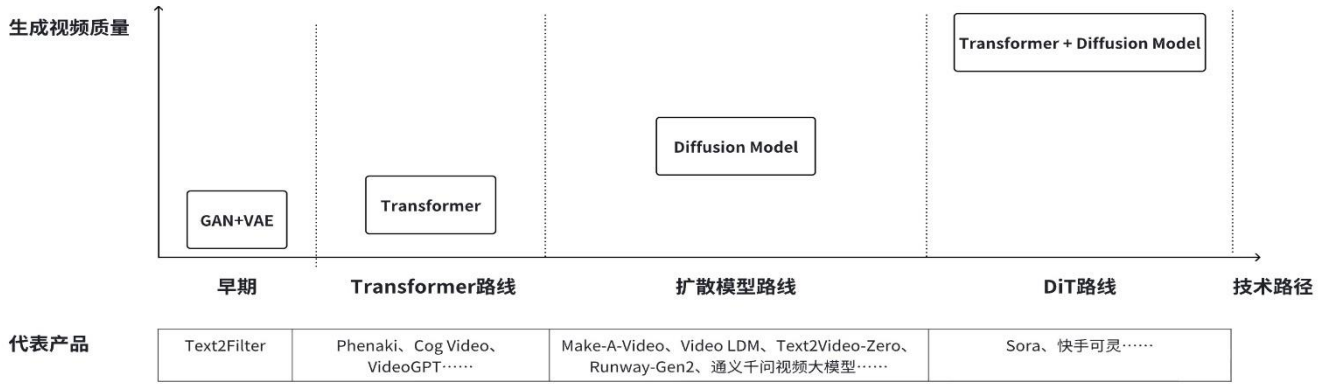
二、市场主流 AI 视频生成技术的迭代路径

图 3：主流文生视频技术的演进路径



资料来源：Carl Vondrick 等《Generating Videos with Scene Dynamics》；Sergey Tulyakov 等《MoCoGAN: Decomposing Motion and Content for Video Generation》；Eiichi Matsumoto 等《Temporal Generative Adversarial Nets with Singular Value Clipping》；Aidan Clark 等《ADVERSARIAL VIDEO GENERATION ON COMPLEX DATASETS》；Chenfei Wu 等《NUWA: Visual Synthesis Pre-training for Neural visUal World creAtion》.；Chenfei Wu 等《GODIVA: Generating Open-DomaIn Videos from nAtural Descriptions》；Wilson Yan 等《VideoGPT: Video Generation using VQ-VAE and Transformers》；Daquan Zhou 等《MagicVideo: Efficient Video Generation With Latent Diffusion Models》；Jonathan Ho 等《IMAGEN VIDEO: HIGH DEFINITION VIDEO GENERATION WITH DIFFUSION MODELS》；Ruben Villegas 等《PHENAKI: VARIABLE LENGTH VIDEO GENERATION FROM OPEN DOMAIN TEXTUAL DESCRIPTIONS》；Uriel Singer 等《MAKE-A-VIDEO: TEXT-TO-VIDEO GENERATION WITHOUT TEXT-VIDEO DATA》；Chenfei Wu 等《NUWA-Infinity: Autoregressive over Autoregressive Generation for Infinite Visual Synthesis》；Wenyi Hong 等《CogVideo: Large-scale Pretraining for Text-to-Video Generation via Transformers》；Sihyun Yu 等《GENERATING VIDEOS WITH DYNAMICS-AWARE IMPLICIT GENERATIVE ADVERSARIAL NETWORKS》；Dan Kondratyuk 等《VideoPoet: A Large Language Model for Zero-Shot Video Generation》；Agrim Guptau 等《Photorealistic Video Generation with Diffusion Models》；Songwei Ge 等《Preserve Your Own Correlation: A Noise Prior for Video Diffusion Models》；Andreas Blattmann 等《Align your Latents: High-Resolution Video Synthesis with Latent Diffusion Models》；Shengming Yin 等《NUWA-XL: Diffusion over Diffusion for eXtremely Long Video Generation》；Weimin Wang 等《MagicVideo-V2: Multi-Stage High-Aesthetic Video Generation》、Runway 公司官网，信达证券研发中心

图 4：AIGC 视频生成的技术演进路径

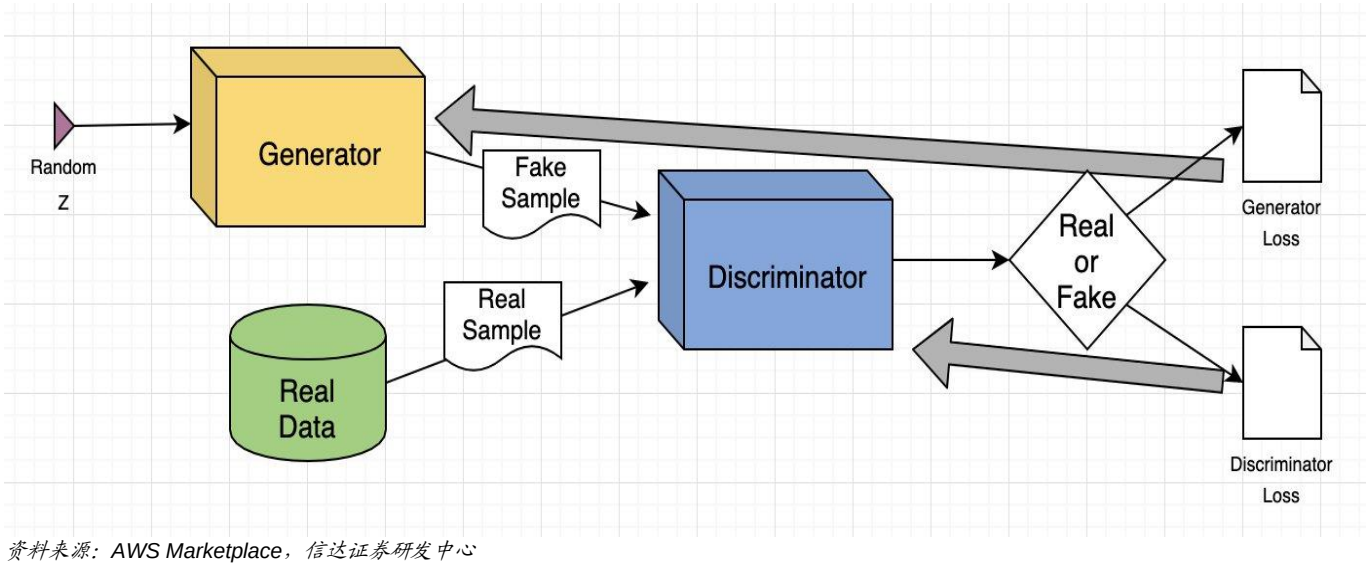


资料来源：信达证券研发中心（注：该图通过图 3 所引用论文总结而来）

1）GAN+VAE

生成对抗网络（Generative Adversarial Networks）核心思想是训练两个网络，生成器（G）和判别器（D）。生成器通过获取输入数据样本并尽可能对其进行修改来生成新数据，试图生成逼真的视频；而判别器尝试预测生成的数据输出是否属于原始数据集，尝试区分真实视频和生成的视频。两个网络通过对抗训练，生成器试图最大化欺骗判别器，而判别器则试图最大化识别生成视频的错误，直到预测网络不再能够区分假数据值和原始数据值。GAN 用于视频生成在 2016 年至 2021 年较为火热，代表模型如 Temporal Generative Adversarial Nets (TGAN) 和 MoCoGAN，它们通过不同的网络架构和训练方法来改进 GAN 在视频生成上的性能。此外，Dual Video Discriminator GAN (DVD-GAN) 通过使用空间和时间判别器的分解来提高视频生成的复杂性和保真度。

图 5：GAN 生成对抗网络运作原理



资料来源：AWS Marketplace，信达证券研发中心

GAN 技术特点如下：1）无需标注数据，可以从未标注的图像中学习生成新的图像或视频；2）多领域应用，可以应用于图像生成、风格迁移、数据增强、超分辨率等多种任务；3）模型灵活，通过改变网络结构，可以适应不同的数据分布和生成任务；4）模型参数小，较为轻便，擅长对单个或多个对象类进行建模。**GAN 作为早期文生视频模型，存在如下缺点**：1）训练过程不稳定，容易出现模式崩溃（mode collapse），即生成器开始生成非

常相似或重复的样本；2）计算资源：训练 GAN 通常需要大量的计算资源和时间；3）对超参数选择敏感，不同的设置可能导致训练结果差异很大。

VAE（Variational Autoencoder 变分自编码器）：对于传统的基本自编码器来说，只能够对原始数据进行压缩，不具备生成能力，基本自编码器给定一张图片生成原始图片，从输入到输出都是确定的，没有任何随机的成分。生成器的初衷实际上是为了生成更多“全新”的数据，而不是为了生成与输入数据“更像”的数据。而变分自用编码器的 Encoder 与 Decoder 在数据流上并不是相连的，不会直接将 Encoder 编码后的结果传递给 Decoder，而是要使得隐式表示满足既定分布。因此，**VAE 引入了隐变量推断,训练过程稳定,但是其生成的图片缺少细节,轮廓模糊; GAN 生成的图像真实清晰,但是训练过程易出现模式崩溃问题。因此，VAE+GAN 的串联融合可以实现数据的自动生成+高质量图像生成的结果。**

2）Transformer 模型

Transformer 是一种先进的神经网络算法，它完全基于注意力机制，不依赖于传统的循环神经网络（RNN）或卷积神经网络（CNN）。Transformer 保留了编码器-解码器的基本结构。编码器将输入序列映射到连续的表示空间，而解码器则基于这些表示生成输出序列。Transformer 模型的自注意力机制，允许序列中的每个元素都与序列中的其他元素进行交互，从而捕捉全局依赖关系；模型还采用多头注意力并行处理，可获取不同空间的信息。

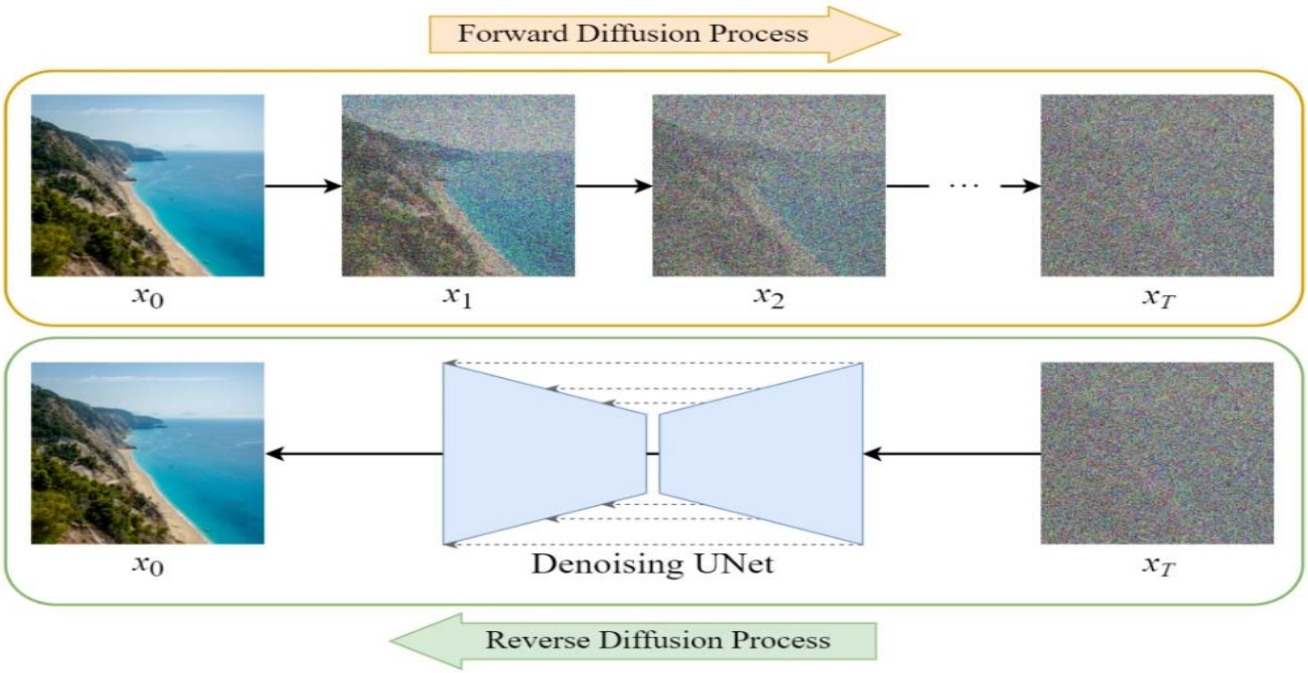
Transformer 模型技术特点如下：1）并行处理序列中的所有元素，这与传统循环神经网络（RNN）相比，大大提高了计算效率；2）可扩展性，能够通过堆叠多个注意力层来增加模型的复杂度和容量；3）泛化能力，除了语言任务，还可以泛化到其他类型的序列建模任务，如图像处理、视频分析等；4）预训练和微调，Transformer 模型通常先在大量数据上进行预训练，再针对特定任务进行微调，使得模型能够快速适应新任务；5）适应长序列数据，在处理诸如语音信号、长时间序列数据等任务具有优势，避免传统模型存在的梯度消失或梯度爆炸问题。

Transformer 存在如下缺点：1）参数效率相对较低，参数数量随输入序列长度的增加而增加，增加了训练时间和成本；2）对输入数据的敏感性较高，模型依赖于输入数据的全局信息进行建模，在处理复杂任务时（如机器翻译、语音识别等），对输入数据的细微变化可能会对模型的输出结果产生较大影响；3）难以处理时空动态变化，模型时基于自注意力机制的静态模型，无法捕捉到时空动态变化的信息，因此在处理视频、时空数据等具有动态变化特性的任务时，需要结合其他技术来提高模型的性能。Transformer 模型在视频生成领域的应用的产品包括 VideoGPT、NUWA、CogVideo、Phenaki 等。这些模型通过结合视觉和语言信息，生成新的视频内容或对现有视频进行操作。它们利用了 Transformer 模型的自注意力机制来处理高维数据，并通过预训练和微调策略来提高性能。此外，这些模型还探索了如何通过多模态学习来提高视频生成的质量和多样性。

3）扩散模型

扩散模型是一种生成模型，通过逐步添加噪声来破坏训练数据，然后通过逆向过程去噪来生成与训练数据相似的新数据。扩散模型分为三大类型：去噪扩散概率模型（DDPM）、基于噪声条件评分的生成模型（SGM）、随机微分方程（SDE），但三种数学框架背后逻辑统一，均为添加噪声后将其去除以生成新样本。

图 6：Diffusion 扩散模型运作原理



资料来源：数据派 THU 公众号，信达证券研发中心

尽管 Transformer 在 Autoregressive Model 中得到广泛应用，但是这种架构在生成式模型中较少采用。比如，作为图像领域生成模型的经典方法，Diffusion Models 却一直使用基于卷积的 U-Net 架构作为骨干网络。随着 Sora、WALT 等基于（Diffusion+Transformer）的探索，国内创业公司如智向未来也在尝试延续这个最新的技术路线，用 Transformer 架构替换掉原来的卷积 U-Net 架构后，生成视频的时长可变、尺寸可变，可以在不同的空间进行建模，同时也可以让视频和图片配对来实现多模态对齐与编码。

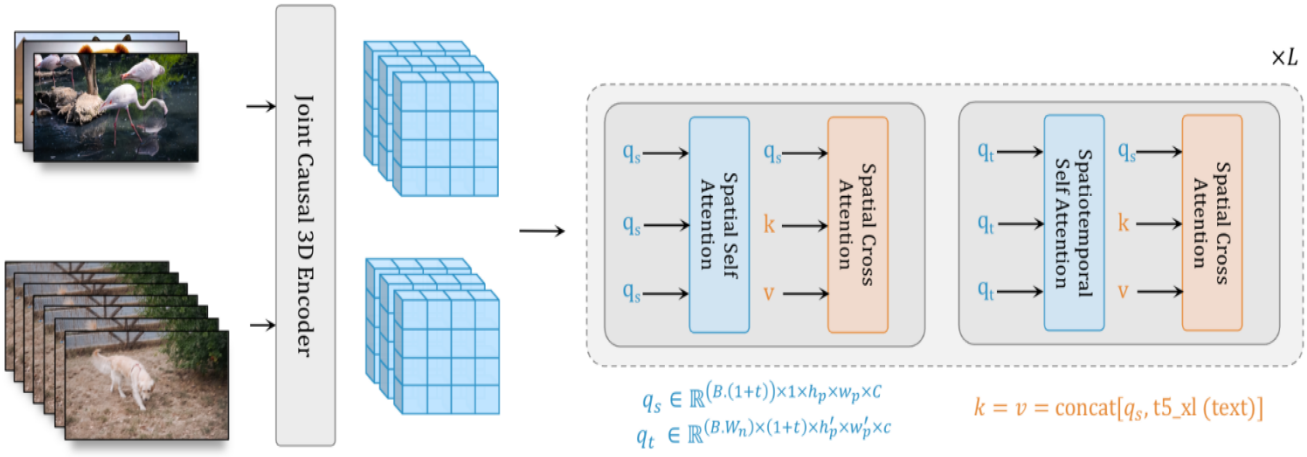
4）DiT（Transformer+Diffusion）

Diffusion Transformer（DiT）模型是从（Scalable Diffusion Models with Transformers, ICCV 2023）中引入。基本上，Diffusion Transformer（DiT）是一个带有变换器（而非 U-Net）的扩散模型，核心思想是利用 Transformer 处理潜在空间中的图像数据块，模拟数据的扩散过程以生成高质量的图像。

W.A.L.T（Window Attention Latent Transformer）

2023 年底，世界知名 AI 科学家李飞飞团队与谷歌合作，推出了视频生成模型 W.A.L.T（Window Attention Latent Transformer）——一个在共享潜在空间中训练图像和视频生成的、基于 Transformer 架构的 Diffusion 扩散模型。**技术迭代主要有两个方向：1）使用因果编码器在统一的潜在空间内联合压缩图像和视频，从而实现跨模态的训练和生成。2）为了提高内存和训练效率，团队使用了为联合空间和时空生成建模量身定制的窗口注意架构。**所以，无需使用无分类器指导，就能在成熟的视频（UCF-101 和 Kinetics-600）和图像（ImageNet）生成基准上实现最先进的性能。最后，团队还为文本到视频生成任务训练了三个模型的级联，包括一个基本的潜在视频扩散模型和两个视频超分辨率扩散模型，以每秒 8 帧的速度生成 512 x 896 分辨率的视频。

图 7：WALT 视频生成模型搭建原理示意图



资料来源：Kihyuk Sohn 等《Photorealistic Video Generation with Diffusion Models》、WALT 视频模型官网，信达证券研发中心

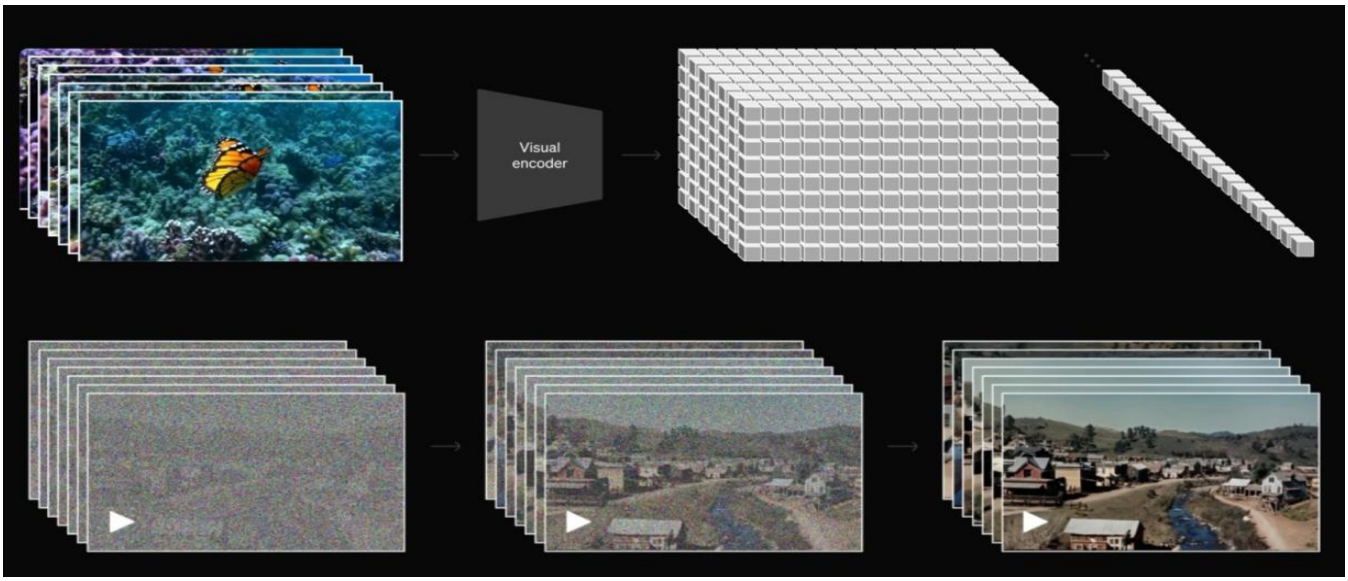
WALT 将图像和视频编码到共享潜在空间中。Transformer 主干使用具有两层窗口限制注意力的块来处理这些潜在空间：**空间层捕获图像和视频中的空间关系，而时空层通过身份注意力掩码模拟视频中的时间动态并传递图像**。文本调节是通过空间交叉注意力完成的。

DiT 模型技术特点如下：1）运用潜在扩散模型，在潜在空间而非像素空间中训练扩散模型，提高了计算效率；2）Patchify 操作，将空间输入转换为一系列 token，每个 token 代表图像中的一个小块；3）条件输入处理，DiT 设计了不同的 Transformer 块变体来处理条件输入（如噪声时间步长、类别标签等）；4）自适应层归一化（adaLN），使用 adaLN 来改善模型性能和计算效率；5）可扩展性：DiT 展示了随着模型大小和输入 token 数量的增加，模型性能（以 FID 衡量）得到提升；6）简化的架构选择，DiT 证明了在扩散模型中，传统的 U-Net 架构并不是必需的，可以被 Transformer 替代。

DiT 模型仍存在以下缺点：1）实现复杂性，虽然 DiT 在理论上简化了架构选择，但 Transformer 的实现可能比 U-Net 更复杂；2）训练稳定性：尽管 DiT 训练稳定，但 Transformer 架构可能需要特定的训练技巧来保持稳定；3）对硬件要求高，虽然 DiT 在计算上更有效率，但 Transformer 模型通常需要大量的内存和计算资源，这可能限制了它们在资源受限的环境中的应用；4）模型泛化能力，DiT 主要在 ImageNet 数据集上进行了评估，其在其他类型的数据和任务上的泛化能力尚未得到验证。

DiT 作为一种新型的扩散模型，通过在潜在空间中使用 Transformer 架构，实现了对图像生成任务的高效和高性能处理。DiT 在 Sora 上表现优秀，Sora 是 OpenAI 发布的爆款视频生成模型，它融合了扩散模型的生成能力和 Transformer 架构的处理能力。受到大规模训练的大型语言模型的启发，Sora 通过在互联网规模的数据上训练，获得了通用化的能力。它**采用基于扩散模型的生成框架，逐步改进噪声样本以产生高保真度的视频输出，并应用 Transformer 架构来处理视频和图像的时空信息，保持物体在三维空间中的连贯性。这种结合生成和变换器优势的方法，使得 Sora 在视频生成和编辑任务中表现出色，能够创造出多样化、高质量的视觉内容。**

图 8：Sora 基于 DiT 模型生成图像视频



资料来源：Sora，信达证券研发中心

表 1：Transformer、Diffusion、DiT 模型的产品梳理

模型类型	模型名称	发布方	发布时间	技术特点及主要功能
Transformer	VideoGPT	Wilson Yan et al.	2021.4	使用 VQ-VAE，并通过 3D 卷积和轴向自注意力机制实现。使用类似 GPT 的架构自回归地对离散潜在表示进行建模。生成架构简单，能生成高保真度视频，尤其是适应动作条件视频。
	NUWA 女娲	微软亚洲研究院	2021.11	采用 3D 变换器编码器-解码器框架，提出 3D 近邻注意力机制简化计算，支持多模态预训练，使用 VQ-GAN 视觉标记 3D tokens，具有零样本能力。在生成图像、视频以及视频预测方面表现优秀。
	CogVideo	清华大学	2022.5	采用多帧率层次化训练策略、双通道注意力机制，灵活文本条件模拟不同帧率视频，顺序生成和递归插值框架使视频生成连贯。对复杂语义的运动理解加强，生成高分辨率、高帧率、高一致性的视频。
	NUWA-Infinity	微软亚洲研究院	2022.7	采用双重自回归生成机制来处理可变尺寸的生成任务，引入 NCP 缓存已生成的相关 patch 来减少计算成本，采用任意方向控制器赋能图像扩展，能生成任意大小高分辨率图像、长时视频、图像动画。
	Phenaki	Google	2022.10	使用因果注意力机制生成可变长度视频，使用预训练的 T5X 来生成文本嵌入，通过双向遮蔽 Transformer 根据文本嵌入生成视频

				token，采用 C-ViViT 编码-解码架构减少 token 数量并在时空一致性表现更好。
	Videopoet	Google	2023.12	仅采用解码器架构能处理多模态输入，支持零样本视频生成；使用双向变换器在标记空间内提高空间分辨率；通过自回归扩展内容来合成长达 10 秒的连贯视频；执行文本、图像、视频编辑到视频的多任务视频生成。
	WALT	Google	2023.12	使用因果编码器联合压缩图像和视频，实现跨模态生成；采用窗口注意力架构，联合空间和时空生成建模；不依赖分类器自由引导可生成视频；通过潜在视频扩散模型和视频超分辨率扩散模型的级联，生成 512×896 分辨率、每秒 8 帧的视频；能根据类别标签、自然语言、过去帧、低分辨率视频生成可控视频。
	Imagen Video	Google	2022.1	采用基础视频扩散模型和用于空间与时间超分辨率扩散模型，采用 v-prediction 参数化避免色彩偏移，应用渐进式蒸馏技术，快速高效采样；使用噪声条件增强来减少级联模型中的域差距，提高样本质量；能生成各种艺术风格和 3D 对象理解的视频，具可控性和对世界知识的理解。
	VideoDiffusion Model	Google	2022.4	从图像和视频数据联合训练减小批量梯度方差；引入条件采样技术，提高空间和时间视频扩展性能；使用特定类型的 3D U-Net 作为扩散模型架构，使时间空间分解；采用因子化的空间-时间注意力机制，能遮蔽模型以在独立图像上运行；使用多种扩散模型采样器；能处理多尺度和多帧视频数据，生成长序列视频。
	Make-A-Video	Meta	2022.9	不需要成对的文本-视频数据进行训练；通过无监督的视频素材学习世界的运动方式；构建在 T2I 模型之上，包括分解全时域 U-Net 和注意力张量，并在空间和时间上近似它们；设计空间-时间管道，通过视频解码器、插值模型、超分辨率模型生成高分辨率、高帧率视频。
	MagicVideo	字节跳动	2022.11	使用 3D U-Net 解码器简化计算；引入帧间轻量适配器，减少对独立 2D 卷积块的需求；采用有向自注意力机制，仅基于所有先前帧计算未来帧的特征；提出 VideoVAE 自编码器，改善像素抖动问题；训练基于扩散的超

Diffusion				分辨率模型，从 256×256 上采样到 1024×1024 的高分辨率。
	Tune-A-Video	新加坡国立大学，腾讯	2022.12	基于预训练的 T2I 扩散模型，使用开放域知识；引入空间时间注意力机制来学习连续运动；使用 DDIM 反演，使生成视频时序一致；只更新注意力块中的投影矩阵而非所有参数，避免对新概念视频生成的阻碍。
	Gen-1	Runway	2023.2	将潜在扩散模型扩展到视频生成，通过将时间层引入到预训练的图像模型中并对图像和视频进行联合训练，无需额外训练和预处理。
	Gen-2	Runway	2023.2	允许使用任意起始帧，通过 I2V 方式生成视频；通过训练模型预测视频下一帧，对视觉世界深入理解；从单个帧的高保真度生成开始，逐步解决视频叙事中的挑战，包括场景、角色和环境的一致性。
	Dreamix	Google	2023.2	采用混合微调方法，结合全时序注意力和时序注意力掩蔽的微调；引入轻量级的帧间适配器，用于调整 I2V 分布；采用有向自注意力机制，捕捉帧间的时序依赖性；提出图像动画框架，转图像为粗糙视频进行编辑。
	NUWA-XL	微软亚洲研究院	2023.3	能够直接在长视频上进行训练，并通过增加深度 m 来轻松扩展到更长的视频；“粗到细”阶段生成，先通过全局扩散模型生成关键帧，再用局部扩散模型递归填充邻近帧之间的内容；支持并行推理，提高长视频生成速度。
	Text2Video-Zero	Picsart AI Research, UT Austin, U of Oregon, UIUC	2023.3	实现零样本学习；在生成帧代码注入运动动力学，能保持全局场景和背景的时间一致性；使用新的跨帧注意力机制保留前景对象的上下文、外观和身份。
	VideoLDM	NVIDIA	2023.4	在潜在空间扩散模型中引入时间维度，将图像生成器转换为视频生成器，实现视频数据的时间对齐；在图像上预训练 LDM，然后在编码的视频上微调生成视频；能够实现高达 1280×2048 分辨率的视频生成。
	PYoCo	NVIDIA	2023.5	提出视频扩散噪声先验，更好地捕捉视频帧之间的内在联系；采用一个由基础模型和三个上采样堆叠组成的级联网络架构；使用了 DEIS 及其随机变体进行样本合成的先进采样技术；小规模模型实现优异性能，从文本嵌入生成高分辨率的视频。

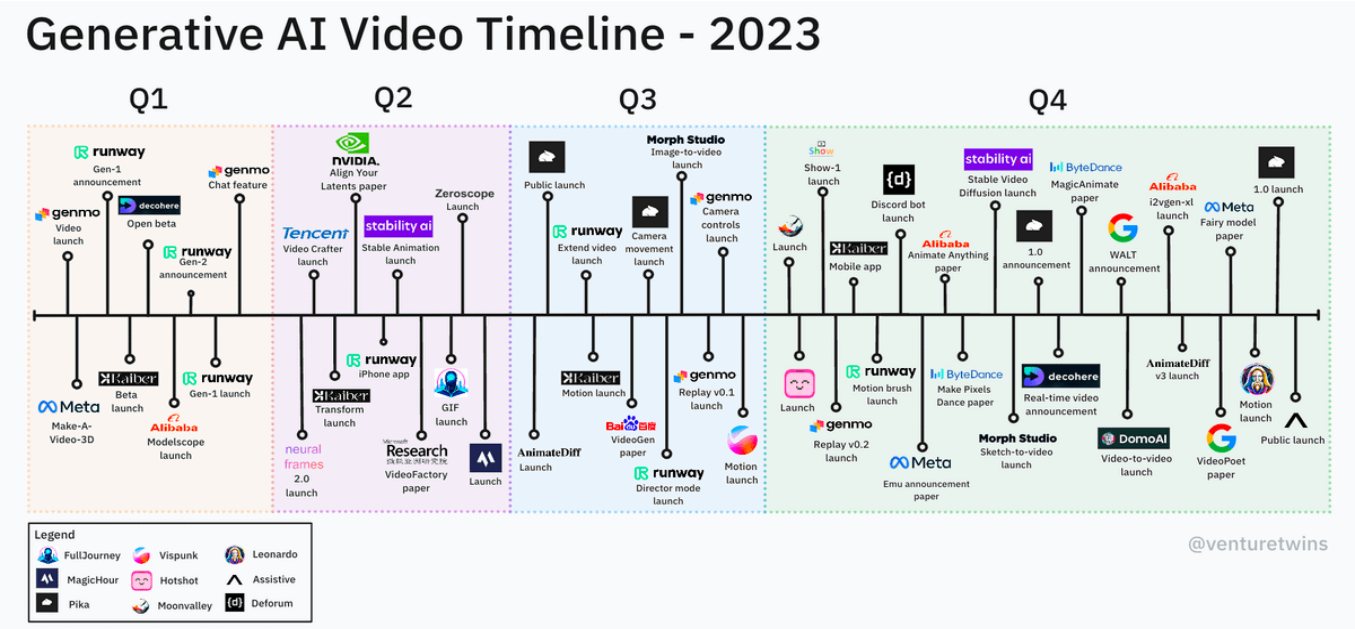
DiT	Sora、可灵等	OpenAI、快手等	2024.2	使用文本条件扩散模型，处理视频和图像的空间时间块；训练了一个网络来降低视觉数据的维度，输入原始视频并输出压缩的潜在表示；能够生成一分钟的高保真视频，能实现视频扩展、视频过渡，输入视频的风格和环境的零样本转换。

资料来源：Carl Vondrick 等《Generating Videos with Scene Dynamics》；Sergey Tulyakov 等《MoCoGAN: Decomposing Motion and Content for Video Generation》;Eiichi Matsumoto 等《Temporal Generative Adversarial Nets with Singular Value Clipping》;Aidan Clark 等《ADVERSARIAL VIDEO GENERATION ON COMPLEX DATASETS》；Chenfei Wu 等《NUWA: Visual Synthesis Pre-training for Neural visUal World creAtion》.；Chenfei Wu 等《GODIVA: Generating Open-DomaIn Videos from nAtural Descriptions》；Wilson Yan 等《VideoGPT: Video Generation using VQ-VAE and Transformers》;Daquan Zhou 等《MagicVideo: Efficient Video Generation With Latent Diffusion Models》;Jonathan Ho 等《IMAGEN VIDEO: HIGH DEFINITION VIDEO GENERATION WITH DIFFUSION MODELS》；Ruben Villegas 等《PHENAKI: VARIABLE LENGTH VIDEO GENERATION FROM OPEN DOMAIN TEXTUAL DESCRIPTIONS》；Uriel Singer 等《MAKE-A-VIDEO: TEXT-TO-VIDEO GENERATION WITHOUT TEXT-VIDEO DATA》;Chenfei Wu 等《NUWA-Infinity: Autoregressive over Autoregressive Generation for Infinite Visual Synthesis》;Wenyi Hong 等《CogVideo: Large-scale Pretraining for Text-to-Video Generation via Transformers》；Sihyun Yu 等《GENERATING VIDEOS WITH DYNAMICS-AWARE IMPLICIT GENERATIVE ADVERSARIAL NETWORKS》；Dan Kondratyuk 等《VideoPoet: A Large Language Model for Zero-Shot Video Generation》；Agrim Guptau 等《Photorealistic Video Generation with Diffusion Models》；Songwei Ge 等《Preserve Your Own Correlation: A Noise Prior for Video Diffusion Models》；Andreas Blattmann 等《Align your Latents: High-Resolution Video Synthesis with Latent Diffusion Models》；Shengming Yin 等《NUWA-XL: Diffusion over Diffusion for eXtremely Long Video Generation》；Weimin Wang 等《MagicVideo-V2: Multi-Stage High-Aesthetic Video Generation》、OpenAI Sora 官网、Runway 公司官网，信达证券研发中心

二、目前海外市场主流的生成式 AI+视频参与者

AI+视频发展以来，技术路径和迭代产品冗杂繁多、功能不一、效果差异，我们选取目前海内外市场主要的生成式视频的参与者：Luma AI（Dream Machine）、Runway （Gen 1-2 & Gen-3 Alpha）、Pika、Sora，集中梳理了其融资历程、产品迭代、核心功能、实测效果比较等多方面，经个别提示词生成视频效果测试，在 Sora 未公开实测情况下，我们认为 Runway Gen-3 Alpha 的视频生成效果，如质量分辨率、生成速度、物体符合物理规律、提示词理解、视频时长等诸多维度上表现均较为优秀。

图 9：2023 年生成式 AI+视频时间表



资料来源：Venture Twins、a16z，信达证券研发中心

（一） Luma AI—Dream Machine

Luma AI 成立于 2021 年，2024 年以其推出的文生视频模型 Dream Machine 而得到全球投资视野的关注，但早期公司仅聚焦在 3D 内容生成，23 年 11 月，Luma AI 在 Discord 服务器上推出了文生 3D 模型 Genie，降低了开发人员的 3D 建模和重建功能的成本，每个场景或者物体的生成只需一美元，创建时间也大幅缩短。公司推出的应用程序 Flythroughs 可以使用户仅使用 AI 和 iPhone 就可创建专业的飞行场景视频，可用于房产中介应用的 3D 环境视频的录制等。融资历程：公司 A 轮融资由 Amplify Partners、Nventures（Nvidia 投资部门）和 General Catalyst 领投，共筹集 2000 万美元；B 轮融资则由硅谷顶级风投公司 Andreessen Horowitz、英伟达领投，筹集 4300 万美元，B 轮估值在 2 亿到 3 亿美元之间。官网显示目前核心团队共 34 人，其中华人 5 位。

Luma AI Dream Machine 是一款由 Luma AI 开发的 AI 视频生成模型，它能够将文本和图像快速转换为高质量、逼真的视频，且具备前后帧输入图片生成连贯视频的功能。在官网的介绍中，该模型具备快速视频生成能力，能够在 120 秒内生成 120 帧视频，可生成具有逼真流畅动作、电影摄影和戏剧效果 5s 镜头，确保视频角色的一致性和物理准确性，适用于创意视频制作、故事讲述、市场营销及教育培训等多种场景。Dream Machine 可以快速将文本和图像制作成高质量视频、是一种高度可扩展且高效的转换器模型，能够生成物理上准确、一致且多变的镜头。

后续迭代的方向主要包括：更长时间、更多角度、飞行连贯性更强、视频内物体编辑、AI 动漫生成等。官网上同时也披露了目前的视频生成面临的难点，诸如：1）视频内物体变形；2）移动僵硬；3）文本错误；4）不符合物理规律；Luma AI 公司致力于继续优化 Dream Machine 的性能，为用户提供更加优质的视频生成服务，并计划将 Dream Machine 技术应用于更多领域，拓展其市场影响力。目前受制于算力和用户高需求，免费用户每天有 20 个视频生成的限额，付费用户在排队中靠前并且每天没有生成数量的上限。

图 10：Luma AI Dream Machine 官网宣传文生视频功能

资料来源：Luma AI Blog，信达证券研发中心

图 11：Luma AI Dream Machine 官网宣传前后帧输入图片生成连贯视频功能

资料来源：Luma AI 官网，信达证券研发中心

我们在实测中发现，如官网所描述的，在生成视频的过程中会遇到例如对象缺失、行动轨迹僵硬、不符合实际物理规律等问题。

图 12：Luma AI Dream Machine 实测演示中会遇到不符合物理规律、物体对象缺失等问题

资料来源：Luma AI 官网，信达证券研发中心

（二） Runway Gen 1-2 & Gen-3 Alpha

Runway 成立 2018 年，总部位于纽约，由 Cristóbal Valenzuela、Alejandro Matamala 和 Anastasis Germanidis 共同创立。公司专注于将艺术与人工智能融合，提供图像和视频编辑工具。自成立以来，Runway 经历了多轮融资，估值迅速增长。其产品包括多种 AI 驱动的创作工具，如 2023 年推出的 Gen-1 和 Gen-2，Runway 仍在不

断创新，2024 年推出新一代视频生成模型 Gen-3 Alpha。据外媒 TechCrunch 报道，近期公司正筹划新一轮融资 4.5 亿美元，估值有望达到 40 亿美元。

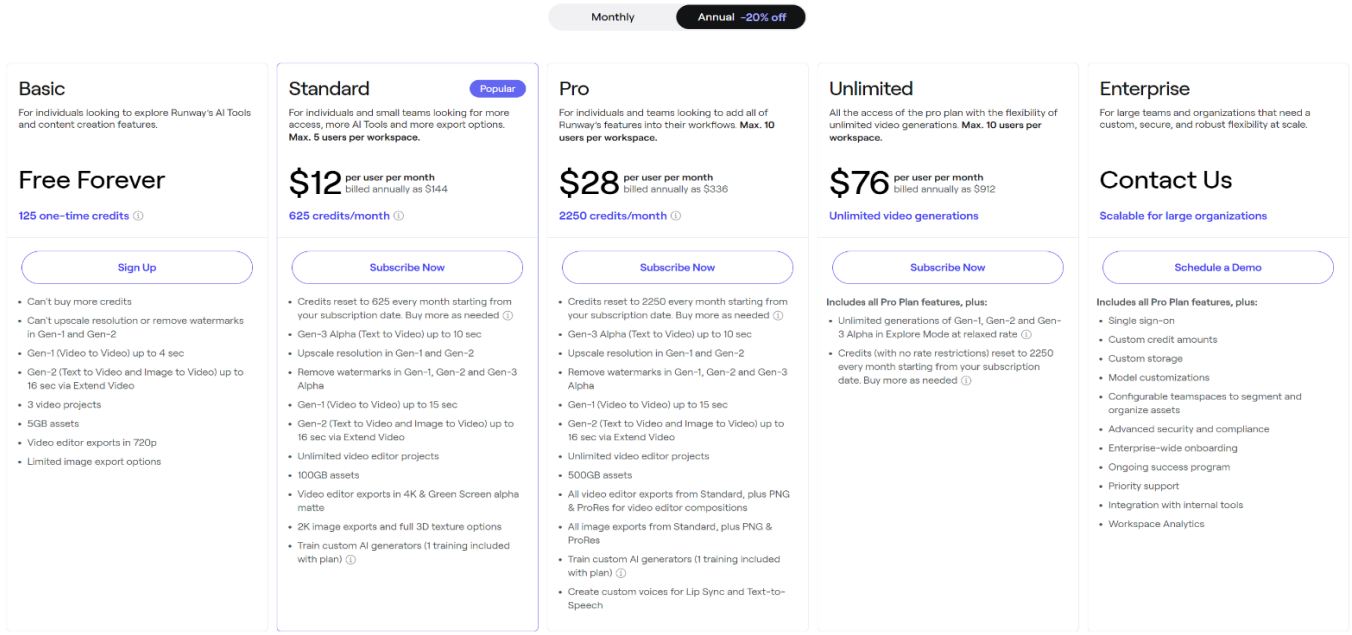
表 2：Runway 历年融资轮次、融资金额及对应估值

时间	融资轮次	融资金额	投资方	估值
2020.12	A 轮	850 万美元	Amplify Partners 领投，Lux Capital 和 Compound Ventures 参投	/
2021.12	B 轮	3500 万美元	Coatue 领投，所有现有投资者均参与其中：Amplify Partners、Lux Ventures 和 Compound	/
2022.12	C 轮	5000 万美元	Felicis 领投，所有现有投资者均参与其中：Amplify Partners、Lux Capital、Coatue 和 Compound	/
2023.06	C+轮	1.41 亿美元	C 轮融资增加 1.41 亿美元，参与的投资者包括谷歌、NVIDIA、Salesforce Ventures 以及现有投资者等	15 亿美元
2024.07	D 轮（据 TechCrunch 报道）	4.5 亿美元	投资机构包括 General Atlantic 等	40 亿美元

资料来源：Runway 官网、The Information、TechCrunch 官网，信达证券研发中心

Runway 不同的定价模式：主要分为永久免费基础版、标准版、高级版、无限制版本和企业级版本服务。永久免费版：用户拥有一次性 125 个 credits 积分，gen-1（视频到视频）上传最长为 4s，gen-2（文生视频和图生视频）通过延长视频功能最长至 16s 等；标准版、高级版和无限制版本的差别在于每月积分的数额、gen-3 的使用、水印的消除、资产库数量、视频质量等方面。

图 13：Runway 产品定价模式



资料来源：Runway 官网、信达证券研发中心

Runway Gen-1（Video to Video）

Gen-1 为视频到视频的模型，即使用文字和图像从现有的视频中生成新的视频，可以实现例如将某个视频转换为完全风格化的动画渲染以及更换现有视频的背景等。首先，选择要用作输入的视频。此视频将决定最终输出的整体构图和动作；其次，选择风格参考，有三种方法可以转换输入视频：选择现有图像、编写文本提示或从 Runway 的样式预设中选择一个；最后，使用结构一致性和提示权重等高级设置来调整样式参考对输入视频的影响程度。在生成之前，可以预览 4 个静态帧以帮助调整设置。Gen-1 最多可以生成 15 秒的视频。在使用 Gen-1 生成视频之前，可以利用上传的视频同时结合自己设置的风格和参数生成免费预览的分镜头脚本，减少多余算力的消耗。

在 Gen-1 官方指导论文《Structure and Content-Guided Video Synthesis with Diffusion Models》中可以知道，在当时的方法中利用视频扩散模型去生成和编辑视频需要在保留现有结构的同时编辑现有素材内容，需要对每个输入进行较为昂贵的重新训练，或者需要跨帧图像编辑。而 Gen-1 提出了一个结构和内容感知模型，该模型可以根据示例图像或文本引导修改视频。编辑完全在推理时执行，无需额外的每个视频的训练或预处理。Gen-1 模型在大规模未配对视频和配对的文本-图像数据集上进行训练。同时，产品展示了通过训练不同细节级别的单目深度估计来控制结构和内容保真度。模型同时在图像和视频上进行训练，这也通过一种新颖的引导方法明确控制了时间一致性。

图 14：Runway Gen-1 视频生视频

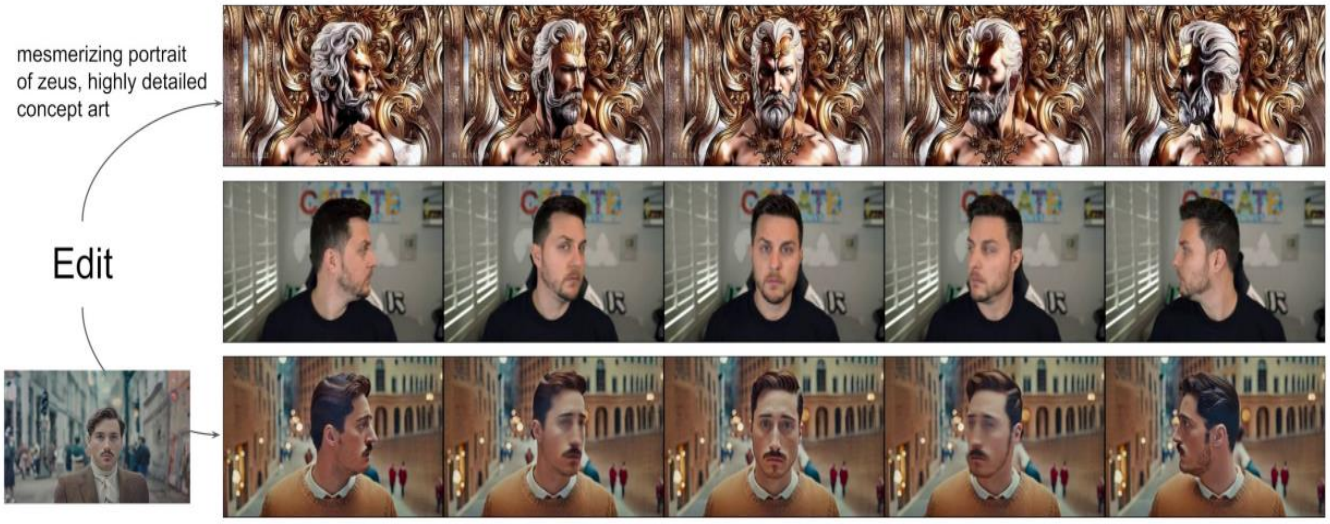


Figure 1. **Guided Video Synthesis** We present an approach based on latent video diffusion models that synthesizes videos (top and bottom) guided by content described through text (top) or images (bottom) while keeping the structure of an input video (middle).

资料来源：Patrick Esser 等《Structure and Content-Guided Video Synthesis with Diffusion Models》，信达证券研发中心

用户通过调节不同的视频风格、风格的变化程度、以及通过图片和文字 prompt 来修改视频。视频的一致性保持较好，但由于是早期的 gen-1 版本，视频分辨率较低。

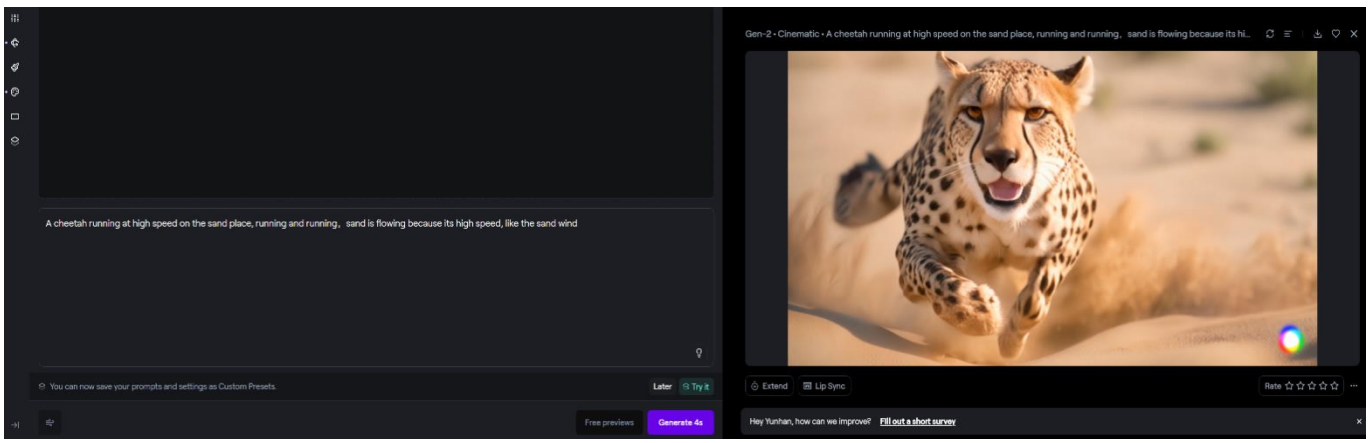
图 15：Runway Gen-1 视频生视频演示（左上为原始视频，右上为预览分镜头脚本，下图为素描风格的视频转换生成）

资料来源：Runway 官网，信达证券研发中心

Runway Gen-2(文生视频和图生视频)

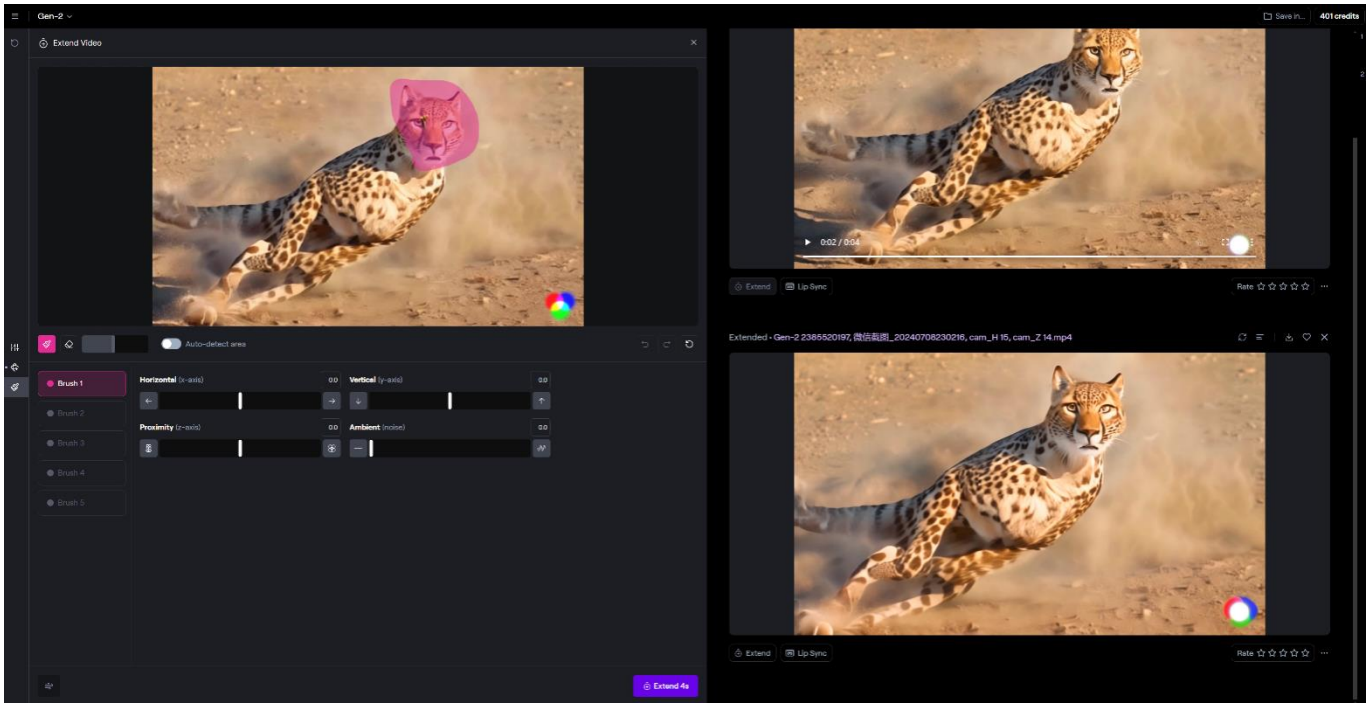
Gen-1 升级为文生视频以及图生视频功能。首先输入文本提示或上传图片；其次可以调整参数设置，可以使用固定种子数、升级和帧插值等高级设置来提高生成的一致性和分辨率；最后设置完成后，点击“生成”即可创建 4 秒的视频。此时，可以选择将镜头延长至总共 16 秒。**除了生成视频之外，Gen-2 为用户提供了更多的视频编辑功能，如运动画笔(为特定区域和主体带来动作和意图的生成)、相机控制（选择相机移动的方向和强度，如缩放、倾斜和平移）、通用运动（控制场景中的一般运动，包括相机和拍摄对象的运动）、延长视频（延长至 16s）、唇形同步（通过添加人物语言，让人物表情富有生命感）**。经实际体验，Gen-2 在文生视频和图生视频的物体物理规律性、视频一致性、分辨率等要素保持相对较好。

图 16：Runway Gen-2 文生视频效果表现较好



资料来源：Runway 官网，信达证券研发中心

图 17：Runway Gen-2 图生视频效果及笔刷功能表现较好

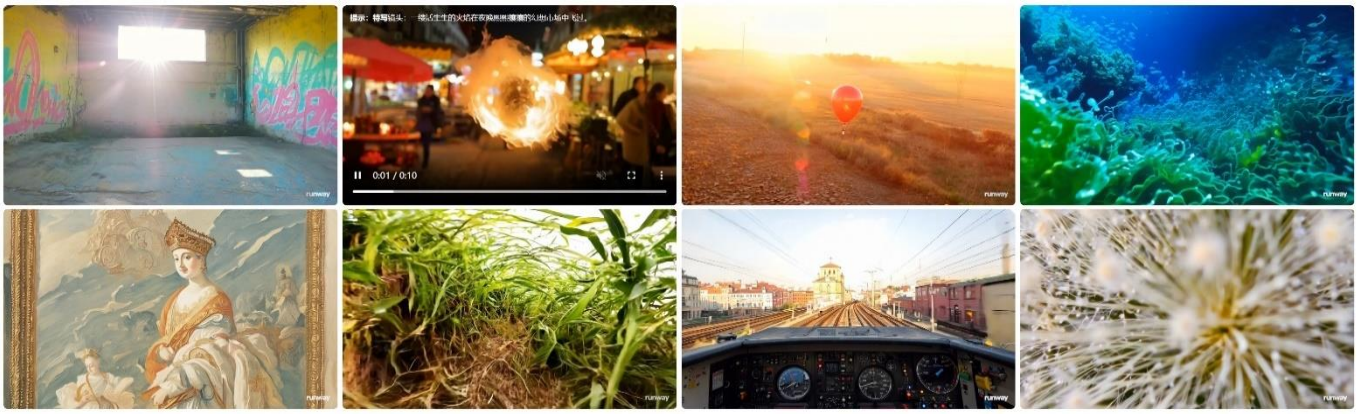


资料来源：Runway 官网，信达证券研发中心

Runway Gen-3 Alpha

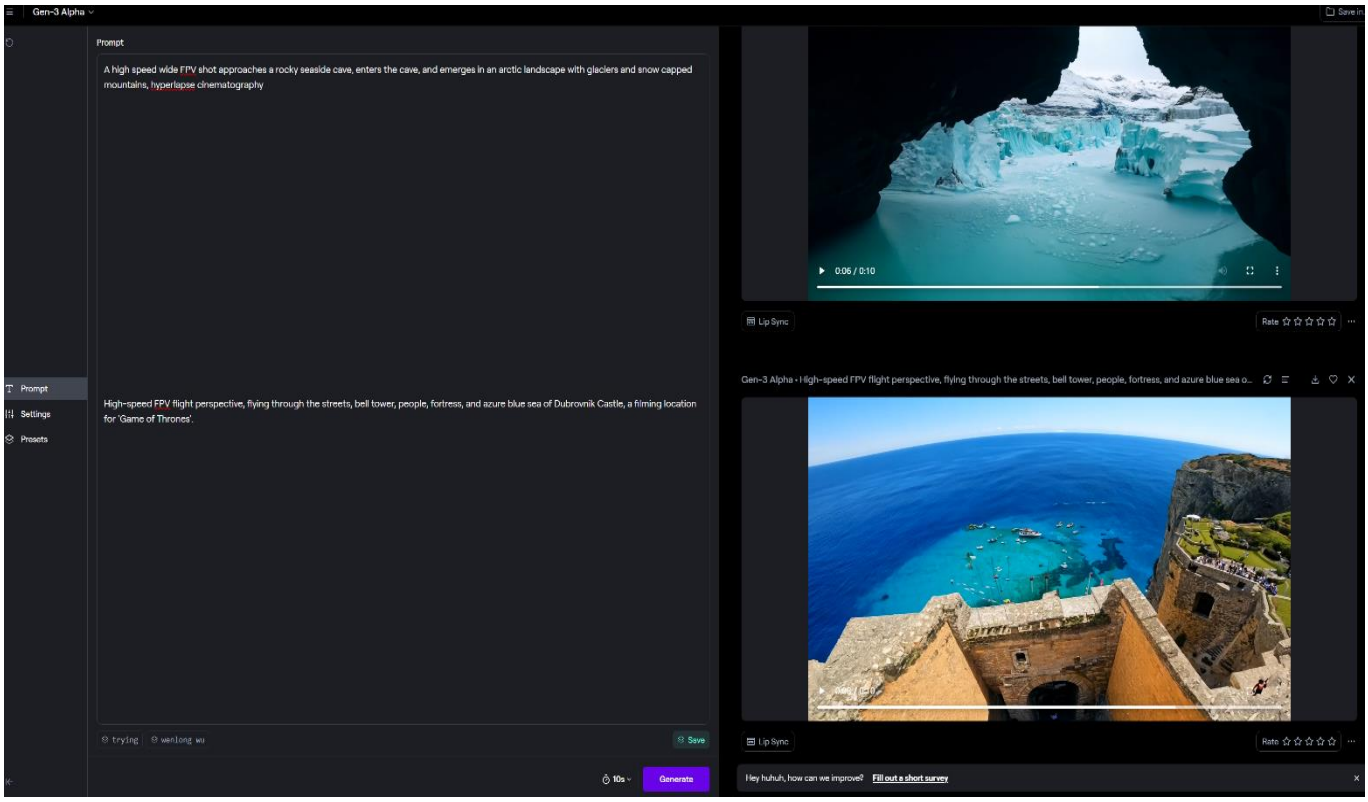
2024 年 6 月 17 日，Runway 推出了第三代视频生成模型 Gen-3 Alpha，与 Gen-2 相比，它在保真度、一致性和运动方面有了重大改进。Gen-3 Alpha 经过视频和图像联合训练，将为 Runway 的文本转视频、图像转视频和文本转图像工具、现有控制模式（如运动画笔、高级相机控制、导演模式）以及即将推出的工具提供支持，以更精细地控制结构、风格和运动。Gen-3 Alpha 将发布一套新的保护措施，包括全新改进的内部视觉审核系统和 C2PA 出处标准。1）细粒度的时间控制：Gen-3 Alpha 已接受过高度描述性、时间密集的字幕的训练，能够实现富有想象力的过渡和场景中元素的精确关键帧；2）逼真的人类角色创造：Gen-3 Alpha 擅长创造具有多种动作、手势和情感的富有表现力的人类角色，从而开启新的故事讲述机会；3）可诠释各种风格和电影术语；4）支持行业定制。Runway Gen-3 Alpha 暂时没有免费版本使用，目前收费标准为 144 美金/年，用户可以选择 5s/10s 的视频生成时长。**综合体验后，我们发现，Gen-3 Alpha 对提示词的理解、视频生成的质量（720p）、生成所需时长、视角等方面均表现较为出色，已然达到了行业头部水准**。

图 18：Runway Gen-3 Alpha 通过运动画笔、高级相机控制、导演模式可以更精细控制结构、风格和运动



资料来源：Runway 官网，信达证券研发中心

图 19：Runway Gen-3 Alpha 两端提示词测试，效果较强



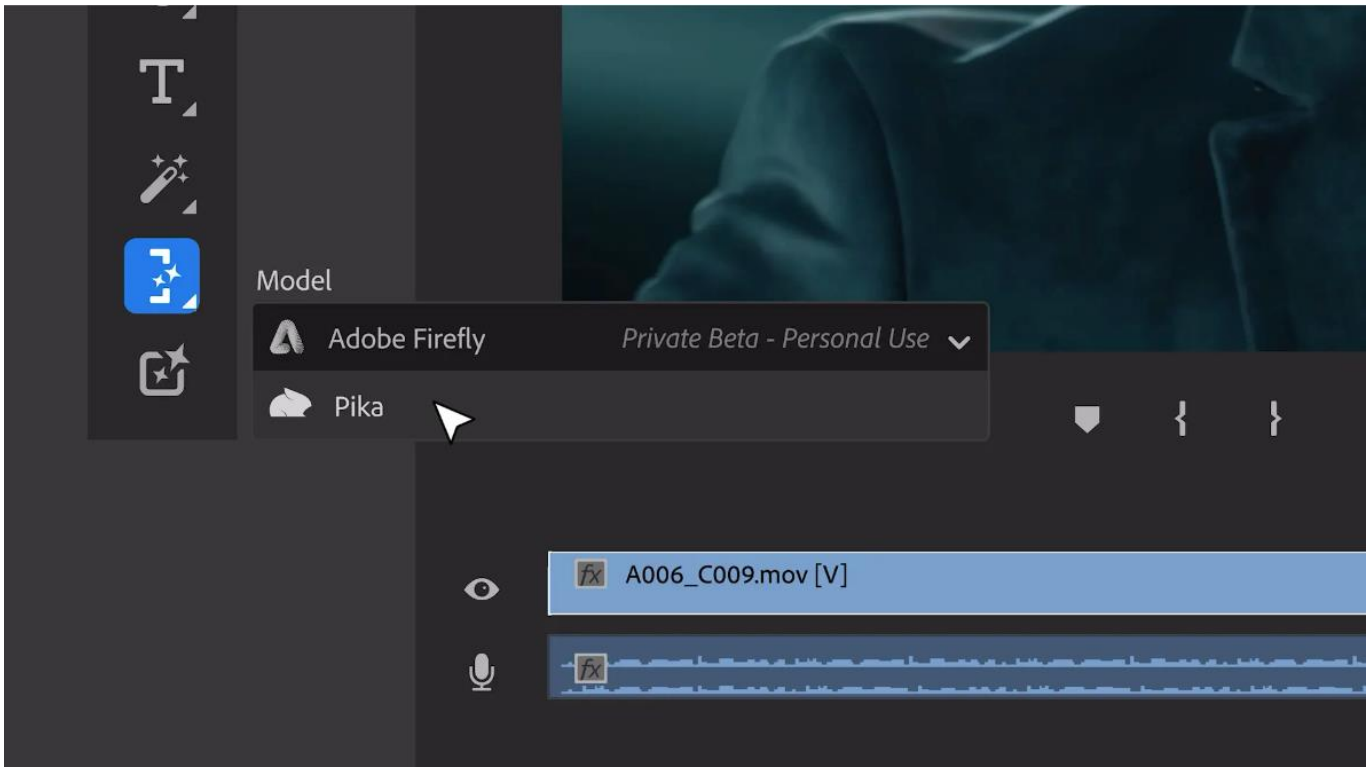
资料来源：Runway 官网，信达证券研发中心

（三） Pika

Pika 是一家利用人工智能重新设计整个视频制作和编辑体验的公司。虽然其他平台专注于为专业人士和企业提供支持，但 Pika 让所有创作者都能创作出高质量的视频，在不到 6 个月的测试期内，Pika 已经帮助超过 50 万人实现了创意构想。Pika 由斯坦福大学 AI Lab 的博士生郭文景和孟辰霖于 2023 年 4 月创立。2023 年 7 月，Pika 开始内测，推出文本生成视频功能；2023 年 9 月，推出/animate 功能，进一步巩固领先地位。Lightspeed 领投 Pika 的 3500 万美元 A 轮融资。在前期三轮融资中筹集了 5500 万美元。2024 年 6 月，Pika 进行了 8000 万美元的 B 轮融资，使公司的总融资额达到 1.35 亿美元。在 Discord 上进行了秘密发布，发布了 1.0 版模型和 Web 应用，推出了多个首次面市的功能，公司团队也从 3 人增加到了 13 人。在订阅模式中，公司同样采用了四种计划：基础版、标准版、无限值版、高级版。其中基础版免费使用，有 credits 限制；标准版本（每年 96 美金）、无限制版（每年 336 美金）的 credits 数量增加，延长 4s 视频时长、无水印等；高级版（每年 696 美金）对于 credits、视频生成时长及其他 AI 功能的使用基本无任何限制。

2023 年 11 月，发布首款 AI 视频生成产品 Pika 1.0，引起业界轰动。Pika 1.0 使任何人都可以：只需输入即可凭空生成高质量视频；将视频延长至任意长度（每次添加 4 秒，无限次添加到任何剪辑）；通过修复即时修改任何视频的某个方面；通过外画功能将视频扩展至任意内容或宽高比；甚至调整摄像机的移动。海外 AI 设计巨头公司 Adobe 在 Document Cloud 和 Digital Experience 产品中与第三方 AI 模型合作，现在正在探索在 Creative Cloud 中添加非 Adobe AI 模型。此前 Adobe 展示了一些早期的“预览”，展示了专业视频编辑未来如何利用 Premiere Pro 中集成的 Runway 或 Open AI Sora 视频生成模型来生成 B-roll 以编辑到项目中，或者如何使用 Firefly 或第三方模型（如 Pika）和 Generative Extend 工具在镜头末尾添加几秒钟。

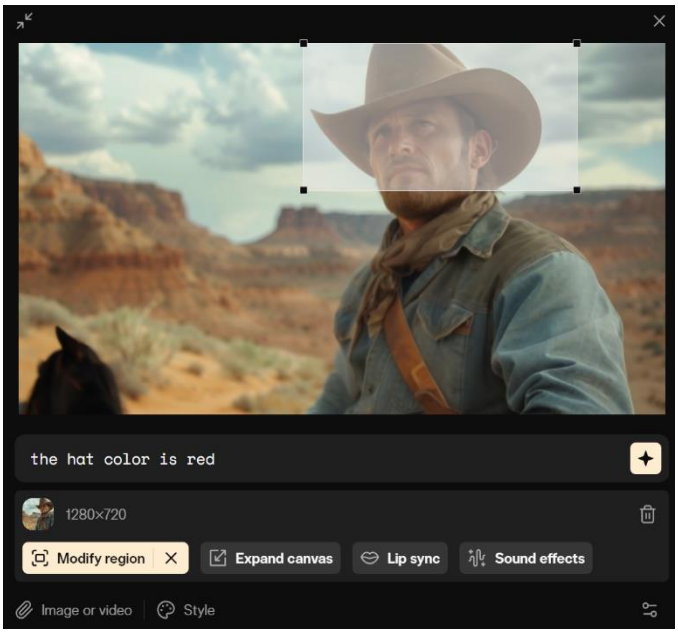
图 20：Adobe 产品中引入第三方视频模型 Pika 优化用户体验

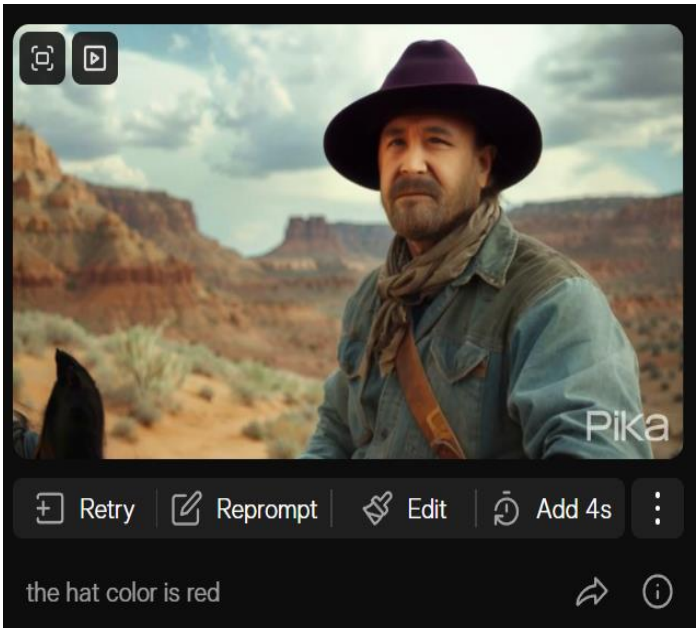


资料来源：Adobe Blog，信达证券研发中心

Pika 目前在文生视频的基础上能实现的功能包括：通过提示词修改指定区域框、更改视频风格、更改视频尺寸、丰富人物面部表情以及通过文字生成音乐给增添音效。实测下来，在保证基本效果的前提下，产品更加符合用户的使用习惯，细节打磨的更多。

图 21：Pika 文生视频界面及视频编辑核心功能





资料来源：Pika 官网，信达证券研发中心

（四） OpenAI Sora

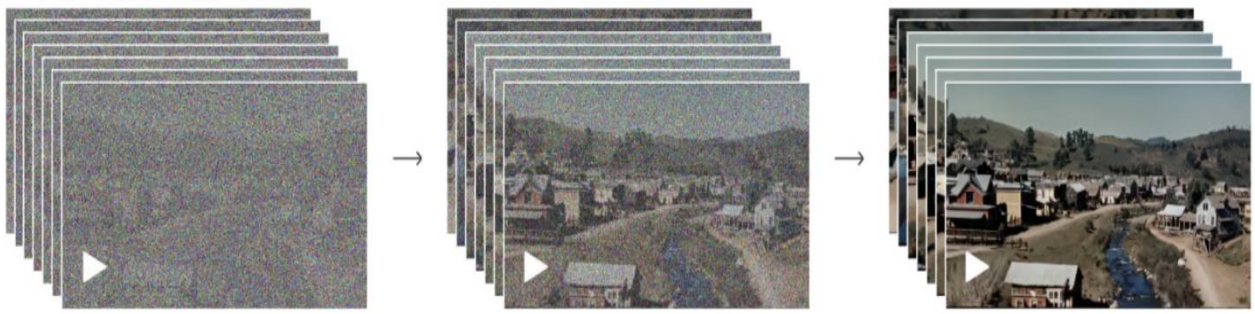
2024 年 2 月 16 日，OpenAI 在官网发布了创新性文生视频模型—Sora。从官网展示的 Sora 生成视频来看，在生成视频质量、分辨率、文本语义还原、视频动作一致性、可控性、细节、色彩等方面表现良好，并且最长可以生成 1 分钟的视频。至此，ChatGPT 已经具备了文本、图像、视频、音频 4 大多模态功能。继 Runway、Pika、谷歌和 Meta 之后，OpenAI 正式加入到这场 AI 视频生成领域“战争”当中。

图 22：Sora 合成的 60 秒视频

资料来源：OpenAI 官网，信达证券研发中心

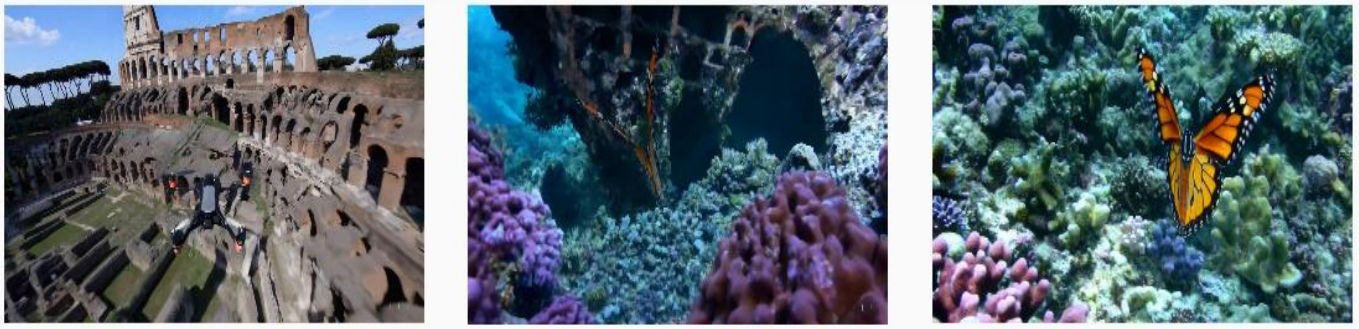
Sora（Transormer+Diffusion，DiT 架构）是一种扩散模型，主要通过静态噪音的视频开始视频生成，然后通过多个步骤去除噪音，最后转换为视频。同时，Sora 采用与 GPT 模型类似的 Transformer 架构，使用了 DALL-E 3 中的重述技术，能够精准还原用户的文本提示语义。Sora 的功能除了文本生成视频之外，还包括根据图像生成视频、对图像进行动画处理、提取视频中的元素、扩展或填充缺失的帧。

图 23：OpenAI 扩散模型过程



资料来源：OpenAI 官网，信达证券研发中心

图 24：Sora 可进行多个视频的组合



资料来源：OpenAI 官网，信达证券研发中心

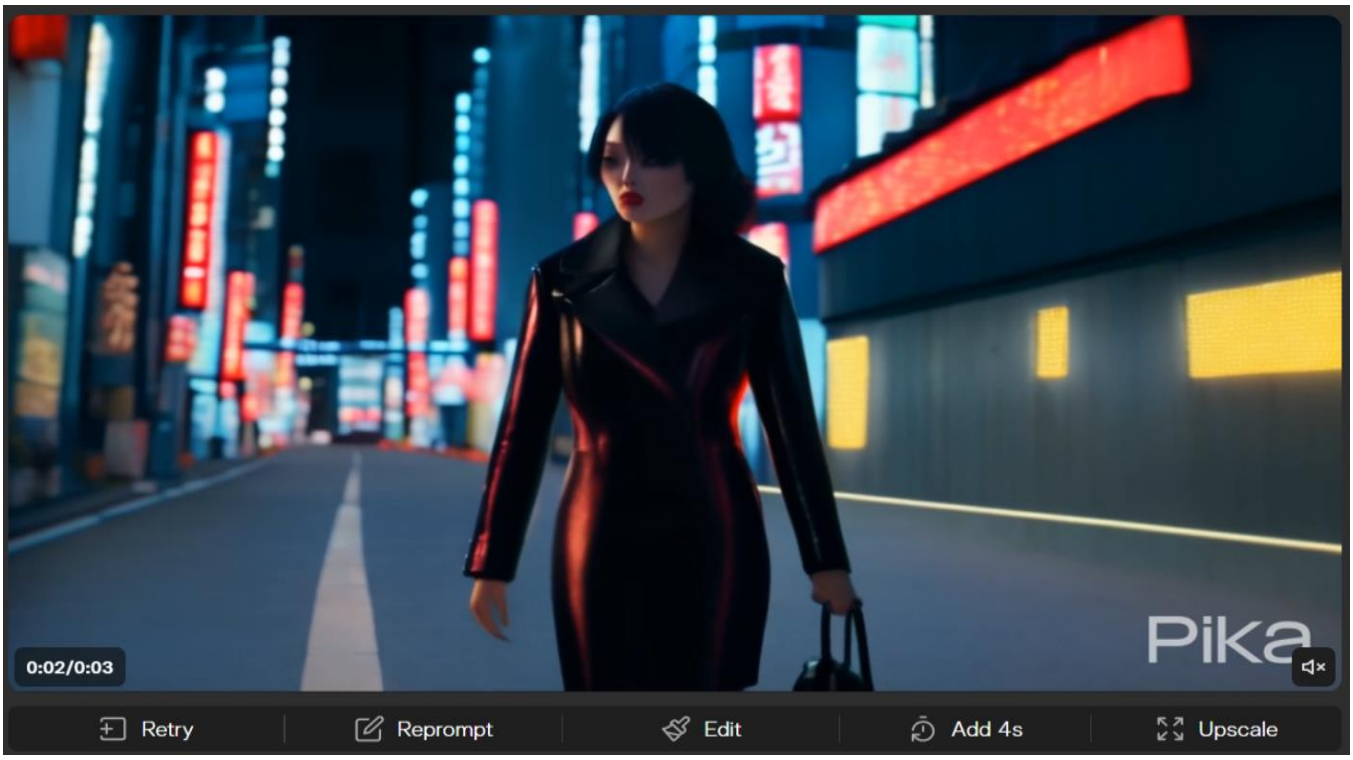
Sora 可以对宽屏 1920x1080p 视频、垂直 1080x1920 视频以及介于两者之间的所有视频进行采样。这样，Sora 就可以直接以原始纵横比为不同设备创建内容。还让用户能够快速制作较小尺寸的内容原型，然后再以全分辨率生成内容。Sora 还可以通过其他输入进行提示，例如预先存在的图像或视频。此功能使 Sora 能够执行各种图像和视频编辑任务 - 创建完美循环的视频、为静态图像制作动画、向前或向后延长视频时间等。此外，Sora 还能保持较长视频的连贯性和对象持久性，Sora 有时还能模拟以简单的方式影响世界状态的行为，例如，画家可以在画布上留下新的笔触，或者一个人吃汉堡时留下的咬痕。Sora 还能够模拟人工过程，比如视频游戏。过往很多生成式视频技术都是通过各种技术对视频数据进行生成模型建模，比如循环网络、生成对抗网络、自回归 Transformer 和扩散模型等方法。它们往往只关注于特定类型的视觉数据、较短的视频或者固定尺寸的视频。而 **Sora 是一种通用的视觉数据模型，能够生成各种持续时间、宽高比和分辨率的视频和图片，甚至长达一分钟的高清视频，对影视的宣传片、短视频切片、动画电影的降本增效具备里程碑意义**。

我们利用 Sora 官网一段知名的 AI 生成视频的提示词来进行横向同类比较，包括 Luma AI、Runway Gen-3 Alpha、Pika 和 Sora 关于生成效果各方面的对比。相同的 Prompt 提示词：“A stylish woman walks down a Tokyo street filled with warm glowing neon and animated city signage. She wears a black leather jacket, a long red dress, and black boots, and carries a black purse. She wears sunglasses and red lipstick. She walks confidently and casually. The street is damp and reflective, creating a mirror effect of the colorful lights. Many pedestrians walk about.”

图 25：Luma AI Dream Machine 生成效果（电影质感，略微不符合物理规律）

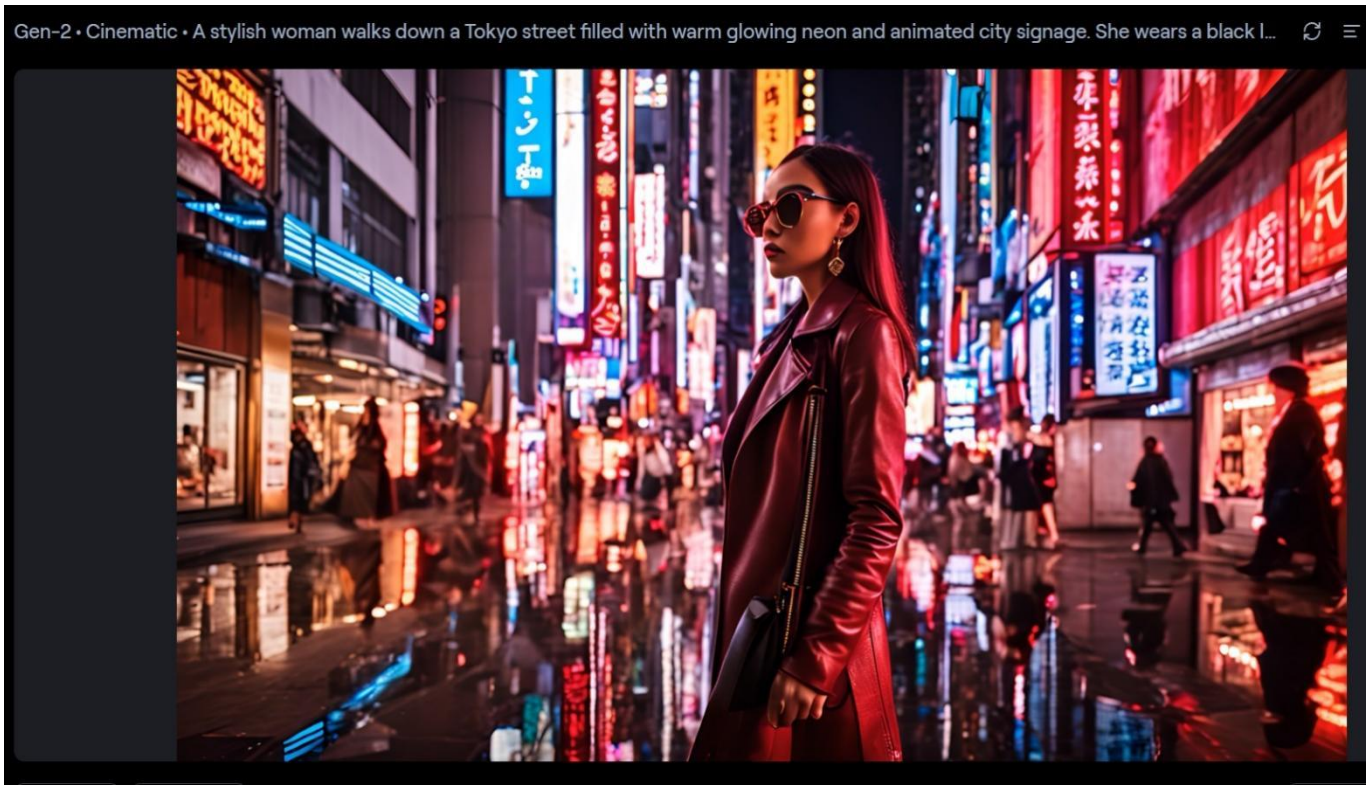
资料来源：Luma AI 官网，信达证券研发中心

图 26：Pika 生成效果（提示词理解、画面质感等方面有差距）



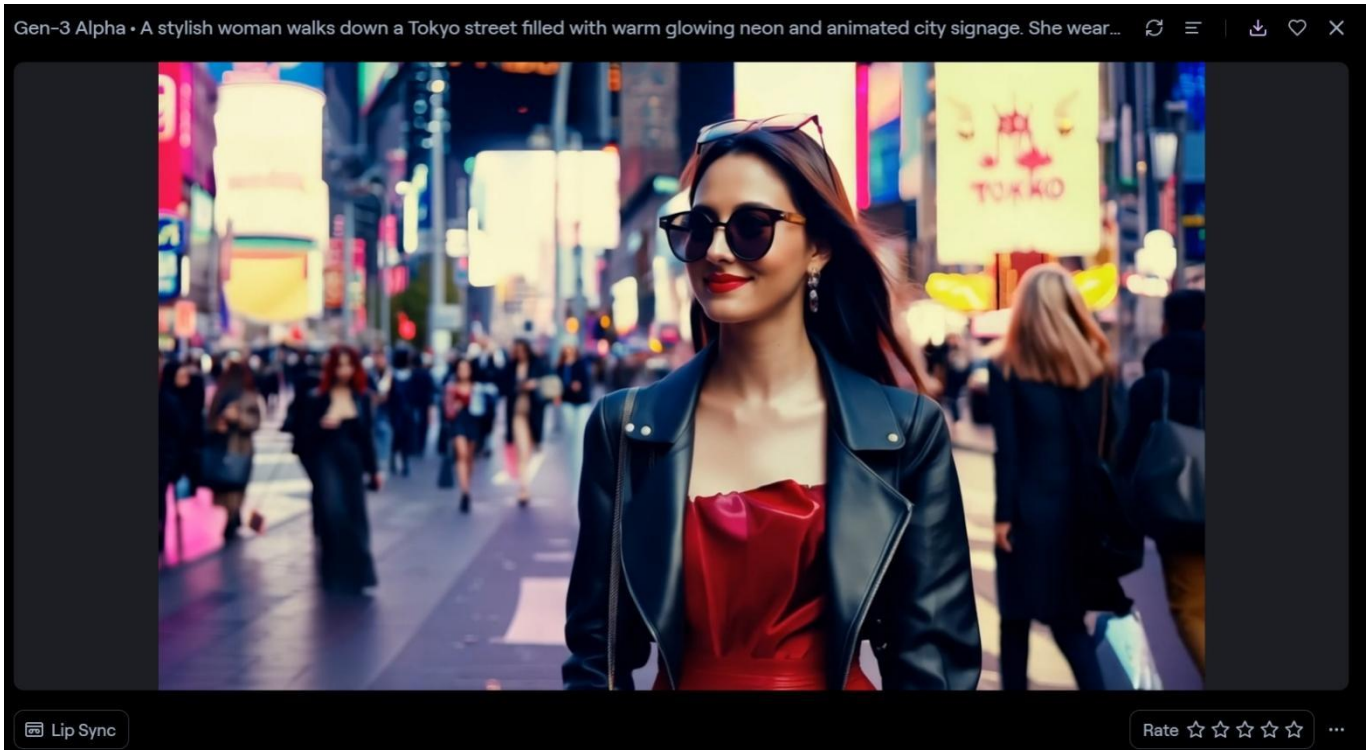
资料来源：Pika 官网，信达证券研发中心

图 27：Runway Gen-2 生成效果（主角没有跟随镜头移动）



资料来源：Runway 官网，信达证券研发中心

图 28：Runway Gen-3 Alpha 生成效果（各方面表现均优秀）



资料来源：Runway 官网，信达证券研发中心

表 3：Luma AI、Pika、Runway Gen-3 Alpha、Sora（暂未实测）关于以上相同提示词生成视频的效果多维度比较

	Luma AI	Pika	Runway Gen-3 Alpha	Sora
上线时间	2024.06	2023.11	2024.06	/
实测综合效果	中	低	高	暂未对外开放测试
分辨率	中	低	720p 高	/
生成时长/单次延长时间	5s/5s	3s/4s	可选 5s/10s	最长 60s
物理规律	中	低	高	/
提示词理解	高	低	高	
生成速度	中	高	高	
其他主要能力	提示词加强、延长时间、首尾帧图片生成等	提示词修改局部区域、改编视频画幅、人物添加表情视频、添加音效等	提示词长度无限制、给人物添加表情视频等	

产品定价	免费用户每月可生成 30 条视频； 标准版$23.99/月 Pro 高级版$79.99/月 Premier 最高级版399.99/月	免费用户初始 250 积分，10 积分可生成 3s 视频； 标准版$8/月 Unlimited 无限值版$28/月 Pro 高级版$58/月	免费用户初始 125 积分 标准版$12/月 Pro 高级版$28/月 Unlimited 无限制版$76/月 企业级定制详询	
最新融资金额	4300 万美元	8000 万美元	据外媒 The Information 报道为 4.5 亿美元	
估值情况	2-3 亿美元	4.7 亿美元	40 亿美元	
24.06 全渠道应用下载量	367,908	/	65,388	
24.04-06 网站拥挤度加总	21.28M	5.844M	16.01M	
24.06 平均月活用户数	549,871	半年达到 500,000 用户	322,691	
ARR	/	/	2500 万美元	
估值指数=估值/ARR	假设 55 万月活，付费率 10%，平均 arpu30 美金/月，则月收入为 165 万美元，假设年收入为 500 万美元，则 2.5 亿美元/500 万美元=50x	/	40 亿美元/2500 万美元=160x	
单活跃用户估值指数	2.5 亿美元/55 万=454.5	4.7 亿美元/50 万=940	40 亿美元/32 万=12500	

资料来源：Sensortower、Similarweb、各公司官网，信达证券研发中心（仅代表以上提示词生成视频横向比较，仅代表信达证券预测）

表 4：海内外视频生成产品单视频所需价格比较（1 美元=7.28 人民币）

	Luma AI	Pika	Runway Gen-3 Alpha	快手可灵	剪映即梦	爱诗科技 Pixverse V2
虚拟道具	/	credits 积分	credits 积分	灵感值； 1 元人民币=10 灵感值	积分； 10.87 人民币=100 积分	credits 积分

免费用户	10 个视频生成	250 初始积分，每日 30 积分	无免费版	66 个（24h 过期）	60 积分（24h 过期）	100 初始积分，每天 50 积分
生成耗时	5 分钟	15 分钟+	60s 生成 5s 的 720p 视频	2-5 分钟	1 分钟	2-5 分钟
单次视频时长	5s	3s	5s/10s	5s	3/6/9/12s	5s
单个视频生成消耗单位虚拟道具数量	付费会员没有生成视频数量限制	10 credits	625 积分=125s gen2 视频	10 个灵感值	3 积分	15/30 积分
年基础会员费用	287.9 美元/年	96 美元/年，每月获得 700 积分+每天 30 积分，共 1600 积分	144 美元/年，每月获得 625 积分	限时基础黄金会员 396 元/年，每月获得 660 灵感值	659 元/年，每月获得 2020 积分，每天赠送 60 积分，共 3820 积分	48 美元/年，每月获得 1000 积分，每天获得 50 积分，共 2500 积分
会员每月可生成视频数量	150 个	1600/10=160 个	125/5=25 个 gen-2 视频	660/10=66 个	3820/3=1273 个	2500/15=167 个
单条视频生成所需价格	**0.16 美元（1.17 人民币）**	**0.05 美元（0.364 人民币）**	**0.48 美元（gen2，3.49 人民币）**	**0.5 元人民币**	**0.04 元人民币**	**0.02 美元（0.174 人民币）**

资料来源：Runway、Luma AI、Pika、可灵、即梦、Pixverse AI 官网，信达证券研发中心

三、目前国内市场主流的生成式 AI+视频参与者

快手—可灵 AI（Diffusion Transformer 架构）

快手的大模型能力涵盖了包括大语言模型、文生图大模型、视频生成大模型、音频大模型、多模态大模型等核心技术方向，并基于快手丰富的业务场景，将生成式 AI 与多模态内容理解、短视频 /直播创作、社交互动、商业化 AIGC、创新应用等业务形态深度结合。可灵大模型的更新迭代速度较快，当视频生成效果接近图形渲染和视频拍摄时，有望对游戏、动画、泛视频行业带来新的机遇，有望促进视频平台生态繁荣。

1）**自研“快意大模型”（KuaiYii）**。13B、66B、175B 三种参数规模，将大模型应用于短视频场景下。

2）**可图大模型（KOLORS）**。由快手大模型团队自研打造的文生图大模型，具备强大的图像生成能力，能够基于开放式文本生成风格多样、画质精美、创意十足的绘画作品。“可图”主打三大核心特性：深入的中文特色理解、长文本复杂语义理解及对齐人类审美的精美画质，让用户低门槛创造高质量图像。

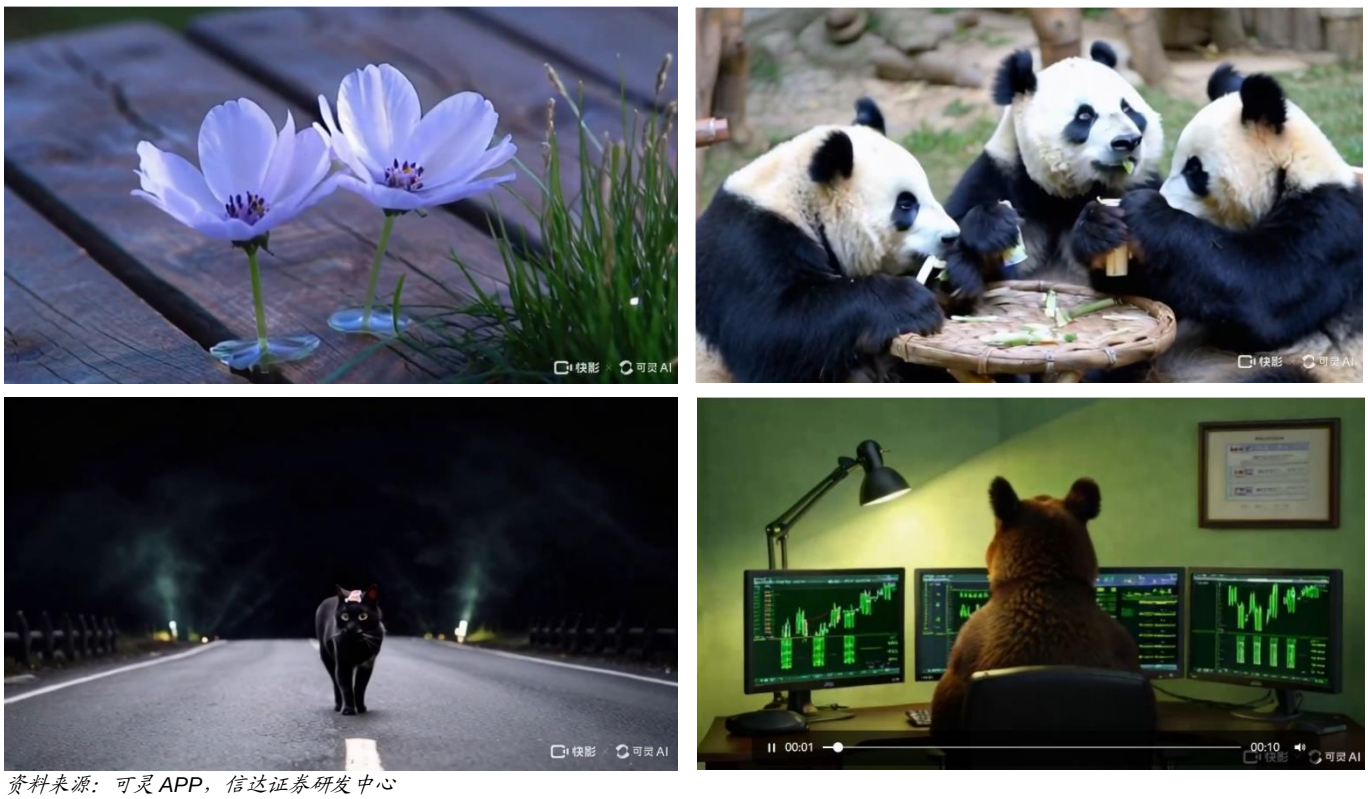
3）**可灵视频生成大模型**。2024 年 6 月 6 日，快手大模型团队自研打造了视频生成大模型—可灵，具备强大的视频生成能力，让用户可以轻松高效地完成艺术视频创作，**包含文生视频能力、图生视频能力及视频续写能力，后续有望上线视频编辑功能。可灵视频模型的重点方向在于：大幅度的合理运动符合物理规律、长达 2 分钟的视频生成能力帧率且达到 30fps、模拟物理世界特性、强大的概念组合能力、电影级别的画面、支持自由的输出视频高宽比。在 2024 年世界人工智能大会上，快手可灵 AI 产品宣布全新升级：高清画质、首尾帧控制、单次生成 10s、Web 端上线、镜头控制**。

图 29：快手大模型产品矩阵及可灵 AI 产品功能升级

资料来源：世界人工智能大会公众号，可图大模型公众号，信达证券研发中心

可灵 AI 经过我们长时间测试跟踪，APP 端的视频生成效果十分出色，无论是在提示词理解、物理规律控制、画质分辨率、生成速度时长、产品使用容易度和产品迭代升级速度上均表现较为亮眼，是国内视频生成大模型产品的头部参与者。**在 APP 端，用户可以选择参数设置：视频时长 5s/10s、高性能（生成速度更快，生成等待时长 4 分钟）或者高表现（画面质量更佳，生成等待时长 10 分钟，目前每天有 3 次机会）、视频比例（16：9、9：16、1：1）**。举例来看，下图左上的提示词：“木头上长出了两朵奇特的透明塑料花，花瓣闪闪发光，花瓣是淡紫色的，花瓣被风吹动 旁边有一棵草在摇曳，氛围光照”。左下图的提示词：“氛围光照，抽象背景，黑猫警长在光怪陆离的路上行走”。右上图提示词：“高清画质，四只带着墨镜的大熊猫在围着一个用竹子编织的桌子周围打扑克牌，同时悠闲的吃着竹子，喝着汽水。”

图 30：快手可灵文生视频

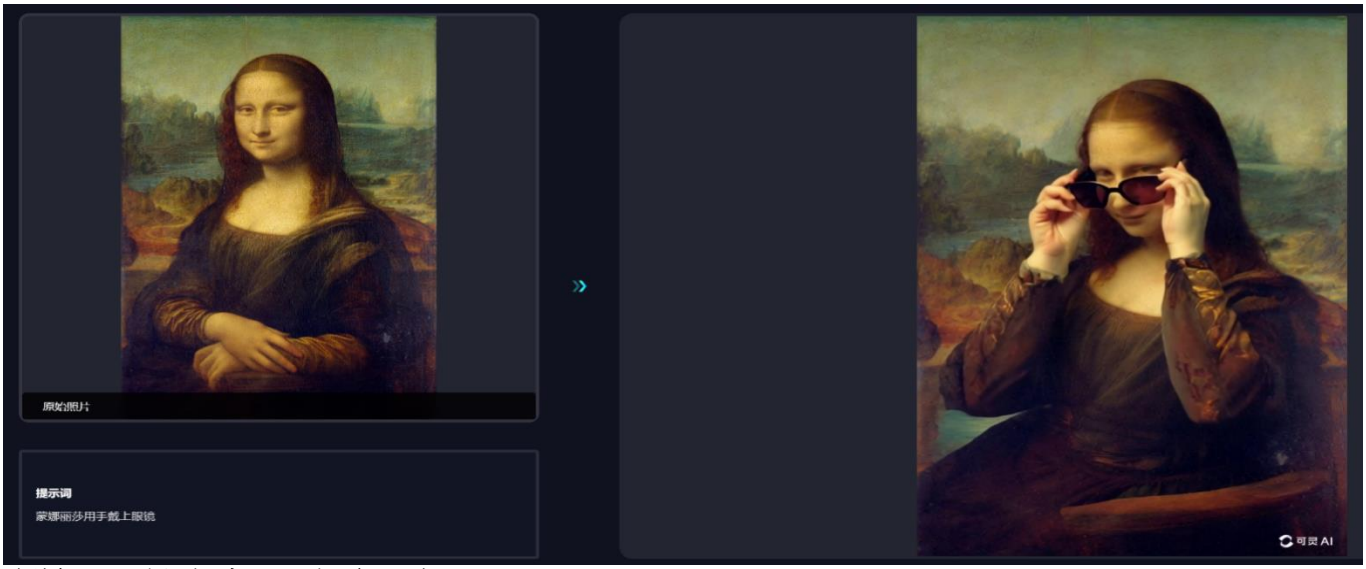


资料来源：可灵 APP，信达证券研发中心

2024 年 7 月 24 日起，可灵 AI 全面开放内测，正式上线了会员体系。用户每日可获得免费灵感值 66，24h 内过期，1 元人民币=10 灵感值。在为期 7 天的会员充值活动中，会员全线五折，其中包括非会员（登录每日赠送灵感值）、黄金会员（396 元/年，每月获得 660 灵感值，约生成 3300 张图片或 66 个高性能视频，包含去水印、高质量视频生成、视频延长、运镜升级功能）、铂金会员（1596 元/年，每月获得 3000 灵感值，约生成 15000 张图片或 300 个高性能视频，包含去水印、高质量视频生成、视频延长、运镜升级功能、新功能优先体验）、钻石会员（3996 元/年，每月获得 8000 灵感值，同样包含上述增值功能）。

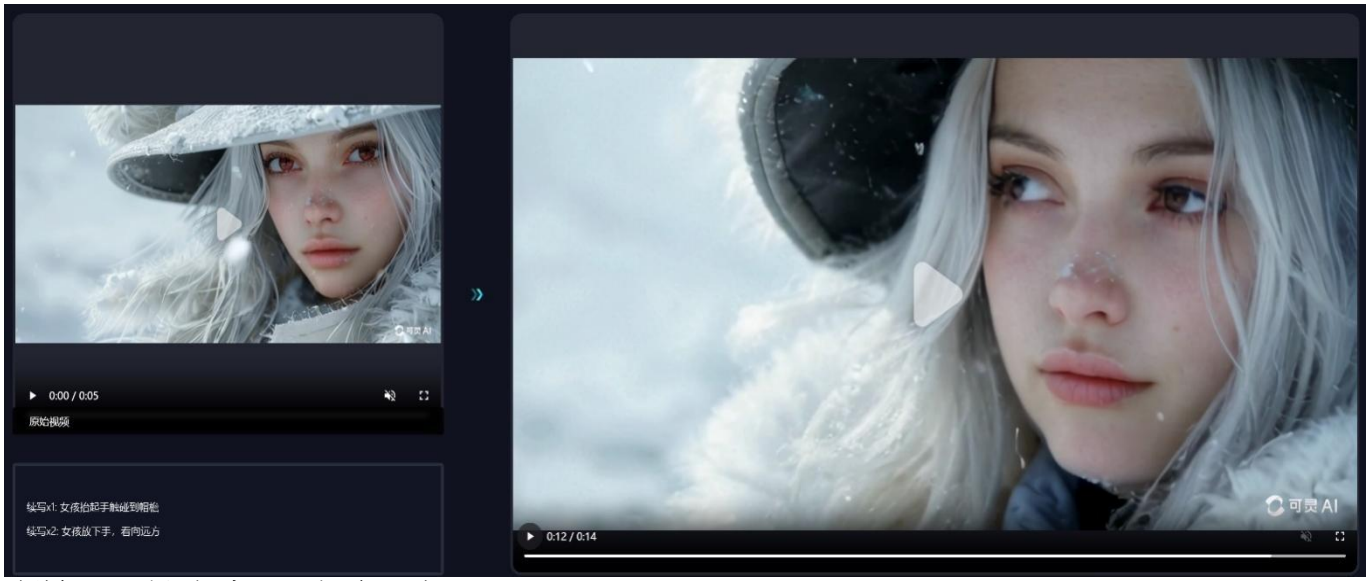
可灵图生视频功能：可灵图生视频模型以卓越的图像理解能力为基础，将静态图像转化为生动的 5 秒精彩视频。配上创作者不同的文本输入，即生成多种多样的运动效果。

图 31：快手可灵图生视频



资料来源：可灵大模型官网，信达证券研发中心

图 32：视频续写功能



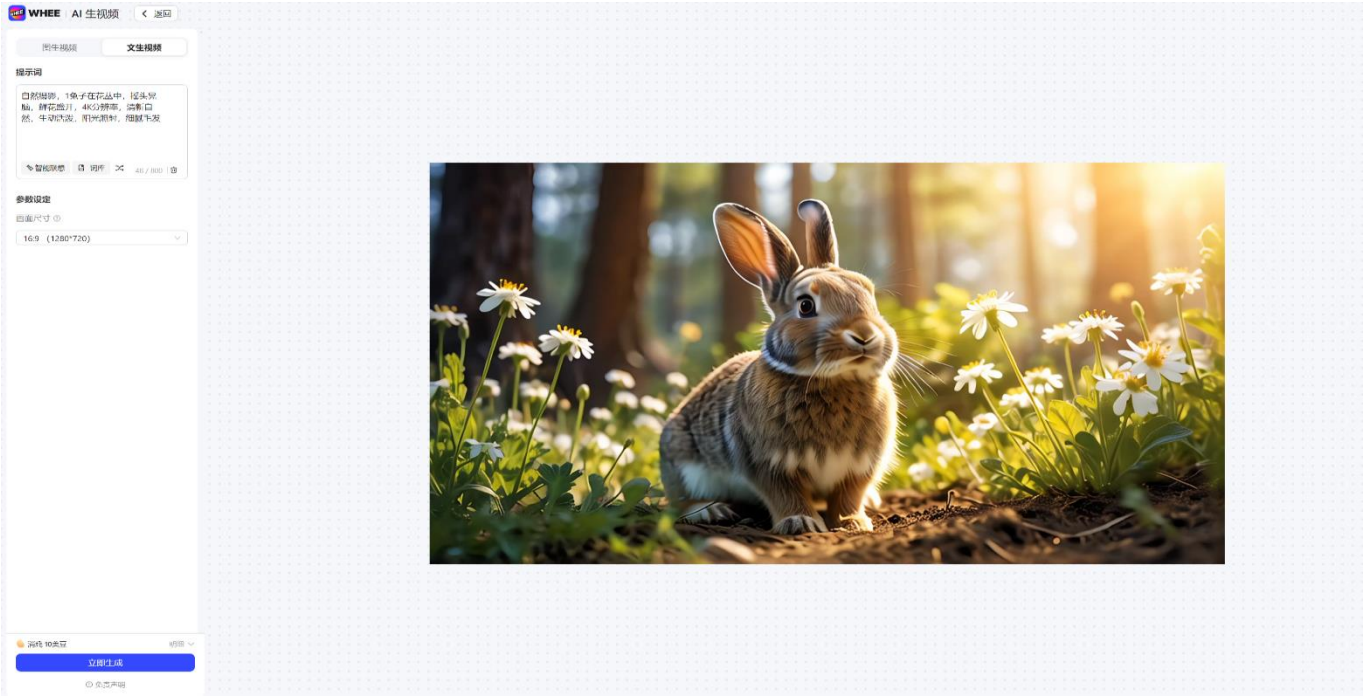
资料来源：可灵大模型官网，信达证券研发中心

美图 MiracleVision4.0 AI 视频

2023 年 12 月，美图公司发布自研 AI 视觉大模型 MiracleVision 4.0 版本，主打 AI 设计与 AI 视频。新增了文生

视频、图生视频、视频运镜、视频生视频四大能力。目前，MiracleVision 的 AI 视频能力已能融入行业工作流，尤其是电商和广告行业。MiracleVision4.0 于 2024 年 1 月陆续上线至美图旗下产品。目前生成一次视频需要消耗 10 美豆，实际测验下来看，其对提示词的理解、物体的像素质量、物理规律、动作的自然效果，尤其是对人物和物体的细节处理上较为优秀，例如动物的毛发帧数。图生视频功能：让图片也动起来。从景深变化到细节动作捕捉，MiracleVision 可以轻松生成。非常的自然流畅。图生视频的基础上，MiracleVision 支持视频运镜。提供了推、拉、摇、移等八种电影级运镜模式，让用户能够轻松模拟专业的镜头运动。后续有望更新视频生视频功能，导入一段视频，再加上不同的提示词，就能获得卡通、科幻、像素风，羊毛毡等不同的艺术风格。

图 33：美图 Whee AI 生视频功能



资料来源：Whee 官网，信达证券研发中心

PixVerse 爱诗科技

爱诗科技 Alsphere 成立于 2023 年 4 月，海外版产品 PixVerse 于 2024 年 1 月正式上线，目前已是全球用户量较大的国产 AI 视频生成产品，上线 88 天，PixVerse 视频生成量已达一千万次。公司早期完成数千万人民币天使轮融资，2024 年 3 月公司完成亿级人民币 A1 轮融资，国内一线投资机构达晨财智领投。创始人王长虎博士深耕计算机视觉与人工智能领域 20 年，带领字节跳动视觉技术团队在巨量规模的用户数据下，解决了多个视觉领域的世界级难题，并从 0 到 1 参与抖音与 Tik Tok 等国民级视觉产品的建设和发展，公司团队成员来自清华、北大、中科院等顶级学府，曾任职于字节、微软亚洲研究院、快手、腾讯的核心技术团队。基于“数据、算法和工程”三大要素，解决“准确性”和“一致性”，用更少资源取得更优效果。公司致力于通过——“融合内容理解与生成；融合文字、图片、视频等多模态”的双融合技术路径，搭建世界一流的 AIGC 视觉多模态大模型。

2024 年 1 月，正式发布文生视频产品 PixVerse 网页版，PixVerse 产品页面月访问量超百万。

2024 年 2 月 18 日，根据《互联网信息服务深度合成管理规定》，国家互联网信息办公室公开发布第四批境内深度合成服务算法备案信息公告，爱诗科技视频生成算法成功通过备案。

2024 年 5 月 17 日，智源研究院举办大模型评测发布会，文生视频模型评测结果显示，爱诗科技旗下产品 PixVerse 位列全球 TOP3，在国内同类型产品中表现最佳。

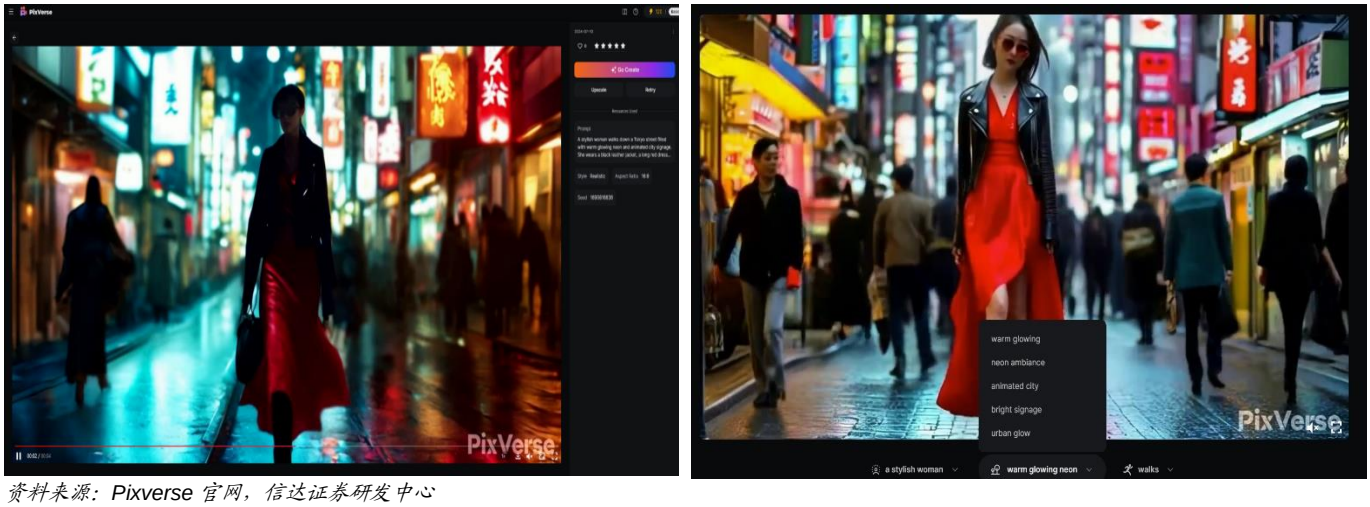
2024 年 5 月 31 日，PixVerse 正式上线 Magic Brush 运动笔刷功能。在图生视频过程中，用户可通过涂抹区域和绘制轨迹，精确控制视频元素运动方式移动，甚至丰富多样的整体动效。

2024 年 6 月 5 日，国内首张 AI 音乐专辑——GxTxPx（伟大科技的造物）正式发布，部分单曲已在网易云平台上线，视频大部分由爱诗科技旗下产品 PixVerse 制作完成。

2024 年 7 月 24 日，爱诗科技正式发布视频生成产品 PixVerse V2，全球同步开放。采用 Diffusion+Transformer（DiT）基础架构，在保证一致性的前提下，一次生成多个视频片段，可实现单片段 8 秒，和多片段 40 秒的视频生成。在人物一致性上布局较深，支持一键生成 1-5 段连续的视频内容，且片段之间会保持主体形象、画面风格和场景元素的一致。PixVerse V2 还支持对生成结果进行二次编辑，通过智能识别内容和自动联想功能，用户可以灵活替换调整视频主体、动作、风格和运镜，进一步丰富创作的可能性。

我们使用了其海外版产品 Pixverse V2 进行测试，V2 版本较 V1 效果优化较好。Pixverse V2 目前可根据多场景生成人物一致性较强的多个镜头视频，同时还可对生成的视频进行风格、人物编辑。

图 34：Pixverse 文生视频（左图为 V1，右图为 V2）

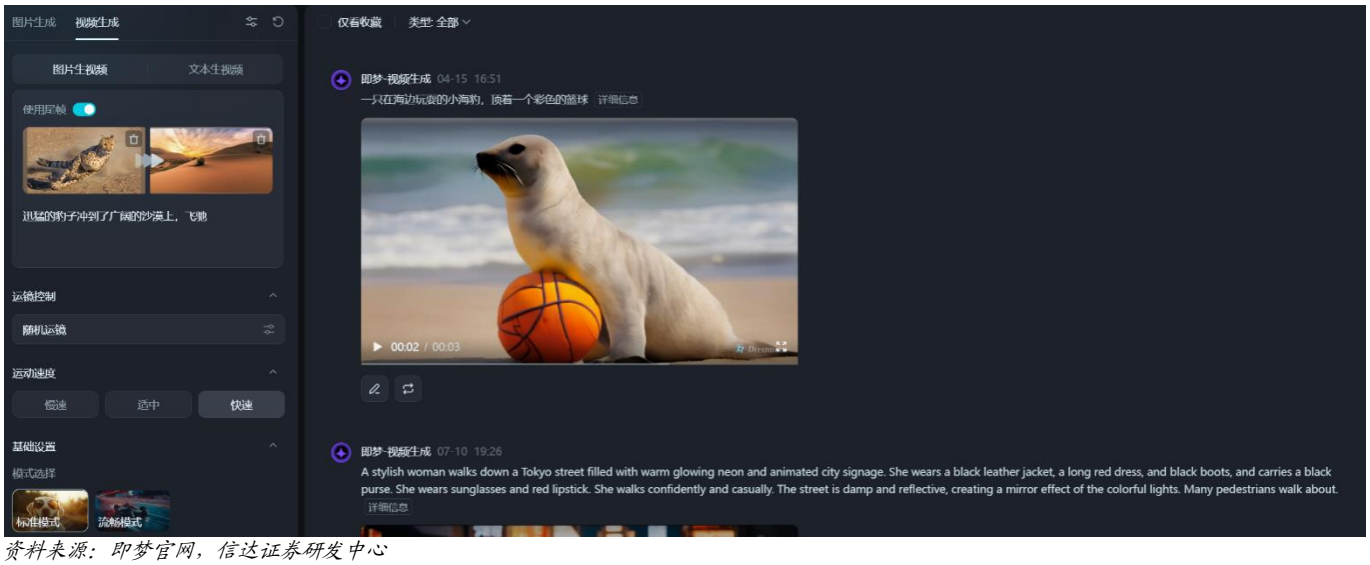


资料来源：Pixverse 官网，信达证券研发中心

即梦 Dreamina（字节剪映）

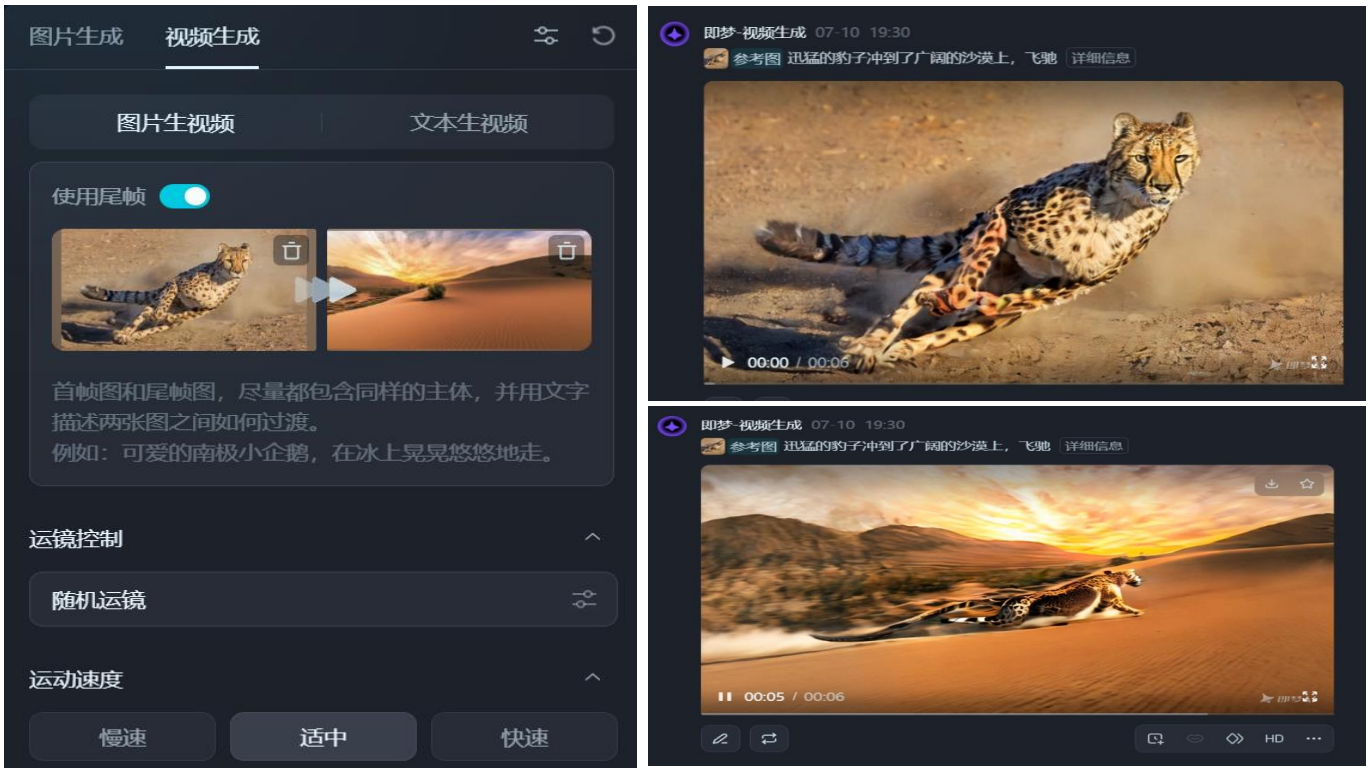
2024 年 5 月，字节剪映旗下针对 AI 创作产品 Dreamina 正式更名为中文“即梦”，AI 作图和 AI 视频生成功能已经上线，用户可输入文案或者图片，即可得到视频动态效果连贯性强、流畅自然的视频片段。创新打造首帧照片和尾帧照片输入方式，增强视频生成的可控性，支持中文提示词创作，把握语义。2024 年 6 月 17 日，上海国际电影节期间，由抖音、博纳影业 AIGMS 制作中心联合出品的 AIGC 科幻短剧集《三星堆：未来启示录》亮相“博纳 25 周年'向新而生'发布会”。即梦 AI 作为《三星堆：未来启示录》首席 AI 技术支持方，借助包括 AIGC 剧本创作、概念及分镜设计、图像到视频转换、视频编辑和媒体内容增强等十种 AIGC 技术，重新为古老 IP 注入新故事、开发新内容。**在产品使用界面，即梦添加了更多用户可控的细节功能，例如运镜控制的种类中，可自行选择移动方向、摇镜方向、旋转角度、变焦程度、幅度大小等，省去用户提示词中复杂的表述；用户还可自行选择运动速度、标准/流畅模式、生成时长和视频比例等，UI 界面更容易被用户接受，简单易行。**

图 35：即梦视频生成功能页面



资料来源：即梦官网，信达证券研发中心

图 36：即梦首尾帧土图生视频



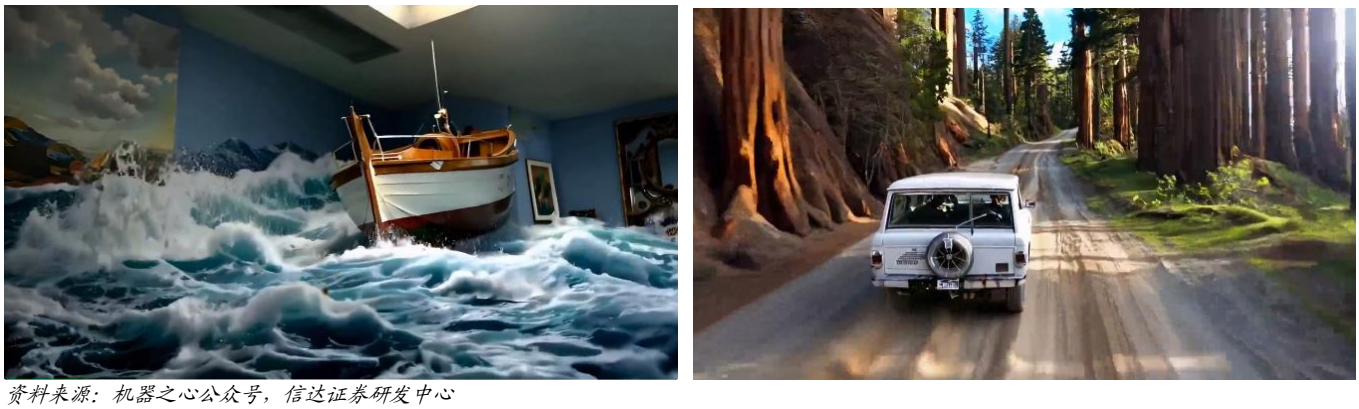
资料来源：即梦官网，信达证券研发中心

清华 Vidu

2024 年 4 月 27 日，在中关村论坛未来人工智能先锋论坛上，生数科技联合清华大学发布了具有“长时长、高一致性、高动态性”性能标签的视频大模型 Vidu，可根据文本描述直接生成长达 16 秒、分辨率达 1080P 的高清视频内容。“高一致性”是团队强调的重点方向。当前国内视频大模型的生成视频时长大多为 4 秒左右，Vidu 则可实现一次性生成 16 秒的视频时长。同时，视频画面能保持连贯流畅，随着镜头移动，人物和场景在时间、空间中能保持高一致性。在动态性方面，Vidu 的动态镜头在推、拉、移之外，开始涉及一段画面中远景、近景、中景、特写等镜头的切换，以及直接生成长镜头、追焦和转场效果。技术路线上，Vidu 采用的是自研 U-ViT 架构，与 Sora 一样是 Diffusion 和 Transformer 的融合架构。这种架构不采用插帧的多步骤处理方式来生成视频，而是

通过单一步骤“端到端”直接生成内容，从文本到视频的转换是直接、连续的。

图 37：Vidu 官方宣传生成视频（左图提示词：画室里的一艘船驶向镜头）

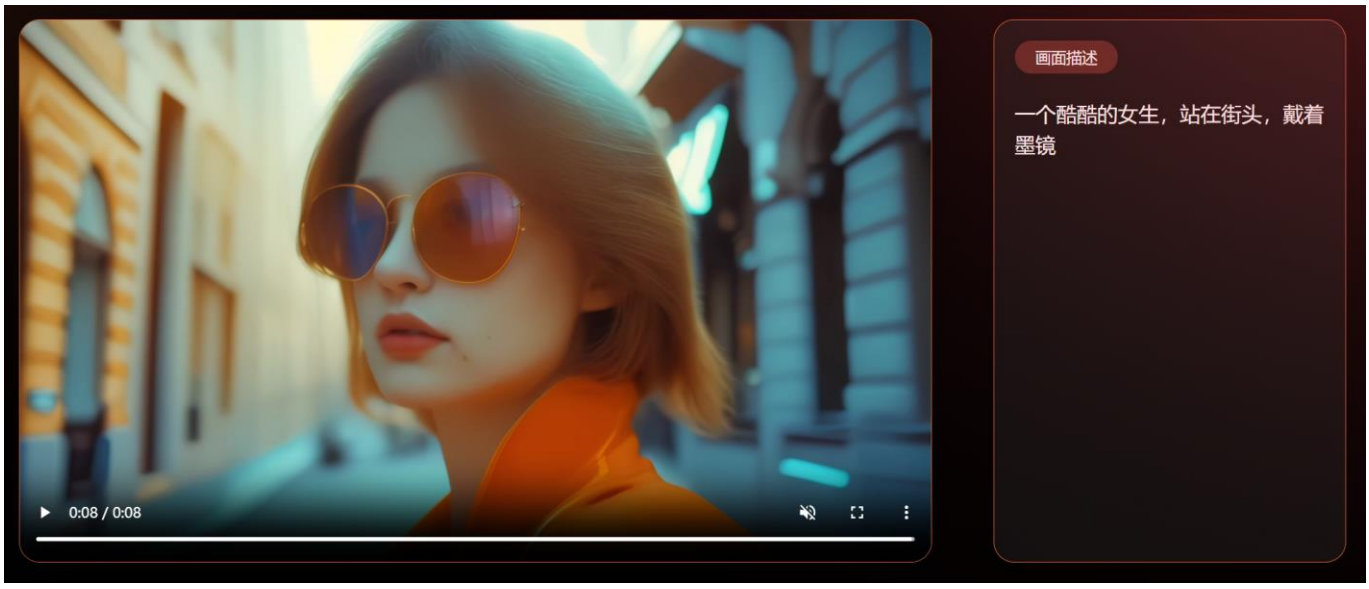


资料来源：机器之心公众号，信达证券研发中心

七火山科技 Etna

2024 年 1 月 16 日，超讯通信与七火山 Seven Volcanoes 签署投资合作协议。自 2023 年成立以来，Seven Volcanoes 一直致力于机器学习算法和深度神经网络技术的研究。2024 年 3 月 7 日，七火山 Etna 模型正式发布，Etna 模型采用最新的神经网络架构，融合了 Transformer 模型的强大语义理解能力，以及 Diffusion 模型的高效内容生成策略，旨在通过高度精确的文本到视频转换，目前暂未对外开放功能测试。

图 38：Etna 宣传用的文生视频效果



资料来源：七火山官网，信达证券研发中心

四、从 AI 生成到 AI 工作流，一站式视频生成+剪辑+故事创作有望成核心方向

一站式 AI 视频生成&剪辑&UGC 创作有望解决市场一直在质疑的“AI+视频没有实质作用问题”。

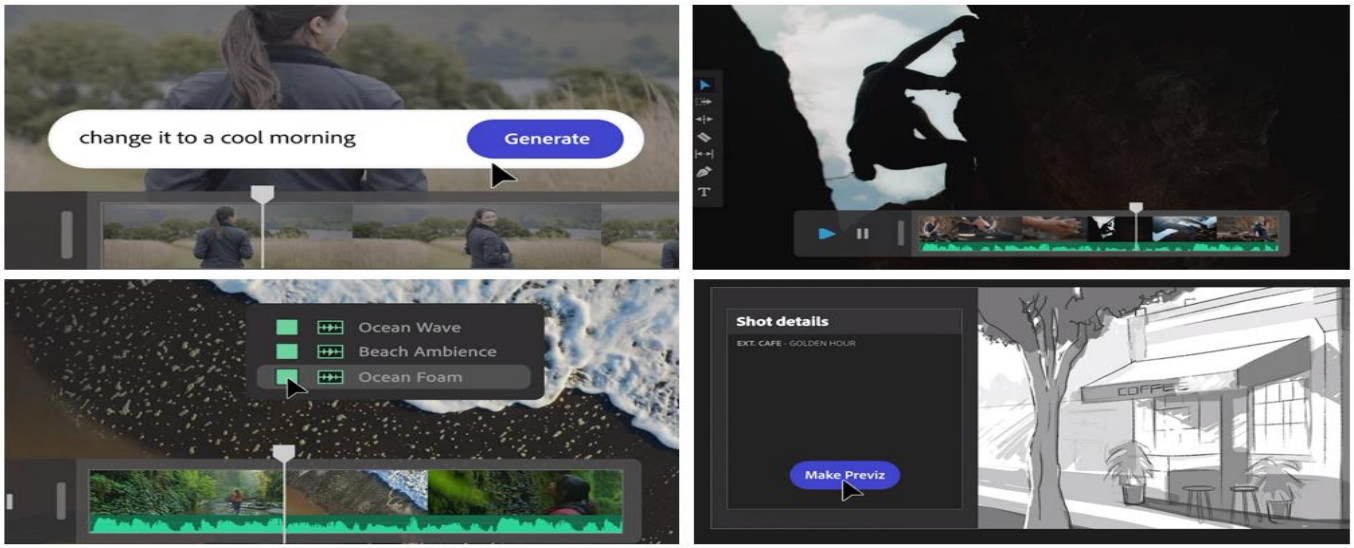
随着生成式 AI 自身大模型技术的迭代、算法的优化、视频数据质量和数量的提升，生成式 AI+视频的发展、竞争正逐渐激烈化。我们认为，在 AI 视频生成领域，底层技术迭代是行业持续发展的前提，**但在迭代技术的同时，我们需要深入思考下，后续技术应用的方向、衍生出哪些商业模式、什么类型的公司会最终受益于生成式 AI+视频的技术红利。目前，AI+视频大概率用于创意设计、创意生成，直接用于 ToB 商业化较少。追溯原因，我们发**

现目前主流 AI 视频工具还处在视频生成竞争的阶段，且大多数为单一功能产品。在视频生成之后，诸如准确的提示词生成、修改视频片段、添加字幕、脚本生成、转场衔接、背景音乐添加等众多细节功能暂未集成，因此现今阶段还需要多种不同的视频创作工具串联使用才能达到直接输出可商业化视频的效果，环节繁琐、多工具之间的格式也可能存在不兼容的可能性，给用户带来使用上的不便。因此我们认为，后续需要持续关注能够一站式提供视频生成+编辑等功能的企业，了解用户痛点，打磨产品细节，才能真正将技术用于生产工作、娱乐等众多环节，带来商业化变现的潜在空间。目前我们可以看到，除了主流公司例如 Sora、Luma AI、Pika、Runway 在积极迭代视频生成能力之外，有一些企业如 Adobe、Heygen、Capitions.AI、OpusClip、快手可灵、字节剪映等诸多工具已经在尝试在 AI 视频剪辑方向发力。

Adobe Firefly& Adobe Express

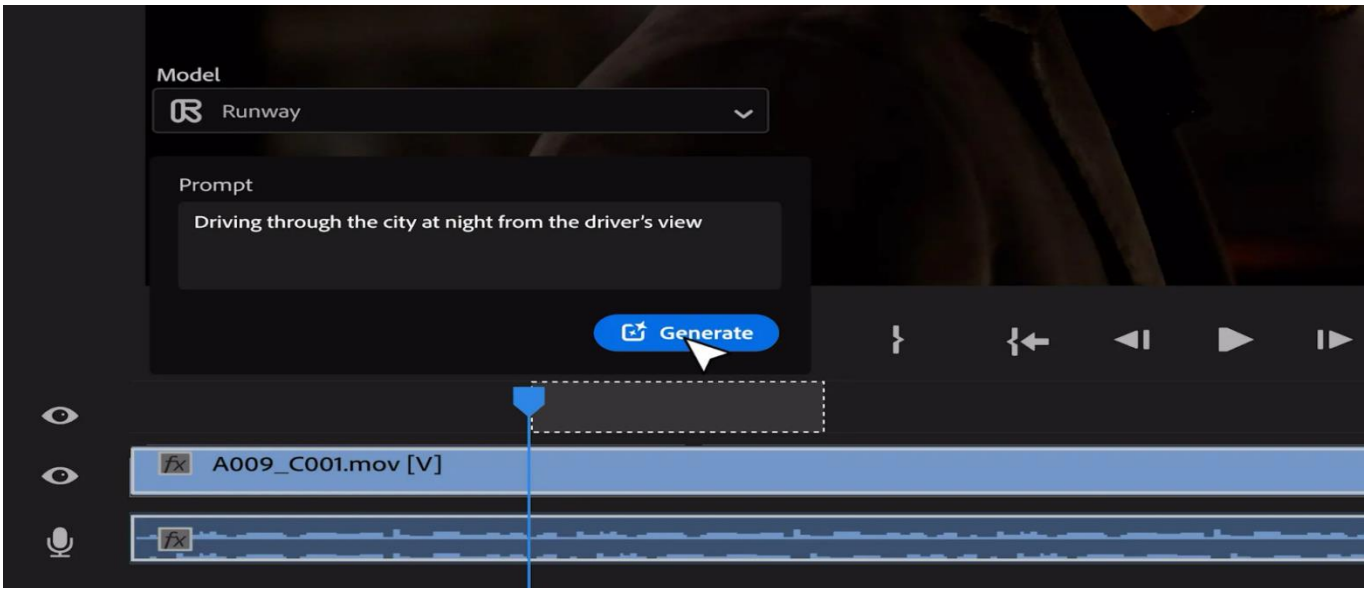
2023 年 4 月，Adobe 发布了一篇关于音频、视频、动画和动态图形设计的功能展示的官方 blog。主要可实现的功能包括：视频编辑、添加背景音乐和效果、脚本字幕文字的自动生成匹配、根据文字分镜头展示、从草图生成动画等和音视频相关的 AI 功能等。

图 39：后续 Firefly 关于多模态音频、视频方向上的功能展望



资料来源：Adobe 公司官网，信达证券研发中心

图 40：Adobe Firefly 集成第三方大模型如 Runway、OpenAI Sora 用于视频剪辑



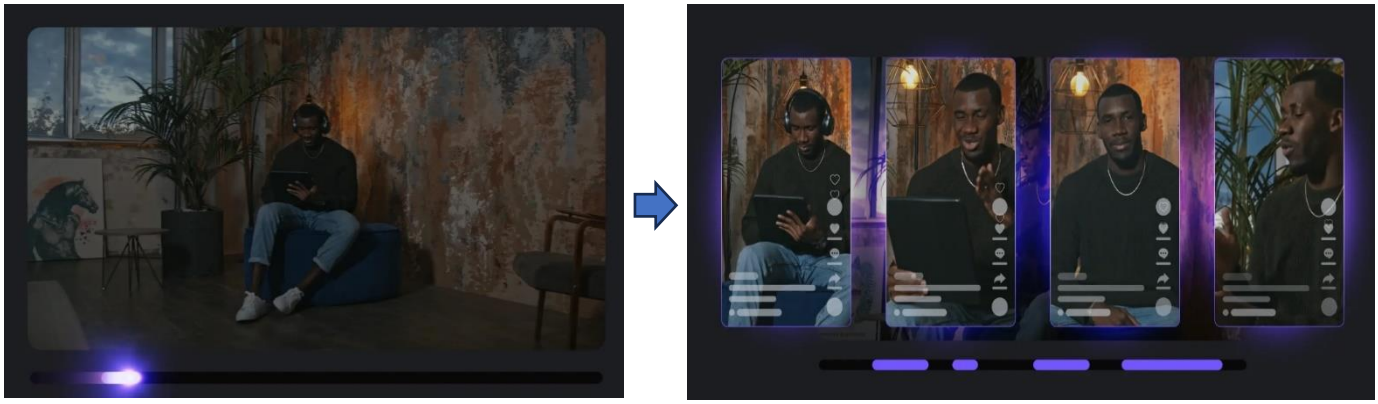
资料来源：Adobe 公司官网，信达证券研发中心

Capitions.AI（AI 字幕、AI 长短视频剪辑）

公司主要产品为 AI 视频编辑器，它将 AI 的能力几乎应用到了整个视频编辑的每个环节。公司产品的全球创作者数量达到 1000 多万，推出了一系列全球首创的生成功能，用户每月制作超过 300 万个视频。2024 年 7 月 9 日，Captions 已筹集 6000 万美元的 C 轮融资，由 Index Ventures 领投，现有投资者 Kleiner Perkins、Sequoia Capital 和 Andreessen Horowitz 也参与其中。新投资者包括 Adobe Ventures、HubSpot Ventures 和 Jared Leto。此次融资使公司筹集的总资本超过 1 亿美元，公司估值为 5 亿美元。AI 会分析素材，并在最合适的时间插入自定义图形、缩放、音乐、音效、过渡和动态背景，所有这些都根据内容进行个性化设置。使用 AI Edit，用户无需从空白画布开始。从三种视频编辑风格中选择一种 Impact、Cinematic 和 Paper，更多风格即将推出。

AI 剪辑生成器：可以使用户利用 AI 将一部长视频变成十部短视频，挑选更多适合在 Reels、TikTok 上播放的短片，达到省时的同时使得传播效果最广。

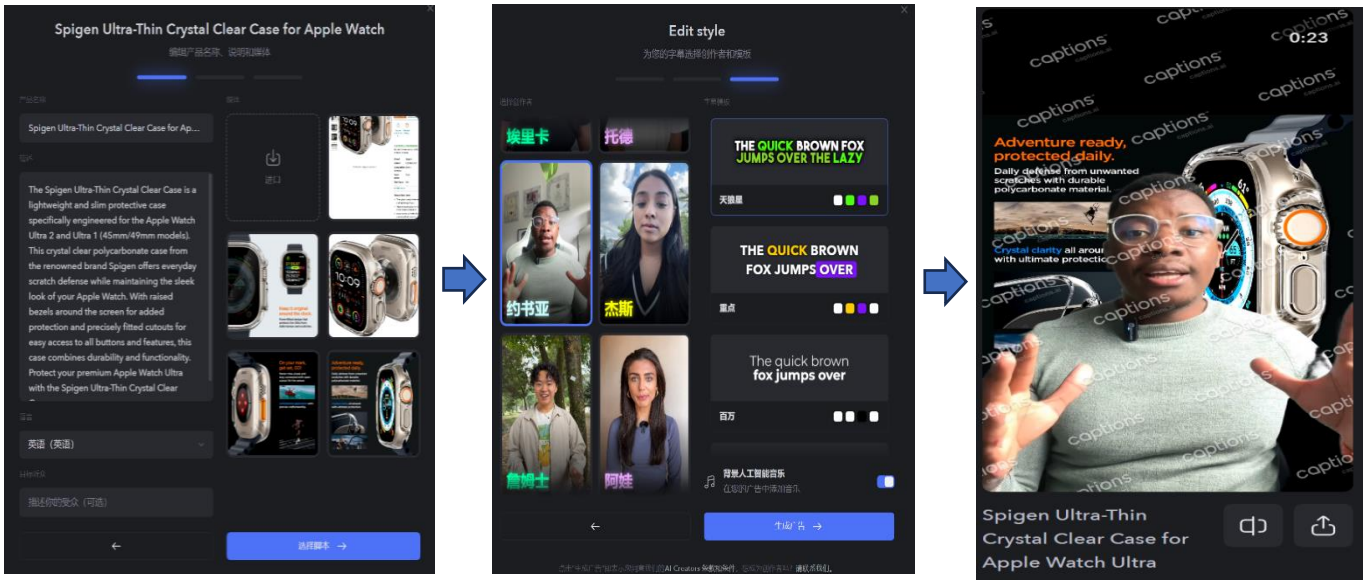
图 41：Captions AI Shorts 功能



资料来源：Captions 公司官网，信达证券研发中心

AI 广告生成器：通过 AI Creators Ads 只需要输入产品链接或者描述，几秒钟即可创建数十个视频广告。输入产品链接或脚本，实现 UGC 广告的效果。下图为导入一个亚马逊链接，选择广告数字人以及对应字幕，即可生成用户想要的广告，一站式分发到各个平台。

图 42：Captions AI AD Creator 功能



资料来源：Captions 公司官网，信达证券研发中心

阿里达摩院“寻光”视频创作平台

2024 年 7 月，在世界人工智能大会上阿里巴巴达摩院最新发布了 AIGC 产品 —— 寻光视频创作平台，旨在提升视频制作效率，解决视频后期编辑问题，通过简易的分镜头组织形式和丰富的视频编辑能力，让用户实现对视频内容的精准控制，并保持多个视频中角色和场景的一致性。“寻光”旨在为用户提供一站式的视频创作工具，让用户回归到关注视频内容本身是寻光致力于做的事情。目前主要功能包括：分镜故事板一键创建、定制自己的故事角色、生成具备一致性的角色和场景画面，再利用运镜控制、运动编辑，创作 AI 视频作品。同时，可以使用各类视频编辑功能进行修改，更有图层拆解和融合功能，定制化视频内容，方便用户利用 AI 创作高质量、高一致性的故事视频片段，而非几十秒的创意 AI 视频。

图 43：阿里达摩院“寻光”一站式视频创作平台视频编辑功能

资料来源：寻光官网，信达证券研发中心

图 44：阿里达摩院“寻光”视频素材创作功能

资料来源：寻光官网，信达证券研发中心

美图 MOKI-AI 创作短片

2024 年 6 月 12 日，美图公司举办以“聊聊 AI 工作流”为主题的第三届美图影像节，现场发布 6 款产品，其中包含了 MOKI-用 AI 做短片。MOKI 不做常规的文生视频，而是聚焦在了 AI 短片创作，其中涉及到动画短片、网文短剧、故事绘本和 MV。目前在视频大模型故事成片的难点包括：视觉风格、场景、角色不一致；无法用分镜进行整体把控；角色无法开口说话。针对此的解决方案，美图试图打造 AI 短片工作流提升可控性：1）先做脚本、视觉风格、角色等前期设定；2）用 AI 生成分镜图，分镜图转视频；3）用台词驱动角色开口说话。

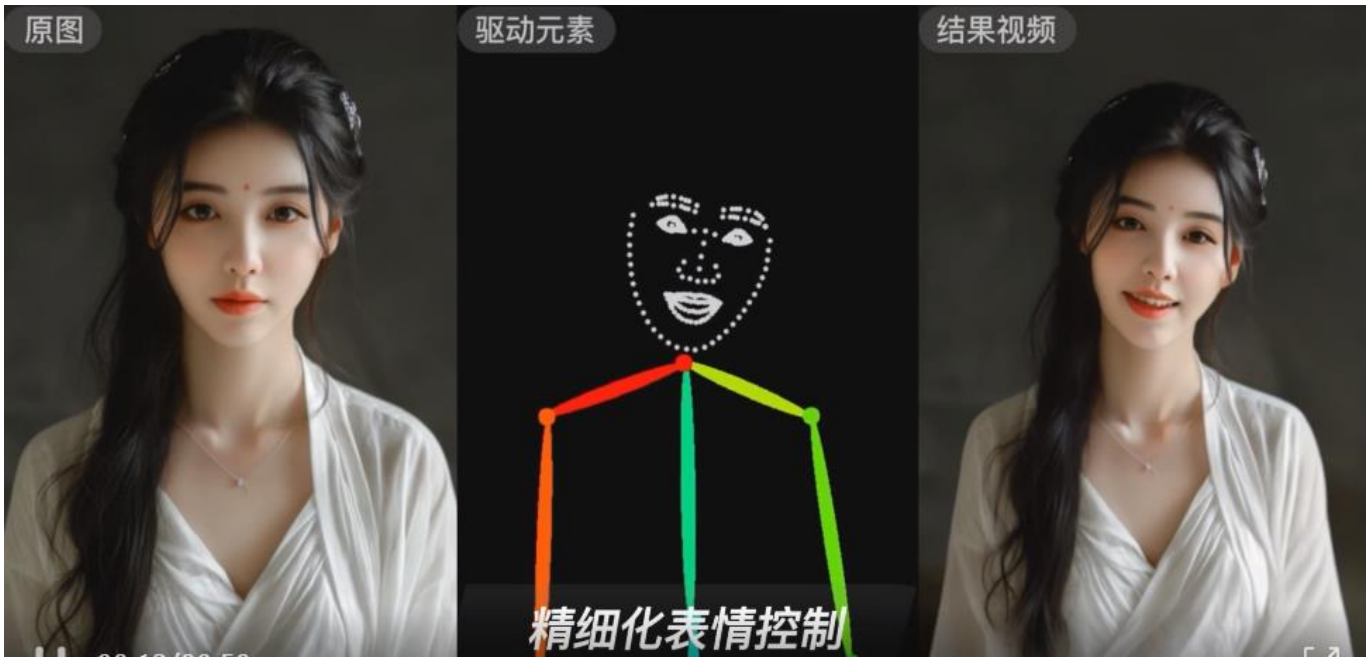
图 45：美图 MOKI AI 短片产品

资料来源：美图公众号，信达证券研发中心

商汤 Vimi——人物视频生成大模型

2024 年 7 月，商汤科技在世界人工智能大会上发布了公司打造的首个可控人物视频生成大模型——Vimi，Vimi 基于商汤日日新大模型的强大能力，仅通过一张任意风格的照片就能生成和目标动作一致的人物类视频，不仅能实现精准的人物表情控制，还可实现在半身区域内控制照片中人物的自然肢体变化。Vimi 具备较强的稳定性，尤其在长视频的情景下，能够稳定保持人物的脸部可控，可生成长达 1 分钟以上的单镜头人物类视频。Vimi 在人物视频场景生成中，可以做到整个的环境都跟着肢体的控制去变化，包括生成合理的头发的抖动。Vimi 相机是 Vimi 可控人物视频大模型体系的第一款 C 端产品，能够满足广大女性用户的娱乐创作需求。

图 46：商汤 Vimi 人物视频生成



资料来源：Vimi 公众号，信达证券研发中心

智向未来（HiDream.ai）—基于自研的 DiT 架构的智象大模型 2.0

智象未来（HiDream.ai），成立于 2023 年 3 月，其自主研发的视觉多模态基础模型实现了不同模态之间的生成转换，支持文生图、文生视频、图生视频和文生 3D，并推出了一站式 AI 图像和视频生成平台—Pixeling 千象。智象大模型 2.0 的整体升级，相较于 1.0 版本在底层架构、训练数据和训练策略上均有质的变化。2023 年 12 月，智象大模型的文生视频打破了 4 秒时长限制，做到了支持 15 秒钟以上的生成时长，同时还支持 4K 画质。**相较于 U-Net，DiT 架构灵活度更高，且能增强图像、视频的生成质量。Sora 的出现更直观地验证了这一点，采用此类架构的扩散模型表现出了天然生成高质量图像和视频的倾向，并在可定制化、生成内容可控性方面具有相对优势。后续上线 AI 分镜头故事创作视频功能：首先输入提示词—分镜头脚本生成—关键帧图片生成—AI 故事视频生成。**

图 47：智象大模型升级 2.0 版本

资料来源：机器之心公众号，信达证券研发中心

图 48：智向未来即将上线一站式分镜头故事创作视频生成功能

资料来源：机器之心公众号，信达证券研发中心

五、AI+视频时代来临，思考哪类公司存在商业化变现的可能性？

（一） 一站式平台型公司：代表性公司—Adobe、美图等

目前在 AI+视频的使用上，大多数尝试的用户仍停留在一段较短时长的视频创意生成阶段，真正用于实际工作效率提升、工作流程替代的较少。原因仍是在于缺少一站式的 AI+视频生成+剪辑产品提供商，目前用户在生成创意视频后，需要自己去多个其他软件产品上调配背景音乐添加、镜头转场、字幕添加、多余镜头删减等，文件格式的适配性也可能存在问题，因此后续随着技术的迭代发展，能够给用户提供一站式工作流的 AI+视频生成、剪辑的平台性公司有望深度受益。如全球绘画、设计领域龙头公司 Adobe，Firefly 集成应用包括以下五个产品：Lightroom、Photoshop、Adobe Express、Illustrator、InDesign，同时 Adobe 产品底层已经集成 Pika 以及后续待正式发布的 OpenAI 视频大模型 Sora，可以使得用户在感受 AI 视频的创意生成的同时，可以直接在 Adobe 的剪辑软件内对视频进行其他环节的修改，达到一站式产出的效果。因此我们认为，一站式 AI 视频生成+剪辑平台型公司后续有望深度受益。

创意设计市场规模预测：根据 Adobe 公司披露数据，2024 年预计 Document Cloud + Creative Cloud + Experience Cloud 三朵云 TAM 总计可达 2050 亿美元，其中 Experience Cloud 市场空间为 1100 亿美元，Creative Cloud 市场空间为 630 亿美元，Document Cloud 市场空间为 320 亿美元。相较其 FY24Q2 创意云收入 31.26 亿美元，主打创意软件设计的 Creative Cloud 的市场空间较大，其 AI 功能的附加值的增长空间也较大。Creative Cloud 包含了其 AI 产品 Firefly、Express，Adobe Firefly 推出 Firefly Image 2 模型，改进了图像生成功能； Adobe Express 同样集成了大量的 AI 功能给创意工作设计、图像领域用户使用。因此，我们假设在创意设计领域 2024 年的市场空间为 630 亿美元，后续有望演变成千亿美元规模以上的市场。

图 49：Adobe Creative Cloud TAM 市场规模预测

资料来源：Adobe 官网，信达证券研发中心

国内视频剪辑软件行业市场规模预测：根据智研瞻产业研究院整理，2020 年中国视频剪辑软件行业市场规模达到了 15.8 亿元人民币，同比增长率为 18.61%。预计未来几年，随着短视频和直播行业的持续火热，视频剪辑软件市场规模将继续保持高速增长，到 2025 年市场规模将达到 34.8 亿元人民币，年复合增长率为 17.8%。因此，**叠加广告营销市场空间、IP 类公司市场空间等，AI+视频市场空间至少为万亿人民币规模，相较目前部分 AI+视频产品的 ARR 仅为百万、千万美元来看，AI+视频成长空间较大，核心还是在于如何把底层技术迭代升级完善的同时，做到一站式 AI+视频生成、剪辑、宣发等环节的强大产品力吸引全球用户来实现商业化，AI+视频的星辰大海远不止于创意视频的生成。**

Adobe 相关创意设计业务收入市占率仍较低，提升空间较大。根据 Adobe 自身业绩指引：Adobe2024 财年目标总收入在 214 亿美元到 215 亿美元之间（上个季度指引：213 亿-215 亿美元）。预计年度新增数字媒体 ARR 约为 19.5 亿美元，数字媒体部门收入在 158 亿美元到 158.5 亿美元之间。数字体验部门收入预计在 53.25 亿美元到 53.75 亿美元之间，数字体验订阅收入在 47.75 亿美元到 48.25 亿美元之间。因此，2024 年预估 Adobe 数字媒体业务营收市占率在 16.6%-16.7%之间，数字体验业务营收市占率在 4.86%，两个主要业务方向的长期营收成长空间广阔，Adobe 在产品 AI 商业化道路才刚刚开始。

表 5：Adobe 数字媒体业务和数字体验业务预估市占率

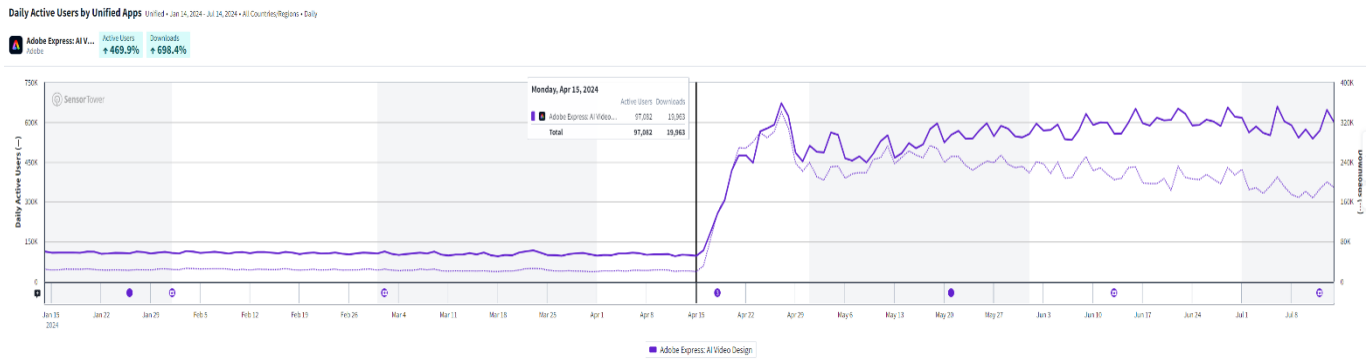
	2024E
2024 年 Adobe 预估创意云+文档云预计市场规模（十亿美元）	95
2024 财年 Adobe 数字媒体部门（创意云+文档云）预计收入（十亿美元）	15.8-15.85
Adobe 数字媒体业务营收市占率	16.6%-16.7%

2024 年数字体验业务预计市场规模（十亿美元）	110
2024 财年 Adobe 数字体验业务预计收入（十亿美元）	5.35
Adobe 数字体验业务营收市占率	**4.86%**

资料来源：Adobe 官网，信达证券研发中心

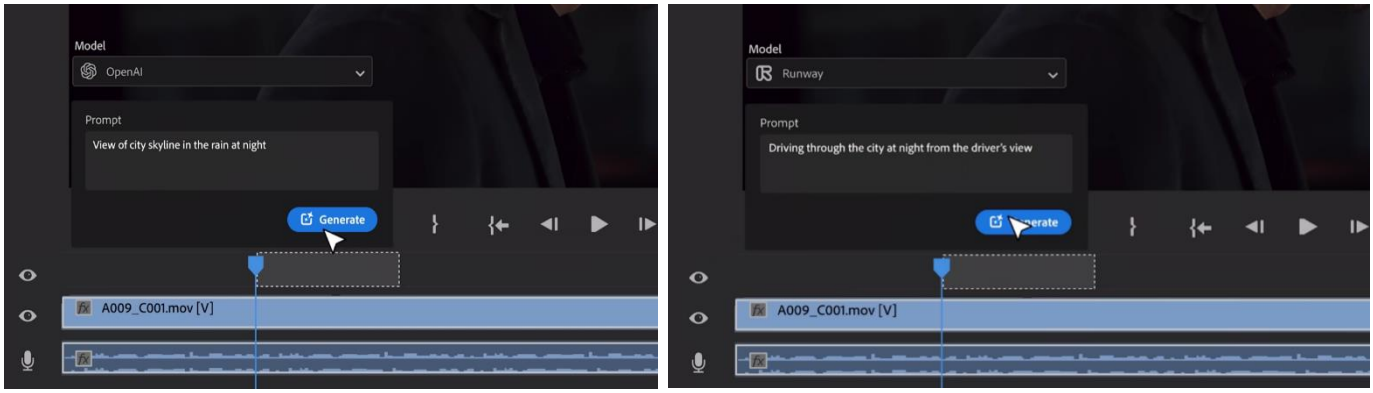
AI 新功能迭代提升 Adobe 产品日活数量，侧面验证用户需求客观存在，只不过市场缺少满足痛点需求 AI 产品。2024 年 4 月 Adobe Express 活跃用户陡然爆发增长，根据第三方 Sensortower 数据，应用日活从 12 万上下提升至 70 万上下并呈现持续提升的趋势，主要原因在于 Adobe 推出全新的 Adobe Express 移动应用程序，具有 Firefly 生成 AI 和 Adobe 创意工具的强大功能，现已在 web 和移动设备上普遍可用。为了满足 TikTok、Instagram 和其他社交内容的爆炸式需求，Adobe Express 可以轻松地在网络和移动设备上创建和协作，释放创造力和生产力，其主要功能包括文本生成图像、生成填充、文本效果、文本到模板、为 Instagram Reels、TikTok 等制作视频等相关生成式 AI 功能，用户数的增长侧面验证了 Adobe 产品在 AI 功能上的迭代准确把握了用户的痛点需求，有望给 Adobe Express 长期收入增长垫定良好基础。Adobe 视频编辑软件 Premiere Pro 定价为$22.99/月，Adobe Express 定价为$9.99/月。在 NAB show 2024 上，Adobe 公司宣布在 24 年内 Premiere Pro 会推出一站式 AI 视频生成剪辑功能，这一变化有望带来 ARR 收入上的增长。

图 50：Adobe Express 在 24 年 4 月迭代 AI 功能后，日活数骤然抬升并稳定提高



资料来源：Sensortower，信达证券研发中心

图 51:Adobe Premiere Pro 引入第三方模型如 Pika、OpenAI、Runway 生成视频片段满足用户一站式视频剪辑需求



资料来源：Adobe Blog，信达证券研发中心

美图

2024 年 6 月 12 日，美图第三届影像节上公布一组数据：“在 AI 驱动下，美图全球 VIP 会员数突破千万”，从 2023 年 6 月 19 日的 719 万提升至 2024 年 6 月 12 日的 1063 万，同比增长幅度+47.8%。美图公司聚焦“生产力和全球化”战略，以 2023 年 6 月推出的美图视觉大模型 MiracleVision(奇想智能)为基石，形成由底层、生态层和应用层构建的 AI 产品生态。2023 年美图实现总收入 27 亿元，同比增长+29.3%。经调整后归母净利润 3.7 亿元，同比增长+233.2%。总收入与净利润增长主要得益于 AI 推动主营业务收入增长，美图用户每天处理数亿份图片和视频，约 83%都用到了泛 AI 功能。2023 年，美图以付费订阅为主的影像与设计产品业务收入 13.3 亿元，同比增长 52.8%；广告业务收入 7.6 亿元，同比增长 20.5%；美业解决方案业务收入 5.7 亿元，同比增长 29.1%。在产品全球化推进中，美图同样步伐加速。AI 正帮助美图公司加速进入全球市场，目前，美图已在全球 195 个国家和地区布局影像产品，美图秀秀、美颜相机、Wink 先后取得多个国家和地区的应用榜单冠军。在 data.ai2024 年 1 月的中国非游戏厂商出海收入排行榜中，美图公司排在第 3 位。

据 QuestMobile 数据，美图秀秀连续 8 年夺得中国图片美化赛道用户规模第一名、美颜相机连续 8 年夺得中国拍照摄影赛道用户规模第一名。影像产品组合的付费订阅用户渗透率持续快速上升，进而推动付费订阅收入大幅增长。截至 2023 年 12 月 31 日，美图公司月活跃用户数达 2.5 亿，同比增长 2.6%。美图付费订阅用户数超 911 万，创历史新高，同比增长 62.3%，付费率仅为 3.64%，ARPU 提升空间较大。

图 52：美图公司底层、生态层、应用层架构

资料来源：美图秀秀桌面版公众号，信达证券研发中心

同海外图像、视频编辑领域龙头公司 Adobe 类似，美图在国内的图像、视频编辑行业的用户较多、认可度较强，在 AI+图像/视频的技术产品迭代上持续发力，同样有望成为一站式图像&视频 AI 加持下的龙头公司，在底层 AI 技术的不断打磨迭代下，逐步应用于旗下所有的产品，有望提升每款产品的用户数和 ARPU 来实现增值创收。

1）Wink 视频剪辑。Wink 在自身拥有大量视频剪辑功能例如画质修复、视频拼接、音频降噪、自动字幕、皮肤细节、画面裁剪等之外，产品希望做到让用户像修图一样来实现修视频的功能，上线了诸如 AI 修复、AI 动漫（可以一键将视频生成动漫风格）、AI 美容、视频美容、AI 一键成片等 AI 功能来满足用户需求。

2）开拍。开拍 APP 主打功能为用 AI 制作口播视频，自动生成 AI 脚本后通过口播剪辑批量化生成口播视频。

3）WHEE。目前已经上线文生视频和图生视频功能，在图生视频的基础上，MiracleVision 支持视频运镜。提供了推、拉、摇、移等八种电影级运镜模式，让用户能够轻松模拟专业的镜头运动。后续有望更新视频生视频功能，导入一段视频，再加上不同的提示词，就能获得卡通、科幻、像素风，羊毛毡等不同的艺术风格。

（二）AI+视频技术头部服务商转型 ToB+ToC 产品类公司：代表性公司—Runway、商汤科技

商汤科技生成式人工智能相关业务在 2023 年的收入获得 200%增长，收入突破 11.8 亿元人民币。公司在国内的生成式人工智能的算力储备、人才储备等维度上均属于第一梯队，公司目前以为 B 端客户提供算力、大模型 API 调用为主，在 AI 技术上迭代发展较快。2024 年 7 月在世界人工智能大会上，商汤科技打造的首个可控人物视频生成大模型——Vimi，以 Vimi 为例来探索商汤在垂直领域细分市场上的 C 端 AI 产品扩张。我们认为，AI 视频生成领域的难点在于创作人物形象的一致性和是否符合世界物理规律上。因暂未拿到实测资格，在 Vimi 微信公众号的介绍中我们看到，Vimi 基于商汤日日新大模型的强大能力，仅通过一张任意风格的照片就能生成和目标动作一致的人物类视频，不仅能实现精准的人物表情控制，还可实现在半身区域内控制照片中人物自然肢体变化，通过已有人物视频、动画、声音、文字等多种元素进行驱动。

Vimi 模型主打在长视频情景下能够稳定保持人物脸部可控，这有望适用于多领域创作。例如能够满足广大女性用户的娱乐创作需求。用户只需上传不同角度的高清人物图片，即可自动生成数字分身和不同风格的写真视频；对于热衷表情包的用户来说，Vimi 通过单张图片即可驱动生成各种趣味的人物表情包，同时还可支持聊天、唱歌、舞动等多种娱乐互动情景，在女性娱乐应用市场中，用领先的 AI 技术打造垂直领域产品，有望打开公司的 ToC 端市场，同时也有望通过大量的用户数据进而反哺 B 端市场客户的使用效果。

图 53：Vimi 在人物一致性功能支持下打造的数字分身打造 AI 视频功能、AI 表情包功能

资料来源：Vimi 公众号，信达证券研发中心

（三）视频剪辑类公司： 代表性公司—快手和抖音

作为国内短视频内容头部公司，均对应推出了其视频剪辑类软件—快手快影和抖音剪映，目前快影已经集成了快手可灵视频大模型的文生视频和图生视频功能，后续有望迭代至 AI 视频剪辑功能；剪映也推出了 AI 创作产品 Dreamina（即梦），同时剪映内部目前也已经上线了诸多 AI 功能，例如一键成片、AI 广告营销等。可以看到，若在视频剪辑领域做到极强产品力，同样有望提升用户付费率，带来商业化变现程度的提升。

快手可灵

快手可灵视频大模型的效果得到广泛的市场认可，**可灵经发布后近三个月，申请体验的用户数量已突破 70 万大关，相比快影的月活数据近 200 万来看，已经有了较高的占比，累计生成的视频作品高达 700 万份**。快手可灵在集团内部的支持下，不管是算力储备、团队人员配置、底层数据都加速了可灵的高质量发布。快手在短视频领域深耕多年，多年的视频数据标签化储备让在可灵在训练阶段的数据端、尤其是视频数据端的优势较为显著。

我们认为，快手可灵作为集团内部较高战略级别的一款产品，未来一定不只是一款面向 C 端的视频生成+剪辑工具，可灵有望赋能快手所有的视频创作者，为现有的快手内容生态提供补充。根据快手大数据研究院数据，2023 年有超过 1.38 亿用户首次在快手平台发布短视频、2023 年坚持 365 天在快手每天发视频的创作者人数高达 61%、2023 年有超过 2200 万创作者在快手平台获得收入、2023 年第三季度快手搜索平均月活跃用户数达到 4.7 亿，从以上数据我们可以看到，快手作为一个超大 DAU 的短视频产品，有着高度活跃的创作者生态氛围，可灵视频大模型的更新迭代有望赋能快手创作者用户，从而给快手平台带来留存率的提升和更多商业化空间的探索。

表 6：快影和剪映产品相关数据

产品	月活用户数	应用近一年净收入	产品定价年订阅费
快影（内嵌可灵）	194.6 万	280 万美元	88 元/年
剪映（包含海外 CapCut 和国内剪映，全渠道）	3.2 亿	1.92 亿美元	499 元/年

资料来源：SensorTower，信达证券研发中心

（四）广告营销类公司：易点天下、蓝色光标、因赛集团、利欧股份等

OpenAI 的视频模型 Sora 一经问世便引起了社会广泛的关注，全球都在探索 AI+视频对各行各业的改变程度，尤其是在广告营销领域。传统广告需要耗费昂贵的拍摄设备、较多的人力支持、后期剪辑的时间成本等等，尤其遇到拍摄难度较高、拍摄场景环境较难情况下，拍摄的成本会极大上升。因此，视频生成模型的迭代有望使得广告制作领域优先受益，目前视频大模型的一次性生成时长较多集中在 20s 以内，同时可进行后续的视频的延长生成，但首先需要解决的便是视频人物一致性和防畸变的问题，已经在逐步改善。另一个重点问题：**目前视频生成所需的时长大多在几分钟，如果更为复杂的提示词在较少的算力支持下，甚至要等几十分钟的时长从而才能得到 20s 以内的 AI 视频，而对于广告营销行业来说，批量化、短时间的视频生成分发是比较重要的环节。我们认为，如果广告视频的生成时长可以控制在 1 分钟、甚至 30s 内，则该节点有望成为 AI 应用在广告领域分水岭。**

AI+短视频营销领域。海外创业公司 Captions.AI 中的 AI 广告生成器功能，仅需要输入商品链接、选择数字人画像，即可自动生成 5 端不同角度的数字人介绍产品的广告短视频切片，一键外发到多个社交媒体平台上，极大提升广告营销类服务商的工作效率。同时，随着 AI+视频的生成时长和画面分辨率的提升，更逼真的长时长广告也有望代替传统的广告服务商，节省拍摄成本。近年来，国内抖音、快手、微信视频号，海外 Facebook、Tiktok、YouTube 等短视频平台崛起，据 Statista 相关数据显示，2030 年全球移动营销市场规模达预计将达到 578.5 亿美元，其中，短视频营销作为数字创意的核心载体，有望成为未来内容生态的主要环节。传统短视频营销内容生成，不仅成本高昂，在制作过程中更会面临定选题、找素材、写脚本、现场拍摄、后期剪辑以及运营发布等多个繁琐的流程环节。

易点天下：2023 年旗下 AIGC 数字营销创作平台 KreadoAI 就开始了 AI+营销的探索和应用，KreadoAI 包含了多模态模型的融合，包括文本生成、图生图、文本生成视频、语音生成等，提供文字到广告创意图片、多语种语音、视频的生成能力。KreadoAI“会说话的照片数字人”功能则应用了图生视频能力，用户只需上传心仪照片或使

用 KreadoAI 提供的文字关键词生成专属的 AI 人物形象，输入一段文字，即可快速生成专业的产品讲解视频、AI 人物口播视频，应用于广告、知识培训等各种创意场景。**在短视频营销领域，易点天下旗下 AIGC 数字营销创作平台 KreadoAI 可为企业提供「AI+」的多场景短视频营销解决方案，目前已覆盖全球 67 个国家、注册用户数 100w+、单月用户访问量超过百万，得到来自巴西、中国、印度、欧美等地区的用户认可**。KreadoAI 的功能覆盖相对全面，能够适用于文案、商品图、视频等常见营销应用领域。在短视频制作方式上，KreadoAI 支持一站式的包揽制作，操作难度低；在降本和增效表现上，KreadoAI 可以将视频制作效率从 12 小时/个缩短至 5 分钟/个，而成本只有真人的 1/100。产品以多语种多人种风格的数字人形象满足跨国营销素材生产需求，并结合投放数据反馈，持续优化视频质量。

蓝色光标：2024 年内 7 月 4 日，蓝色光标与昆仑万维正式宣布达成战略合作，携手打造 AI 营销创新生态。蓝色光标从 2023 年初就确立 All in AI 战略、快速开启一系列 AI 营销落地实践和探索动作。在短短一年多的时间里，创造 300 多个 AI 营销案例、由 AI 驱动的收入达到 1.08 亿元，取得显著成果；并推出营销行业垂直模型“BlueAI”，大量应用于多家知名品牌的营销活动中，大幅提升了广告投放的精准度和效果，收获客户认可。截至 2024 年 5 月，昆仑万维天工 AI 每日活跃用户（DAU）已超过 100 万，位列国内人工智能企业第一梯队。昆仑万维将在合作中开放大模型底层能力与 APP 平台分发资源，基于天工 AI 大模型、AI 搜索、AI 音乐、AI 视频、AI 社交等强大的 AI 技术能力为生成式广告发展提供技术基础。蓝色光标的长期 AI 战略方向：1）越来越多的 AI 收入，从 8-10 亿到追求 30-50 亿，甚至未来的 100 个亿；2）更多 AI Native，持续提高 AI 的占比、浓度和含金量，将人工调优的比例进一步下降；3）过去因行业内卷造成的人才密度下降，有望在 AI 时代得到巨大的改善。公司需要更多的 AI 人才，不仅要放下身段吸引人才加入，也需要培养 AI 的种子型人才；4）持续提高 AI 产品的易用性、与业务的结合度，高度聚焦行业的底层逻辑，结合业务场景不断迭代；5）视频多模态方面，要寻求更大的突破。

因赛集团：公司旗下 AIGC 营销产品—InsightGPT 继 3 月初推出图生视频产品后，再度聚焦 AI 视频创作领域，正在上线文生视频营销应用产品「AI 营销视频」，为 AIGC 营销领域带来新的数智化解决方案。目前官网宣传上线了例如 AI 整合营销、AI 营销创意、AI 电商营销、AI 短视频营销和 AI 营销工具功能。因暂未实测，根据 InsightGPT 官网，用户可以通过输入提示词得到生成的脚本，结合音乐生成、视频生成、人声生成等，整合 AIGC 多模态营销能力，最终得到高质量的 AI 生成视频进而应用到营销领域，提升内容创作效率。

利欧股份：2023 年 9 月，利欧数字率先发布营销领域大模型「利欧归一」，在通用 L0 级语言模型基础之上，结合利欧数字长期积累的大量营销行业知识、投放经验以及对客户需求的深入理解，训练出适配各媒体平台投放工作流的 SEMGPT 专属模型。以 LEO Copy、LEO Diffusion 为例，前者聚焦文案创作功能、后者聚焦图片生成功能，是 LEO AIAD 多模态内容精准控制，最佳优化营销生产力的两项代表性功能。LEO Copy 产品可以让内容创作者只需输入自己的 idea，即可在数秒内一键生成适用于小红书种草、抖音短视频、大众点评测评、信息流标题等特定营销场景、平台、投放渠道所需要的营销创意内容。在 LEO Diffusion 中，AI 会自动帮助设计师完成复杂的 Prompt 编写、基础模型选择、风格模型选择、模型参数设置等工作，设计师出图的速度从原先的平均 6 小时/图提升至 1 分钟/图。

（五） UGC 社区类公司：代表性公司—Bilibili

传统电影级别的镜头的制作成本、制作时长、团队配置、设备配置等均需要耗费大量的人力、物力才能完成，AI+视频的迭代使得大量的 UGC 创作者用户可以按照自己的创意想法，不断生成、修改想要的视频效果，部分生成的效果已经堪比电影级别镜头。当某个 UGC 创作者平台积淀了大量类似的 AI+视频创作者用户后，一个 AI+视

频的开源社区就会形成，带来商业化可能性。**作为国内内容创作者生态知名平台，2023 年第四季度，B 站日均活跃用户超过 1 亿。2023 全年超 300 万 UP 主在 B 站获得收入，同比增长超 30%。2024 年第一季度，B 站日均活跃用户数达 1.02 亿，同比增长 9%。月均活跃用户数创历史新高，达到了 3.41 亿用户日均使用时长 105 分钟，创历史新高，带动超 150 万 UP 主在 B 站获得收入。通过考试“正式会员”数达到了 2.36 亿，正式会员第 12 个月留存率近 80%。大会员付费用户数据为 2190 万，其中超过 80%为年度订阅或自动续订用户**。海量的创作者用户、优良的社区氛围有望带动创新作品的发展。目前，B 站大会员基础连续包年订阅费为 128 元/年，假设 B 站后续能成立较深厚的 AI 视频创作社交社区模块，带来 10%大会员付费数量的提升，仅对大会员付费角度来看，有望带来亿元级别收入上的提升。而长期给 B 站带来持续性贡献的是 B 站新用户的加入和老用户留存率的提升，有望带来广告及其他形式商业化变现的提升。

（六） 视频数据类公司：代表性公司—华策影视、捷成股份、视觉中国、中广天择等

表 7：相关公司提供视频数据用于训练多模态大模型

公司	AI 视频数据相关
捷成股份	国内外流媒体平台、电影视频制作公司均积累了海量视频素材，在前期的生成式 AI+视频的技术迭代发展中，优质的视频数据对于模型质量的训练优化显得至关重要。捷成股份与华为云签署协议共同建设视频大模型。捷成股份凭借十多年来积累的 20 万小时影视视听节目素材和通过数据清洗来为华为云投入高质量数据集，授权华为用于视频大模型训练。华为方面投入基础模型、算力、模型优化与专业服务等。24 年 3 月，捷成自主研发的 AI 智能创作引擎 ChatPV 正式发布，并接入华为云盘古大模型的通用语言解析能力，服务于 AI 视频创作应用。
视觉中国	2023 年 10 月，视觉中国与华为云正式签署关于视觉大模型的合作协议。双方将以华为云盘古大模型为基础打造视觉大模型，共同实现视觉领域高度智能化发展，推进更深层次的内容产业智能转型。视觉中国专注“AI+内容+场景”战略，依托海量优质合规的专有数据、全球创作者生态、数字版权交易场景等核心竞争力，聚焦“以客户为中心”的 AIGC 技术创新，能够在视觉领域为多模态大模型训练提供所需的海量、高质量数据集，助力多模态大模型生态建设。
华策影视	公司现有超 5 万小时正版影视版权库和 150 万分钟（估算约 2 万 TB）高清/超高清的原始拍摄素材，可通过清洗、标注、加工等技术处理后形成版权数据集。上线“AI 视频分析检索功能”，可以对视频中的人、物等特定元素进行自动标签，快速锁定相关素材等。
中广天择	打造 AI 算料综合服务交易平台：在人工智能快速发展的背景下，公司利用自身优势创新业务发展，首先是公司拥有大量优质自有版权的音视频版权数据，其次是公司具备强大的渠道能力，利用现有 500+城市广电合作客户资源开展行业版权数据资源整合，在行业主管部门的支持下，打造中国广电行业优质版权的垂类数据集，在此基础上，积极建设 AI 模型训练的算料综合服务和交易平台。

资料来源：捷成股份公众号、视觉中国公众号、华策影视公告、中广天择公告，信达证券研发中心

（七） IP 类公司：代表性公司—上海电影、汤姆猫、中文在线等

随着视频生成技术的迭代发展，传统 IP 方可以有效利用新技术来改造 IP 的使用，可以将 IP 衍生出更多的流媒体内容，来实现影视、短视频、IP 周边变现。

表 8：IP 类公司可基于 AI+视频开发更多 IP 衍生品

公司	IP 变现相关
上海电影	公司拥有 60 个经典动画+影视 IP。2024 年 2 月 29 日，上海电影在上海影城 SHO 发布 iNEW 新战略，以“iPAi 星球计划”为抓手，结合 AI 主攻 IP 内容焕新和 IP 商业化，打造 AI+IP 在影视行业的全新重点战略布局。三大先导行动：探索中国动画学派 AI 模型、发起全球创造者计划聚焦 AI 在短剧和影视垂直领域的应用开发、IP+AI 赋能商业开发计划来加速 AI 对海量 IP 储备的商业化放量。同时，举办了“全球 AI 电影马拉松大赛”，10 万美金奖金池助力 IP 二创，在全球范围内发掘和寻找优质 AI 影视创投项目，招募全球 AI+影视方向人才，建立 AI 合作生态。
汤姆猫	围绕汤姆猫家族 IP 为核心，线上与线下协同发展的全栖 IP 生态运营商。截至 2023 年底，汤姆猫家族 IP 系列应用在全球范围内的累计下载量已超过 230 亿人次，全球 MAU 最高达 4.7 亿人次。公司汤姆猫家族 IP 系列动画作品已被翻译成 32 种语言，全球累计播放量已超过 1,100 亿次。同时，公司国内研发团队与西湖心辰合作的汤姆猫 AI 讲故事等产品，已初步完成主要功能的测试。公司 AI 硬件团队正研发一款基于生成式人工智能技术的 AI 语音交互陪伴机器人。公司 IP 属性强，深耕 AI 情感陪伴赛道。
中文在线	截至 2023 年底，公司以自有原创内容平台、知名作家、版权机构为正版数字内容来源，累积数字内容资源超 560 万种，网络原创驻站作者 450 余万名；与 600 余家版权机构合作，签约知名作家、畅销书作者 2,000 余位。公司 IP 衍生业务以文学 IP 为核心，向下游延伸进行 IP 衍生开发，着力打造“网文连载+IP 衍生同步开发”的创作模式。截至 2023 年底，公司可用于 AI 大模型训练的数据集已超过 60TB，主要由小说和出版物组成，为公司在有声书、漫画、动漫、视频等多模态领域商业化打下基础。

资料来源：中国基金报、上观新闻、汤姆猫公告、银柿财经、中文在线公告，信达证券研发中心

（八） AI 短剧/AI 短片等方向探索类公司

表 9：部分公司对 AI 短剧/AI 短片方向上的探索

公司	AI 短剧/AI 短片探索相关
美图公司	发布 AI 短片工作流工具—MOKI。在脚本、视觉风格、角色等前期设定完成后，AI 自动生成分镜图并转为视频素材，通过智能剪辑、AI 配乐、AI 音效、自动字幕等功能串联素材并实现成片。
因赛集团	探索 AIGC 技术赋能短剧创作及制作提质增效，未来将适当参与优质 AI 短剧的出品以及 AI 短剧相关技术和应用产品的布局。公司参与出品了由北京华坞科技制作的国内首部 AI 商业微短剧《西西里的美丽传说》。

快手	快手平台每天约有 2.7 亿用户在观看短剧，播放量过亿短剧有 300 多部，有超 10 万创作者进行短剧相关的内容创作。推出“星芒短剧+可灵大模型”创作者孵化计划。2024 年 7 月 12 日，快手首部 AIGC 原创奇幻微短剧《山海奇镜之劈波斩浪》线下看片会正式举行，AI 技术的加持使中国传统神话题材的场景布置和 CG 特效变得更加高效。
柠萌影视	较早布局短剧赛道，旗下打造的精品短剧《二十九》总播放量超 8.3 亿，集均播放量超 4100 万，豆瓣评分 8.1 分，成为了 2023 现象级爆款短剧。
超讯通信	携手 Seven Volcanoes，领航 AI 短剧出海。超讯通信子公司超讯人工智能科技有限公司（以下简称“超讯人工智能”）与 AI 多模态应用公司 Hong Kong Inequation Limited 签署了相关投资协议。通过本次投资，恰好弥补了超讯人工智能在短视频方面的短板，更丰富完善了超讯通信 AI 产品线及服务生态，有效提升了公司竞争力。
博纳影业	由博纳影业 AIGMS 制作中心，联合抖音、即梦 AI 生成式人工智能创作平台，出品并制作的 AI 生成式连续性叙事科幻短剧《三星堆：未来启示录》第一季，揭开一段跨越时空的古文明探险旅程，开启了影视产业和人工智能技术深度融合的全新篇章，探索从 AIGC 生成式短剧集到“AI+实拍长剧集”，到“AI+工业化电影”的三步走模式，形成了影视 IP 开发的“N+2”模式。

资料来源：美图公司公众号、因赛集团官网、影视制作公众号、柠萌影视公众号、超讯通信官网、博纳影业公众号，信达证券研发中心

表 10：相关上市公司估值表（截至 2024.07.24）

分类	证券简称	总市值	归母净利润（百万元）				市盈率		
		(亿元)	23A	24E	25E	26E	24E	25E	26E
一站式平台型	Adobe	17,061	53483.0	59276.0	65823.0	74153.0	28.8	25.9	23.0
	美图公司	101.9	368.3	548.5	787.9	1046.0	18.6	12.9	9.7
技术服务类	商汤	375.2	-6440.0	-3645.0	-2522.0	-1742.0	-	-	-
UGC 社区类	Bilibili	434.2	-4822.32	-1734.0	-143.8	990.0	-	-302.1	43.9
视频剪辑类	快手	1,746.4	6396.0	15734.6	21711.7	27026.2	11.1	8.0	6.5
IP 类公司	阅文集团	237.1	804.9	1249.2	1400.7	1543.0	19.0	16.9	15.4
	上海电影	81.3	127.0	234.6	336.6	426.7	34.7	24.2	19.1
	汤姆猫	116.4	-864.6	200.0	300.0	400.0	58.2	38.8	29.1
	中文在线	142.6	89.4	134.3	172.7	215.1	106.2	82.6	66.3
广告营销类	易点天下	62.2	217.0	287.5	357.7	433.5	21.6	17.4	14.4
	蓝色光标	122.4	116.6	371.6	527.1	639.6	32.9	23.2	19.1
	因赛集团	49.5	41.6	/	/	/	/	/	/
	利欧股份	97.5	1966.0	/	/	/	/	/	/
视频数据训练类	华策影视	117.3	382.2	457.1	519.0	564.7	25.7	22.6	20.8
	视觉中国	75.4	145.6	169.1	197.8	224.1	44.6	38.1	33.6
	捷成股份	93.2	450.0	593.0	650.3	728.3	15.7	14.3	12.8
	中广天择	25.2	-8.7	/	/	/	/	/	/
AI 短剧/短片探索类	博纳影业	58.3	-552.6	314.1	499.0	578.3	18.6	11.7	10.1
	超讯通信	42.1	18.8	111.1	164.8	243.4	37.9	25.5	17.3
	柠萌影视	0.0	213.6	258.1	322.7	372.3	0.0	0.0	0.0
其他	光线传媒	212.4	417.8	1068.1	1212.0	1373.9	19.9	17.5	15.5

万达电影	227.7	912.2	1342.1	1669.2	1939.4	17.0	13.6	11.7
芒果超媒	355.4	3555.7	2037.7	2270.7	2520.7	17.4	15.7	14.1

资料来源：iFind，信达证券研发中心（来源于 iFind 一致预期，Adobe 来源于 Bloomberg 一致预期，取用经调整后净利润指标）

六、风险因素

AI 底层大模型发展不及预期：AI 大模型升级迭代速度减缓，多模态大模型升级不及预期；

AI 视频技术迭代不及预期：AI+视频的算法优化减缓，视频数据语料不足导致产品迭代缓慢；

AI 视频产品付费渗透率提升不及预期：AI+视频产品力较弱，用户付费意愿较低影响公司现金流。

研究团队简介

冯翠婷，信达证券传媒互联网及海外首席分析师，北京大学管理学硕士，香港大学金融学硕士，中山大学管理学学士。2016-2021 年任职于天风证券，覆盖互联网、游戏、广告、电商等多个板块，及元宇宙、体育二级市场研究先行者（首篇报告作者），曾获 21 年东方财富 Choice 金牌分析师第一、Wind 金牌分析师第三、水晶球奖第六、金麒麟第七，20 年 Wind 金牌分析师第一、第一财经第一、金麒麟新锐第三。

凤超，信达证券传媒互联网及海外团队高级研究员，本科和研究生分别毕业于清华大学和法国马赛大学，曾在腾讯担任研发工程师，后任职于知名私募机构，担任互联网行业分析师。目前主要负责海外互联网行业的研究，拥有 5 年的行研经验，对港美股市场和互联网行业有长期的跟踪覆盖。主要关注电商、游戏、本地生活、短视频等领域。

刘旺，信达证券传媒互联网及海外团队高级研究员。北京大学金融学硕士，北京邮电大学计算机硕士，北京邮电大学计算机学士，曾任职于腾讯，一级市场从业 3 年，创业 5 年（人工智能、虚拟数字人等），拥有人工智能、虚拟数字人、互联网等领域的产业经历。

李依韩，信达证券传媒互联网及海外团队研究员。中国农业大学金融硕士，2022 年加入信达证券研发中心，覆盖互联网板块。曾任职于华创证券，所在团队曾入围 2021 年新财富传播与文化类最佳分析师评比，2021 年 21 世纪金牌分析师第四名，2021 年金麒麟奖第五名，2021 年水晶球评比入围。

白云汉，信达证券传媒互联网及海外团队研究员。美国康涅狄格大学金融硕士，曾任职于腾讯系创业公司投资部，一级市场从业 2 年。后任职于私募基金担任研究员，二级市场从业 3 年，覆盖传媒互联网赛道。2023 年加入信达证券研发中心，目前主要专注于美股研究以及结合海外映射对 A 股、港股的覆盖。

分析师声明

负责本报告全部或部分内容的每一位分析师在此申明，本人具有证券投资咨询执业资格，并在中国证券业协会注册登记为证券分析师,以勤勉的职业态度,独立、客观地出具本报告；本报告所表述的所有观点准确反映了分析师本人的研究观点；本人薪酬的任何组成部分不曾与，不与，也将不会与本报告中的具体分析意见或观点直接或间接相关。

免责声明

信达证券股份有限公司(以下简称“信达证券”)具有中国证监会批复的证券投资咨询业务资格。本报告由信达证券制作并发布。

本报告是针对与信达证券签署服务协议的签约客户的专属研究产品，为该类客户进行投资决策时提供辅助和参考，双方对权利与义务均有严格约定。本报告仅提供给上述特定客户，并不面向公众发布。信达证券不会因接收人收到本报告而视其为本公司的当然客户。客户应当认识到有关本报告的电话、短信、邮件提示仅为研究观点的简要沟通，对本报告的参考使用须以本报告的完整版本为准。

本报告是基于信达证券认为可靠的已公开信息编制，但信达证券不保证所载信息的准确性和完整性。本报告所载的意见、评估及预测仅为本报告最初出具日的观点和判断，本报告所指的证券或投资标的的价格、价值及投资收入可能会出现不同程度的波动，涉及证券或投资标的的历史表现不应作为日后表现的保证。在不同时期，或因使用不同假设和标准，采用不同观点和分析方法，致使信达证券发出与本报告所载意见、评估及预测不一致的研究报告，对此信达证券可不发出特别通知。

在任何情况下，本报告中的信息或所表述的意见并不构成对任何人的投资建议，也没有考虑到客户特殊的投资目标、财务状况或需求。客户应考虑本报告中的任何意见或建议是否符合其特定状况，若有必要应寻求专家意见。本报告所载的资料、工具、意见及推测仅供参考，并非作为或被视为出售或购买证券或其他投资标的的邀请或向人做出邀请。

在法律允许的情况下，信达证券或其关联机构可能会持有报告中涉及的公司所发行的证券并进行交易，并可能会为这些公司正在提供或争取提供投资银行业务服务。

评级说明

投资建议的比较标准	股票投资评级	行业投资评级
本报告采用的基准指数 ：沪深 300 指数（以下简称基准）； 时间段：报告发布之日起 6 个月内。	**买入：**股价相对强于基准 15%以上；	**看好：**行业指数超越基准；
	增持：股价相对强于基准 5%～15%；	**中性：**行业指数与基准基本持平；
	持有：股价相对基准波动在±5% 之间；	**看淡：**行业指数弱于基准。
	卖出：股价相对弱于基准 5%以下。	

风险提示

证券市场是一个风险无时不在的市场。投资者在进行证券交易时存在赢利的可能，也存在亏损的风险。建议投资者应当充分深入地了解证券市场蕴含的各项风险并谨慎行事。

本报告中所述证券不一定能在所有的国家和地区向所有类型的投资者销售，投资者应当对本报告中的信息和意见进行独立评估，并应同时考量各自的投资目的、财务状况和特定需求，必要时就法律、商业、财务、税收等方面咨询专业顾问的意见。在任何情况下，信达证券不对任何人因使用本报告中的任何内容所引致的任何损失负任何责任，投资者需自行承担风险。