国元证券 GUOYUAN SECURITIES

传媒行业

投资评级

推荐 维持

AI视频生成：商业化加速，国产厂商表现亮眼

——AI应用系列报告

证券研究报告 2025年6月27日

证券分析师: 李典
邮箱: lidian@gyzq.com.cn
SAC执业资格证书编码:S0020516080001
联系人: 郜子娴
邮箱:gaozixian@gyzq.com.cn

报告要点

- **技术路径：从GAN-Transformer-Diffusion Model-DiT，技术演进推动行业进入可商用阶段：**20世纪90年代以来，AI视频生成持续迭代，经历了多个关键技术的突破，经历了从静态图像序列拼接成视频流-GAN-Transformer-Diffusion Model-DiT的技术演进，生成内容丰富度以及可控性持续提升。Sora验证了Diffusion和Transformer结合的有效性，并带动DiT架构逐渐成为当前重点方向。
- **行业进入商业化发展快轨，根据Fortune Business Insights预计2032年全球规模突破184亿人民币。**根据Fortune Business Insights测算，2024年AI视频生成全球规模为6.15亿美金，预计到2032年达到25.63亿美金，2025至2032年的复合增速为20%。AI视频生成产业规模受到价格以及模型能力的双重驱动。价格方面，目前国内外主流模型的API单秒价格在0.2-1元/秒，根据量子位智库测算，顶级动画电影制作成本约为十万元级/秒，本土动画电影制作成本在万元级/秒，本土TVC视频制作成本在千元级/秒，不论是对比广告视频或者影视剧/电影的单秒成本，AI视频生成均具有极大成本替代优势。另一方面模型能力持续突破，如时长方面Veo2支持理论最高2分钟的4K视频生成，可灵大师版可以通过续写方式将视频延长至3分钟。生成速度方面，以Seedance1.0为例，根据官方数据，对于 5 秒 1080p 分辨率的视频生成任务，实测推理耗时 41.4 秒（基于 L20 测试），效率大大提升。生成质量方面，现阶段AI生成视频常常需要“抽卡”，用户需反复尝试，才能生成相对理想结果，随着产品迭代，“抽卡率”有望降低。丰富度方面，声音内容陆续添加，如Veo3通过将视频像素转化为语义信号，实时生成匹配的视频声音，大大增强了视频生成可用性。下游应用来看，2B下游应用包括影视内容创作、商业广告、电商营销、城市文旅、动画教育、游戏娱乐等，应用场景多元化。2C方面，创意玩法持续拉近产品与C端用户距离，内容持续破圈。
- **竞争格局方面，可灵等国产厂商表现突出。**从模型表现来看，根据Artificial Analysis ，Seedance1.0、Veo2/3、Kling2.0、Sora、MiniMax产品表现居前，国内模型水平达到全球领先水平。从商业化来看，根据The Information，Runway24年12月ARR达到8400万美金，25年底ARR冲击2.65亿美金；可灵2025年3月年化收入运行率突破1亿美金，4月及5月单月付费金额均突破1亿人民币，二者商业化表现突出。在使用量和访问量维度来看，Kling、Sora、Veo、Hailuo、PixVerse等表现居前。
- **投资建议及相关标的。**我们看好AI视频生成行业发展，产品可用性持续提升，商业化进入发展快轨。重点关注：快手（1024.HK）以及美图公司（1357.HK）（未覆盖）。
- **风险提示：**产品推广不及预期的风险；AI技术发展不及预期的风险；监管趋严的风险；算力瓶颈的风险；AI视频技术迭代不及预期的风险；商业化不及预期的风险；AI生成内容的版权风险

1. 技术路径：从GAN-Transformer-Diffusion Model-DiT

2. AI视频生成行业：逐步进入可商用阶段，应用及商业化加速

3. 投资建议及相关标的

4. 风险提示

1.技术路径：从GAN-Transformer-Diffusion Model-DiT

- **技术路径探索，从图像拼接生成-GAN-Transformer-Diffusion Model-DiT：**相比于图片生成视频生成更为复杂，若按帧率24帧/秒，一条5秒的视频就百余张图像合成，同时视频生成还需要考虑时序连贯性、空间一致性等问题。20世纪90年代以来，AI视频生成持续迭代，经历了多个关键技术的突破，经历了从静态图像序列拼接成视频流-GAN-Transformer-Diffusion Model-DiT的技术演进，生成内容丰富度以及可控性持续提升。

图：AI视频生成技术路径探索

	探索期	早期	增长期	未来方向
	GAN	Transformer	Diffusion Model	Diffusion + Transformer
技术节点	• 2014年GAN发表，2016年左右开始用于视频生成 • 2013年VAE出现，可以通过编码器和解码器的机制学习数据的潜在表示	• 2017年Transformer发表，逐步向各个领域渗透，早期在视频生成领域也有尝试 • 2021年Google发布Video Vision Transformer	• 2020年 DDPM论文发表 • 2021年 Latent Diffusion Models论文发表 • 2022年 Video Diffusion Model 论文发表	• 2022年12月Diffusion Transformer论文发表 • 2024年2月OpenAI发布Sora，取得惊艳的生成效果，验证了Diffusion和Transformer结合的有效性
案例	• Text2Filter、TGANs-C，VGAN、TGAN、VideoGPT、MoCoGAN、DVD-GAN、DIGAN • 应用范围有限（只能总特定生成，例如数字），生成分辨率低、生成时长极短	• GODIVA、VideoGPT、Phenaki、Cog Video、NUWA • 早期基于Transformer的视频生成效果有限，在当时的成本比较高	• Stable Video Diffusion、Make-A-Video、Runway Gen1/2、Video LDM • **由于SD的开源激发社区创新，推动扩散模型在视觉生成领域占据主导地位**	• Sora、Videopoet、W.A.L.T • 由于领军企业的示范效应，Diffusion Transformer已经成为视频生成领域的核心路线，领域内其他玩家都开始聚焦DiT并取得进展 • 目前业界对于Sora的架构大多是猜想，从架构大方向上接近，但具体细节上不同模型、产品各异
特点	• **生成多样性差**：生成器-判别器对抗的模型架构决定生成效果逼真但多样性、创造性差，很难泛化 • 生成速度快：GAN可以一次出图，无需做多步推理	• **可扩展性强**：Scale up路线上目前最好的架构 • **生成速度慢、开销大**：需要把图片进行分割再转换成超长序列，计算量随分辨率增加平方级增加，复杂度高	• **生成质量高**：生成效果好，细节丰富清晰，训练要远比GAN稳定 • **显存开销大**：视频生成有大量相关依赖关系的图片同时生成，对计算显存有极高要求	• **可扩展性强**：基于扩散模型的视频模型，核心的构件是卷积神经网络U-Net，随着模型规模scale up增益会放缓甚至消失。DiT（Diffusion Transformer）将传统扩散模型的U-Net替换为Transformer，从而能够实现更强的可拓展性 • **长期来看语言模型路线[1]可成为主要路线，可以**将多种模态融合到一起（由于架构的成熟和优良的扩展性）

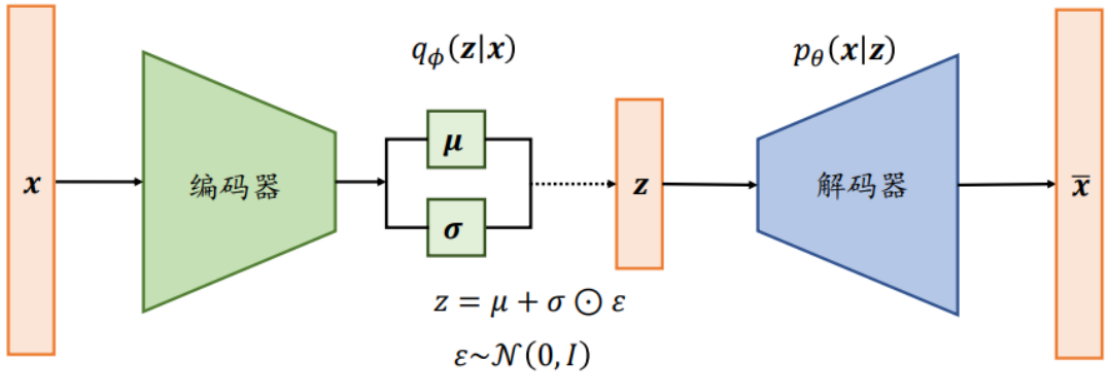
1）在视频生成的语境下，主要指基于Transformer这样的自回归模型

数据来源：量子位《2024年AI视频生成研究报告》，国元证券研究所

1.技术路径：从GAN-Transformer-Diffusion Model-DiT

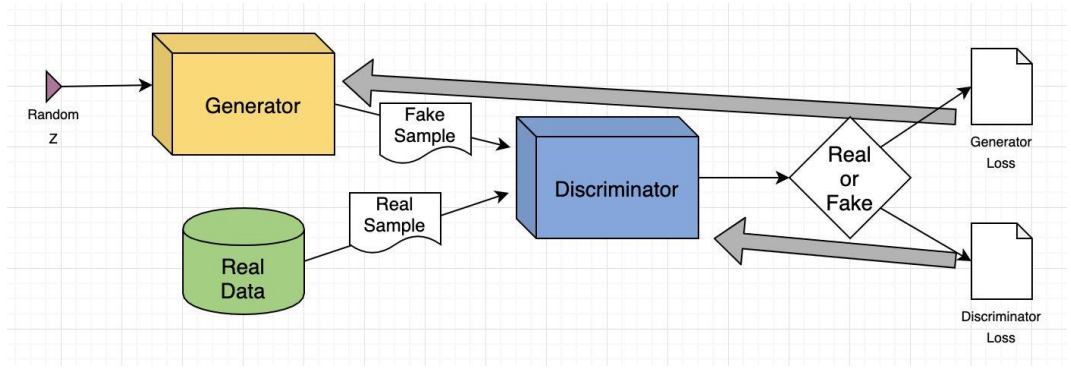
- **VAE（Variational Autoencoder 变分自编码器）**于2013年由Kingma和Welling提出，2015开始被应用于视频预测任务，通过编码器将输入数据映射到潜在空间，然后通过解码器从潜在空间生成数据。VAE生成视频内容存在模糊等问题。
- **GAN（Generative Adversarial Networks,对抗生成网络）**于2014年由Ian Goodfellow等人提出，2016年左右开始用于视频生成，主要由生成器（Generator）和判别器（Discriminator）两部分组成，生成器负责生成数据，判别器负责判断数据的真实性，两者在训练过程中不断竞争，生成器学习产生越来越真实的数据，而判别器学习更好地区分真假数据，通过两者相互博弈来实现高质量图片或是视频的生成。对抗训练提升了生成内容的质量，但存在多样性有限、训练不稳定容易出现梯度消失和模式崩溃等问题。**VAE+GAN**，VAE与GAN常常结合使用，使生成的视频内容更加多样化，并改善生成视频的清晰度和真实性。基于GAN的视频生成产品包括VideoGAN、TGAN等。

图：VAE模型示意图



数据来源：Diederik P Kingma, Max Welling《Auto-Encoding Variational Bayes》，阿里云开发者社区，国元证券研究所

图：GAN模型示意图

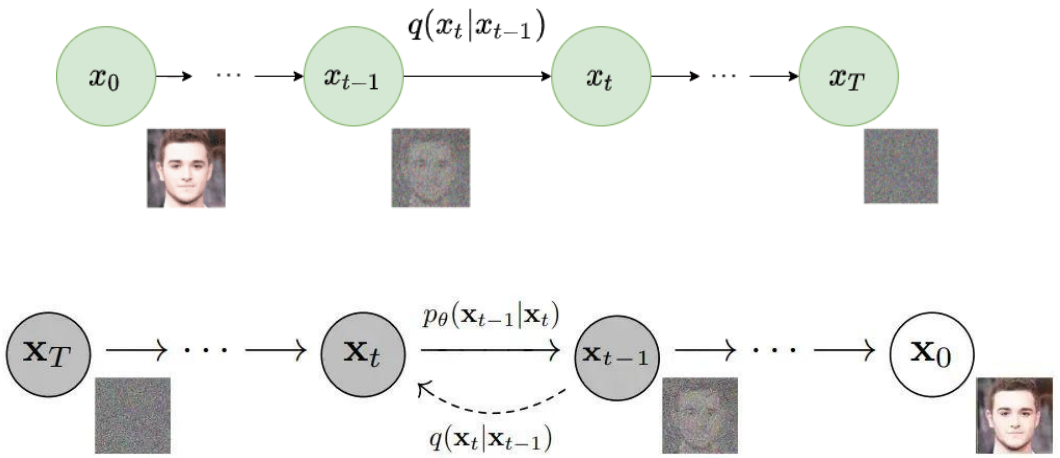


数据来源：Ian J. Goodfellow等《Generative Adversarial Networks》，AWS，国元证券研究所

1.技术路径：从GAN-Transformer-Diffusion Model-DiT

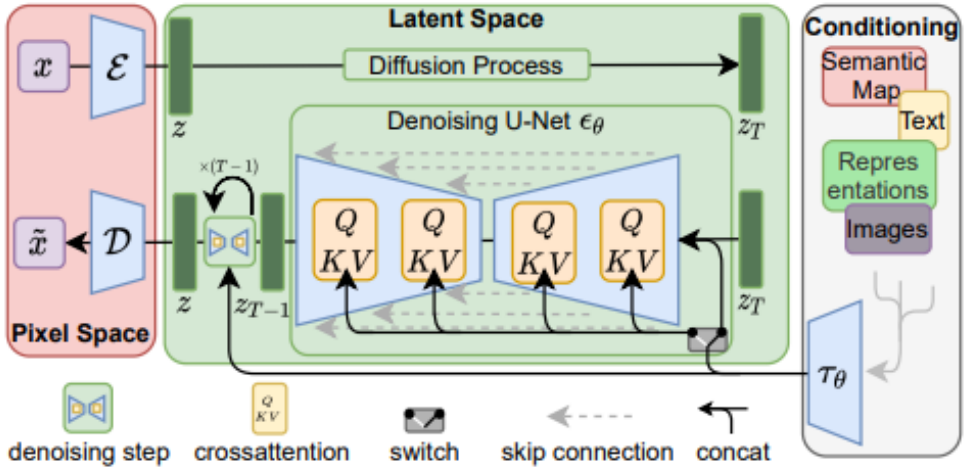
- **Transformer:** Transformer架构于2017年由Google团队提出并逐渐应用于视频生成领域。2020年ViT论文发表，2021Google发表了Video Vision Transformer。Transformer架构更好的建模像素点之间的上下文关系，提升视觉内容生成质量，缺点在于高昂的成本和较慢的生成速度。代表性产品包括VideoGPT、Phenaki等。
- **Diffusion Model：** 扩散模型灵感源于热力学，由一个逐级添加高斯噪声的前向扩散过程和一个逐级预测并消除噪声的反向过程组成。2020年Ho等人的论文《Denoising diffusion probabilistic models》(DDPM) 发表成为关键节点，2021年Latent Diffusion Models论文发表，将扩散空间从图像空间转移到潜空间（Latent Space），大幅提升计算效率。 2022年Video Diffusion Model论文发表，通过加入时序注意力层的方式将扩散模型中的二维U-Net扩展至三维，将扩算模型应用于视频生成任务。另外，基于LDM的开源系列模型Stable Diffusion推出，进一步推动扩散模型在视觉生成领域的主导作用。

图：扩散模型中前向扩散过程以及反向扩散过程



数据来源：Ho等《Denoising Diffusion Probabilistic Models》，《How diffusion models work: the math from scratch》，国元证券研究所

图：LDM模型示意图

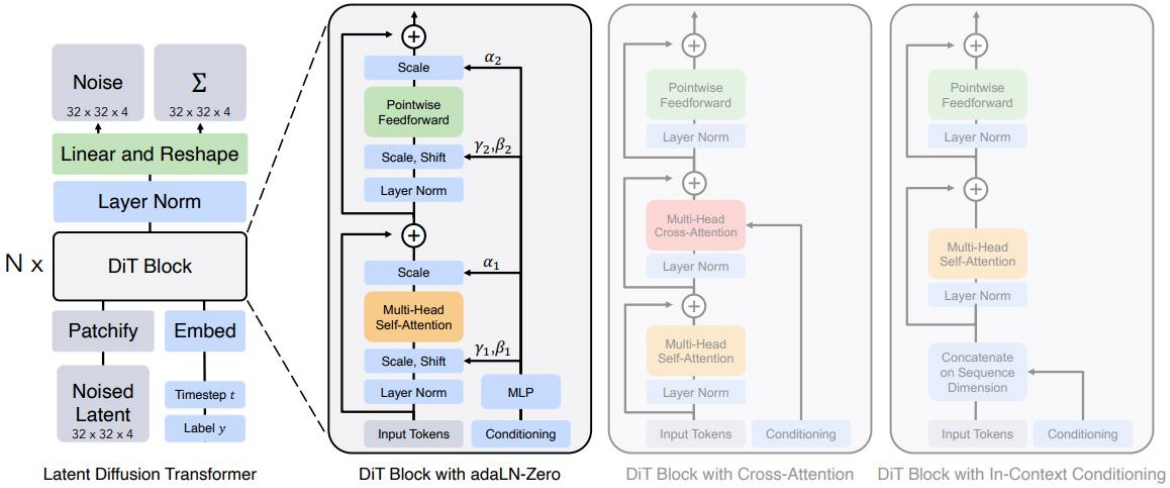


数据来源：Robin Rombach等《High-Resolution Image Synthesis with Latent Diffusion Models》，国元证券研究所

1.技术路径：从GAN-Transformer-Diffusion Model-DiT

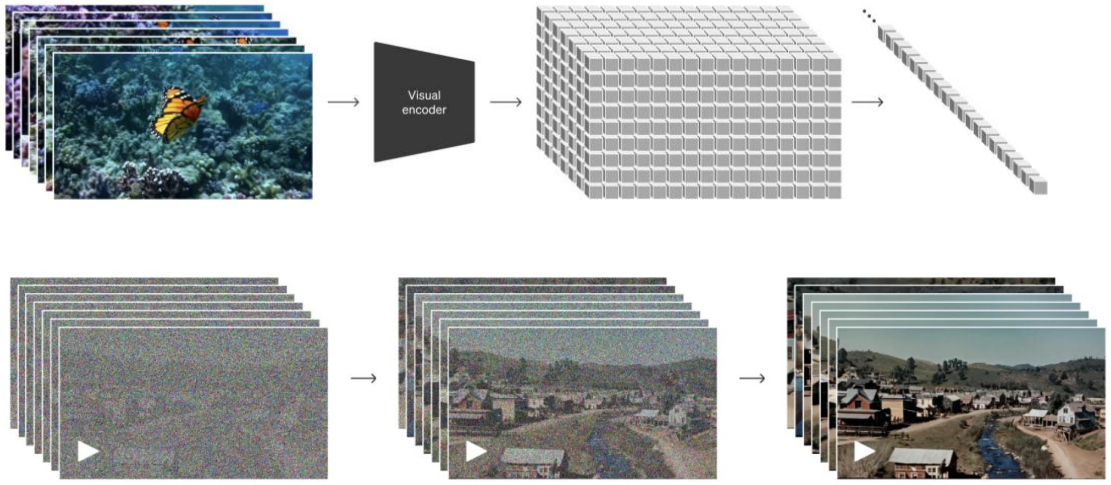
- **DiT：**2022年12月《Scalable Diffusion Models with Transformers》论文发表，提出将传统扩散模型的U-Net替换为Transformer，从而提升传统扩散模型的可扩展性。采用AdaLN-Zero（自适应归一化）注入条件信息（文本/图像/轨迹），取代交叉注意力。2024年2月OpenAI发布Sora，验证了Diffusion和Transformer结合的有效性，并带动DiT架构成为重点方向。根据Sora发布的技术文章，Sora模型先将视频压缩到低维潜在空间中，然后将表示分解为时空patch，从而实现了视频的“patch化”，采用DiT架构，在多个领域展示了显著的扩展性能。同时对比过去的视频生成模型，Sora生成视频时长显著提升（60秒），对于自然语言和物理世界规律有了更强的理解能力。

图：DiT模型示意图



数据来源：William Peebles等《Scalable Diffusion Models with Transformers》，国元证券研究所

图：Sora的模型情况



数据来源：OpenAI《Video generation models as world simulators》，国元证券研究所

1.技术路径：从GAN-Transformer-Diffusion Model-DiT

- **Sora为行业带来深刻启发，不少厂商跟进DiT架构，并在此之上进行创新。**以可灵为例，可灵使用了DiT架构，同时对模型中的隐空间编/解码、时序建模等模块进行了升维处理。Seedance1.0引入精确描述模型提升数据多样性与可用性，统一高效的预训练框架实现多镜头切换与多模态输入，后训练构建复合奖励系统提升画面生动性、稳定性和美感。
- **行业围绕提升效率和可控性以及长视频生成等目标持续迭代。**比如扩散模型的优化策略和模型蒸馏、扩散模型的混合专家（MoE）架构、引入人类反馈的强化学习、分层生成等，另外除了DiT架构之外，其他几种混合架构研究也在持续推进（右表）。

表：部分采用DiT架构的模型

产品名称	介绍
可灵	DiT架构，快手还对模型中的隐空间编/解码、时序建模等模块进行了升维处理。 （1）在隐空间编/解码上，快手自研了3D VAE网络，实现时空同步压缩，获得了较高的重建质量，在训练性能和效果之间取得了极佳平衡。 （2）在时序信息建模上，快手设计了一款计算高效的全注意力机制作为时空建模模块。该方法可以巧妙地融合时间与空间的信息，能对视频数据进行综合分析和处理，可精准捕捉到视频帧内的局部空间特征以及跨帧的时间动态特征，从而更全面地理解和再现视频中的运动信息。因此，从快速移动的物体、剧烈变化的场景，到复杂的人物动作，可灵都能精确捕捉，生成的视频内容动态性十足，同时具有很高的物理世界真实感。
Seedance	采用DiT架构，引入精确描述模型提升数据多样性与可用性，统一高效的预训练框架实现多镜头切换与多模态输入，后训练构建复合奖励系统提升画面生动性、稳定性和美感。 在模型架构和训练策略上的优化包括： （1）构建了解耦空间层和时间层的扩散 Transformer 模型，空间层在单帧内部执行注意力聚合，而时间层则专注于跨帧的注意力计算。 （2）多镜头多模态旋转位置编码。 （3）统一的任务框架，不同生成任务（如文本到图像、文本生视频和图像生视频）之间相互学习，同时，部署侧只需一个模型就可实现多种任务，降低了模型应用门槛。
腾讯混元	DiT架构，并在架构设计上进行多处升级，包括引入多模态大语言模型作为文本编码器、基于自研Scaling Law的全注意力DiT、自研3D VAE等。

数据来源：可灵官网，字节跳动Seed微信公众号，华尔街见闻，国元证券研究所

表：部分混合架构模型情况

混合模型架构	介绍
Transformer +3D卷积网络	一些研究工作探索了将Transformer与3D卷积网络结合的架构。3D卷积网络擅长捕捉视频的时空特征，而Transformer能够处理长距离依赖关系。这种混合架构能够同时利用两者优势，提高视频生成的连贯性和真实感
自回归模型 +Transformer	自回归模型如PixelRNN或PixelCNN在像素级别上生成数据，而Transformer则能够处理序列数据。将两者结合可以生成高质量的视频内容，同时保持对长距离依赖的有效建模
VAE+扩散模型	在DiT的基础上，学者们进一步探索了将变分自编码器与扩散模型结合的架构。这种混合模型首先使用VAE将高分辨率的视频编码到低维潜在空间，然后在潜在空间上应用扩散模型进行生成，从而在保持高生成质量的同时减少计算成本
GAN+扩散模型	GAN在图像生成领域取得了巨大成功，而扩散模型则在样本质量上展现出了优势。将两者结合的混合架构旨在结合GAN的稳定性和扩散模型的高质量样本生成能力
TimeSformer	TimeSformer是一种专门为视频理解任务设计的Transformer架构，它通过将时间维度融入到标准的Transformer架构中，有效地处理视频数据。这种架构可以用于视频生成任务，生成具有时空一致性的内容
交叉注意力机制	在Transformer架构中引入交叉注意力机制，可以使模型在生成视频时更好地利用额外的上下文信息，如文本描述或对象标签，从而生成与这些上下文信息更加一致的视频内容

数据来源：量子位《量子位：2024年度AI十大趋势报告》，国元证券研究所

1. 技术路径：从GAN-Transformer-Diffusion Model-DiT

2. AI视频生成行业：逐步进入可商用阶段，应用及商业化加速

2.1 驱动因素：价格和性能共同驱动行业成长

2.2 行业应用：B端应用场景多元化，C端内容持续破圈

2.3 产品及竞争格局：可灵等国产厂商表现突出

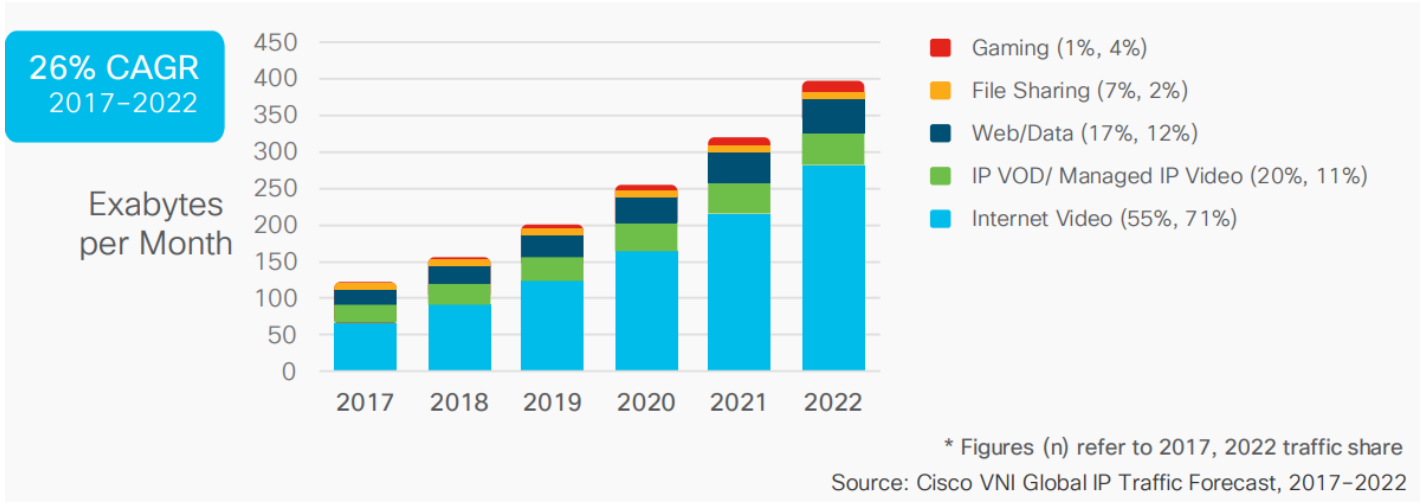
3. 投资建议及相关标的

4. 风险提示

2.AI视频行业：逐步进入可商用阶段，应用及商业化加速

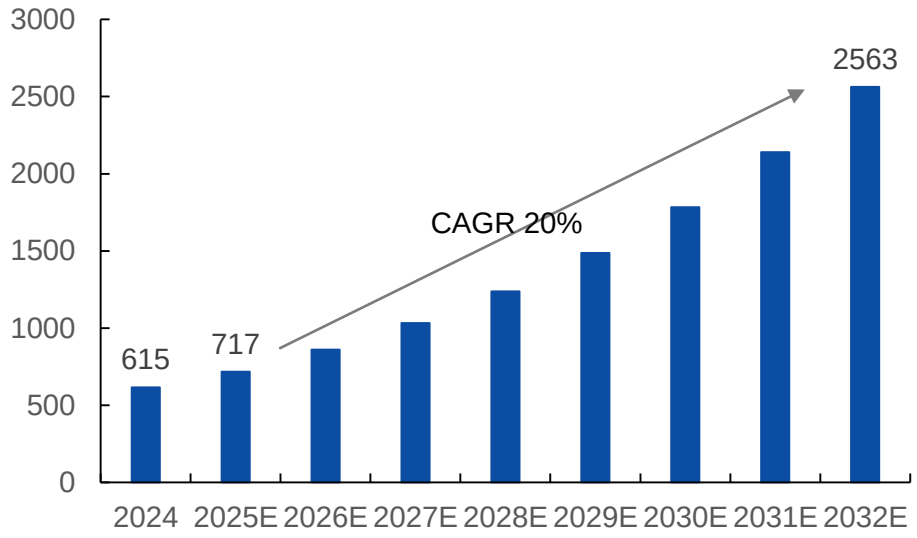
- **全球视频内容占互联网流量的比重持续攀升预计超8成**，根据Cisco的预测，2022年视频内容占互联网流量的比重预计由2017年的75%增加至82%，而中国方面，Cisco预测2020年中国视频内容流量占比达到87%。
- **根据Fortune Business Insights，预计到2032年AI视频生成全球规模达到25.63亿美金，约合184亿人民币。**根据Fortune Business Insights测算，2024年AI视频生成全球规模为6.15亿美金，预计2025年达到7.17亿美金，同比增长17%，预计到2032年达到25.63亿美金，2025至2032年的复合增速为20%。分应用来看，AI视频生成下游应用包含2B及2C两端，根据Fortune Business Insights的统计，目前AI视频生成下游需求最大的应用方向为市场营销及广告，而社交媒体的应用增长最快。

图：全球视频内容流量占比



数据来源：Cisco《Visual Networking Index: Forecast and Trends, 2017–2022》，国元证券研究所

图：AI视频生成全球规模（百万，美元）

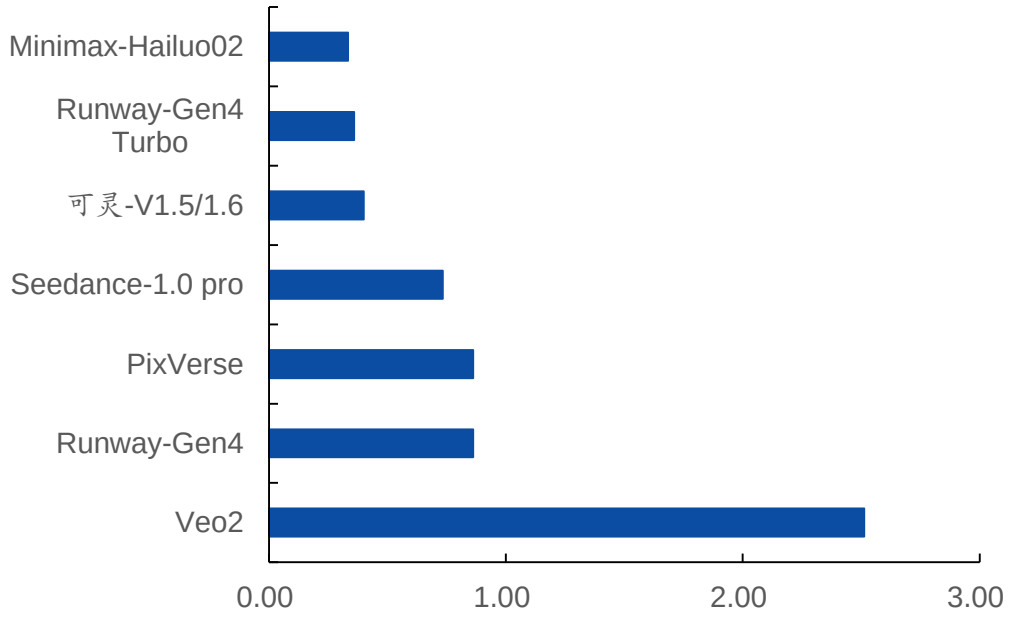


数据来源：Fortune Business Insights，国元证券研究所

2.1驱动因素：价格和性能共同驱动行业成长

- **AI视频生成产业规模受到价格以及模型能力的双重驱动。**
- **价格方面，目前国内外主流模型的API单秒价格在0.2-1元/秒。**目前主流产品采取积分制，大部分支持API接入，价格方面目前国内外主流模型的API单秒价格在0.2-1元/秒，相对价格较高的Veo2生成价格为0.35美元/秒（约2.51元/秒），根据量子位智库测算，顶级动画电影制作成本约为十万元级/秒，本土动画电影制作成本在万元级/秒，本土TVC视频制作成本在千元级/秒，不论是对比广告视频或者影视剧/电影的单秒成本，AI视频生成均具有极大成本替代优势。

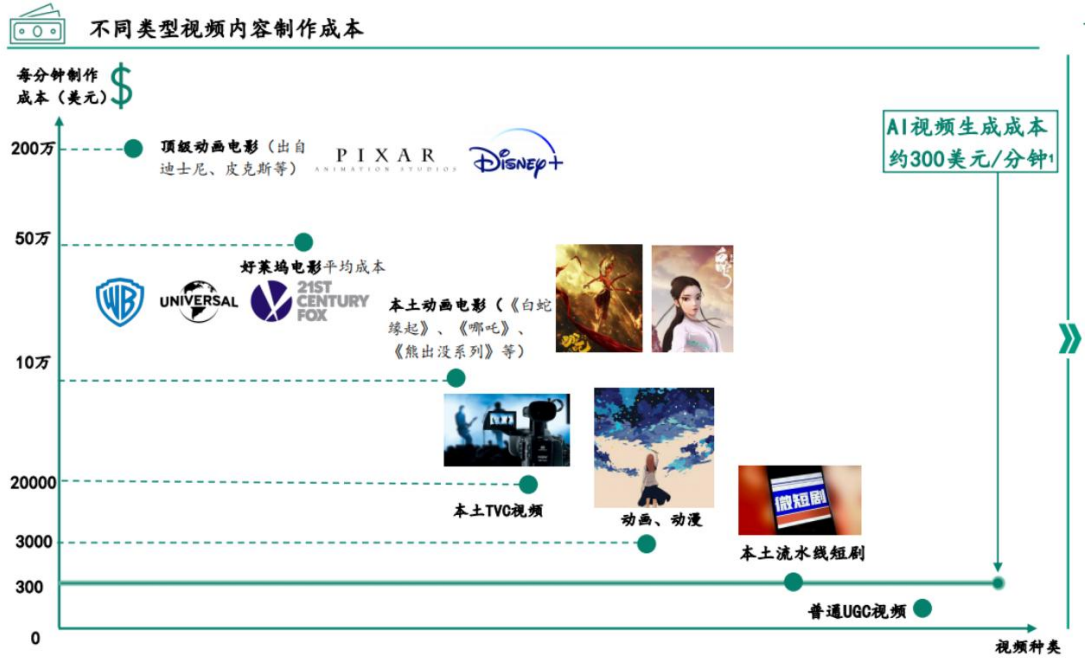
图：部分AI视频生成产品API单秒价格情况（元/秒）



数据来源：各公司官网，国元证券研究所

注：Vidu2.0选取720P 4s时长对应价格计算，Hailuo2.0采用768P 6s时长对应价格计算，可灵采用720p 5s对应价格计算，PixVerse采用720p 5s时长对应价格计算

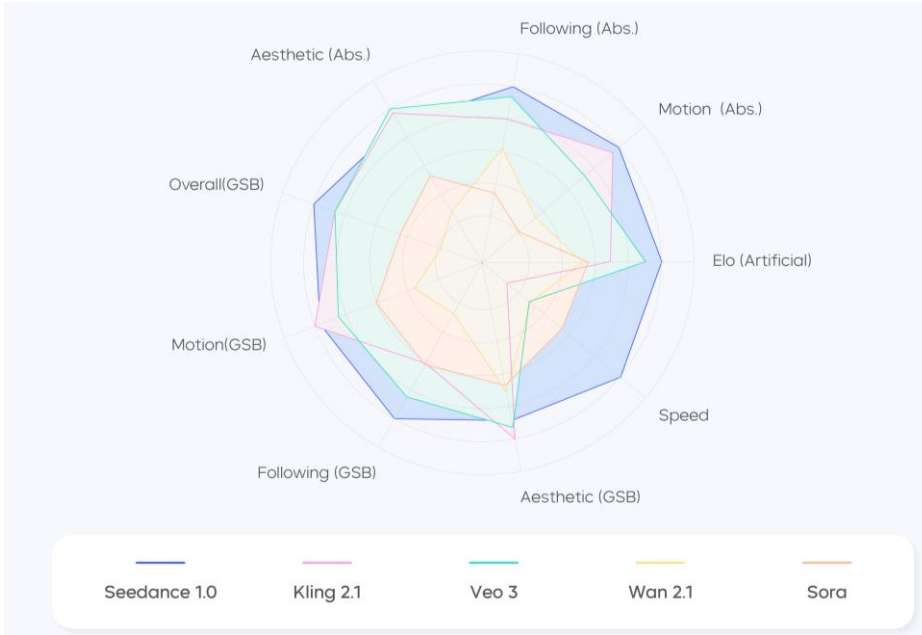
图：不同类型视频内容制作成本对比

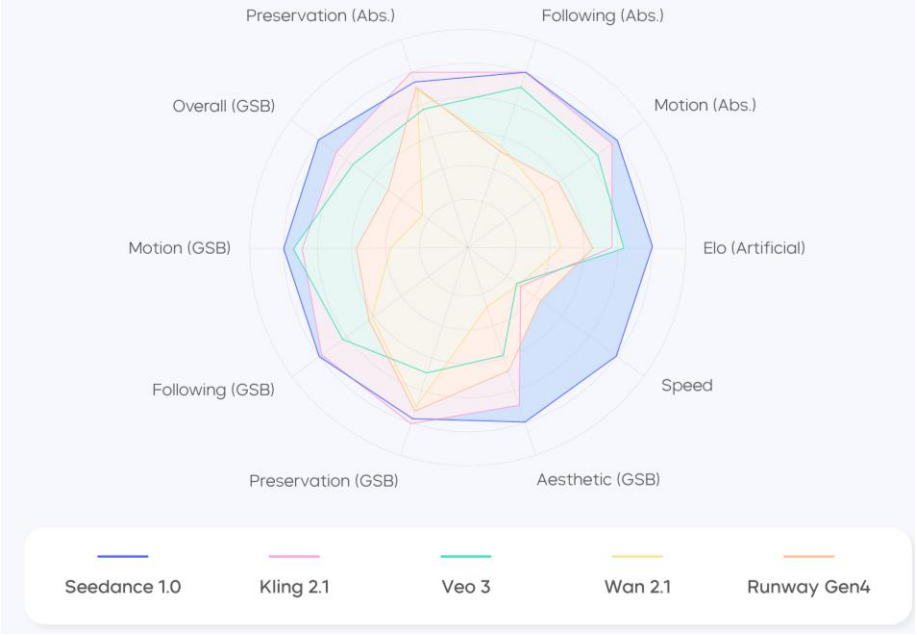


数据来源：量子位智库《AI视频生成研究报告》，国元证券研究所

- **模型能力提升推动可用性。**视频大模型主要能力评估标准包括生成效率、内容语义的一致性、视频质量（包括时空一致性、物理规则遵循、叙事连贯性、长序列稳定性、单帧图像的生成质量等）、动态表现、画面美学等。**时长方面，**当前主流的生成时长为5-10秒，随着模型迭代时长上限逐渐突破，Veo2支持理论最高2分钟的4K视频生成，可灵大师版可以通过续写方式将视频延长至3分钟，时长突破有望进一步满足叙事需求扩展下游应用。**生成速度方面，**随着硬件性能提升以及模型迭代，整体生成速度大幅提高，以Seedance1.0为例，根据官方数据，对于 5 秒 1080p 分辨率的视频生成任务，实测推理耗时 41.4 秒（基于 L20 测试），效率大大提升。**生成质量方面，**现阶段AI生成视频常常需要“抽卡”，用户需反复尝试，才能生成相对理想结果，随着产品迭代，“抽卡率”有望降低。**丰富度方面，**声音内容陆续添加，如Veo3通过将视频像素转化为语义信号，实时生成匹配的视频声音，大大增强了视频生成可用性。

图：SeedVideoBench-1.0多维度评测





数据来源：字节跳动Seed官网，国元证券研究所

2.2行业应用：B端应用场景多元化，C端内容持续破圈

- **2B下游应用包括影视内容创作、商业广告、电商营销、城市文旅、动画教育、游戏娱乐等。**以电商服装行业为例，由于高频上新的需求，视频内容成本高企，AI视频内容生成极大解决该痛点。影视短片方面，AI短剧应用逐渐成熟，比如快手星芒短剧于24年7月13日推出国内首部AIGC原创奇幻微短剧《山海奇镜之劈波斩浪》，导演陈坤曾在接受《综艺报》的采访中表示，玄幻和科幻是传统影视的两大痛点，CG特效成本高、周期长，AI可以有效缩短制作周期、降低制作成本，效果可以达到传统特效的70%到80%。根据短剧自习室数据，该剧累计播放量突破5000万。
- **2C方面，创意玩法持续拉近产品与C端用户距离，内容持续破圈。**比如可灵的“毛绒”特效、PixVerse的“毒液”特效等，引发在社媒上广泛传播。

表：快手可灵应用场景相关内容

应用场景	介绍
影视短片	依托卓越的视频图文生成能力及强大的创作者生态，可为客户定制风格多元的影视短片。创意驱动、AI赋能，让更多好故事被看到。
商业广告	结合行业特性及品牌调性产出优质广告素材，支持丰富营销玩法。奇趣灵感，AI为商业化素材提供更多可能性。
电商营销	智能生成商品展示视频，快速制作适配不同场景的高质量内容，提升用户体验与转化率，助力电商营销更精准、更高效。
城市文旅	智能生成城市特色文化与景观，打造沉浸式文旅宣传视频，吸引游客探索城市魅力，加速文旅经济增长。
动画教育	通过生成生动的教育视频和互动内容，帮助深入理解复杂概念，增强学习兴趣与参与度，使教育更加高效和有趣。
音乐MV	帮助AI创作者快速实现创意，智能生成丰富的视觉效果和创意场景，增强观众的沉浸体验。
游戏娱乐	智能生成高质量角色、场景与特效，提升游戏开发效率。快速制作吸引人的宣传视频，生动展示游戏特点，助力游戏的发布与推广。

数据来源：快手可灵官网，国元证券研究所

2.3产品及竞争格局：可灵等国产厂商表现突出

- **从模型表现来看，Seedance1.0、Veo2/3、Kling2.0、Sora、MiniMax产品表现居前，国内模型水平达到全球领先水平。**根据Artificial Analysis最新AI视频模型表现排行，文生视频方面，字节跳动的Seedance1.0、Google的Veo3/Veo2、以及快手Kling2.0、OpenAI的Sora表现前五；在图生视频方面字节跳动的Seedance1.0、MiniMax的Hailuo02、Google的Veo3以及快手Kling2.0/1.6排名前五。
- **从商业化来看，**根据The Information，Runway24年12月ARR（Annual Recurring Revenue）达到8400万美金，25年底冲击2.65亿美金，可灵2025年3月年化收入运行率突破1亿美金，4月及5月单月付费金额均突破1亿人民币。

表：Artificial Analysis 文生视频及图生视频排行榜（统计时间：2025年6月19日）

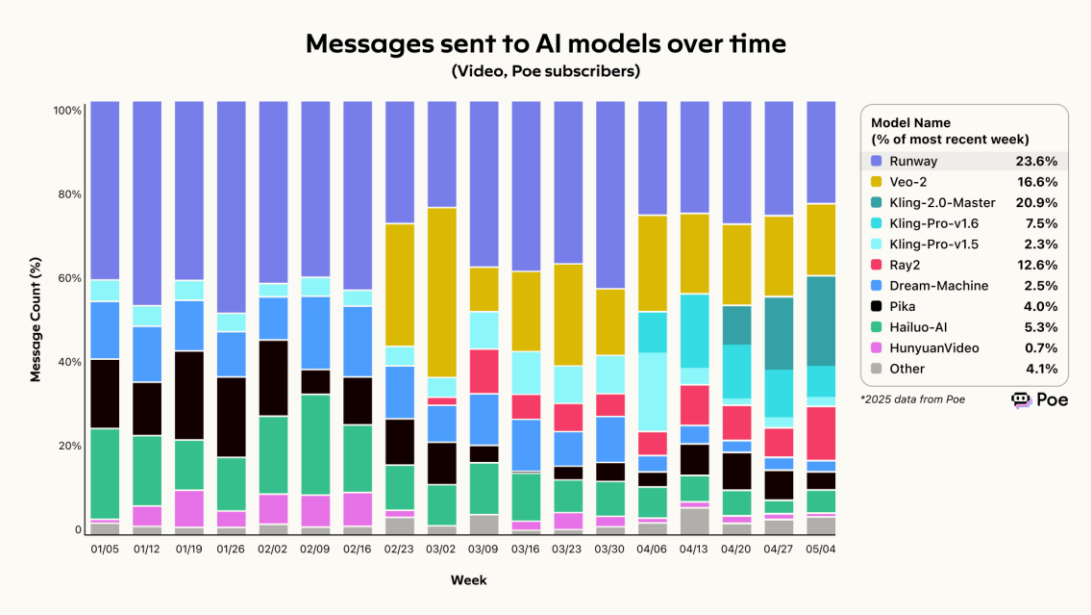
文生视频			图生视频		
排名	公司	模型	排名	公司	模型
1	ByteDance	Seedance 1.0	1	ByteDance	Seedance 1.0
2	Google	Veo 3 Preview (No Audio)	2	MiniMax	Hailuo 02 (0616)
3	Google	Veo 2	3	Google	Veo 3 Preview (No Audio)
4	Kuaishou	Kling 2.0	4	Kuaishou	Kling 2.0
5	OpenAI	Sora	5	Kuaishou	Kling 1.6 (Pro)
6	Kuaishou	Kling 1.5 (Pro)	6	Runway	Runway Gen 4
7	MiniMax	T2V-01	7	Google	Veo 2
8	Pika	Pika 2.0	8	MiniMax	I2V-01-Director
9	Alibaba	Wan 2.1 14B	9	Runway	Runway Gen 3 Alpha Turbo
10	Kuaishou	Kling 1.6 (Standard)	10	Alibaba	Wan 2.1 14B
11	Kuaishou	Kling 1.6 (Pro)	11	Pika Art	Pika 2.2
12	MiniMax	T2V-01-Director	12	Runway	Runway Gen 3 Alpha
13	Tencent	Hunyuan Video (Open Source)	13	OpenAI	Sora

数据来源：Artificial Analysis，国元证券研究所

2.3产品及竞争格局：可灵等国产厂商表现突出

- **使用量份额角度来看，Kling表现突出。**根据Poe发布的数据，截止2025年5月4日前一周使用份额情况，Kling市场份额达到了30%，其中Kling2.0大师版4月底推出后仅三周便占据了21%的市场份额，Runway使用量份额排名第二达到23.6%，Veo-2位列第三，份额为16.6%。
- **从web端访问量及APP端MAU来看，**根据AI产品榜统计的5月AI视频产品web端访问量来看，Sora、Kling、Loom、Hailuo、PixVerse访问量排名前五。移动端来看，即梦5月MAU达到3065万，PixVerse达到1998万，分别排进5月全球AI APP MAU的第11位以及第23位。

图：AI视频生成使用量情况



数据来源：Poe《2025春季人工智能模型使用趋势》2025.5.13，国元证券研究所

表：5月AI视频Web端排行榜

排名	名称	Web访问量
1	Sora | ChatGPT	63.34M
2	Kling ai | 快手	17.21M
3	Loom	16.45M
4	Hailuo AI | MiniMax	14.74M
5	PixVerse	10.78M
6	runway	7.61M
7	Sora.com | OpenAI	7.18M
8	HeyGen	6.54M
9	Vidnoz AI	5.49M
10	Pollo	4.09M
11	VIVA	3.46M
12	Vidu 海外|生数科技	3.41M
13	Viggle	3.18M
14	Luma AI	2.61M
15	Synthesia	2.5M
16	Pika Art	2.21M
17	Animaker	2.17M
18	Fliki	1.96M
19	Clipfly	1.81M
20	virbo	1.68M

数据来源：AI产品榜，国元证券研究所

1. 技术路径：从GAN-Transformer-Diffusion Model-DiT

2. AI视频生成行业：逐步进入可商用阶段，应用及商业化加速

3. 投资建议及相关标的

3.1 快手——可灵产品表现亮眼，商业化加速

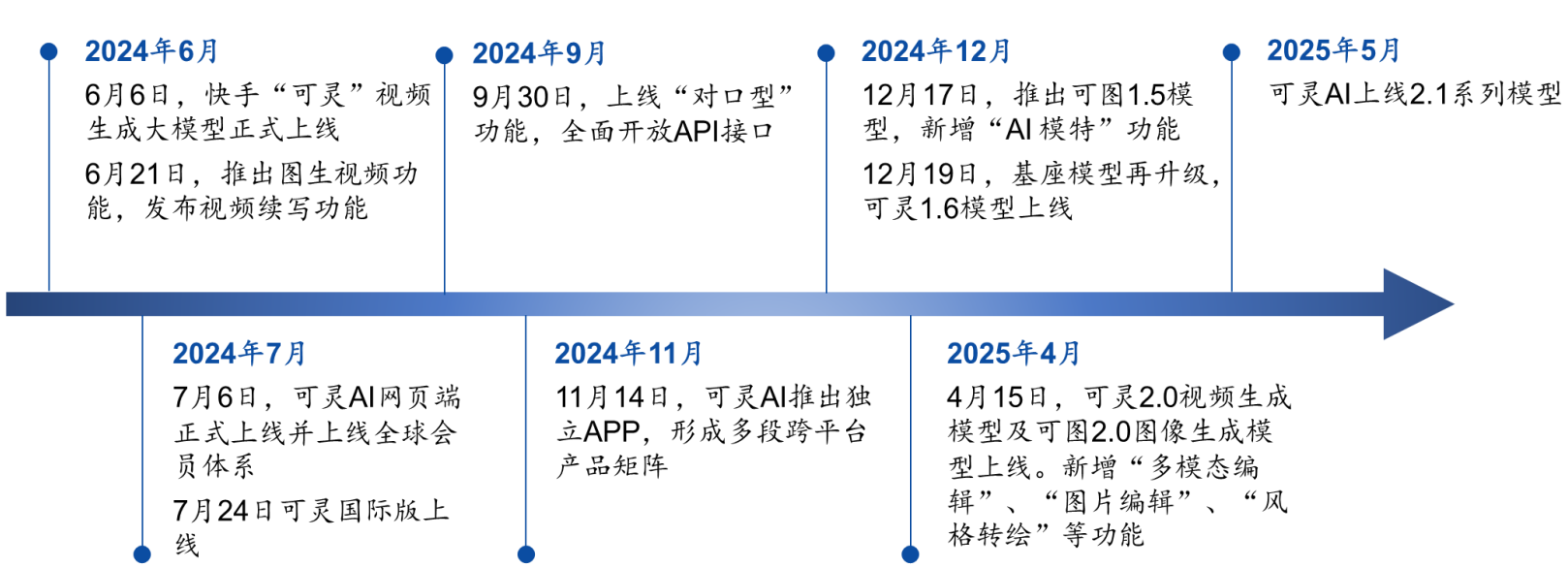
3.2 美图公司（未覆盖）——AI赋能，与阿里深化合作

4. 风险提示

3.1快手——可灵产品表现亮眼，商业化加速

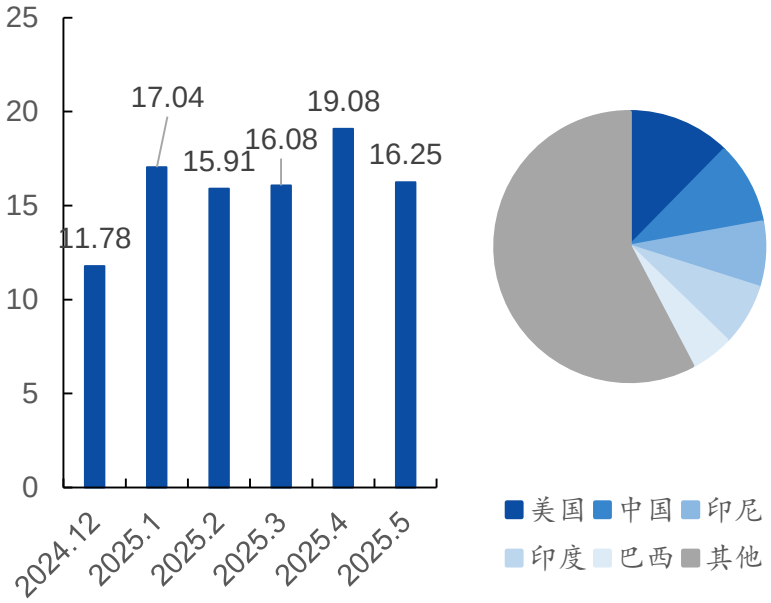
- 快手可灵（Kling）于2024年6月发布，采用DiT架构并对模型中的隐空间编/解码、时序建模等模块进行了升维处理。2024年7月可灵国际版上线，2024年12月可灵1.6模型发布，25年4月15日可灵大模型2.0版本及可图大模型2.0版本正式面向全球发布，5月2.1系列模型上线。截至目前，可灵AI已累计完成超20次迭代，全球用户规模已突破2200万，拥有来自世界各地的超1.5万开发者和企业客户，累计生成1.68亿个视频及3.44亿张图片素材。
- 根据SimilaWeb，Klingai.com5月访问量达到1625万次，访问流量分布来看，美国、中国、印尼、印度、巴西为5月访问量前五的国家。

图：可灵AI发展和迭代情况



数据来源：官方微信公众号，国元证券研究所

图：月度web端（klingai.com）访问量（百万次）以及2025年5月流量分布情况（%）

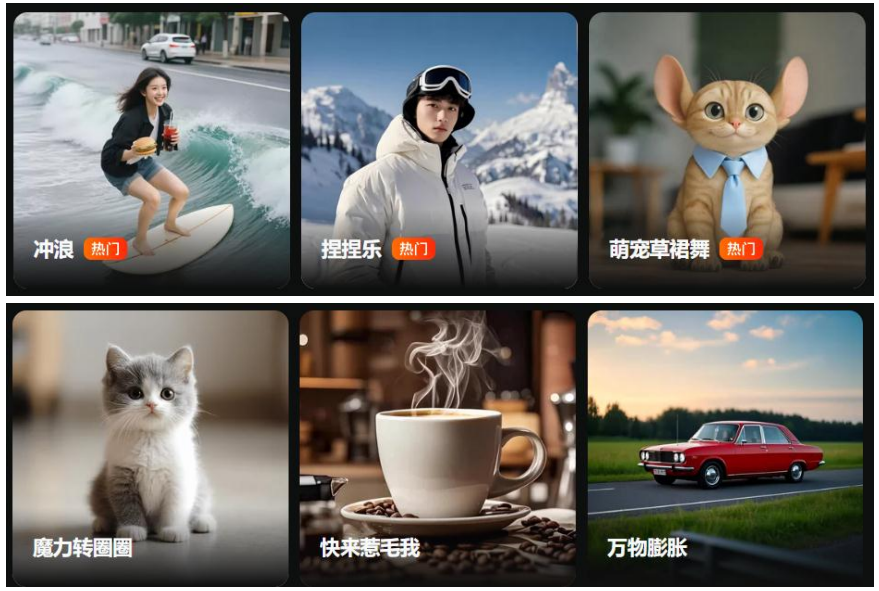


数据来源：SimilarWeb，国元证券研究所

3.1快手——可灵产品表现亮眼，商业化加速

- **产品及功能方面，**目前可灵具有可灵网页版、可灵APP、可灵小程序以及快影APP四个产品形态。具有运镜控制、首尾帧能力、运动笔刷、对口型等进阶功能。对AI生成后的视频可续写4-5秒，支持多次续写（最长3分钟），实现视频延长。同时可灵支持多图参考，可以上传最多四张参考图，内容可以是人物、动物、场景、服饰等各种主体或背景图片，帮助生成更加可控的AI创意视频。创意特效方面，“捏捏乐”、“快来惹毛我”、“老照片动起来”助力可灵在社媒传播破圈。
- **商业化方面，**2025年2月可灵上线后累计收入超1亿人民币，25年Q1可灵收入超1.5亿人民币，2025年3月年化收入运行率突破1亿美金，月度付费金额在4月和5月均突破1亿人民币，营收增速和营收水平均位居全球视频生成大模型产品和独立AI产品前列。
- **我们预计公司2025-2027年经调整净利润200.77/246.20/287.21亿元，维持公司“买入”评级。**

图：可灵的部分创意特效



数据来源：可灵AI，国元证券研究所

表：快手盈利预测

财务数据和估值	2023	2024	2025E	2026E	2027E
营业收入(百万元)	113,470	126,898	141,907	156,661	170,402
收入同比（%）	20.48	11.83	11.83	10.40	8.77
归母净利润(百万元)	6,399	15,344	17,213	21,463	25,293
经调整归母净利润(百万元)	10,271	17,716	20,077	24,620	28,721
每股收益（元）	1.49	3.56	4.00	4.98	5.87
经调整每股收益（元）	2.38	4.11	4.66	5.71	6.67
经调整市盈率(P/E)	20.31	9.28	11.65	9.50	8.14

数据来源：Wind，国元证券研究所
注：PE采用2025年6月25日收盘价

3.2美图公司（未覆盖）——AI赋能，与阿里深化合作

- **AI赋能业务，加速全球化，深入生产力场景打造第三曲线。**美图公司成立于2008年，致力于打造优秀的影像与设计产品。公司目前C端产品包括美图秀秀、美颜相机、Wink等，B端生产力场景主要产品包括美图设计室、开拍、WHEE、MOKI等，并通过自研美图奇想大模型+采购+开源方式获取AI模型能力持续赋能业务。C端方面，美图秀秀全球累计10亿用户，在图片美化赛道长期保持用户规模第一。美颜相机用户规模及月活跃用户数保持行业领先。AI视频工具Wink于2022年3月独立推出，目前全球月活突破3000万。24年“AI发型”、“AI换装”等AI功能带动海外用户增长。截至2024年底，美图公司全球月活跃用户数达到2.66亿，中国内地以外的国家和地区月活跃用户数为9451万，同比增长21.7%。生产力场景方面，目前月活用户数2219万，同比增长25.6%，涵盖商业摄影、商业设计、专业视频编辑等环节。其中美图设计室截止24年底订阅用户113万，单产品收入约2亿元，聚焦商品营销各个环节物流设计；开拍则聚焦于商品口播，推出一年多月活跃用户数突破100万。5月20日，美图公司宣布与阿里巴巴签署2.5亿美元可转债协议，双方将在电商和模型开发领域展开深入合作。阿里巴巴将在旗下的电商平台优先推广美图AI电商工具，同时协助美图开发以数据驱动的电商生图及生视频的新工具及新功能，美图公司承诺未来三年向战略投资者阿里巴巴采购不低于5.6亿元的云服务。
- **根据Wind一致预期，预计2025-2027公司实现营业收入42.43/51.69/61.56亿元，归母净利润为8.10/10.92/14.58亿元。**

表：美图公司模型层、生态层、应用层一览

数据来源：美图公司官方微信公号，国元证券研究所

表：美图公司盈利预测（采用Wind一致预期）

财务数据和估值	2023	2024	2025E	2026E	2027E
营业收入(百万元)	2,695.74	3,340.76	4,252.94	5,168.64	6,156.10
同比（%）	29.27	23.93	27.30	21.53	19.10
归母净利润(百万元)	378.29	805.18	810.32	1,091.45	1,458.44
同比（%）	301.83	112.84	0.64	34.69	33.63
每股收益（元）	0.08	0.18	0.18	0.24	0.32
PE	38.62	15.56	42.96	31.89	23.87

数据来源：Wind，国元证券研究所
注：PE采用2025年6月25日收盘价，盈利预测采用Wind 90天一致预期

1. 技术路径：从GAN-Transformer-Diffusion Model-DiT

2. AI视频生成行业：逐步进入可商用阶段，应用及商业化加速

3. 投资建议及相关标的

4. 风险提示

4.风险提示

- 产品推广不及预期的风险
- AI技术发展不及预期的风险
- 监管趋严的风险
- 算力瓶颈的风险
- AI视频技术迭代不及预期的风险
- 商业化不及预期的风险
- AI生成内容的版权风险

投资评级说明

（1）公司评级定义		（2）行业评级定义	
买入	股价涨幅优于基准指数 15%以上	推荐	行业指数表现优于基准指数 10%以上
增持	股价涨幅相对基准指数介于 5%与 15%之间	中性	行业指数表现相对基准指数介于-10%~10%之间
持有	股价涨幅相对基准指数介于-5%与 5%之间	回避	行业指数表现劣于基准指数 10%以上
卖出	股价涨幅劣于基准指数 5%以上		

备注：评级标准为报告发布日后的6个月内公司股价（或行业指数）相对同期基准指数的相对市场表现，其中A股市场基准为沪深300指数，香港市场基准为恒生指数，美国市场基准为标普500指数或纳斯达克指数，新三板基准指数为三板成指（针对协议转让标的）或三板做市指数（针对做市转让标的），北交所基准指数为北证50指数。

分析师声明

作者具有中国证券业协会授予的证券投资咨询执业资格或相当的专业胜任能力，以勤勉的职业态度，独立、客观地出具本报告。本人承诺报告所采用的数据均来自合规渠道，分析逻辑基于作者的职业操守和专业能力，本报告清晰准确地反映了本人的研究观点并通过合理判断得出结论，结论不受任何第三方的授意、影响，特此声明。

证券投资咨询业务的说明

根据中国证监会颁发的《经营证券业务许可证》（Z23834000），国元证券股份有限公司具备中国证监会核准的证券投资咨询业务资格。证券投资咨询业务是指取得监管部门颁发的相关资格的机构及其咨询人员为证券投资者或客户提供证券投资的相关信息、分析、预测或建议，并直接或间接收取服务费用的活动。证券研究报告是证券投资咨询业务的一种基本形式，指证券公司、证券投资咨询机构对证券及证券相关产品的价值、市场走势或者相关影响因素进行分析，形成证券估值、投资评级等投资分析意见，制作证券研究报告，并向客户发布的行为。

法律声明

本报告由国元证券股份有限公司（以下简称“本公司”）在中华人民共和国境内（台湾、香港、澳门地区除外）发布，仅供本公司的客户使用。本公司不会因接收人收到本报告而视其为客户。若国元证券以外的金融机构或任何第三方机构发送本报告，则由该金融机构或第三方机构独自为此发送行为负责。本报告不构成国元证券向发送本报告的金融机构或第三方机构之客户提供的投资建议，国元证券及其员工亦不为上述金融机构或第三方机构之客户因使用本报告或报告载述的内容引起的直接或连带损失承担任何责任。本报告是基于本公司认为可靠的已公开信息，但本公司不保证该等信息的准确性或完整性。本报告所载的信息、资料、分析工具、意见及推测只提供给客户作参考之用，并非作为或被视为出售或购买证券或其他投资标的的投资建议或要约邀请。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波动。在不同时期，本公司可发出与本报告所载资料、意见及推测不一致的报告。本公司建议客户应考虑本报告的任何意见或建议是否符合其特定状况，以及（若有必要）咨询独立投资顾问。在法律许可的情况下，本公司及其所属关联机构可能会持有本报告中所提到的公司所发行的证券头寸并进行交易，还可能为这些公司提供或争取投资银行业务服务或其他服务，上述交易与服务可能与本报告中的意见与建议存在不一致的决策。

免责条款

本报告是为特定客户和其他专业人士提供的参考资料。文中所有内容均代表个人观点。本公司力求报告内容的准确可靠，但并不对报告内容及所引用资料的准确性和完整性作出任何承诺和保证。本公司不会承担因使用本报告而产生的法律责任。本报告版权归国元证券所有，未经授权不得复印、转发或向特定读者群以外的人士传阅，如需引用或转载本报告，务必与本公司研究所联系并获得许可。网址：www.gyzq.com.cn

国元证券研究所

合肥	上海	北京
地址：安徽省合肥市梅山路 18 号安徽国际金融中心 A 座国元证券	地址：上海市浦东新区民生路 1199 号证大五道口广场 16 楼国元证券	地址：北京市东城区东直门外大街 46 号天恒大厦 A 座 21 层国元证券
邮编：230000	邮编：200135	邮编：100027