仅供内部参考，请勿外传！

传媒

发布时间：2026-07-27

证券研究报告 / 行业深度报告

AI 视频生成技术成熟，内容产出产业升级

---传媒行业深度研究报告

优于大势

上次评级:优于大势

报告摘要：

AI 视频生成是生成式 AI 领域最具挑战性也最具变革性的方向之一。与文本和图像生成不同，视频生成需要同时建模空间信息（每一帧的视觉内容）和时间信息（帧与帧之间的运动关系），还要保证两者的一致性和连贯性。

2025 年，AI 视频生成技术进入了全面成熟阶段，技术路线已经基本收敛到 DiT 架构，模型能力在生成质量、时长、可控性和生成速度等方面都有了显著提升。同时，音视频联合生成、多模态输入、长视频生成等技术也取得了重大突破。如今的视频生成技术已经走向成熟，开始向理解物理规律、实时交互和多模态融合的“世界模型”演进。

在 AI 视频生成领域，中国团队表现出色，已经形成了与国际巨头分庭抗礼的局面，头部阵营形成“大厂巨头+明星初创”的格局。字节即梦 Seedance 2.0、快手可灵 Kling3.0 等先进模型在视频时长和物理真实感方面具有顶尖水平，直接推动了 AI 视频生成技术在影视、广告、游戏等传媒行业的大规模工业化应用。

AI 视频生成技术的应用已经带来传媒多行业的变革。

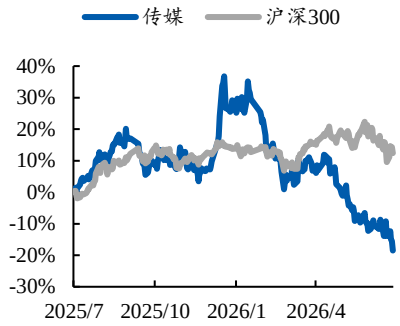
影视方面，AI 漫剧凭借拟真性要求低、与视频生成契合度最高的优势率先实现商业闭环：2025 年抖音端原生漫剧上线超 6 万部、市场播放量超 700 亿次，2026 年市场规模预计达 400 亿元、较 2025 年增长 138%；产业已构建“上游 IP 开放—中游 AI 工具—下游平台分发”的协同生态。行业正从无序扩张转向高质量发展，并升级为 IP 开发的前端入口——通过低成本漫剧验证市场反馈后，优质 IP 向真人剧、游戏、衍生品逐级延伸，实现价值最大化。技术上 AI 仿真人短剧是漫剧第五阶段，其播放量占比从 2026 年 1 月的 5%跃升至 3 月的 40%，已成为 AI 剧增量的绝对主力，预计 2030 年市场规模近 687 亿元，华策影视《绯色之契》等传统影视公司的入局有望在精品化维度建立壁垒。长影视领域 AI 仍以辅助工具为主，动画电影率先落地，华策《太平年》特效场景效率提升 8—10 倍，博纳《三星堆：未来往事》成为首部获“龙标”的 AI 深度参与电影。

广告方面，AI 从单点工具升级为全链路基础设施，素材生成成本曲线下移、个性化上限提升，2030 年 AIGC 广告营销规模有望达 1500 亿元，蓝色光标 AI 驱动收入已增至 37.25 亿元；数字人是商业化最先跑通的场景，直播电商“成本降至真人 1/10、转化率不降反升”的逻辑已获京东、百度大规模验证。

游戏方面，AI 改造沿“美术资产—买量素材—运营服务—玩家体验”四层展开，前两层已有三七互娱等公司可验证的财务证据，网易、巨人网络已在智能 NPC 与 AI 原生玩法上先行探索；AI 原生游戏将沿“降本增效—可感知功能—原生玩法—生成式开放世界”逐级演进，行业整体仍处第一、二阶段之间。

风险提示：监管与内容合规风险；技术迭代与落地不及预期风险；行业竞争加剧；商业化与盈利兑现风险；AI 广告舆情与品牌风险；新业务探索不确定性风险。

历史收益率曲线



涨跌幅（%）	1M	3M	12M
绝对收益	-10%	-24%	-18%
相关收益	-4%	-21%	-31%

行业数据

成分股数量（只）	171
总市值（亿）	8704
流通市值（亿）	7727
市盈率（倍）	123.19
市净率（倍）	2.37
成分股总营收（亿）	
成分股总净利润（亿）	
成分股资产负债率（%）	259.18

相关报告

《AI 真人剧：奇点已至，星辰大海》

--20260526

《IOS 调降佣金率，游戏渠道利润率改善》

--20260316

《高股息防御，AI+IP 打开全新增长曲线》

--20260310

证券分析师：赵明扬
执业证书编号：S0550525070009
zhaomy@nesc.cn

研究助理：刘昱亨
执业证书编号：S0550123100008
liuyh2@nesc.cn

仅供内部参考，请勿外传！

目 录

1. AI 视频生成的原理和技术发展 4
1.1. 所有视频技术都建立在图像生成的突破之上 4
1.2. 从动图到真实世界模型 6
1.2.1. 早期探索阶段：GAN/VAE 主导的“动图时代” 6
1.2.2. 技术突破阶段：Transformer 与扩散模型争锋 9
1.2.3. 能力飞跃阶段：DiT 架构大一统与 Sora 时刻 12
1.2.4. 技术成熟阶段：向真实物理世界演进 14
1.2.5. 视频生成模型的未来：实时性、真实性以及成本控制的提升 16
1.3. 中国模型团队快速崛起，商业化进展迅速 17
2. AI 视频将引发传媒多行业的技术革命 18
2.1. 影视：漫剧&短剧-长影视 18
2.1.1. AI 短剧：从漫剧开始的成本革命 18
2.1.2. AI 仿真人短剧：技术已成熟，正向主流内容形态跃升 20
2.1.3. AI 长影视：以辅助工具为主，动画电影率先落地，真人电影仍处起步阶段 23
2.2. 广告：效率革命与精准营销 24
2.2.1. 素材生成与个性化广告：成本曲线下移与个性化上限提升 25
2.2.2. 数字人：商业化最先跑通的场景 27
2.3. 游戏：自演化 AI 原生游戏的基础 28
2.3.1. 从美术资产生成到 AI 原生游戏 29
2.3.2. AI 原生游戏的自演化：游戏玩法的变革性升级 30
3. 风险提示 31

图表目录

图 1：AI 图像生成模型发展时间线 5
图 2：AI 图像/视频生成模型核心组件及工作流程 6
图 3：GAN 网络的基本结构 7
图 4：扩散网络的加噪、去噪过程 10
图 5：ViT 模型示意图 11
图 6：DiT 网络架构示意图 13
图 7：AI 视频模型的工业应用场景（AI 生图） 15
图 8：2023-2025 年短视频端漫剧播放量 19
图 9：AI 漫剧产业生态 19
图 10：AI 漫剧作为 IP 开发的前端入口 20
图 11：2026 年 Q1 各类型 AI 剧播放量分布 21
图 12：2026 年 3 月各类型在播 AI 剧数量分布 21
图 13：2026 年 Q1 各类型 AI 剧上新数量占比 22
图 14：2026 年 3 月上新 AI 剧各类型占比 22
图 15：2026-2030 年国内 AI 真人短剧市场规模预测 22
图 16：《绯色之契》预告片截图 23
图 17：太平年中利用可灵生成乌鸦食肉糜场景 23
图 18：太平年中利用可灵生成暴风雨海战场景 23
图 19：《三星堆：未来往事》已获得国家电影局"龙标" 24
图 20：2023-2027 年中国互联网广告市场规模变化 24
图 21：宁德时代个人定制 AI 广告 26
图 22：可口可乐 AI 广告 26
图 23：数字人产业生态 28

仅供内部参考，请勿外传！

图 24：AI 对游戏制作的四层改造 29
图 25：不同 AI 价值导向场景企业覆盖数量 30
图 26：《Whispers from the Star》游戏 PV 截图 31

表 1：AI 图像生成技术发展历程 4
表 2：AI 视频生成技术发展历程 5
表 3：AI 视频生成技术各阶段能力对比 16
表 4：AI 视频生成头部厂商 17

仅供内部参考，请勿外传！

1. AI 视频生成的原理和技术发展

AI 视频生成是生成式 AI 领域最具挑战性也最具变革性的方向之一。与文本和图像生成不同，视频生成需要同时建模空间信息（每一帧的视觉内容）和时间信息（帧与帧之间的运动关系），还要保证两者的一致性和连贯性。本章将按照时间顺序，系统梳理 AI 视频生成技术从早期探索到如今工业化应用的完整发展历程，以及每个阶段的核心技术原理、代表性模型、技术效果和商业化应用，并重点阐述了图像生成技术对视频生成的奠基作用和技术传承关系。

本章主要对 AI 视频生成技术做一个全面的介绍，包括技术发展的时间线、各项关键技术的原理、各阶段达成的技术指标，以及 AI 视频生成的产业链分析。我们将在 1.2 节详细介绍 AI 视频生成技术四个发展阶段的演进过程，并尝试用更加通俗易懂的方式介绍其关键技术原理。我们将在 1.3 节对 AI 视频生成技术的各项指标进行对比总结和未来的展望。我们将在 1.4 节对当前开展 AI 视频生成业务的企业进行梳理，并对其产品进行对比。

1.1. 所有视频技术都建立在图像生成的突破之上

AI 视频生成本质是**“连续图像生成+时序一致性建模”**，每一次视频技术的重大革命都源于图像生成技术的先验突破。视频生成技术的核心组件多数都来自图像生成的技术积累，两者的发展呈现出明显的“图像先行，视频跟进”的规律，视频生成技术通常滞后于图像生成技术约 1-2 年。

图像生成技术经历了四个清晰的发展阶段，每个阶段都为后续视频生成技术的突破奠定了坚实基础：

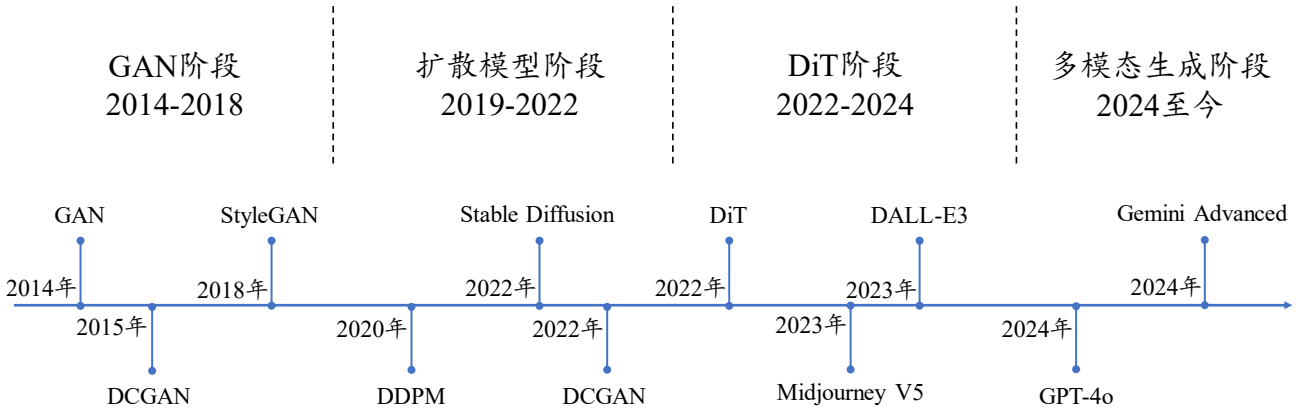
表 1：AI 图像生成技术发展历程

时间区间	发展阶段	核心技术路线	里程碑模型	对视频生成的奠基意义
2014-2018	GAN	生成对抗网络	GAN(2014)、DCGAN(2015)、StyleGAN(2018)	证明了 AI 生成逼真图像的可能性，奠定了生成式 AI 的基础
2019-2022	扩散模型	扩散模型+U-Net	DDPM(2020)、Stable Diffusion(2022)、Midjourney V3(2022)	解决了 GAN 训练不稳定和模式崩溃问题，为视频扩散模型铺平道路
2022-2024	DiT	扩散 Transformer	DiT(2022)、Midjourney V5(2023)、DALL-E 3(2023)	证明了 Transformer 在生成领域的可扩展性，直接催生了 Sora 等视频模型
2024 至今	多模态生成	统一多模态 DiT	GPT-4o(2024)、Gemini Advanced(2024)	实现了文本、图像、音频的统一建模，推动音视频联合生成

数据来源：东北证券

仅供内部参考，请勿外传！

图 1：AI 图像生成模型发展时间线



数据来源：东北证券

图像生成技术为视频生成提供了几乎所有的核心技术组件和方法论：

- **骨干网络架构：**从 2D 卷积到 3D 卷积，从时空 U-Net 到扩散 Transformer，所有骨干网络都先在图像领域验证成功后才被扩展到视频领域。VAE 编码器/解码器更是直接复用图像生成的成熟组件，大幅降低了视频模型的开发难度。
- **训练方法与技巧：**对抗训练、扩散训练、自回归训练等核心训练范式都源于图像生成。分类器自由引导（CFG）、无分类器引导等提升生成质量的关键技术也都是先在图像生成上成熟，再直接迁移到视频领域。
- **数据利用方式：**联合图像-视频训练已成为行业标准，利用海量图像数据弥补视频数据的不足。图像预训练权重作为视频模型的初始化参数，能够大幅降低训练成本和时间。
- **推理优化技术：**蒸馏、量化、注意力优化等加速技术都先在图像生成上得到验证，再应用到视频生成。目前视频生成广泛使用的蒸馏技术，直接继承自图像生成的 LCM 蒸馏技术。

AI 视频生成技术的发展与图像生成技术紧密相关，同样形成了比较清晰的四个发展阶段：

表 2：AI 视频生成技术发展历程

时间	发展阶段	核心技术路线	视频里程碑事件	对应图像生成里程碑	行业意义
2016-2020	早期萌芽探索阶段	GAN/VAE 主导	VGAN(2016) DeepFake(2017)	GAN(2014) StyleGAN(2018)	证明 AI 生成视频的可能性
2020-2023	技术突破阶段	扩散模型+Transformer	Runway Gen(2023) Pika 1.0(2024)	ViT(2020) Stable Diffusion(2022)	文生视频从实验室走向商业化
2023-2024	架构统一能力飞跃阶段	DiT：扩散 Transformer	Sora(2024) Kling1.0(2024)	DiT(2022) Midjourney V5(2023)	确立行业标准，生成能力质变
2025至今	真实世界技术成熟阶段	多模态融合	Sora2(2025) Kling3.0(2026)	GPT-4o(2024) Midjourney V7(2025)	进入工业化大规模应用阶段

数据来源：东北证券

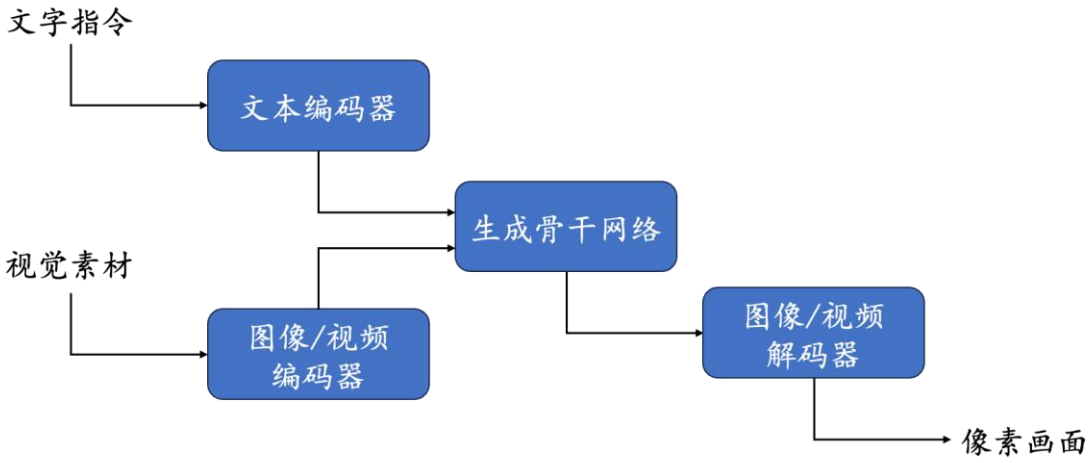
仅供内部参考，请勿外传！

一个完整的 AI 视频生成模型，是多个模块协同工作的结果，并依赖一套科学的体系来评估其优劣。

技术架构组件：

- 文本编码器：理解你的文字指令，常用 CLIP 或 T5 模型。
- 图像/视频编码器：将视觉数据压缩到高效的潜空间（Latent Space） 表示。
- 生成骨干网络：这是模型的核心，负责在潜空间中进行“创作”，主流架构是 DiT（Diffusion Transformer）。
- 图像/视频解码器：将潜空间中的表示解码，还原成我们肉眼可见的像素画面。

图 2：AI 图像/视频生成模型核心组件及工作流程



数据来源：东北证券

评估指标：

- 自动化指标：FVD 用于量化生成视频与真实视频在分布上的距离；CLIPSIM 评估视频与文本描述的相关性。新指标如 World Consistency Score（WCS）则开始关注视频在物理世界中的时空一致性。
- 人工评估：自动化指标有时无法全面反映人的主观感受。因此，通过人工对画质、运动自然度和图文一致性进行打分（MOS），是评估模型效果的金标准。

1.2. 从动图到真实世界模型

本节我们将根据时间线，对 AI 视频生成技术的各发展阶段进行介绍，并尝试用通俗易懂的方式介绍其关键技术的原理及效果。

1.2.1. 早期探索阶段：GAN/VAE 主导的“动图时代”

这一阶段的核心特征是将在图像生成领域取得成功的生成对抗网络（GAN）和变分自编码器（VAE）扩展到视频领域。研究人员主要通过“逐帧生成+时序建模”的思路来解决视频生成问题。

- GAN：通过一个生成器和一个判别器相互博弈来提升生成质量。代表模型有 TGAN 和 MoCoGAN 等。
- VAE：通过学习数据的潜在空间表示来生成新的数据。

仅供内部参考，请勿外传！

由于计算资源和模型架构的限制，这一阶段生成的视频分辨率极低、时长很短（通常只有几秒），并且运动模式单一，存在严重的模糊和时间不一致性问题。这个时期由于算力和模型架构实际上远未达到视频生成的需要，仅能作为视频生成可能性的早期探索阶段，相关的技术更多起到论证 AI 视频可行性的作用。

技术效果：

- 能够生成 3-5 秒的低分辨率视频（通常为 64×64 或 128×128 像素）。
- 可以生成简单场景下的单一物体运动，如飘动的旗帜、旋转的物体。
- 部分模型能够生成人脸表情变化等简单动态。

主要局限性：

- 继承了 GAN 的所有缺陷：训练极不稳定，模式崩溃问题在视频领域更加严重。
- 时序一致性差，容易出现物体突然消失、变形或运动不连贯的情况。
- 无法处理复杂场景和长视频生成。
- 计算成本极高，即使生成低分辨率视频也需要大量 GPU 资源。

代表性产品：此阶段主要为学术探索，通过如 Text2Filter、TGANs-C 等论文展示了概念的可行性，尚未形成商业产品。唯一有影响力的应用是 DeepFake 换脸技术（2017 年出现），它基于 GAN 架构，能够将一个人的脸替换到另一个人的视频中，虽然引发了巨大的伦理争议，但也证明了 AI 在视频生成方面的潜力。

1.2.1.1. 生成对抗网络-GAN

GAN 由 Ian Goodfellow 于 2014 年提出，其核心思想源于博弈论中的零和博弈。它由两个相互竞争的神经网络组成：生成器(Generator)和判别器(Discriminator)。生成器(G)和判别器(D)的基本规则如下：

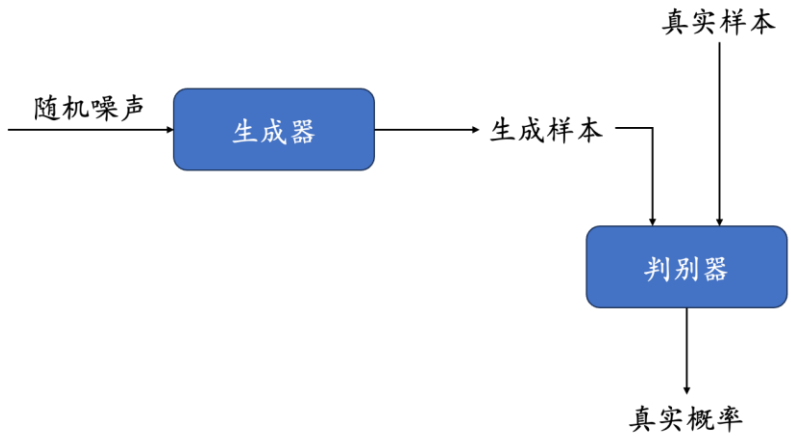
生成器(G)：

- 输入：从简单分布（如高斯分布）中采样的随机噪声向量 z。
- 输出：模拟真实数据分布的假样本 G(z)（如图像、视频）。
- 目标：尽可能生成逼真的样本，“欺骗”判别器。

判别器(D)：

- 输入：真实样本 x 或生成样本 G(z)。
- 输出：0 到 1 之间的概率值，表示输入是真实样本的概率。
- 目标：准确区分真实样本和生成样本。

图 3：GAN 网络的基本结构



数据来源：《Generative Adversarial Nets》、东北证券

仅供内部参考，请勿外传！

GAN 的训练过程是生成器和判别器之间的动态对抗过程。在每一次训练迭代中，分别对两者进行训练：一方面训练生成器(G)的样本生成能力，使其能够生成更逼真的样本；同时训练判别器(D)的识别能力，使其能准确区分真实样本和生成器生成的样本。具体地训练过程如下：

1） 训练判别器时，固定生成器参数。分别输入真实样本和生成样本，通过反向传播更新判别器参数，使判别器对真实样本的输出更接近 1，对生成样本的输出更接近 0。
2） 训练生成器时，固定判别器参数。输入随机噪声，生成假样本，通过反向传播更新生成器参数，让判别器对假样本的输出接近 1。

GAN 模型最根本的目标还是“生成”，“对抗”只是用于帮助训练“生成”能力的手段。因此，上述的两个步骤应当不断重复，直到生成器生成的样本足够逼真，判别器无法准确区分（准确率约为 50%）。

GAN 模型相较于更早的生成式模型，具有以下优点：1）生成样本质量高，细节丰富；2）无需显式建模数据分布，通过对抗学习隐式学习；3）生成速度快（单步生成）；4）应用场景广泛，可用于图像、视频、音频等多种模态。

“生成+对抗”的创新性思想给后续的生成式模型，尤其是无监督学习模型和自监督学习模型提供了非常关键的灵感。但 GAN 仍具有 1）训练不稳定：生成器和判别器的平衡难以维持；2）模式崩溃：生成器只生成少数几种样本，多样性不足；3）难以评估：没有统一的评估指标；4）可解释性差：黑盒模型，难以理解生成过程等缺点。因此 GAN 模型本身的生成能力并不能达到显著可用的程度。

1.2.1.2. GAN 在视频生成领域的应用

视频生成与图像生成的本质区别在于增加了时间维度，需要同时建模空间信息和时间信息，并保证帧与帧之间的时序一致性。将 GAN 应用到视频生成面临的最大挑战就是如何有效建模时序依赖关系，为此各国研究者进行了多种尝试。

VedioGAN(2016)是首个基于 GAN 的视频生成模型，其核心思想是直接将图像 GAN 扩展到视频领域，使用 3D 卷积同时处理空间和时间维度。VGAN 的视频生成效果并不理想，3D 卷积计算量过大，难以生成高分辨率视频，生成内容的时序一致性较差，物体运动不连贯，且只能生成简单场景下的单一物体运动。

TGAN（2017）在 VGAN 的基础上将视频生成分解为时间生成和图像生成两个独立的过程，使用两个生成器分别处理。时间生成器（G0）将输入的噪声生成为时间步的特征值，再由图像生成器（G1）将时间步特征转化为对应的视频帧。判别器（D）依然使用 3D 卷积，对整个视频序列进行判别。TGAN 明确分离了时间和空间建模，提高了时序一致性，生成的视频比 VGAN 更加流畅。

MoCoGAN（2017）将视频分解为静态内容和动态运动两个独立的部分分别建模。潜在空间被划分为内容子空间和运动子空间，内容向量在整个视频序列中保持不变，决定了视频的静态内容（如物体、背景），运动向量随时间变化，决定了视频的动态变化（如物体运动、相机移动），使用循环神经网络（RNN）建模运动向量的时序依赖关系。MoCoGAN 能够实现更精细的控制，固定内容改变运动或固定运动改变内容，生成的视频时序一致性也更好。

DVD-GAN（2018）使用两个独立的判别器分别判别空间一致性和时间一致性，降低计算复杂度，提高生成质量。空间判别器对单个视频帧进行判别，保证每一帧的图像质量；时间判别器对视频帧序列进行判别，保证帧与帧之间的运动连贯性；生成

仅供内部参考，请勿外传！

器使用多尺度架构，先生成低分辨率视频，再逐步上采样到高分辨率。DVD-GAN 通过双判别器架构提高计算效率的同时，对空间质量和时间连贯性进行了优化，能够生成分辨率更高（256×256）的视频。

尽管 GANs 在视频生成方面取得了一定进展，但模型架构本身仍然存在严重的局限性：1）训练极不稳定：GAN 本身的训练不稳定问题在视频领域更加严重；2）模式崩溃：生成的视频往往缺乏多样性，只能生成少数几种场景和运动；3）长视频生成困难：随着视频时长增加，时序一致性迅速下降；4）物理逻辑错误：经常出现违反物理规律的现象，如物体漂浮、穿模等；5）可扩展性差：难以通过增加参数量和数据量持续提升效果。

1.2.1.3. 变分自编码器-VAE

VAE 由 Diederik P. Kingma 和 Max Welling 于 2013 年提出，是一种基于变分推断的生成式自编码器。它在传统自编码器的基础上引入了概率建模思想，使其具备生成全新样本的能力。传统自编码器将输入数据编码为潜在空间中的一个固定点，然后解码重建输入，这种确定性的映射导致无法通过随机采样生成全新样本。

VAE 将编码器从确定性映射改为概率分布映射，通过编码器（E）将高维数据压缩到低维潜在空间，通过解码器（D）将高斯分布中采样的潜在变量重建为数据样本，即将低维潜在变量映射回高维数据空间，采样后重建样本与原始样本之间的差异为模型训练损失。VAE 模型通过学习潜在变量遵循的概率分布，基于该分布进行采样生成新的类似样本。

于 GAN 相比，VAE 的优点在于：1）训练稳定：基于似然的训练，没有 GAN 的对抗平衡问题；2）有明确的潜在空间：潜在空间是连续且结构化的，便于进行插值和操作；3）可解释性强：可以通过分析潜在变量理解数据的生成因素 4）数学基础扎实：基于变分推断，有严格的理论支撑。

与 GAN 类似，将 VAE 应用到视频生成也需要解决时序建模的问题。各国研究者提出了时空联合压缩、结合 RNN 建模时序依赖、引入向量量化技术等方法解决 VAE 模型的时序建模问题，但仍具有较大的局限性：

- 生成样本模糊：传统 VAE 生成的视频往往缺乏细节，不够清晰
- 长程依赖建模能力弱：难以生成超过 10 秒的连贯视频
- 计算复杂度高：3D 卷积和 Transformer 的计算量都很大
- 生成速度慢：自回归生成需要逐帧预测，速度较慢

尽管 GAN 和 VAE 在早期视频生成中取得了一定进展，但它们的固有缺陷限制了进一步发展。2020 年之后，扩散模型逐渐取代 GAN 和 VAE 成为生成式 AI 的主流技术路线。

1.2.2. 技术突破阶段：Transformer 与扩散模型争锋

2020 年是 AI 生成领域的转折点，扩散模型（Diffusion Model）在图像生成任务上超越了 GAN，展现出更好的训练稳定性和生成质量。同时，Transformer 架构在自然语言处理领域的巨大成功也启发了研究人员将其应用于图像和视频生成。2020-2021 年，研究的重点在图像生成上，为视频模型奠定了关键基础。2022-2023 年，图像生成的成功经验被迅速迁移到视频领域，各大公司纷纷推出自己的视频生成模型。这一阶段的视频生成模型形成了两条技术路线：

- **扩散模型路线：**包括 Meta 的 Make-A-Video、英伟达的 Video LDM、Runway 的 Gen-1/Gen-2 等，通过在图像扩散模型中加入时间层来建模帧间关系。

仅供内部参考，请勿外传！

- **自回归（Transformer）路线：**以谷歌的 Phenaki 为代表，采用类似于语言模型的方式，根据之前的帧和文本提示预测下一帧。

技术效果：
- 生成质量显著提升，视频分辨率从 128×128 提升到 1280×768。
- 时序一致性大幅改善，物体运动更加连贯，减少了突然变形和消失的问题。
- 文本理解能力增强，能够更好地理解复杂的文本描述，生成符合语义的视频。
- 生成时长从 3-5 秒提升到 10-15 秒。

主要局限性：
- 物理逻辑错误频发，经常出现违反物理规律的现象，如物体漂浮、穿模等。
- 人物生成效果差，人脸容易扭曲，手部和手指问题尤为严重。
- 生成速度慢，生成一段 10 秒视频需要几分钟甚至几十分钟。
- 长视频生成困难，超过 15 秒后，视频内容容易变得混乱或重复。

代表性产品与应用：这一阶段开始出现一些早期的商业化尝试：
- Runway Gen-1（2023 年 2 月）：基于视频到视频的生成模式，可以将现有视频转换为不同风格。
- Runway Gen-2（2023 年 3 月）：首个面向公众的文生视频产品，支持生成 4 秒 720p 视频。
- Pika 1.0（2023 年 4 月）：以 Discord 机器人形式提供服务，生成速度快，风格多样。

1.2.2.1. 扩散模型-Diffusion Model

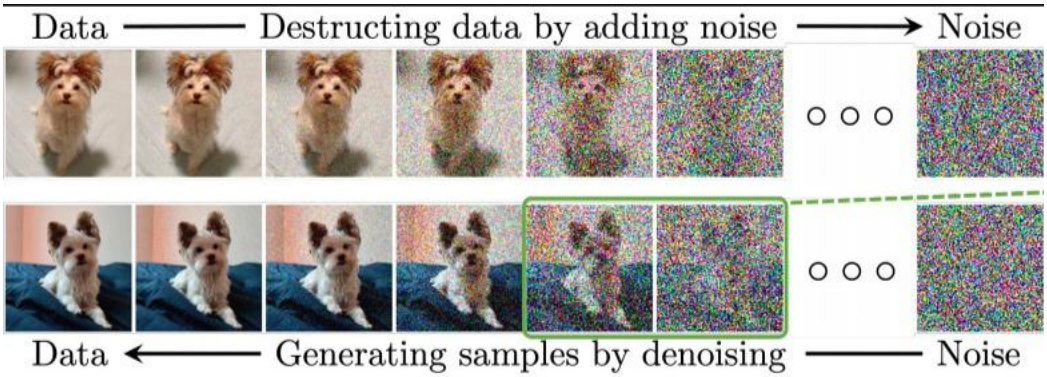
扩散模型是一种非对抗式的生成式 AI 范式，核心思想是“逐步加噪、逆向去噪”，彻底解决了 GAN 训练不稳定、模式崩溃的痛点。

扩散模型的生成过程分为两个对称的物理过程：

- **前向扩散过程：**向真实数据中逐步、微小地添加高斯噪声，经过上千步后，原始数据完全被噪声淹没，变成符合标准正态分布的纯噪声。该前向过程是一个固定的马尔可夫链，无需训练，完全由数学公式定义，数学上可以直接推导任意步数 t 的带噪数据，无需逐步迭代，大幅提升了训练效率。

- **逆向去噪过程：**训练一个神经网络（通常称为“去噪网络”），学习从纯噪声中逐步预测并去除噪声，最终恢复出清晰的真实数据。生成时只需输入随机噪声，经过逐步去噪即可得到全新样本。

图 4：扩散网络的加噪、去噪过程



数据来源：CSDN、东北证券

仅供内部参考，请勿外传！

理论上，由于前向扩散即加噪声的过程是固定的，加噪声的逆操作也应是固定的。理想情况下，我们希望去噪操作就等于加噪声逆操作。然而，加噪声的逆操作不太可能从理论上求得，我们只能用一个神经网络去拟合它，扩散模型的训练就是将这个逆向去噪过程神经网络拟合至加噪声的逆操作的过程。

和 GAN 的“一步生成、对抗博弈”、VAE 的“编码重构、概率建模”相比，扩散模型的训练目标更简单、过程更稳定，生成质量与多样性的均衡性最优。纯扩散模型只能无条件生成随机样本，为了实现“文本生成视频”等可控生成，需要引入条件引导：交叉注意力机制：将文本编码后的特征向量，通过交叉注意力注入去噪网络的每一层，让生成过程贴合文本语义；无分类器引导（CFG）：同时训练有条件和无条件两个分支，推理时放大两者的差异，显著提升生成质量与文本对齐度。除文本外，参考图像、参考视频、深度图、骨骼动作等都可以作为条件注入。

对比 GAN/VAE，扩散模型的核心优势在于：

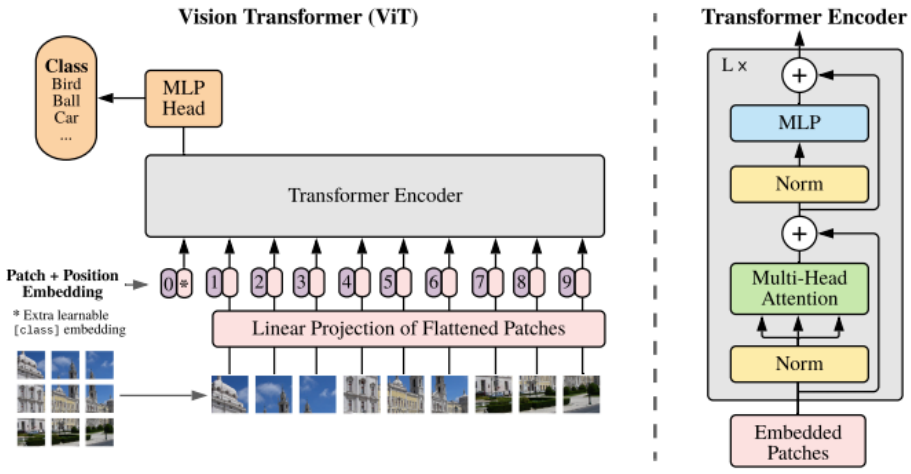
- 训练极度稳定：没有对抗博弈，损失函数单调下降，不存在模式崩溃问题；
- 生成质量与多样性均衡：既可以生成高度逼真的细节，又能保证样本丰富多样；
- 可扩展性强：模型参数量、数据量越大，生成效果持续提升，符合大模型缩放定律；
- 可控性好：支持多种条件灵活引导，便于实现可控生成。

1.2.2.2. 视觉 Transformer-ViT

ViT 是 Transformer 架构从自然语言处理向计算机视觉领域的迁移，它打破了 CNN（卷积神经网络）对视觉任务的垄断，凭借全局建模能力和强可扩展性，成为当前大模型时代的视觉主干网络。CNN 的核心是通过卷积核在图像上滑动，提取局部特征；而 ViT 彻底抛弃了卷积归纳偏置，将图像转化为和文本一样的一维 token 序列，直接用标准 Transformer 处理，即**图像序列化。**

ViT 将一张图像切分为若干个固定大小的小块（Patch，通常为 16×16 像素），每个小块通过线性投影，转化为一个特征向量（即 Transformer 中的一个 token），并为每个 token 添加位置编码，保留原始的空间位置信息，最后将 token 序列输入多层 Transformer 编码器，完成全局特征建模。

图 5：ViT 模型示意图



数据来源：《AN IMAGE IS WORTH 16X16 WORDS: TRANSFORMERS FOR IMAGE RECOGNITION AT SCALE》、东北证券

仅供内部参考，请勿外传！

ViT 的核心思想在于将图像输入转换为一维特征向量后，用 Transformer 架构进行处理，因此 ViT 模型的核心组件以及优势均源于 Transformer 架构：

- **补丁嵌入（Patch Embedding）**：完成“图像→序列”的转换，将二维像素块映射为一维特征向量，是 ViT 的输入接口。
- **位置编码（Positional Embedding）**：Transformer 本身不具备位置感知能力，需要手动为每个 token 添加位置信息，让模型知道每个小块在原图中的空间坐标。
- **多头自注意力（MSA）**：ViT 的核心，让每个 token 都能和图像中所有其他 token 计算关联度，实现全局感受野。相比 CNN 的局部感受野，它天然擅长捕捉长距离的依赖关系（比如人物全身的比例、物体间的空间关联）。
- **前馈网络（FFN）+层归一化+残差连接**：标准 Transformer 块，通过多层堆叠提升模型表达能力。

ViT 最大的意义在于实现了视觉任务的“大模型化”：它和 NLP 大模型共享完全一致的架构范式，能够通过堆参数量、堆数据持续提升能力，为后续多模态统一建模、视频大模型奠定了架构基础。

1.2.3. 能力飞跃阶段：DiT 架构大一统与 Sora 时刻

早期扩散模型的去噪网络普遍采用 CNN 架构的 U-Net，但 U-Net 的长程依赖短板和可扩展性瓶颈，逐渐成为生成质量进一步提升的瓶颈。2022 年 DiT（Diffusion Transformer）提出，用 ViT 替换了 U-Net 主干，证明 Transformer 在生成领域优于 U-Net，开启了生成式视觉的“大模型时代”。DiT 模型的核心贡献是首次证明纯 Transformer 架构可以完全替代传统扩散模型中的卷积 U-Net 主干，且具备极佳的可扩展性——**模型参数量与计算量越大，生成质量持续提升，严格遵循大模型缩放定律**。这一发现彻底打破了 U-Net 的性能天花板，为后续视频大模型、多模态生成模型奠定了核心架构基础。2024 年 2 月 OpenAI 发布的 Sora 模型是 AI 视频生成领域的里程碑事件，模型能力实现了质的飞跃。Sora 彻底改变了人们对 AI 视频生成能力的认知，其采用的 DiT 架构也成为行业的事实标准，引领了后续所有主流视频模型的发展方向。

技术效果：
- 生成时长突破，从 15 秒提升到 60 秒，部分场景甚至更长。
- 物理真实感大幅提升，能够准确模拟重力、碰撞、流体等常见物理现象。
- 3D 一致性增强，支持复杂的相机运动，物体在 3D 空间中保持一致性。
- 多镜头叙事能力，能够在单个视频中实现多个镜头的自然切换。
- 复杂场景生成能力，可以生成包含多个物体、复杂交互和丰富背景的场景。

主要局限性：
- 精细结构仍有错误，手部、手指等精细结构仍然容易出错。
- 长视频逻辑一致性不足，超过 1 分钟后内容容易出现重复或逻辑混乱。
- 生成成本极高，生成一段 60 秒视频需要大量 GPU 资源和时间。
- 可控性不足，用户对生成过程的精细控制能力有限。

代表性产品与应用：这一阶段 AI 视频生成产品开始大规模商业化：
- Runway Gen-3（2024 年 6 月）：专业级视频生成平台，引入了更精细的时序控制，提供丰富的编辑工具。
- Pika 1.5（2024 年 10 月）：以创意风格和快速生成为特色，深受创作者喜爱。

仅供内部参考，请勿外传！

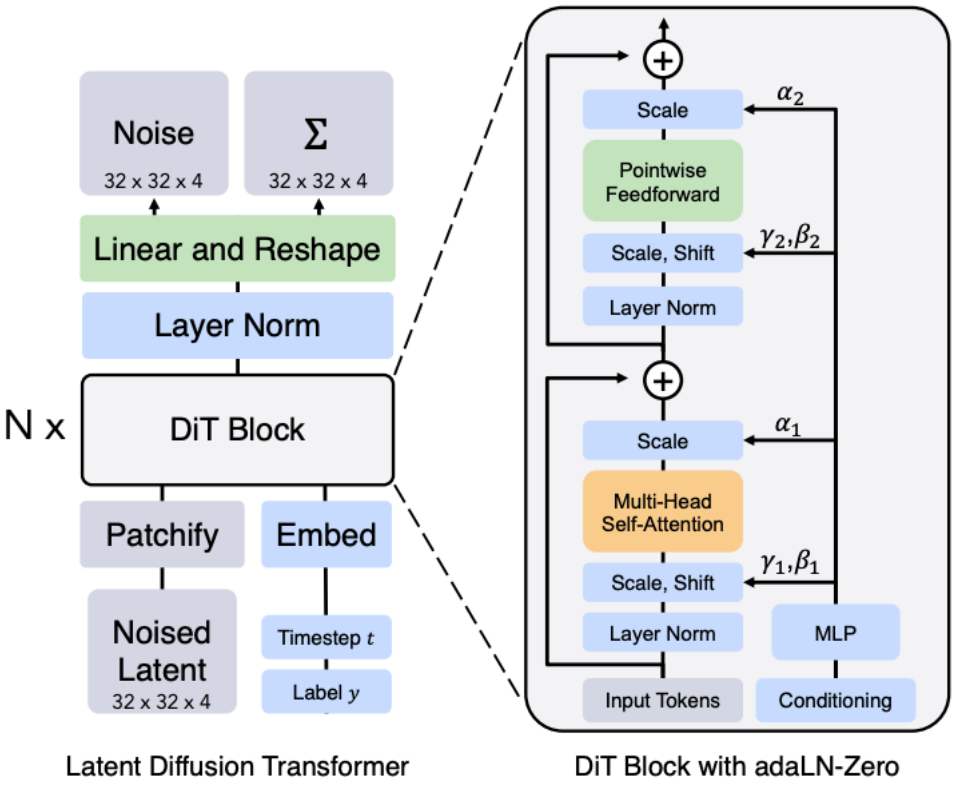
- 快手可灵（2024 年 6 月）：国内用户量最大的文生视频平台之一，以长视频生成能力著称。
- 字节即梦（2024 年 5 月）：与剪映深度集成，提供一站式视频创作体验。
- MiniMax 海螺（2024 年 8 月）：支持音视频联合生成。

1.2.3.1. 扩散 Transformer-DiT

DiT 模型的整体流程和传统扩散模型一致，仅将去噪网络的主干从 U-Net 替换成了 Transformer，具体变化如下：

- 输入处理：带噪的潜在表示经过补丁嵌入，转化为 token 序列；
- 条件注入：扩散时间步、文本条件等信息，通过自适应层归一化（adaLN-Zero）或交叉注意力注入到每个 Transformer 块中；
- 主干网络：多层 Transformer 块堆叠，通过自注意力完成时空全局建模；
- 输出头：将 token 序列还原为空间形状，输出预测的噪声。

图 6：DiT 网络架构示意图



数据来源：《Scalable Diffusion Models with Transformers》、东北证券

DiT 架构的核心是标准的 ViT 风格 Transformer 块（多头自注意力+前馈网络+层归一化+残差连接），最大的设计差异在于条件注入方式——扩散模型需要将时间步、文本/类别条件融入到每一层网络中。论文作者对比了四种条件注入方案，最终 adaLN-Zero（零初始化自适应层归一化）效果最优，即上图所示的 Transformer 块。具体的条件注入过程如下：

> 先将时间步 t 和条件 c 拼接后投影，生成缩放（scale）和偏移（shift）参数。将参数注入到每个 Transformer 块的层归一化中，控制特征的分布。采用零初始化策略：训练初始时，条件注入的权重为 0，Transformer 块等价于恒等映射，大幅提升训练稳定性。

仅供内部参考，请勿外传！

DiT 架构的提出对图像/视频生成及多模态大模型的发展具有重要价值。

- **范式级架构替换：**首次证明纯 Transformer 可以完全替代 U-Net 作为扩散模型主干，且在大参数量下效果全面超越卷积架构，将生成式视觉正式带入大模型时代。
- **验证扩散缩放定律：**证明扩散模型同样遵循“参数量-数据量-性能”的缩放规律，为后续模型迭代指明了明确的技术路线。
- **统一多模态架构底座：**Transformer 是 NLP、语音、视觉通用的架构，DiT 的出现让生成式视觉与大语言模型可以复用同一套技术栈，极大降低了多模态融合的成本。
- **工程化友好：**DiT 架构简单统一，只有一种标准 Transformer 块，便于硬件优化、分布式训练和推理加速，工业落地效率远高于复杂的 U-Net。

1.2.4. 技术成熟阶段：向真实物理世界演进

2024 年 GPT-4o 发布，实现多模态统一建模；图像生成进入实时生成时代。2025 年，AI 视频生成技术进入了全面成熟阶段，技术路线已经基本收敛到 DiT 架构，模型能力在生成质量、时长、可控性和生成速度等方面都有了显著提升。同时，音视频联合生成、多模态输入、长视频生成等技术也取得了重大突破。

如今的视频生成技术已经走向成熟，开始向理解物理规律、实时交互和多模态融合的“世界模型”演进。最新的视频生成技术在视频长度、可控性、效率、音视频联合生成等方向的能力都获得了极大的提升：

- **视频长度：**
  通过滑动窗口注意力和分块生成技术，最长生成时长从 1 分钟提升到 5 分钟以上。多镜头叙事能力增强，能够生成有完整故事情节的短片。
- **可控性：**
  为了满足专业创作需求，模型的可控性得到了极大提升，包括运动控制（手绘运动轨迹、关键帧控制、相机运动控制）、角色控制（上传参考图像保持角色一致性）、风格控制和局部编辑。
- **效率优化：**
  通过蒸馏技术将原本需要几十步的扩散过程蒸馏到 8 步甚至更少，结合稀疏注意力和量化技术，大幅提升了生成速度，降低了硬件要求。
- **音视频联合生成：**
  新一代模型普遍支持原生音频生成，能够生成与视频内容同步的对话、音效和环境音。部分模型还支持空间音频，声音会随着物体的位置和运动而变化。

仅供内部参考，请勿外传！

目前主流的 AI 视频模型已经达到了以下技术水平：

- **分辨率：**主流模型支持 1080p，高端模型支持 4K。
- **时长：**普通模型支持 15-30 秒，高端模型支持 1-5 分钟。
- **生成速度：**实时生成成为可能，部分模型达到 10fps 以上。
- **可用率：**专业模型的生成可用率超过 90%，大幅降低了创作成本。
- **物理真实感：**大多数常见物理现象都能准确模拟。
- **人物生成：**人脸和手部生成质量大幅提升，基本告别“恐怖谷”效应。

基于以上能力水平，AI 视频模型已经能够在影视、广告、游戏等行业的多种场景下实现工业化应用，本报告后续章节会对具体应用场景进行分析。报告内难以直接展示模型的视频生成结果，故使用图片生成结果简要展示清晰度及真实程度。图 7 为豆包 Seedream 4.5 图像生成模型的生成结果。

图 7：AI 视频模型的工业应用场景（AI 生图）



数据来源：豆包 Seedream-4.5、东北证券

仅供内部参考，请勿外传！

1.2.5. 视频生成模型的未来：实时性、真实性以及成本控制的提升

表 3：AI 视频生成技术各阶段能力对比

发展阶段	分辨率	视频时长	动作连贯性与物理质量	核心缺陷
早期萌芽探索阶段	极低（约 64×64）	3 秒以内	运动模式单一，严重模糊，时间一致性极差	画面几乎无可辨识的主体，无实用价值
技术突破阶段	低（约 256×256）	5-10 秒	画质明显提升，但主体一致性差，运动常违反物理规律，画面易“畸变”、“抽搐”	画质较粗糙，角色外貌易突变，无法维持逻辑性长镜头
架构统一能力飞跃阶段	480P-1080P	约 60 秒	开始具备初步物理规律模拟能力(光影、遮挡)，运动流畅度大幅提高	复杂交互和长视频的逻辑一致性仍有破绽
真实世界技术成熟阶段	1080P-4K/8K	分钟级以上	运动与物理一致性显著增强，可同步生成原生音频，向理解三维世界演进	长程交互、记忆与可控性仍在探索，计算成本极高

数据整理：东北证券

虽然当下的 AI 视频生成模型已经在多项技术指标上均已得到了长足的进步，在视频细节和生成成本上仍有提升空间。同时，生成内容的版权以及潜在的安全问题也需要业界探索出更完善的解决方案。

- **长视频叙事能力：**5 分钟以上的长视频生成仍然存在内容重复和逻辑混乱的问题，距离完整的视频/影片生成仍有较大差距
- **精细可控性：**满足专业创作者对镜头语言、光影效果、角色表演等方面的精确控制需求仍然是一个挑战
- **计算成本：**训练和推理仍然需要大量 GPU 资源，限制了大规模应用和普及
- **版权与伦理问题：**训练数据的版权归属和深度伪造技术的滥用风险亟待解决，需要建立完善的法律法规和行业规范

伴随着技术能力与产业环境的逐渐成熟，AI 视频生成将从“单点内容生成工具”向“全链路创作系统”升级，延伸出多个具备长期价值的核心发展方向。

- **全流程自动化：**从剧本创作、分镜设计到成片输出的端到端 AI 化，实现“输入剧本，输出电影”甚至“输入灵感，输出电影”的终极创作模式。
- **实时交互式生成：**支持游戏、VR/AR 等实时应用场景，用户可以实时与生成的内容进行交互。
- **3D 立体视频生成：**直接生成双目立体视频和 VR 内容，为元宇宙提供内容支撑。
- **通用世界模型：**视频生成模型将进化为能够理解和模拟真实世界的通用 AI，具备推理、规划和决策能力，为机器人等物理 AI 应用场景提供智能支持。

仅供内部参考，请勿外传！

1.3. 中国模型团队快速崛起，商业化进展迅速

在 AI 视频生成领域，中国团队表现出色，已经形成了与国际巨头分庭抗礼的局面，头部阵营形成“大厂巨头+明星初创”的格局。截至 2026 年 6 月，AI 视频生成领域的国际头部玩家仅剩 Google 一家，OpenAI 已逐步关停 Sora 的 C 端服务，基本退出主流消费市场竞争，Runway 自 2025 年 12 月发布 Gen-4.5 模型后已半年未更新，疑似逐渐掉队。与之相反的是，中国团队在 AI 视频生成领域竞争十分激烈，模型保持较高的更新频率，模型的性能也不断提高。综合 Artificial Analysis、LM Arena 以及 SuperCLUE 各自最新的三份榜单，以下厂商处于 AI 视频生成领域第一梯队：

表 4：AI 视频生成头部厂商

厂商	国别	最新模型	发布时间	最新模型性能排名
Google	美国	gemini-omni-flash	2026.5	1
字节即梦	中国	dreamina-seedance-2.0	2026.3	2
快手可灵	中国	Kling-3.0	2026.2	3
阿里巴巴	中国	happyhorse-1.0	2026.4	3
生数科技	中国	Vidu Q3	2026.1	5
爱诗科技	中国	PixVerse V6	2026.3	5

数据来源：Artificial Analysis、LM Arena、SuperCLUE、东北证券

其中，Google 于 2026 年 5 月 19 日发布的 Gemini Omni Flash 作为 Gemini Omni 系列的首款模型，将此前独立的视频（Veo）、图像（Nano Banana）、音乐（Lyria）及交互（Genie）生成与理解能力融合进单一模型框架，首次使模型同时具备了理解世界和生成世界的能力。Gemini Omni Flash 发布后强势登顶视频生成榜单，当前版本专注于视频生成与编辑能力，根据官方路线图，未来该系列模型将逐步扩展，支持图像、音频等更多输出形式。

国内厂商中，字节即梦综合能力最强，Seedance 2.0 自发布后各视频生成榜单均为帮手，且具有生态闭环的优势。即梦依托剪映的庞大用户基础，实现了“生成-编辑-发布”的全链路无缝衔接，是国内工业化应用程度最高的模型。快手可灵在视频时长和物理真实感方面具有顶尖水平，且国际化推进速度较快，海外用户占比持续提升。阿里巴巴 2026 年 4 月凭借 HappyHorse 模型杀入第一梯队，此前的开源模型通义万相 Wan 系列模型深度绑定电商生态，在电商场景具有深度积累。

爱诗科技和生数科技两家均为 AI 视频生成领域的初创企业。生数科技是技术派创业公司代表，在生成速度与成本上具有优势，其原创 U-ViT 架构带来了极致的生成效率。爱诗科技则为工程化产品驱动路线的创业公司代表，优势在创意短视频及社交领域，C 端用户增长快。

除以上第一梯队厂商以外，国内多数大模型厂商均在视频生成领域有所布局，如 MiniMax 海螺系列，昆仑万维 SkyReels，智谱 CogVideo、腾讯混元、阶跃星辰等。

仅供内部参考，请勿外传！

2. AI 视频将引发传媒多行业的技术革命

2.1. 影视：漫剧&短剧-长影视

在影视行业，动漫剧由于本身拟真性要求相对较低，与视频生成契合度最高，已率先实现商业闭环。据 DataEye 研究院《2025 年漫剧数据报告》，2025 年仅抖音端原生漫剧上线数即突破 6 万部，市场整体播放量超 700 亿次。2025 年以前漫剧市场的作品供给量不足千部，AI 视频技术的大规模应用为漫剧市场提供了 60 倍的供给量增长。

短剧方面，AI 生成的拟真短剧内容接受度在快速提升，2026 年以来已经逐步进入量产阶段。但由于短剧的制作周期以及成本限制，AI 生成剧在真实性与实拍剧仍有一定的差别。人物体相似度接近人类但存在细微差异时会产生恐怖谷效应，因此在短剧场景中，拟真人短剧当前在科幻、魔幻、神话志怪等和现实差距较大的题材更加适配。

长剧与电影方面，CG 特效等高价值环节有望成为首批替代场景。但鉴于 CG 特效仍为影视工业中技术门槛最高模块之一，具备专业团队与预算的头部厂商短期内仍可能优先采用成熟 CG 方案，并逐使用 AI 替换/调整原先的工作流程。相较而言，中小型影视团队或更可能依托 AI 提升特效制作水平，率先受益于技术平权红利。

2.1.1. AI 短剧：从漫剧开始的成本革命

2.1.1.1. AI 漫剧：AI 视频生成首个大规模应用场景

AI 技术帮助漫剧有效跨越了传统影视制作的特效成本限制，实现了内容形态的降维打击。其核心优势体现在以下方面：

填补题材空缺：仙侠修真、玄幻架空、赛博朋克等因特效成本过高而难以真人化的网文与漫画 IP，可通过 AI 低成本快速完成视频化，极大丰富了内容供给。

技术适配度高：漫剧非写实的风格化特征对画面精度和物理逻辑的容错空间更大，不仅适配当前 AI 多模态技术的能力边界，还能有效避开“恐怖谷效应”，提供更顺滑的视觉体验

AI 漫剧市场规模倍速增长，2026 年前 5 个月已超 220 亿元。在 AI 技术的深度赋能下，漫剧市场正迎来爆发式增长。根据 DataEye 的统计和预测，仅 2026 年前 5 个月，综合考虑广告消耗、非效果广告模式以及资本投入等情况，预计国内 AI 剧漫剧市场规模已超 220 亿元。考虑到赛道监管趋严、竞争激烈，不少企业重心转向出海等因素，2026 年全年国内 AI 剧漫剧市场规模保守估计将达 400 亿，较 2025 年增长 138%。

2025 年市场整体播放量超 700 亿，抖音端占比近七成。2024 年起，抖音及快手平台端原生漫剧内容开始兴起，部分动漫厂商开始布局端原生漫剧，主要以表情包/沙雕漫剧和小说漫为主。2025 年市场整体播放量超 700 亿，仅抖音端原生漫剧的播放量就达到了整体漫剧市场近 70%。

仅供内部参考，请勿外传！

图 8：2023-2025 年短视频端漫剧播放量



数据来源：DataEye、东北证券

产业生态：当前，漫剧产业已构建起上中下游协同发展的“AI+IP”双轮驱动生态体系，AI 技术与头部平台 IP 开放形成强大合力。该生态体系主要由以下环节构成：

（1） 上游 IP 供给：阅文集团开放 10 万部精品 IP 并设立亿元创作基金，知乎、番茄小说等平台同步开放海量版权资源，为漫剧提供优质故事蓝本。

（2） 中游内容制作：字节跳动、快手等大厂及专业公司提供全链路 AI 工具支持，如 Seedance 2.0、可灵 3.0 等模型大幅降低了复杂多模态漫剧的创作门槛。

（3） 下游渠道分发：内容通过抖音、快手、B 站等平台广泛分发，并依托用户付费（IAP）或广告（IAA）模式实现商业变现，形成完整的产业闭环。

图 9：AI 漫剧产业生态



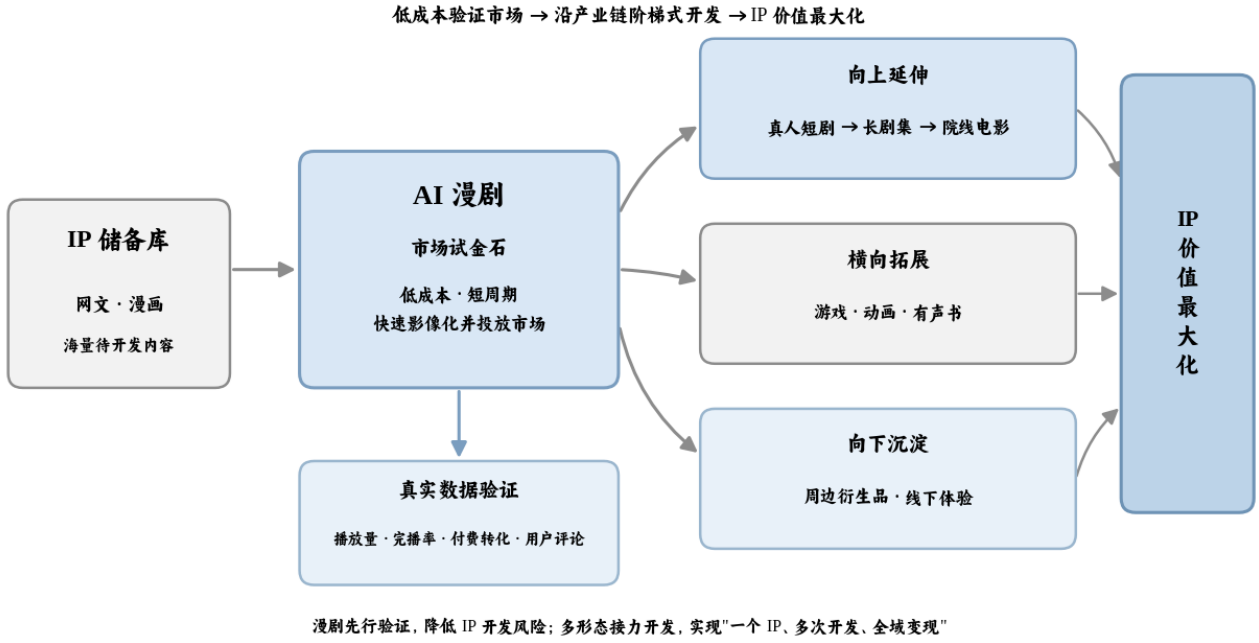
数据来源：艾媒咨询、东北证券

仅供内部参考，请勿外传！

流量红利开始退潮，AI 漫剧正在从无序扩张向高质量发展转变。AI 漫剧市场规模迅速膨胀的同时，题材同质化严重，部分作品粗制滥造、剧情逻辑混乱，抄袭搬运、擦边引流等问题也趋于严重。双重因素推动漫剧市场的转变：一方面，监管与平台治理逐步收紧，内容审核标准日趋明确，低质、侵权内容的生存空间被持续压缩；另一方面，市场供给趋于饱和，观众的审美快速升级，单纯依靠猎奇和爽点已难以留住用户。

AI 漫剧的未来：从单一变现环节升级为 IP 开发的前端入口。IP 开发往往是一条“高投入、长周期、高风险”的路径，一部真人剧或一款游戏动辄数千万甚至上亿元的投入，一旦市场反响不及预期，损失难以挽回。AI 漫剧可以对 IP 进行低成本验证，充当“市场试金石”，可以用极小的代价，将储备的网文、漫画 IP 快速影像化并投放市场，通过真实数据精准判断一个 IP 的市场潜力。经过漫剧验证的优质 IP，可以沿产业链条逐级延伸，向上开发为真人短剧、长剧集乃至院线电影，横向拓展至游戏、动画、有声书等形态，向下沉淀为周边衍生品、线下体验等消费场景。

图 10：AI 漫剧作为 IP 开发的前端入口



数据来源：东北证券

2.1.2. AI 仿真人短剧：技术已成熟，正向主流内容形态跃升

AI 漫剧在 2025 年内完成了四代形态的快速迭代。1.0 沙雕漫（50-150 元/分钟）以表情包+搞笑旁白为主，门槛极低但无付费转化；2.0 动态漫（150-300 元/分钟）对静态漫画做简单动态处理，流量高但追更粘性弱；3.0 小说剧（200-500 元/分钟）以小说旁白+AI 简单场景为核心，本质是“听书视频化”，适配付费点播；4.0 AI 漫剧（精品级 1000-3000 元/分钟）实现 AI 动画/真人画风融合、强叙事连续性，具备完整世界观，采用“免费广告+阶梯付费+IP 衍生”混合变现。形态演进反映了行业从“轻量化引流”向“精品化留存”的价值升级路径。

从技术上看，AI 仿真人短剧可以归结为 AI 漫剧的第五发展阶段，即由 AI 生成的以数字人/真人为角色形象的微短剧。我们认为 AI 仿真人短剧将是影视行业下一个快速爆发的赛道，这是由技术进步和市场上限共同决定的。

仅供内部参考，请勿外传！

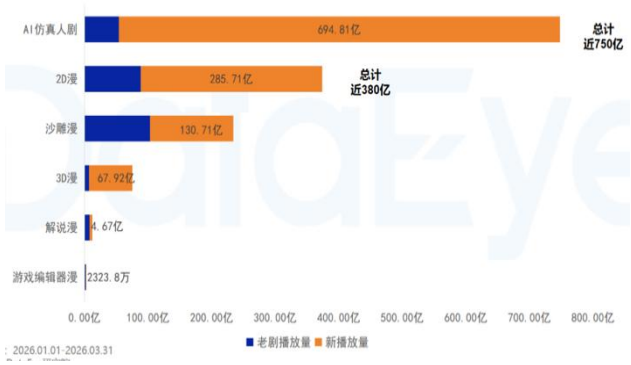
从技术进步的维度看，AI 仿真人短剧的爆发条件正在加速成熟。过去几年，人物形象的一致性难题逐步被攻克，数字角色可以在数十集篇幅内保持面容、神态、服装的稳定；口型与语音的精准对齐、微表情与肢体动作的自然生成，让“恐怖谷”效应快速消退；多镜头叙事、运镜调度和光影质感也在向专业影视水准逼近。

从市场上限的维度看，仿真人形态决定了这条赛道的天花板远高于前几个阶段。AI 漫剧的前四代形态，无论画风如何演进，本质上仍服务于二次元审美偏好的受众，用户规模和付费能力存在明确边界。而仿真人短剧直接切入了真人微短剧所验证过的庞大市场。当内容形象从“漫画脸”换成“真人脸”，目标受众便从垂直人群扩展至全年龄层的泛娱乐用户，单用户价值、广告承载能力、品牌合作空间乃至出海潜力都将同步跃升。从播放数据上看，漫剧界里程碑之作《斩仙台下，我震惊了诸神》，单平台播放量破 10 亿，放在真人短剧里只能排到 60 名开外，体量根本无法与动辄 30 亿的头部剧抗衡。

从产业格局的维度看，这一转变还将重塑影视行业的生产关系。真人短剧行业长期受制于演员片酬、场地档期、制作班底等重资产要素，而仿真人短剧将这些要素全面数字化：数字演员不知疲倦、档期无限、形象可控，场景可以任意生成，拍摄不再受物理条件约束。这不仅意味着成本的下降，更意味着内容生产从“项目制”走向“流水线制”，从劳动密集走向算力密集。率先建立数字演员库、打通生成管线的团队，将拥有类似游戏引擎之于游戏行业的底层优势。

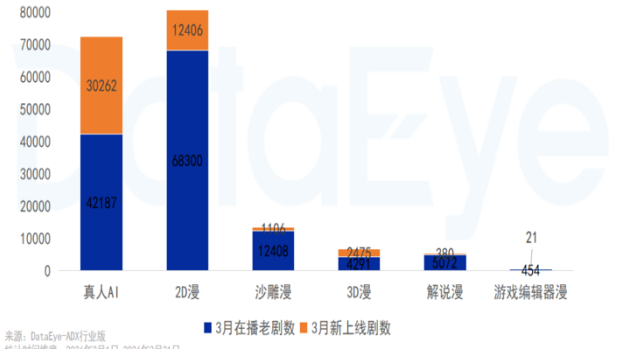
AI 仿真人短剧播放量及占比快速提升，2026 年 1 月至 3 月由 5%跃升至 40%。根据 DataEye 数据，2026 年 Q1 AI 短剧总播放量近 1300 亿，3 月份环比 2 月份播放量增长 78.9%，增速主要由 AI 仿真人剧贡献；其中 AI 仿真人剧实现近 750 亿总计播放量，占比约 58%，位居细分类别第一。分月份看，AI 剧细分品类中 2026 年 1-3 月 AI 仿真人剧播放量占比大幅提升，分别为 5%、30%、40%，2D+3D 漫剧占比分别为 71%、55%、49%，沙雕漫为 15%、10%、13%。

图 11：2026 年 Q1 各类型 AI 剧播放量分布



数据来源：DataEye，东北证券

图 12：2026 年 3 月各类型在播 AI 剧数量分布

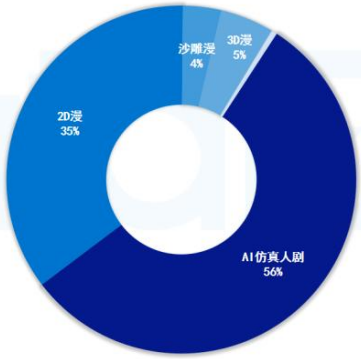


数据来源：DataEye，东北证券

从作品上新数量及新老剧播放占比看，2026 年 Q1 AI 仿真人剧已成 AI 剧增量绝对主力。2026 年 Q1 共发布 AI 剧超 10 万部，其中 AI 仿真人剧占比 56%；新剧贡献播放量占比高达 92.88%，单季度播放量即超过去年全年所有 AI 剧总和。其中 3 月 AI 仿真人漫剧中新剧占比达 41.77%，而在播的新剧播放量贡献比达 75.07%，验证了作品质量的快速提升下对用户吸引力增加。

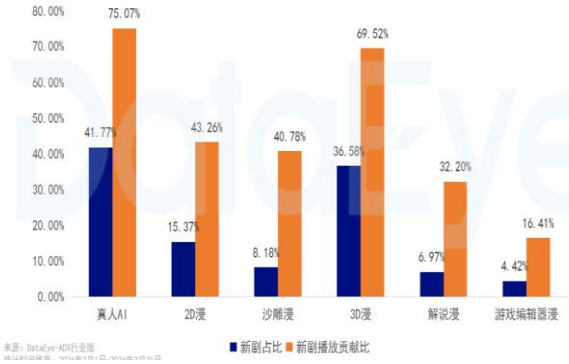
仅供内部参考，请勿外传！

图 13：2026 年 Q1 各类型 AI 剧上新数量占比



数据来源：DataEye，东北证券

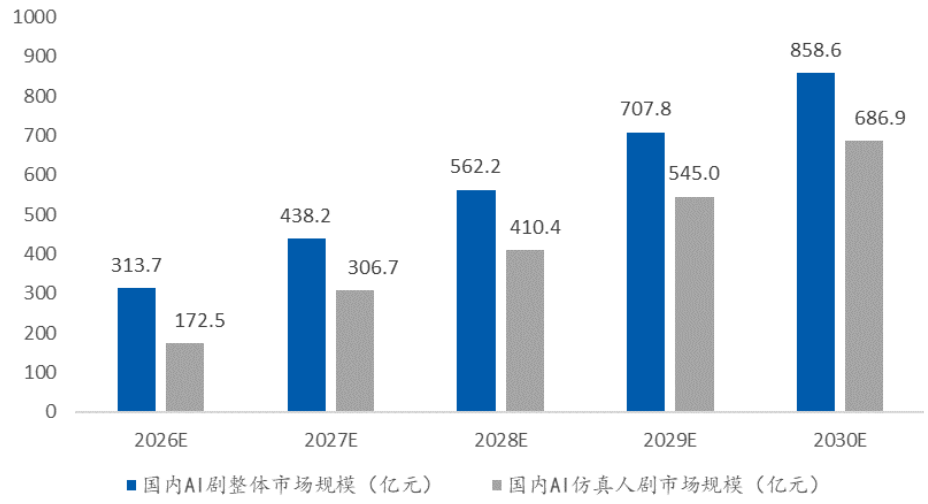
图 14：2026 年 3 月上新 AI 剧各类型占比



数据来源：DataEye，东北证券

根据 2026 年 3 月 AI 仿真人短剧播放量占比（40%）及产业趋势，我们预计 2026 全年占比水平将达到 55%，2030 年将至 80%。根据《2025-2026 年中国 AI 漫剧行业趋势白皮书》，我国 AI 短剧市场规模将由 2026 年全年 313.7 亿元增长至 2030 年 858.6 亿元，据此推测我国 AI 仿真人剧市场规模 2026 年预计实现 173 亿元，2030 年或将达到近 687 亿元。

图 15：2026-2030 年国内 AI 真人短剧市场规模预测



数据来源：《2025-2026 年中国 AI 漫剧行业趋势白皮书》，东北证券

传统影视公司切入 AI 真人剧赛道，有望在精品化维度建立差异化壁垒。如华策影视原创 AI 短剧《绯色之契》预告片已于 2026 年 4 月 29 日上线，该片由公司旗下华策环球星核工作室（Huace Global Nova Studio）全 AI 制作完成，以高概念叙事与电影级质感拓展未来影像表达。长期来看，传统影视公司凭借长周期内容工业化经验、经典 IP 储备及主流渠道合规优势切入 AI 真人剧赛道，有望在精品化维度建立差异化壁垒。

仅供内部参考，请勿外传！

图 16：《绯色之契》预告片截图



数据来源：公司微博、东北证券

2.1.3. AI 长影视：以辅助工具为主，动画电影率先落地，真人电影仍处起步阶段

观众对电影、电视剧的质量要求相较短剧更高，因此目前 AI 在电影、电视剧，尤其是真人内容领域的应用仍以辅助工具为主，主要包括 AI 换脸、辅助大成本场景及特效生成等。业内已有案例包括博纳影业出品的奇幻动作电影《传说》利用 AI 技术重现 27 岁的成龙形象，奈飞出品的阿根廷科幻剧集《El Eternauta》利用 AI 技术生成布宜诺斯艾利斯建筑倒塌的灾难画面等。

相较于真人电影对写实性的极高要求，动画电影的风格化表达为 AI 生成提供了容错空间，AI 生成更易实现其对角色面部表情、动作设计等内容的创作。因此，目前 AI 动画电影相较于 AI 真人电影发展更成熟。

头部影视公司除积极布局短剧业务外，已在 AI 赋能影视大制作上做出尝试：

华策影视长剧《太平年》中观众印象深刻的乌鸦食肉糜、暴风雨海战等场景均由可灵 AI 生成，后者在保障画面的专业级水准效率的基础上制作提升 8-10 倍。根据中国视听大数据数据，《太平年》单集最高收视率 3.2%，收视总规模达 11.4 亿人次，创 CVB 有收视记录（2018 年）以来古装历史类电视剧集收视率最高纪录。

图 17：太平年中利用可灵生成乌鸦食肉糜场景



数据来源：东西文娱，东北证券

图 18：太平年中利用可灵生成暴风雨海战场景



数据来源：东西文娱，东北证券

仅供内部参考，请勿外传！

博纳影业出品制作国内首部东方超写实灵智共创长篇电影《三星堆：未来往事》，现已正式获得国家电影局颁发的“龙标”，为我国首部获准院线上映的 AI 深度参与制作的电影。此外博纳影业储备《中华志怪之罗刹海市》多部 AI+电影作品。

图 19：《三星堆：未来往事》已获得国家电影局"龙标"



数据来源：中漫文化，东北证券

2.2. 广告：效率革命与精准营销

广告营销是 AI 视频生成技术商业化兑现最直接的行业分母。据 QuestMobile 2026 年 3 月发布的《2025 年中国营销市场年度报告·市场篇》，其统计口径下中国互联网广告市场规模由 2023 年的 7146.1 亿元增长至 2025 年的 7930.8 亿元，并预计 2026 年、2027 年将分别达到 8362.6 亿元、8847.6 亿元。互联网广告大盘已由流量扩张期进入稳健增长期，行业增长动能正切换至技术与内容供给侧。据中关村互动营销实验室、秒针营销科学院与北京师范大学新闻传播学院 2026 年 1 月联合发布的《2025 中国互联网广告营销趋势报告》，2025 年视频类信息流广告同比增速高达 18.85%，为所有广告形式中增长最快的品类，视频化是大盘内部最显著的增量方向，这为 AI 视频生成提供了天然的渗透场景。

图 20：2023-2027 年中国互联网广告市场规模变化



数据来源：QuestMobile、东北证券

仅供内部参考，请勿外传！

AI 对广告行业的渗透正从单点工具升级为全链路基础设施。《2025 中国互联网广告营销趋势报告》将“AI 技术渗透”列为重塑品牌营销底层逻辑的三大核心驱动力之首，2026 年行业核心驱动力将彻底从流量红利转向 AI 技术全链路渗透。

市场规模测算上，据量子位智库《中国 AIGC 广告营销产业全景报告》，2023 年中国生成式 AI 在广告营销行业中的业务营收占比尚不足 1%，市场规模约 50 亿元。预计未来每年增速或达 60%以上，2030 年 AIGC 广告营销业务营收规模有望达 1500 亿元。

以行业龙头蓝色光标为例，其 AI 驱动收入由 2024 年的 12 亿元增至 2025 年的 37.25 亿元，占总营收比重由约 2%（12 亿元/607.97 亿元）提升至 5.42%（37.25 亿元/686.93 亿元）。综合看，AI 在广告行业的收入渗透率目前仍处于低个位数水平，但增速显著高于大盘，渗透率抬升将是广告行业的核心结构性变量

2.2.1. 素材生成与个性化广告：成本曲线下移与个性化上限提升

素材生产是 AIGC 在广告营销链路中落地最深的环节。量子位智库预测，未来 5 年内 70%的营销内容可由 AI 自动生成并迭代，内容生产将由存量市场转变为增量市场。

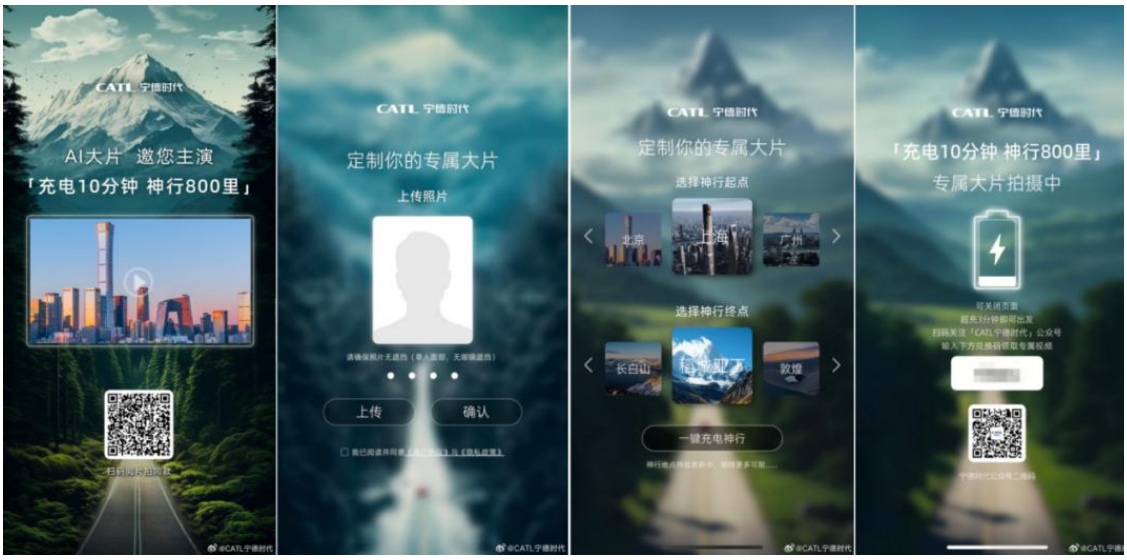
营销服务商侧，2024 年 9 月，蓝色光标与快手可灵 AI 签署战略合作协议，其自研 AI 行业模型 Blue AI 心影创作平台首批接入可灵 AI 视频大模型 API；心影平台为蓝色光标自研的 AI 视频理解、生成和剪辑平台，已在商品推广、APP 推广、短剧推广、视频品效评估等多类效果投放场景中提供智能视频创作服务。

内容公司侧，中文在线依托自研“逍遥大模型”的文本生成及理解能力、结合多模态 AI 视频生成技术，打造了覆盖智能广告投放全流程的逍遥 AI Agent 系统：该系统构建了本地化广告素材智能生产流程，通过提取内容关键文字信息生成图片、音频、视频等多模态广告创意素材，实现内容、素材与受众的精准适配，直接服务于短剧等数字内容的出海买量。

在批量生成的基础上，生成式 AI 把广告个性化从“精准定向一类人”推进到“只为一个人打造广告”。代表性案例是 2023 年 11 月广州车展期间蓝色光标为宁德时代发布的广告片——全球第一部由用户自己选择剧情、自己出演的 AIGC 广告片：用户只需选择性别、脸型、旅行起止地点并上传照片，约三分钟即可获得一部由 AI 生成、本人主演的个性化驾驶大片

仅供内部参考，请勿外传！

图 21：宁德时代个人定制 AI 广告



数据来源：广告门、东北证券

从投放效果看，据广告门案例页披露，该 AI TVC 小程序通过裂变而来的用户占总 UV 的 80%，为行业均值的 2.5 倍以上；超过 25%的用户参与互动，为行业平均互动率的 3-5 倍；参与互动用户最终加粉留资的留存率达 90%，效率较传统页面广告从浏览到留资提升近 10 倍。

国际头部品牌已将 AI 视频生成用于正式广告片的规模化制作，效果获得第三方创意评估背书，但技术短板与舆情风险同步暴露。2025 年 11 月，可口可乐发布 2025 年节日营销广告，其中 1995 年经典广告《节日将至》（Holidays are Coming）的 AI 重制版由 5 名 AI 专家在 30 天内利用 OpenAI Sora、谷歌 Veo 3 和 Luma AI 等工具、从 7 万条视频片段中提炼完成，后期进行微调；创意效果评估机构 System1 给予该广告满分 5.9 分，认为其具备推动品牌长期增长的潜力。

图 22：可口可乐 AI 广告



数据来源：新浪财经、东北证券

仅供内部参考，请勿外传！

综合来看，素材生成环节的成本曲线下移与个性化上限提升，正推动广告行业竞争要素从媒介采买规模转向“AI 内容产能+投放数据闭环”，具备自研模型与 Agent 平台的服务商有望获得超额收益。

2.2.2. 数字人：商业化最先跑通的场景

AI 数字人是 AI 视频生成与多模态大模型技术商业化落地最为成熟的场景之一，目前已从泛娱乐的“流量符号”演进为直播电商、品牌营销、金融客服、政务服务等领域的生产力工具。据艾媒咨询 2025 年 11 月发布的《2025 年中国数字人产业发展报告》，2024 年中国数字人核心市场规模为 339.2 亿元，预计 2030 年将达到 935.6 亿元；2024 年数字人带动市场规模为 4785.3 亿元，预计 2030 年为 10468.6 亿元。据艾媒咨询 2026 年 2 月发布的数据，中国数字人核心市场规模从 2017 年的 8.1 亿元增长至 2025 年的 480.6 亿元，年均增长率达 66.60%，2026 年预计进一步扩大至 572.9 亿元。上述口径为“广义数字人”，覆盖虚拟主播、虚拟偶像、数字员工等核心环节及对关联产业的带动效应。

若采用狭义口径，据 IDC 发布的《中国 2024 年 AI 数字人市场份额》报告，2024 年中国 AI 数字人市场规模约 41.2 亿元（按 PaaS/SaaS 年度收入计），同比增长 85.3%，预计 2029 年将达到 250.5 亿元，2024—2029 年复合增速 43.5%，其中百度以 9.8%的市场份额位居行业第一。

直播电商：数字人商业化最先跑通的场景，降本增效逻辑已获验证。数字人直播电商已进入规模化商用阶段，头部平台披露的数据验证了“成本降至真人约 1/10、转化率不降反升”的商业逻辑。

互联网大厂中，京东方面，据《证券日报》2025 年 9 月报道，京东于 2024 年春节试水数字人直播，言犀数字人春节期间累计带货 GMV 超 4000 万元；2024 年 4 月，刘强东数字分身“采销东哥”首播，开播不足半小时，京东家电家居、京东超市两大采销直播间即吸引超 1000 万次观看。据凤凰网 2024 年 6 月报道，京东在当年 618 媒体开放日披露，言犀数字人直播成本仅为真人成本的 1/10，累计服务品牌超 5000 家，带动 GMV 超百亿元。2025 年 618 期间，京东数字人直播已实现“高商业可用”：成本仅为真人 1/10，可 7×24 小时全天开播，提升直播间转化率 30%，直播表现超过 80%真人主播。据快科技 2026 年 1 月报道，2025 年以来京东数字人累计带动 GMV 数百亿元，累计服务超 5 万家品牌；2025 年 12 月京东官宣数字人直播免费开放后，10 天内报名商家数即突破 1 万家。

百度方面，据新华网 2025 年 6 月报道，过去一年百度慧播星数字人主播累计超过 10 万，直播转化率平均提升 31%，开播成本降低 80%。据经济观察网 2026 年 1 月报道，截至目前慧播星已服务超 10 万数字人主播，覆盖电商、教育、法律、健康等 30 余个行业，2025 年百度双 11 期间 83%的商家使用了其数字人直播服务；实践数据显示，应用慧播星可帮助商家运营成本平均下降 80%，GMV 平均提升 62%；2025 年 6 月 15 日，业内首个超头主播数字人直播间“罗永浩数字人”首秀 GMV 突破 5500 万元，下单用户增长 230%，人均观看时长提升 30%。

虚拟代言人与品牌 IP：从“流量造星”回归理性，IP 运营能力成为分水岭。虚拟代言人赛道经历了从资本热捧到理性回归的完整周期。

2021 年虚拟美妆数字人“柳夜熙”在抖音上线 48 小时即斩获百万粉丝，点燃产业圈；但其制作方创壹科技 CEO 梁子康透露，公司在推出“柳夜熙”之前，仅半年多在研发人员与技术设备等方面的投入就达百万元，第一条“柳夜熙”短视频成本达数十万元。此后企业纷纷跟进推出自研虚拟代言人，屈臣氏推出“屈晨曦”，花西子发布“花西子虚拟人”，招商银行上线“AI 小招”。

仅供内部参考，请勿外传！

主流媒体 IP 化运营则提供了相对成功的样本：据艾媒咨询《2025 年中国数字人产业发展报告》案例研究，央视网 3D 超写实数字虚拟主播“小 C”作为总台体系内首个常态化运营的超写实数字人，已推出 200 余条内容，社交平台粉丝超 200 万，全网传播量达亿级。我们认为，早期超写实虚拟代言人“高投入、难持续”的教训表明，虚拟 IP 的商业价值取决于持续内容运营与场景落地能力，而非形象本身的技术精度；随着生成式 AI 将数字人制作成本大幅摊薄，虚拟代言人的竞争重心正从“造星”转向对现有 IP 的运营打造。

图 23：数字人产业生态



数据来源：艾媒咨询、东北证券

2.3. 游戏：自演化 AI 原生游戏的基础

游戏是 AI 视频生成技术商业化落地最具确定性的下游场景之一，其分母足够大且仍在扩张。据中国音像与数字出版协会游戏出版工作委员会（游戏工委）2025 年 12 月 19 日在 2025 年度中国游戏产业年会上发布的《2025 年中国游戏产业报告》，2025 年国内游戏市场实际销售收入达 3507.89 亿元，同比增长 7.68%，游戏用户规模达 6.83 亿，同比增长 1.35%，两项指标均创历史新高。

结构上看，自研游戏国内市场实际销售收入 2910.95 亿元，同比增长 11.64%；自研游戏海外市场实际销售收入 204.55 亿美元，同比增长 10.23%，规模已连续六年超过千亿元人民币，出海已成为行业确定性第二曲线；小程序游戏市场收入 535.35 亿元，同比大幅增长 34.39%，是增速最快的细分板块。

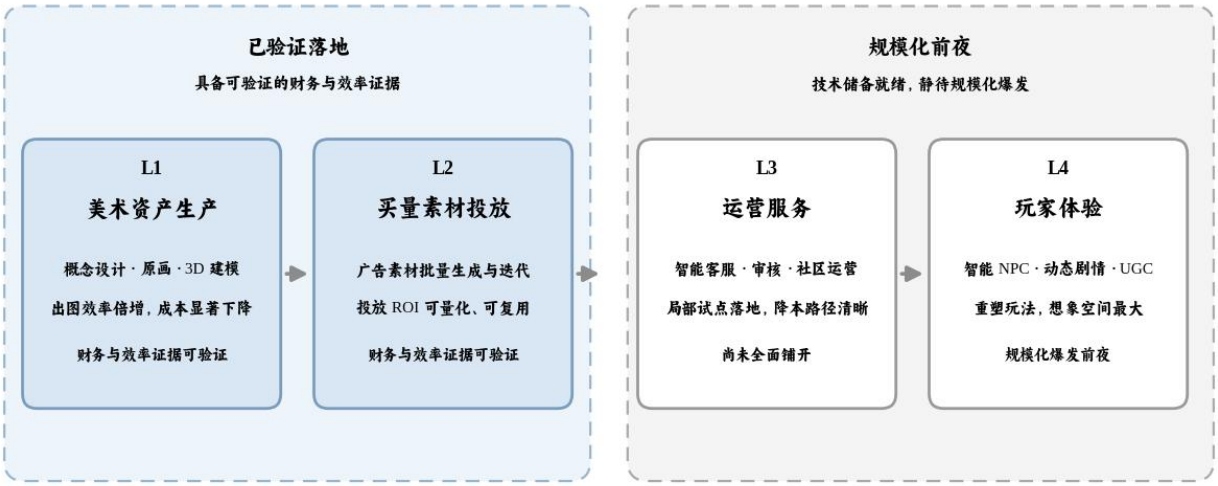
在一个超过 3500 亿元、用户接近 7 亿的成熟存量市场中，**内容生产与获客投放是游戏厂商最大的两项可变成本**，AI 对这两端的渗透直接决定其利润弹性，这也是 AI 视频生成在游戏行业率先规模化的产业逻辑。

仅供内部参考，请勿外传！

2.3.1. 从美术资产生成到 AI 原生游戏

AI 对游戏制作的改造沿“美术资产生产—买量素材投放—运营服务—玩家体验”四个层次逐级展开，前两层已具备可验证的财务与效率证据，后两层仍处规模化前夜。

图 24：AI 对游戏制作的四层改造



数据来源：东北证券

美术资产层：AI 渗透最深、证据最充分的环节。据三七互娱 2025 年年度报告，公司在发行环节使用 AI 技术辅助生成的 2D 美术资产占比已超过 80%，2D 图片每季度产出超过 50 万张，AI-3D 辅助资产生成占比超过 30%。效率维度，三七互娱精细人物立绘的产出由传统人工方式的 8—10 人天压缩至 AI 方式的 2.5 人天。恺英网络投资的公司自研的虚幻引擎 AIGC 工具“已经能够批量形成高精度的人物和模型渲染图”，“能将创作时间从两个月降低到半个小时”。完美世界游戏 2025 年已将 AI 相关技术应用于智能 NPC、场景建模、AI 绘画、AI 剧情、AI 配音等方面，其中端游《诛仙世界》接入英伟达技术，借助生成式 AI 实现音频到面部动画的直接转换，使角色情感表达更加细腻自然、开发效率提升。

平台化供给方面，腾讯混元游戏视觉生成平台为业内首个工业级 AIGC 游戏内容生产引擎，其角色多视图生成耗时由传统的 12 小时压缩至几分钟至 30 分钟，主体一致性达 99%，已应用于多个头部游戏项目，助力美术团队效率提升 300%。

买量素材层：AI 视频对游戏发行成本结构的重塑最为直接。视频素材是效果广告买量的核心物料，AI 生成直接压缩素材成本与投放人力。三七互娱智能投放平台“量子”系统投入使用后，游戏产品上线全天广告所耗时间从原来的 6 小时缩短到 30 分钟，效率提升 10 倍以上，部分地区智能投放占比已超过 60%。据三七互娱 2025 年年度报告，公司由 AI 深度参与生成的游戏广告素材视频占比已超过 70%，“量子”与“宙斯”平台 AI 辅助投放广告比例达总投放的 50%，AI 图片扩散在海外平台消耗占比达 20%。

运营层：AI 成为长线运营的效率基础设施。据三七互娱 2025 年年度报告，公司旗下全部游戏均已接入 AI 智能客服，解决准确率达 90%；AI 本地化翻译覆盖旗下所有海外游戏、覆盖 18 个语种，单语种最高准确率可达 95%；用于知识产权保护的素材侵权检测系统上线后 5 个月内，识别及获取外部侵权核心线索效率提升约 9 倍。

仅供内部参考，请勿外传！

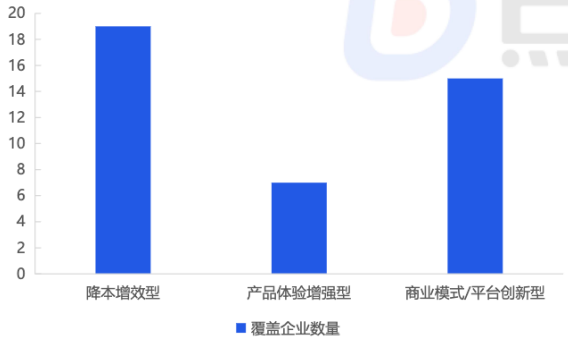
行业层面，点点数据报告亦将客服、审核、风控、推荐、社区治理等运营环节列为 AI 率先渗透的方向，典型案例包括世纪华通智能客服、心动公司 TapTap 社区 AI 与腾讯 AI 风控工具等。

玩家体验层：从产品后台走向玩家可感知，仍处早期探索。2025 年已有少数上市游戏公司出现较具体的玩家可感知 AI 产品化探索：网易披露 AI 驱动 NPC 可加入玩家游戏活动并模拟互动、表情和肢体语言；巨人网络披露 AI“残局挑战”及 2025 年 3 月上线的“内鬼挑战”，通过大模型能力强化 AI 玩家的推理与拟真交互；游族网络推出 AI 玩伴“小游酱”，探索智能交互、专属助理和情感陪伴。我们认为，体验型 AI 一旦在智能 NPC、AI 陪玩、动态剧情等方向形成稳定玩法，将改变游戏内容供给方式与长线运营逻辑，是 AI 在游戏行业从“降本工具”走向“体验资产”的关键一跃。

2.3.2. AI 原生游戏的自演化：游戏玩法的变革性升级

玩家可感知的 AI 功能已开始落地，但整体仍处早期阶段。当前游戏上市公司的 AI 应用仍明显偏向“降本增效”，AI 最先进入的是任务可拆分、流程可标准化、效果可量化的后台环节；产品体验增强型 AI 仅覆盖 7 家公司，玩家可感知的 AI 功能与玩法仍处于早期探索阶段。

图 25：不同 AI 价值导向场景企业覆盖数量



类型	覆盖企业数量	定义	典型场景	价值判断
降本增效型	19家	AI帮助企业降低生产、运营、投放、客服等成本，或提升同等成本下的产出效率	AI美术、文案生成、代码辅助、自动化测试、翻译、智能客服、广告素材、智能投放、审核风控	最广泛、最现实，但不能直接等同于已形成财务降本
产品体验增强型	7家	AI成为玩家或用户可感知的产品能力，提升互动性、个性化、内容丰富度或社区体验	智能NPC、AI角色交互、AI陪玩、个性化推荐、AI社区工具	覆盖较少，且很多仍停留在探索或定性描述
商业模式/平台创新型	15家	AI形成独立产品、平台服务、云/算力服务、工具订阅或新收入模式	腾讯元宝/混元/云AI，金山办公AI，网龙教育AI，世纪华通AI云数据	目前与游戏关联较弱

数据来源：点点数据、东北证券

从具体公司实践看，头部厂商已出现较具体的产品化探索。网易在年报中披露，其 AI 驱动 NPC 可加入玩家游戏活动，并模拟互动、表情和肢体语言；据网易伏羲官网 2025 年 1 月刊载的介绍，《逆水寒》手游于 2024 年 11 月上线“AI 大模型竞技场”玩法，阿里通义、百度文心、MiniMax abab、月之暗面 Kimi 与字节豆包五大国产大模型以九位智能 NPC 形象入驻游戏，玩家亦可借助“自捏江湖友人”功能自行创建具备性格、情商与记忆设定的智能 NPC，上述 AI NPC 可自主游历江湖、结交好友并迭代进化。巨人网络旗下社交推理游戏《太空杀》上线 AI 原生玩法“残局对决”，联合阿里通义千问、腾讯混元、字节豆包三大国产大模型，打造游戏行业首个“多用户与 AI 智能体混合对抗”的竞技机制；该产品此前已陆续推出“AI 推理小剧场”、“AI 残局挑战”、“侦探剧场”、“内鬼挑战”等多个 AI 原生玩法，相关模式累计参与玩家达数百万，对局总量达数千万。另据腾讯云 2025 年 4 月披露，《太空杀》接入腾讯混元 Turbo S 大模型后，2 亿注册玩家可与超 700 万 AI 角色同台竞技，接入后一个月内 AI 玩家累计参与近 90 万场对局。游族网络则选择 AI 陪伴作为切入点，公司推出的智能游戏玩伴“小游酱”及智能 NPC 通过多模型协作技术兼顾情感互动与游戏辅助功能，使玩家能够在交互过程中“养成”专属 AI 陪玩。

前沿探索层面，AI 原生游戏已从上市公司的玩法改良延伸至创业团队的原生产品。米哈游创始人蔡浩宇创办的 AGI 公司 Anuttacon 公布首款 AI 游戏《Whispers from

仅供内部参考，请勿外传！

the Star》并发布预告片，该游戏为融入大语言模型的科幻背景沉浸式互动叙事作品，以语音交互为核心玩法，主角 Stella 的对话由 AI 实时生成，玩家可通过实时文字、语音消息及视频通话引导其在外星球脱险。相较于 Krafton 等厂商“在传统玩法中加入 AI 赋能”的路径，《Whispers from the Star》在形态上更倾向于“原生 AI 的游戏”，即以 AI 实时交互为底层机制重构玩法本身。

图 26：《Whispers from the Star》游戏 PV 截图



数据来源：游戏 PV 截图、东北证券

更远期看，随着实时世界模型与视频生成技术的演进，电影与游戏的边界可能进一步消融：一段可实时交互的生成式影像即构成一个开放世界，NPC 可由大模型实时生成并与玩家建立长期记忆关系，普通玩家亦可借由生成式工具直接创作游戏内容，UGC 有望成为游戏内容供给的重要来源。我们认为，AI 原生游戏的演进路径将沿“后台降本增效—玩家可感知 AI 功能—AI 原生玩法—生成式开放世界”逐级展开，当前行业整体仍处于第一、二阶段之间，随着 AI 模型多模态能力的逐渐增强，三、四阶段的游戏玩法将有变革性升级。

3. 风险提示

监管与内容合规风险。文档指出漫剧赛道监管趋严、平台治理收紧，内容审核标准日趋明确；行业曾存在抄袭搬运、擦边引流、侵权等问题。AI 生成内容还涉及肖像权、著作权归属及内容标注、备案等要求，若监管进一步加码或出现侵权纠纷，相关公司内容上线与变现节奏可能受到冲击。

技术迭代与影视落地不及预期风险。AI 仿真人短剧与实拍剧在真实性上仍有差距；长影视领域 CG 特效替代技术门槛高，头部厂商短期仍优先采用成熟 CG 方案，AI 渗透进度可能慢于预期。

短剧行业竞争加剧与流量红利退潮风险。供给趋于饱和、题材同质化严重、观众审美快速升级，单纯依靠猎奇和爽点已难以留住用户。若相关公司无法完成从“轻量化引流”向“精品化留存”的转型，增长逻辑可能弱化。

商业化与盈利兑现风险。数字人虚拟代言人曾有“高投入、难持续”的教训，商业价值取决于持续内容运营能力而非技术精度；短剧出海依赖投流买量，成本占比高、

仅供内部参考，请勿外传！

盈利模型尚待验证。AI 收入渗透率虽快速提升，但绝对占比仍处低个位数，业绩弹性兑现需要时间。

AI 广告舆情与品牌风险。品牌 AI 广告在获得创意评估背书的同时技术短板与舆情风险同步暴露，AI 生成内容一旦引发公众对“真实性”“伦理”的负面舆情，可能波及品牌方与制作方。

游戏新业务探索不确定性风险。AI 原生游戏、玩家体验层 AI（智能 NPC、AI 陪玩、动态剧情）仍处早期探索阶段，覆盖公司数量有限，能否形成稳定玩法和可持续商业模式存在较大不确定性。

仅供内部参考，请勿外传！

研究团队简介：

赵明扬：北京大学硕士（网络传播新媒体方向）；曾任行业主管部委负责产业研究，科技领域投资（VC PE）、投行(FA)、实业（上市公司投资总监）十余年，所投企业 A 股、美股、港股皆有上市。2025 年 1 月加入东北证券，现任传媒互联网行业首席分析师。

刘昱亨：北京大学计算机硕士，北京航空航天大学工学学士。2023 年加入东北证券，现任东北证券传媒互联网行业研究助理。

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，并在中国证券业协会注册登记为证券分析师。本报告遵循合规、客观、专业、审慎的制作原则，所采用数据、资料的来源合法合规，文字阐述反映了作者的真实观点，报告结论未受任何第三方的授意或影响，特此声明。

投资评级说明

股票投资评级说明	买入	未来 6 个月内，股价涨幅超越市场基准 15%以上。	投资评级中所涉及的市场基准： A 股市场以沪深 300 指数为市场基准，新三板市场以三板成指（针对协议转让标的）或三板做市指数（针对做市转让标的）为市场基准；香港市场以摩根士丹利中国指数为市场基准；美国市场以纳斯达克综合指数或标普 500 指数为市场基准。
	增持	未来 6 个月内，股价涨幅超越市场基准 5%至 15%之间。	
	中性	未来 6 个月内，股价涨幅介于市场基准-5%至 5%之间。	
	减持	未来 6 个月内，股价涨幅落后市场基准 5%至 15%之间。	
	卖出	未来 6 个月内，股价涨幅落后市场基准 15%以上。	
行业投资评级说明	优于大势	未来 6 个月内，行业指数的收益超越市场基准。	
	同步大势	未来 6 个月内，行业指数的收益与市场基准持平。	
	落后大势	未来 6 个月内，行业指数的收益落后于市场基准。	

仅供内部参考，请勿外传！

重要声明

本报告由东北证券股份有限公司（以下称“本公司”）制作并仅向本公司客户发布，本公司不会因任何机构或个人接收到本报告而视其为本公司的当然客户。

本公司具有中国证监会核准的证券投资咨询业务资格。

本报告中的信息均来源于公开资料，本公司对这些信息的准确性和完整性不作任何保证。报告中的内容和意见仅反映本公司于发布本报告当日的判断，不保证所包含的内容和意见不发生变化。

本报告仅供参考，并不构成对所述证券买卖的出价或征价。在任何情况下，本报告中的信息或所表述的意见均不构成对任何人的证券买卖建议。本公司及其雇员不承诺投资者一定获利，不与投资者分享投资收益，在任何情况下，我公司及其雇员对任何人使用本报告及其内容所引发的任何直接或间接损失概不负责。

本公司或其关联机构可能会持有本报告中涉及到的公司所发行的证券头寸并进行交易，并在法律许可的情况下不进行披露；可能为这些公司提供或争取提供投资银行业务、财务顾问等相关服务。

本报告版权归本公司所有。未经本公司书面许可，任何机构和个人不得以任何形式翻版、复制、发表或引用。如征得本公司同意进行引用、刊发的，须在本公司允许的范围内使用，并注明本报告的发布人和发布日期，提示使用本报告的风险。

若本公司客户（以下称“该客户”）向第三方发送本报告，则由该客户独自为此发送行为负责。提醒通过此途径获得本报告的投资者注意，本公司不对通过此种途径获得本报告所引起的任何损失承担任何责任。

东北证券股份有限公司

网址：http://www.nesc.cn 电话：95360,400-600-0686 研究所公众号：dbzqyanjiusuo

地址	邮编
中国吉林省长春市生态大街 6666 号	130119
中国北京市西城区锦什坊街 28 号恒奥中心 D 座	100033
中国上海市浦东新区杨高南路 799 号	200127
中国深圳市福田区福中三路 1006 号诺德中心 34D	518038
中国广东省广州市天河区冼村街道黄埔大道西 122 号之二星辉中心 15 楼	510630