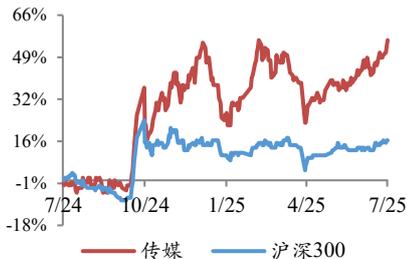


# AI 系列专题跟踪——视频及图像生成模型

行业评级：增持

报告日期：2025-07-15

## 行业指数与沪深 300 走势比较



分析师：金荣

执业证书号：S0010521080002

邮箱：jinrong@hazq.com

## 相关报告

1.OpenAI 开放深度研究模型 API 权限，稳定币概念或持续发酵 2025-06-29

2.GPT-5 预计夏季发布，关注华为概念行情 2025-06-22

## 主要观点：

### ● 生成式 AI 模型开源闭源并行发展，模型版本持续迭代

从生成式 AI 模型出现至今，国内外的 AI 视频图像模型不断迭代，整体上呈现开源与闭源生态并行的发展趋势。海外及国内 AI 大厂（如 Google、Adobe、OpenAI、字节跳动）在闭源模型竞争上日趋激烈，不断在各自领域加速渗透大型商业化市场，例如广告、影视制作等场景；开源模型的落地更多集中在降低中小开发者门槛，在游戏、短视频等交互场景中推动技术普惠化。

目前国内外主流的生成式 AI 模型版本主要为：Gen-4 Turbo、可灵 2.1、即梦 AI 3.0、Midjourney V7 等。整体在智能化水平高速提升的背景下，AI 驱动的创新落地成为各厂商竞争点，例如将生成式 AI 与搜索、广告内容、办公等多个领域进行融合，加快模型在 B 端和 C 端的落地。

### ● 生成式 AI 在影视行业内全链路布局，多维度提升影视质量

生成式 AI 已经初步在影视行业落地，例如在动画短片和部分特效场景中。AI 视频生成已经具备复杂场景生成能力，在保证质量的同时可以实现分镜，真实物理规律以及角色一致性等。从落地的场景角度，生成式 AI 在剧本生成，角色/场景建模，动画生成，后期配音，剪辑调色等环节均有不同程度的应用。

我们认为在影视领域，算力资源和一致性为目前主要的限制因素。原因在于生成更长的视频意味需要更多的计算资源，或者在资源有限的情况下，由于模型自身的技术原因，为了保证质量，视频的时长会存在压缩的情况，这也就对视频生成的前后一致性有了更高的要求。

### ● 助力游戏内容生成，游戏内赋能交互场景

市场上游戏类别多样化，因此对于研发效率和研发质量有着更高要求，游戏内容也从传统的固定交互逐步走向个性化互动模式。生成式 AI 在此过程中可以从图像、文本、音频、3D 等方面进行画面及场景生成。

在内容生成上，生成式 AI 可以从美术，背景内容，玩法上进行对应生成。美术包括角色的建模和对应的贴图动画和局内特效；背景内容包括背景生成，剧情生成等；玩法包括关卡难度，道具属性调整等。在玩家游玩的交互层面，生成式 AI 可以通过 NPC 的个性化交互进行辅助游玩，包括但不限于玩法介绍，装备介绍，局内攻略等。同时在局内场景上也可以实现智能识别并动态响应。

### ● 生成式 AI 未来趋势

在影视领域，未来生成式 AI 会不断地在精度、质量和时长三个维度进行创新迭代。首先在精度方面，要在复杂提示词和参数的条件下精准生成和控制镜头角度，运动轨迹和背景布局等。目前主流的生成工具对于相对复杂的指令响应有限。未来条件指令输入的方式多模态化，文字+图片生成视频方式会被逐步采用。

在质量上，主要痛点还是会出现不符合真实物理规律的情况，例如“局部合理，整体荒谬”情况，稳定性有待提高。

时长上，当前主流的生成模型生成时长集中在半分钟内，例如可灵 2.0 最长可以在 1080p 分辨率下生成 2 分钟视频。限制时长主要与模型复杂程度有限、计算资源有限有关。

#### ● 投资建议

国内生成式 AI 行业内百花齐放，各领域厂商均有布局。在游戏的制作与应用方面，腾讯凭借自身技术积累，赋能多个核心环节的创作生成，全线覆盖影视内容制作的前、中、后期，实现降本增效，创意生成。同时公司开源混元 3D 生成模型 2.0 版本，并同步推出混元 3D AI 创作引擎，对游戏角色的建模进行解耦，加快内容的制作。阿里巴巴的通义万相模型在业内领先，目前开源至 Wan2.1-VACE 版本，可以同时支持文生视频、图像参考视频生成、视频重绘、视频局部编辑、视频背景延展以及视频时长延展等全系列基础生成和编辑能力。同时阿里在生成模型上进行多任务能力探索，包括：图像参考；视频重绘；局部编辑等。快手可灵大模型从 23 年推出开始，先后经历了 6 个版本迭代至可灵 2.1，在画质质量、动态质量、美学表现、运动合理性以及语义理解等方面不断提升。自研模型和自身运营视频平台打开了商业化落地的想象空间，形成广告等内容的自产自销，从而推动内循环广告健康增长。建议关注在模型侧持续投入布局 and 商业化落地初具成效的相关公司，模型侧包括：腾讯控股、阿里巴巴-W、快手-W、美图公司等，游戏端包括：网易-S、恺英网络、三七互娱、吉比特、神州泰岳、巨人网络等，影视端包括：光线传媒、万达电影、上海电影、中文在线、昆仑万维等，广告端包括：分众传媒，蓝色光标、省广集团、易点天下等。

#### ● 风险提示

版本迭代不及预期风险，相关政策风险，商业化不及预期风险等。

## 正文目录

1 AI 视频及图像生成模型未来展望	8
1.1 AI 视频图像模型技术发展趋势	8
1.2 AI 视频图像模型未来应用场景展望	11
2 RUNWAY	15
2.1 GEN 系列模型迭代更新，内容生成技术不断精进	15
2.2 GEN 系列模型应用场景案例及核心技术	16
2.3 RUNWAY 产品价格模式	22
3 可灵 AI	23
3.1 可灵 AI B 端 C 端同步发展	23
3.2 可灵 AI 商业化模式及运营数据	27
3.3 近期更新	29
4 即梦 AI	30
4.1 即梦 AI 模型版本迭代、核心功能及技术	30
4.2 即梦 AI 模型应用场景、生成效果及同行业产品对比	32
4.3 即梦 AI 产品价格模式	35
5 MIDJOURNEY	36
5.1 MIDJOURNEY 版本迭代、优化情况及效果图比较	36
5.2 MIDJOURNEY 核心技术及应用场景	37
5.3 MIDJOURNEY 产品价格模式	39
6 STABLE DIFFUSION	40
6.1 STABLE DIFFUSION 版本迭代与变体及核心功能	40
6.2 STABLE DIFFUSION 核心技术及应用场景	42
6.3 STABLE DIFFUSION 产品价格模式	45
7 豆包	45
7.1 豆包大模型关键更新迭代、主要功能变化	45
7.2 豆包大模型核心技术及应用场景	47
7.3 豆包视觉理解模型、图片及视频生成模型产品价格模式	48
8 海艺 AI-SEAART AI	50
8.1 SEAART AI 版本迭代、核心技术及合作情况	50
8.2 SEAART AI 用户分层及应用场景	51
8.3 SEAART AI 产品价格模式	53
9 LIBLIBAI	54

9.1 LIBLIBAI 概况、融资情况及核心功能 .....	54
9.2 LIBLIBAI X 星流 STAR-3 ALPHA 大模型功能及应用场景 .....	55
9.3 全球第一款专业设计 AI AGENT: LOVART .....	58
9.4 LIBLIBAI 产品价格模式 .....	59
10 美图 AI AGENT ROBONEO .....	60
11 投资建议 .....	62
11.1 影视行业应用 .....	62
11.2 游戏行业应用 .....	64
11.3 相关公司 .....	65
风险提示: .....	66

## 图表目录

图表 1 2024 年 11 月至 2025 年 4 月 AI 视频图像模型发展图	8
图表 2 近期 3D 生成模型及核心技术介绍	9
图表 3 CUBE 3D 生成 3D 模型效果图	10
图表 4 UNI-3DAR 生成 3D 模型效果图	10
图表 5 KISS3DGEN 生成 3D 模型效果图	10
图表 6 TRIPOSG 和 TRIPOSF 生成 3D 模型效果图	11
图表 7 《WHISPERS FROM THE STAR》游戏	11
图表 8 自然选择《EVE》游戏场景	11
图表 9 AI 陪伴产品《EVE》主要功能	12
图表 10 上海电影 X 即梦 AI“未来影像计划·AI 动画创作周”-1	12
图表 11 上海电影 X 即梦 AI“未来影像计划·AI 动画创作周”-2	13
图表 12 天猫 AI 联动明星、企业共创年画	13
图表 13 康师傅 AI 写春联	13
图表 14 可口可乐春节数字人新年祝福	14
图表 15 伊利《千年江南》短片	14
图表 16 二手玫瑰、京东家电《我要开花 AIGC 版》	14
图表 17 《我要开花 AIGC 版》 workflow 节选	14
图表 18 武汉警官职业学院新校区 (AI 辅助生成图)	15
图表 19 鄂州花湖机场 BIM 设计图	15
图表 20 LOVART 设计一款包	15
图表 21 LOVART 设计一款宠物食品品牌	15
图表 22 RUNWAY GEN 系列模型迭代情况、主要特点及应用场景	16
图表 23 GEN-1 模型应用示例：使用几个包装盒视频，生成工厂视频	17
图表 24 GEN-2 模型应用示例：文本生成视频 (TEXT TO VIDEO)	17
图表 25 GEN-2 模型应用示例：图片生成视频 (IMAGE TO VIDEO)	17
图表 26 GEN-2 模型应用示例：文本+图片生成视频 (TEXT + IMAGE TO VIDEO)	18
图表 27 GEN-2 模型应用示例：故事板 (STORYBOARD)	18
图表 28 GEN-2 模型应用示例：渲染 (RENDER)	18
图表 29 GEN-3 ALPHA 模型应用示例：精细的时间控制	19
图表 30 GEN-3 ALPHA 模型应用示例：逼真的人类角色	19
图表 31 GEN-3 ALPHA 模型应用示例：艺术家创作	19
图表 32 GEN-3 ALPHA 模型应用示例：行业定制	20
图表 33 GEN-4 模型应用示例：《THE LONELY LITTLE FLAME》	20
图表 34 GEN-4 模型应用示例：《NEW YORK IS A ZOO》	20
图表 35 GEN-4 模型应用示例：《THE HERD》	21
图表 36 RUNWAY GEN-4 自制短片故事内容及技术	21
图表 37 GEN-4 TURBO 模型应用示例：	21
图表 38 RUNWAY GEN 系列模型核心技术	22
图表 39 RUNWAY C 端用户月度订阅价格	22
图表 40 可灵 AI WEB 端界面	23
图表 41 可灵 AI APP 端界面	23

图表 42 可灵 AI 版本迭代情况	24
图表 43 可灵 AI 运营数据	25
图表 44 “2025 年 AI 视频生成企业 TOP20” 榜单	25
图表 45 VIT 架构	26
图表 46 256x256 图像下架构评测	26
图表 47 可灵会员价格	27
图表 48 可灵视频生成资源包	27
图表 49 可灵图片生成资源包	27
图表 50 可灵 AI WEB 端全球访问量	28
图表 51 可灵 AI APP 端国内下载量	28
图表 52 CONCEPTMASTER 多主体生成示例	29
图表 53 文生视频模型胜负比	30
图表 54 图生视频模型胜负比	30
图表 55 即梦 AI 模型迭代及核心功能突破情况	31
图表 56 即梦 AI 核心技术	31
图表 57 即梦 3.0 对齐比较 (ALIGNMENT COMPARISON)	32
图表 58 即梦 3.0 结构比较 (STRUCTURE COMPARISON)	33
图表 59 即梦 3.0 美学比较 (AESTHETIC COMPARISON)	33
图表 60 即梦 3.0 设计比较 (DESIGN COMPARISON)	33
图表 61 即梦 3.0 文本渲染比较 (TEXT RENDERING)	34
图表 62 即梦 3.0 超现实主义肖像比较 (PHOTOREALISTIC PORTRAIT)	34
图表 63 即梦 3.0 超现实主义肖像效果图	35
图表 64 即梦 AI 用户月度订阅价格及积分情况	35
图表 65 MIDJOURNEY 模型版本迭代、优化情况及效果图比较	36
图表 66 MIDJOURNEY V7 模型核心技术亮点	37
图表 67 MIDJOURNEY 应用场景：绘画创作	38
图表 68 MIDJOURNEY 应用场景：家居设计和装修	38
图表 69 MIDJOURNEY 应用场景：电商海报宣传	38
图表 70 MIDJOURNEY 应用场景：LOGO 和徽章设计	39
图表 71 MIDJOURNEY 应用场景：插画、漫画和油画	39
图表 72 MIDJOURNEY 产品订阅价格及说明	40
图表 73 STABLE DIFFUSION 版本迭代与变体情况及性能介绍	40
图表 74 STABLE DIFFUSION 核心功能及技术实现	41
图表 75 STABLE DIFFUSION 用户类型及核心需求	42
图表 76 STABLE DIFFUSION 工作流程及 U-NET 网络+SCHEDULE 算法迭代去噪过程	42
图表 77 STABLE DIFFUSION 包含模型及核心技术	43
图表 78 STABLE DIFFUSION 应用场景：电商运营	43
图表 79 STABLE DIFFUSION 应用场景：设计师/插画师	44
图表 80 STABLE DIFFUSION 应用场景：游戏开发	44
图表 81 STABLE DIFFUSION 应用场景：AI 研究	44
图表 82 STABLE DIFFUSION B 端及 C 端产品订阅价格及说明	45
图表 83 豆包大模型更新迭代情况	46
图表 84 豆包产品功能更新情况	47
图表 85 豆包大模型家族九款模型	48

图表 86 豆包九款模型主要功能及应用场景简介	48
图表 87 豆包视觉理解模型收费价格	49
图表 88 视觉理解 TOKEN 用量说明	49
图表 89 豆包视频生成模型收费价格	49
图表 90 豆包图片生成模型收费价格	49
图表 91 SEAART AI 核心技术优势	50
图表 92 SEAART AI 与亚马逊云科技、阿里云、华为云技术合作情况	51
图表 93 海艺互娱基于亚马逊云科技服务于海外用户的技术架构示意图	51
图表 94 SEAART AI 落地页功能界面：免费 AI 艺术生成器与创意社区、广场	52
图表 95 COMFYUI 核心功能介绍	52
图表 96 SEAART AI 游戏内容生成效果图：像素游戏场景生成器、猫狗斗争游戏场景	53
图表 97 SEAART AI 产品订阅价格及说明	53
图表 98 LIBLIBAI 2023 年至今融资情况	54
图表 99 LIBLIBAI 核心功能介绍	54
图表 100 STAR-3 ALPHA 大模型与 MIDJOURNEYV6.1、DALLE-3、FLUX.1DEV 评测比较	55
图表 101 STAR-3 ALPHA 大模型功能特性	55
图表 102 STAR-3 ALPHA 应用场景：电商与产品展示	56
图表 103 STAR-3 ALPHA 应用场景：广告与市场推广	56
图表 104 STAR-3 ALPHA 应用场景：插画与艺术创作	57
图表 105 STAR-3 ALPHA 应用场景：室内设计与建筑可视化	57
图表 106 STAR-3 ALPHA 应用场景：社交媒体内容创作	57
图表 107 LOVART 一站式调用 GPT IMAGE-1 等模型	58
图表 108 LOVART 智能任务分解流程	58
图表 109 LOVART 设计 PUPU 品牌包并生成海报	59
图表 110 LOVART 生成宠物零食广告图	59
图表 111 LIBLIBAI B 端及 C 端产品订阅价格及说明	60
图表 112 ROBONEO 网页版首页	60
图表 113 ROBONEO 产品交互品牌设计——“设计一个科技感 LOGO，主色为蓝色，搭配简约线条”分析过程	61
图表 114 ROBONEO 产品交互品牌设计——输出成果图	62
图表 115 SORA 生成视频出现椅子漂浮情况	63
图表 116 《海上女王郑一嫂》剧照	63
图表 117 不同 AI 技术适用场景	64
图表 118 生成式 AI 在游戏开发过程中应用架构图	65

# 1 AI 视频及图像生成模型未来展望

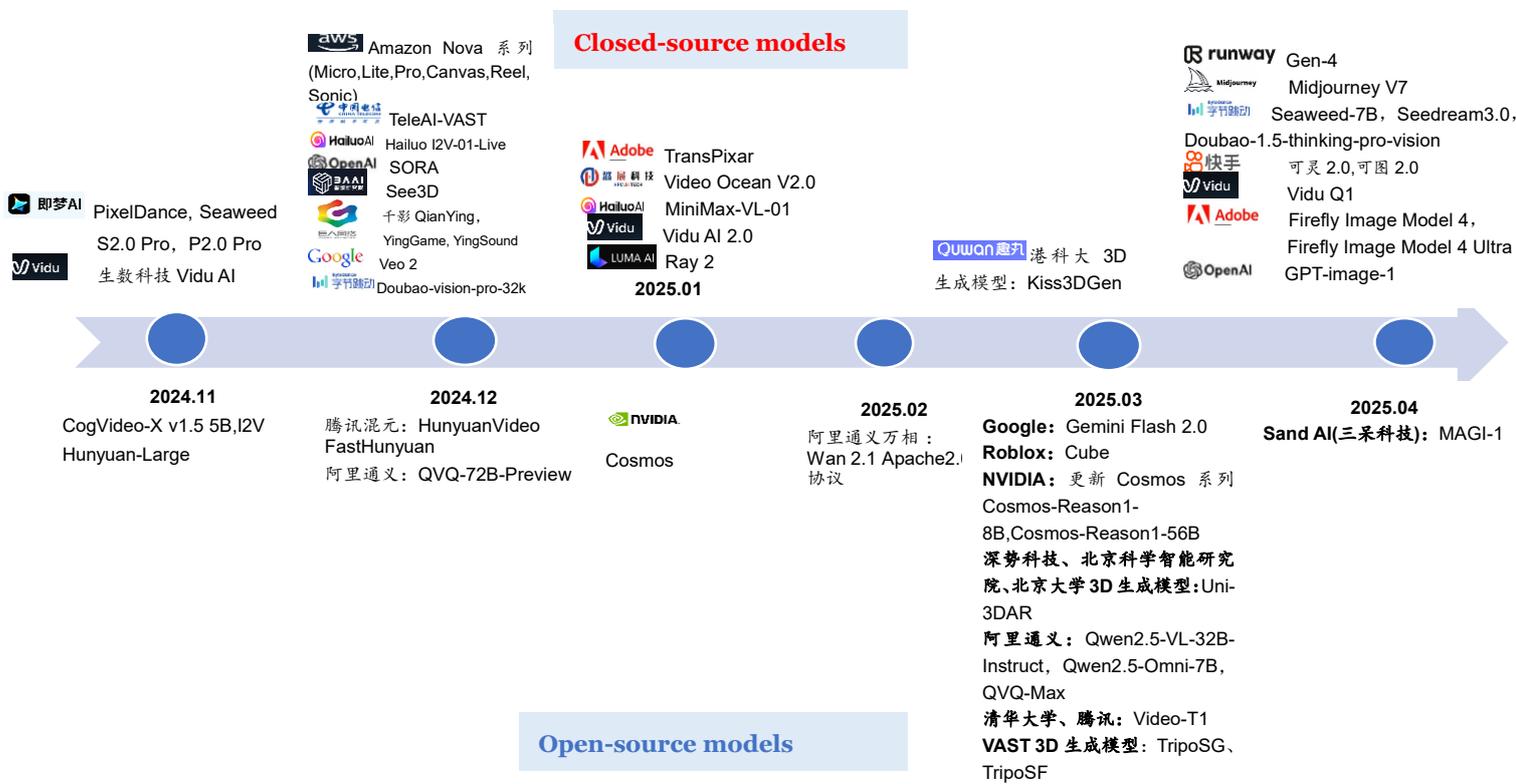
## 1.1 AI 视频图像模型技术发展趋势

2024 年 11 月至今，国内外 AI 视频图像模型不断迭代。按照模型源代码及内部细节是否公开，可分为闭源模型和开源模型。目前海外头部 AI 大模型厂商已发布多款新版本视频及图像生成模型，包括 OpenAI GPT-image-1、Runway Gen-4、Midjourney V7、Google Veo 2、Adobe TransPillar 等，国内快手发布可灵 2.1 及可图 2.0，字节跳动 Seaweed-7B 和 Seedream3.0、阿里通义 QVQ-72B-Preview、腾讯混元 HunyuanVideo 和 FastHunyuan 等。

未来 AI 视频图像模型技术发展将聚焦于 3D 生成模型、多模态融合、参数性能提升，例如 Roblox Cube 3D 模型、深势科技 Uni-3DAR、趣丸科技 Kiss3DGen、VAST TripoSG 和 TripoSF 等聚焦 3D 生成，推动影视游戏内容生成、场景设计及 3D 建模自动化；Owen2.5-Omni-7B、Gemini Flash 2.0 等模型强调多模态融合，文本、视频、图像联合训练，支持更灵活的跨模态生成；腾讯混元 HunyuanVideo、阿里通义 QVQ-72B-Preview 模型参数规模迅速扩大，计算能力及复杂场景处理性能不断提升。

开源与闭源生态并行：海外及国内 AI 巨头（如 Google、Adobe、OpenAI、字节跳动）闭源模型竞争日趋激烈，加速渗透大型商业化市场如广告、影视制作等；开源模型降低中小开发者门槛，在游戏、短视频等交互场景中推动技术普惠化。

图表 1 2024 年 11 月至 2025 年 4 月 AI 视频图像模型发展图



资料来源：阿里官网，字节跳动官网，谷歌官网，英伟达，OpenAI，Adobe，Runway，快手官网等，华安证券研究所

3D 生成模型未来或成为重要竞争领域，推动影视、游戏内容生成、科学研究领

**域 3D 建模自动化。**今年以来，在传统 AI 视频生成模型版本技术更迭的同时，众多 AI 技术厂商开始发布基于 3D 对象的生成模型：

1) 3 月 18 日，Roblox 发布开源 3D 生成模型 Cube，主要用于为游戏开发者及时构建 3D 对象和场景，在开源 Cube 3D 中，可以直接从文本和未来的图像输入生成 3D 模型和环境，生成的对象与当今游戏引擎完全兼容，核心技术突破在于 3D 标记化；2) 3 月 25 日，深势科技、北京科学智能研究院及北京大学发布 3D 生成模型 Uni-3DAR，未来可应用于构建通用科学智能体，核心技术是统一微观与宏观 3D 结构、统一生成和理解的自回归框架，后期可以引入多模态信息并结合大语言模型与科学文献知识，为构建通用科学智能体提供理论基础；3) 3 月 26 日，香港科技大学（广州）与趣丸科技推出 3D 生成框架 Kiss3DGen，可应用于游戏开发和影视制作中角色及场景设计、VR 和 AR 内容创作等，核心技术包括重新利用 2D 扩散模型、3D Bundle Image、法线图增强等；4) 3 月 28 日，VAST 发布开源 3D 生成模型 TripoSG 和 TripoSF，TripoSG 用于基础 3D 生成，TripoSF 专注于高分辨率三维重建和生成任务，未来应用场景包括影视与游戏创作、3D 打印与制造业等，核心技术是全新 3D 表示法 SparseFlex、视锥体感知训练策略及动态拓扑支持与开放生态。

图表 2 近期 3D 生成模型及核心技术介绍

发布时间	发布主体	模型名称	核心技术
3 月 18 日	Roblox	Cube 3D	<b>3D 标记化：</b> 将文本和图像与 3D 形状联系起来，允许将 3D 对象表示为标记，使用自回归 transformers 架构提供可扩展性和多模态兼容性，初始阶段创作者能够根据文本提示生成 3D 对象，未来 Roblox 计划让创作者基于多模态输入整个场景。
3 月 25 日	深势科技、北京科学智能研究院、北京大学	Uni-3DAR	<b>1) 统一微观与宏观 3D 结构 (Compressed Spatial Tokens)：</b> 层次化、由粗到细的 token 化方法，将 3D 结构转化为一维 token 序列，实现数据的高效压缩和统一表示，既适用于微观也适用于宏观 3D 结构建模； <b>2) 统一生成和理解的自回归框架 (Masked Next-Token Prediction)：</b> 构建统一的自回归框架使 3D 结构生成和理解任务能够在单一模型内同时进行，使不同任务对应 token 在模型内部彼此独立、互不干扰，支持联合训练，自回归特性便于将其他模态数据统一到单个模型。
3 月 26 日	香港科技大学（广州）、趣丸科技	Kiss3DGen	<b>1) 重新利用 2D 扩散模型：</b> 将 3D 生成问题转化为 2D 图像生成问题，通过微调预训练的 2D 扩散模型，基于强大的 2D 生成能力避免从头训练复杂的 3D 生成模型； <b>2) 3D Bundle Image：</b> 由多视图图像及其对应的法线图组成拼贴表示，从 2D 图像中提取足够的信息重建完整的 3D 模型； <b>3) 法线图增强：</b> 通过捕捉物体表面的方向信息，帮助模型重建更精确的 3D 几何形状。
3 月 28 日	VAST	TripoSG、TripoSF	<b>1) 全新 3D 表示法 SparseFlex：</b> 通过稀疏体素结构优化内存占用，支持高达 1024 <sup>3</sup> 分辨率的建模精度，仅需在物体表面附近存储体素信息，既能捕捉外部纹理（如布料的单面结构），又能生成内部细节（如座椅、驾驶舱机械构造）； <b>2) 视锥体感知训练策略：</b> 借鉴实时渲染中的视锥体剔除技术，在训练时仅激活相机视角内的体素，显著降低计算开销； <b>3) 动态拓扑支持与开放生态：</b> 支持任意拓扑结构，可处理开放表面（如布料、叶片）和封闭实体，并兼容基于渲染损失的直接优化流程。

资料来源：Roblox, Uni-3DAR, ComfyUI Wiki, Kiss3DGen, AI 工具集, 华安证券研究所

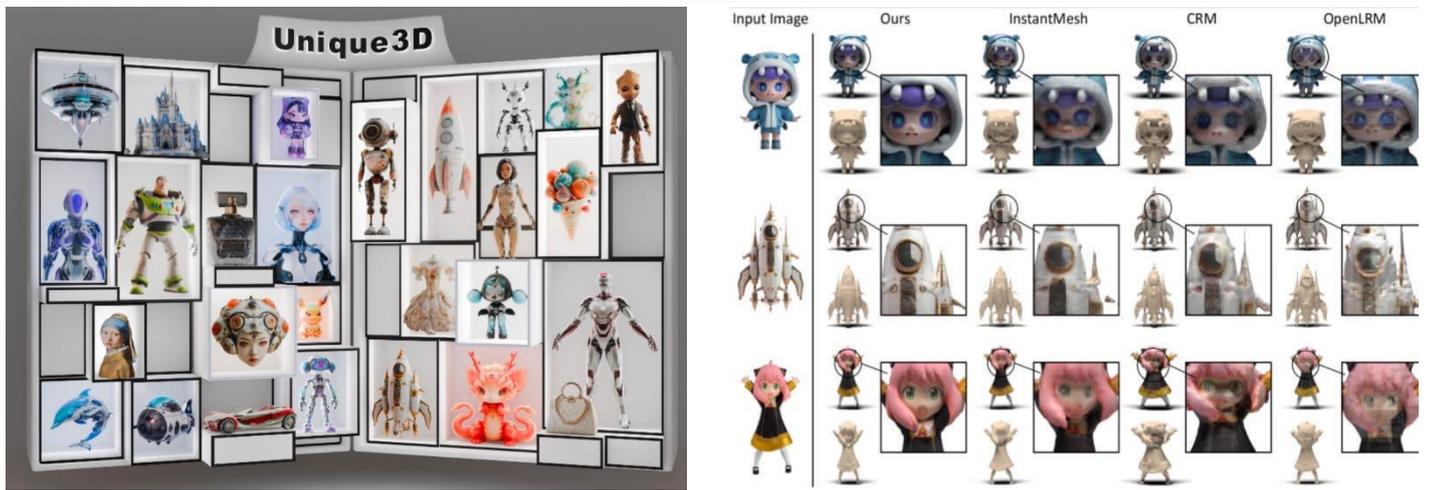
**3D 生成模型应用雏形初现，游戏与 3D 打印领域将成为未来焦点。**1 月 21 日腾讯混元宣布开源 3D 生成大模型 2.0 版本，支持文字、图像生成 3D 资产，几何结构更加精细，纹理色彩更加丰富，目前 3D 生成大模型已应用于腾讯游戏业务中；VAST 与拓竹、纵维立方等 3D 打印行业领先厂商合作，引入 AI 自动建模功能，大幅提高生成效率并降低人工建模成本，通过 AI 重构 3D 创作流程实现技术突破，应用领域覆盖消费级市场和工业制造，在手办潮玩、课程教育、定制化设计等细分领域应用雏形初现。

图表 3 Cube 3D 生成 3D 模型效果图



资料来源: Roblox, 华安证券研究所

图表 4 Uni-3DAR 生成 3D 模型效果图



资料来源: Uni-3DAR, 华安证券研究所

图表 5 Kiss3DGen 生成 3D 模型效果图



资料来源: Kiss3DGen, 华安证券研究所

图表 6 TripoSG 和 TripoSF 生成 3D 模型效果图



资料来源：VAST，华安证券研究所

## 1.2 AI 视频图像模型未来应用场景展望

在游戏领域，AI 陪伴为目前 AI 模型主流应用场景。1) 3 月 15 日，米哈游创始人蔡浩宇新公司 Anuttacon 发布新款 AI 游戏《Whispers from the Star》，同时针对美国地区玩家 iOS 平台开启封闭测试招募。游戏核心机制是通过实时对话推动剧情发展，玩家任务是帮助主角 Stella 在外星星球生存并找到回家的路，在此过程中 Stella 对话由实时生成，根据玩家输入 Stella 反应、情绪和动作会发生变化，玩家也可以通过视频、语音和文字多模态输入与 Stella 互动，每次选择都可能影响故事走向和 Stella 命运，实现玩家自定义的开放剧情，更加具备沉浸式游戏体验。2) 恺英网络投资的自然选择公司推出首款 3DAI 伴侣游戏产品《EVE》，内容层面通过游戏级别的剧情设计提供丰富的交互体验，是基于 AI 技术的虚拟恋爱陪伴产品；技术层面通过自研 AI 对话模型 Vibe 深度理解用户输入，专为情感陪伴设计并满足用户在情感交流中的需求，自研 AI 记忆模型 Echo 记录用户细微需求和过往互动细节，为用户提供更个性化服务和深度互动。恺英网络作为《EVE》产品的投资方，未来将在 AI 大模型和 AI 交互应用两大板块持续进行研发投入，持续深化 AI 技术在游戏行业应用。

图表 7 《Whispers from the Star》游戏



资料来源：Whispers from the Star 官网，华安证券研究所

图表 8 自然选择《EVE》游戏场景



资料来源：AIHub，华安证券研究所

图表 9 AI 陪伴产品《EVE》主要功能

主要功能	功能介绍
AI 对话模型 (Vibe)	能够深度理解用户的输入, 提供如真人般的自然交流体验, 专为情感陪伴设计, 旨在满足用户在情感交流中的需求, 给用户带来温暖的陪伴。
AI 记忆模型 (Echo)	是一个强大的记忆系统, 能够记录用户的细微需求和过往互动细节, 通过这种记忆功能, EVE 能够为用户提供更个性化的服务和更深度的互动。
3D 互动与角色定制	允许用户定制专属的 3D 虚拟形象, 用户可以像“捏崽”一样为 AI 伴侣创建独一无二的外观和特征, 增强个性化体验, “定制+高度互动”的 3D 人物动画使得用户更容易与 EVE 建立深度的虚拟情感连接。
剧情与实时推荐系统	结合了实时内容推荐系统和游戏级别的剧情设计, 提供丰富的交互体验, 用户可以参与到精心设计的剧情中, 体验由 AI 推动的故事进展, 使得陪伴感更加生动有趣。

资料来源: AIHub, 华安证券研究所

**“AI+影视动画”合作成果初现, AI 技术赋能传统经典 IP。**“未来影像计划·AI 动画创作周”由上海电影和即梦 AI 于 3 月共同发起, 设立三大评选赛道: 1) 以《大闹天宫》为代表的中国经典角色形象; 2) 《宝莲灯》为代表的中国经典动画故事; 3) 《鹿铃》、《山水情》为代表的中式美学风格。上海电影与即梦 AI 未来将以“科技赋能内容, 创新驱动未来”为核心, 从 AI 动漫短剧共创、AI 青年创客培育、AI 内容城市展厅三个维度逐步落地双方的战略合作。上海电影与即梦 AI 将继续深度探索“AI+影视动画”, 推进 AI 技术在影视内容生成、IP 运营、宣发模式等全链条的创新应用。

图表 10 上海电影 X 即梦 AI“未来影像计划·AI 动画创作周”-1



资料来源: 上影元, 华安证券研究所

图表 11 上海电影 X 即梦 AI“未来影像计划·AI 动画创作周”-2



资料来源：上影元，华安证券研究所

**AI 图像及视频生成持续赋能广告营销领域，创意营销引领潮流。** AI 视频及图像生成模型已应用于广告营销行业，2024 年龙年春节各大厂商抓住春节契机，通过 AI 图像及视频生成赋能传统经典 IP，如生肖、春联等，广告营销创意不断更新，从 AI 生成平面海报到多媒体互动，对消费者实现精准营销甚至用户共创。1) 天猫联动 20 多位明星与多个热门 IP，发起 AI 共创年画活动，用户通过 AI 互动在明星或 IP 制作的年画添加自己的一笔，创作出带有个人创意的年画，“明星联动+个性定制”的互动模式吸引大量粉丝参与，持续助力品牌营销；2) 康师傅 AI 写春联并引入定制数字人形象，用户可以生成带有个人形象的海报；3) 可口可乐春节互动营销，引入 AI 视频生成技术，用户不仅可以生成个人数字形象，还能选择个性化语音生成动画视频，说出新年祝福；4) 伊利短片《千年江南》，通过 AI 视频生成技术实现场景快速切换，特效丝滑且有质感；5) 二手玫瑰与京东家电合作《我要开花 AIGC 版》，使用 AI 视频生成技术制作明星数字形象，配合独特的主题视觉以及辨识度极高的音乐。

图表 12 天猫 AI 联动明星、企业共创年画

图表 13 康师傅 AI 写春联



资料来源：数英 DIGITALING，华安证券研究所

资料来源：数英 DIGITALING，华安证券研究所

图表 14 可口可乐春节数字人新年祝福



资料来源：数英 DIGITALING，华安证券研究所

图表 15 伊利《千年江南》短片



资料来源：数英 DIGITALING，华安证券研究所

图表 16 二手玫瑰、京东家电《我要开花 AIGC 版》



资料来源：数英 DIGITALING，华安证券研究所

图表 17 《我要开花 AIGC 版》 workflow 节选



资料来源：数英 DIGITALING，华安证券研究所

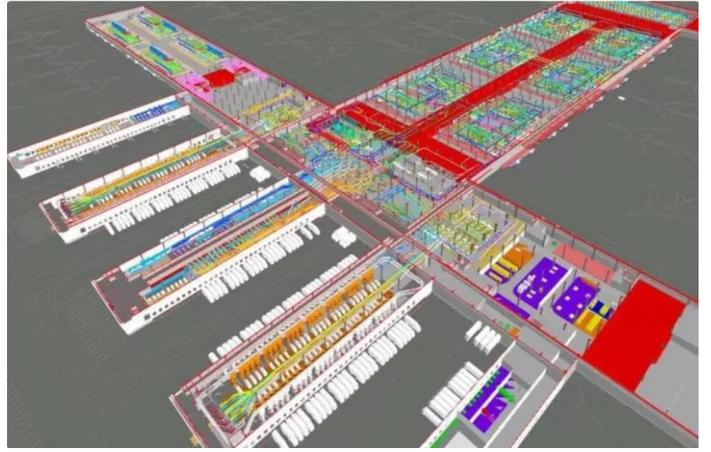
“AI+设计”探索新质生产力，AI 建筑设计、品牌包装设计方兴未艾。去年以来，国内建筑设计研究院已将 AI 辅助设计软件应用于建筑设计领域，包括校园新校区设计、建筑外观设计等；2025 年 5 月，LiblibAI 发布全球首个设计类 AI 智能体（AI Agent）Lovart，创始人陈冕曾任字节跳动剪辑全球商业化负责人；聚焦品牌设计，从 LOGO、海报到品牌 VI 系统，Lovart 支持一站式生成；根据用户提供产品图和创意方向完成广告与视频制作；支持文创与个性化内容创作。1) 中南建筑设计院设计师通过“Giant AI”软件（建筑创意具现软件）勾画，使得需要设计师数日渲染修改的复杂建筑场景图在 4 小时内即可辅助设计人员实现设计快速成型，目前 Giant AI 辅助设计已应用于武汉警官职业学院新校区、天门石家河遗址博物馆、陕西省政务和公安大数据中心等省内外多个项目；鄂州花湖机场是我国首个采用 BIM 模型搭建、深度应用数字化建设的机场，机场运用“数字孪生”理念，通过一套模型实现了项目的全生命周期管理，实现设计的“图模一致”到建造的“物模一致”；2) Lovart 支持从创意拆解到专业交付的全链路设计，给出具体英文提示即可设计出一款前卫、时尚的包包；同理给出具体风格要求及建议即可设计一款宠物食品品牌，如名称为“Billy”、卷毛贵宾犬、充满趣味的线描插画、手绘涂鸦风格并保留笔触感。

图表 18 武汉警官职业学院新校区 (AI 辅助生成图)



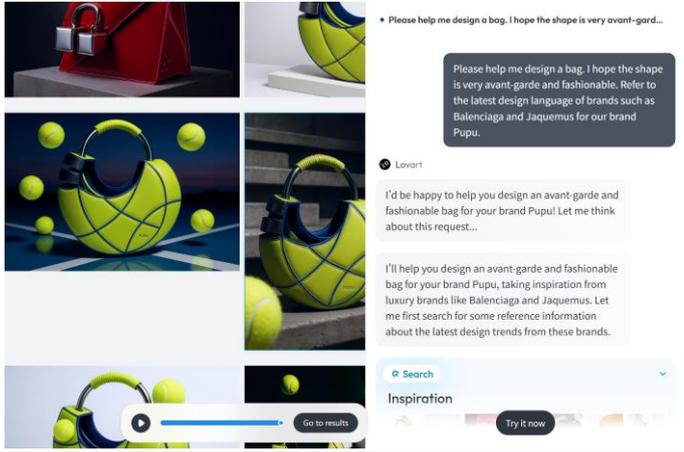
资料来源：极目新闻，华安证券研究所

图表 19 鄂州花湖机场 BIM 设计图



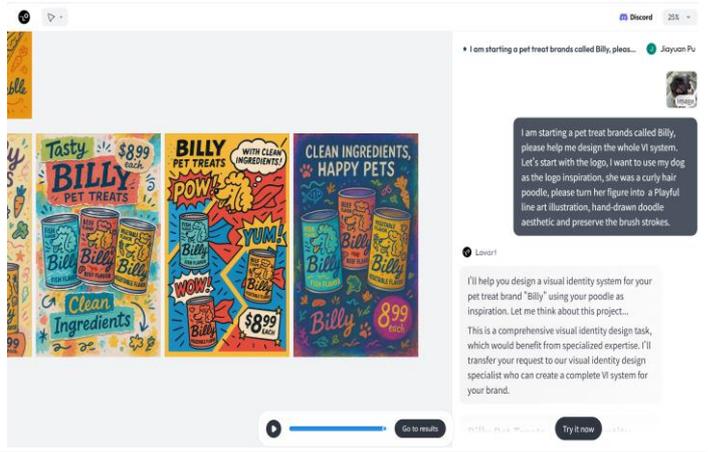
资料来源：极目新闻，华安证券研究所

图表 20 Lovart 设计一款包



资料来源：Lovart 官网，华安证券研究所

图表 21 Lovart 设计一款宠物食品品牌



资料来源：Lovart 官网，华安证券研究所

## 2 Runway

### 2.1 Gen 系列模型迭代更新，内容生成技术不断精进

2023 年 2 月，Runway 发布首个 AI 编辑模型 Gen-1，可以在原视频的基础上编辑出想要的视频，使用扩散模型的结构和内容来引导视频合成 (Video to Video)，依赖原始视频素材；2023 年 3 月，Runway 提出首个多模态视频 workflow 模型 Gen-2，区别于 Gen-1 的关键点在于 Gen-2 使用文本驱动视频生成 (Text to Video)，直接由文字生成视频，摆脱原始视频素材依赖；2024 年 6 月，Runway 发布 Gen-3 Alpha，通过大规模多模态训练基础设施，显著提升视频保真度、一致性和动态表现；2025 年 4 月，Runway 发布 Gen-4 和 Gen-4 Turbo 模型，全新 AI 视频模型能够更好地维护连贯的现实环境，并在解读用户提示方面表现更佳。

图表 22 Runway Gen 系列模型迭代情况、主要特点及应用场景

发布时间	模型名称	主要特点	应用场景
2023 年 2 月	Gen-1	应用文本 prompt 或参考图像指定的任何风格， <b>将现有视频转化为新视频</b> ，是一种结构可控和内容感知引导的视频扩散模型，能够对已有视频进行风格转换、内容编辑等操作。	使用几个包装盒视频，根据视频形状特点，可以生成一个工厂的视频。
2023 年 3 月	Gen-2	从头开始直接生成视频的技术突破，在原有技术的基础上增加了 <b>文本生成视频、图片生成视频以及文字和图片共同生成视频</b> 的新功能。	派拉蒙和迪士尼等顶级电影制片公司，可用于改进特效制作和多语言配音等方面，如为静态图像添加动态元素，实现口型同步等。
2024 年 6 月	Gen-3 Alpha	<p><b>高保真视频生成</b>：能够生成长达 10 秒的视频片段，视觉层面高度清晰细节，动态表现逼真；</p> <p><b>多模态输入支持</b>：支持多种输入方式，文本描述、静态图像或现有视频转化为视频内容；</p> <p><b>精细时间控制</b>：通过训练学习大量具有时间序列信息的描述性字幕，使其能够精确控制视频中的时间流动和关键帧设置；</p> <p><b>逼真人物角色生成</b>：能够生成具有丰富表情、动作和情感反应的人类角色；</p> <p><b>高级控制模式</b>：提供多种高级控制工具，<b>运动画笔 (Motion Brush)</b> 允许用户手动调整视频中运动轨迹；<b>高级相机控制 (Advanced Camera Control)</b> 提供对摄像机视角和运动的精细调节；</p> <p><b>全新视觉审核系统</b>：系统遵循 C2PA (Content Credentials) 标准，对视频内容自动审核。</p>	<p><b>电影和电视制作</b>：前期制作中可以帮助导演和制片人预览场景和故事情节；</p> <p><b>广告创意</b>：广告行业可以快速制作吸引人的广告视频；</p> <p><b>游戏开发</b>：游戏设计中生成游戏内的动画和过场动画；</p> <p><b>社交媒体内容创作</b>：内容创作者可以生成独特视频内容，吸引关注；</p> <p><b>企业宣传</b>：企业可以制作高质量宣传视频，展示产品特点、企业文化或服务优势；</p> <p><b>艺术创作</b>：艺术家和设计师可以探索新的艺术表现形式，创作出独特的视觉作品。</p>
2025 年 4 月	Gen-4	<p><b>世界一致性</b>：跨场景保持人物、物体和环境的一致性，无需额外训练；</p> <p><b>参考图能力</b>：仅需要一张参考图生成一致的角色或物体，适应不同光线和场景；</p> <p><b>场景覆盖</b>：提供参考图和描述后，支持从任意角度重建场景，满足多视角需求；</p> <p><b>生成式视觉特效 (GVFX)</b>：快速生成高质量特效，与实拍和传统特效无缝融合。</p>	<p><b>影视制作</b>：能够生成运动轨迹极其逼真、充满动态感的视频内容，为影视制作带来了全新的可能性； 例如：Runway 团队用 Gen-4 制作了短片《The Lonely Little Flame》、《New York is a Zoo》和《The Herd》</p> <p><b>游戏开发</b>：高度一致性和动态运动捕捉能力，能够快速生成高质量的和视频内容并缩短周期；</p> <p><b>广告宣传</b>：能够快速生成符合用户需求的广告视频；</p> <p><b>社交媒体内容制作</b>：帮助创作者快速生成各种风格的短视频。</p>
2025 年 4 月	Gen-4 Turbo	<p><b>更快的视频生成</b>：运行速度比前代模型最高提升 3 倍，30 秒即可生成 10 秒的视频，同时保持输出的动态和保真度；</p> <p><b>提高成本效益</b>：每秒视频仅消耗 5 个积分，而 Gen-4 每秒视频最多消耗 12 个积分，积分成本远低于前代产品，满足预算有限客户需求。</p>	<p><b>独立创作者</b>：快速生成短片素材；</p> <p><b>大型制片公司</b>：高效验证视觉创意。</p>

资料来源：Runway, AIGC 视频，华安证券研究所整理

## 2.2 Gen 系列模型应用场景案例及核心技术

Runway Gen-1 模型核心在于需要依赖提供的文本或参考图像，指定任何风格后将现有视频转化为新视频。技术端 Gen-1 提出了可控的结构和内容感知的视频扩散模型，通过大规模无字幕视频和配对文本-图像数据上进行训练。可以将输入视频转换为与用户提供的图像或文本提示相匹配的效果，同时灵活控制模型对视频结构的控制力度，通过自定义的指导方法来调整推理过程，以控制生成视频与原视频的

时间一致性。以下可以使用几个包装盒的参考视频，指定 Gen-1 生成视频：

图表 23 Gen-1 模型应用示例：使用几个包装盒视频，生成工厂视频



资料来源：AIGC 视频，华安证券研究所

Runway Gen-2 模型技术端通过结合深度学习与生成对抗网络 (GAN)，在原有 Gen-1 基础上增加多项功能。以下应用场景包括文本生成视频 (Text to Video)、图片生成视频 (Image to Video)、文本+图片生成视频 (Text + Image to Video)、故事板 (Storyboard)、渲染 (Render) 等。

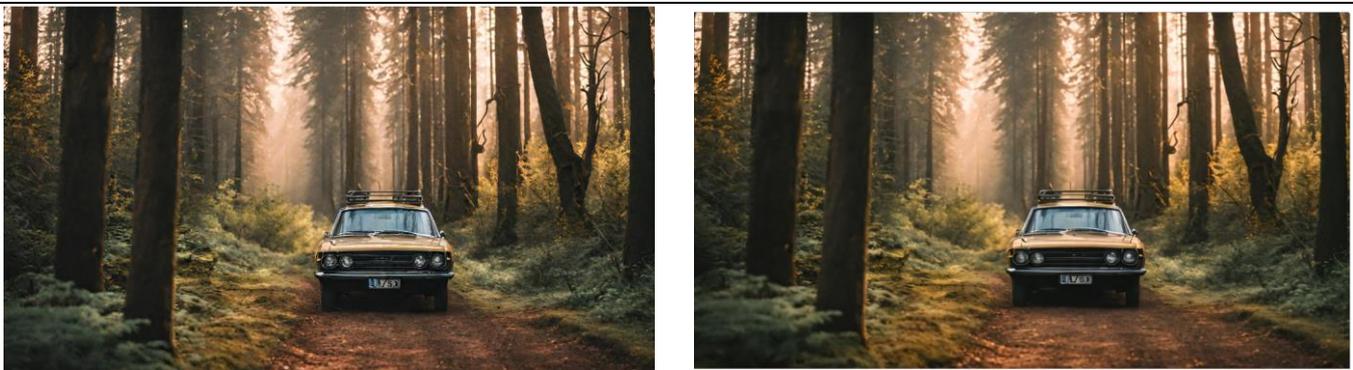
图表 24 Gen-2 模型应用示例：文本生成视频 (Text to Video)



提示：傍晚的阳光从纽约市阁楼的窗户里探出头来

资料来源：Runway 官网，华安证券研究所

图表 25 Gen-2 模型应用示例：图片生成视频 (Image to Video)



提示图片：使用驾驶图像生成视频 (变体模式)

输出视频

资料来源：Runway 官网，华安证券研究所

图表 26 Gen-2 模型应用示例：文本+图片生成视频（Text + Image to Video）



提示文字：一个男人走在街上的低角度镜头，被周围酒吧的霓虹灯照亮

输出视频

资料来源：Runway 官网，华安证券研究所

图表 27 Gen-2 模型应用示例：故事板（Storyboard）

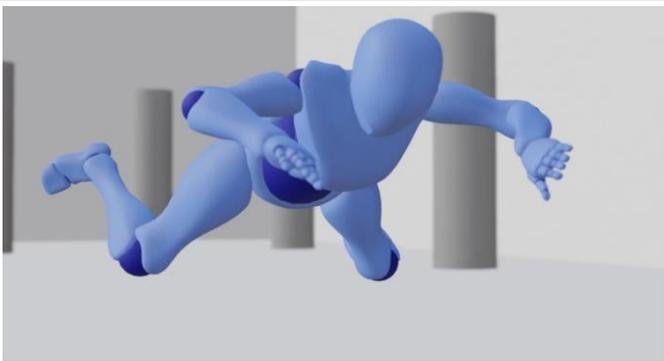


输入视频：转换为完全风格化和动画化的渲染

输出视频

资料来源：Runway 官网，华安证券研究所

图表 28 Gen-2 模型应用示例：渲染（Render）



输入视频：应用输入图像或提示，将无纹理的渲染转换为逼真的输出

输出视频

资料来源：Runway 官网，华安证券研究所

Runway Gen-3 Alpha 模型技术端在全新基础设施上进行大规模多模态训练，保真度、一致性和运动表现上有重大提升，同时推出多种高级控制工具并带有保护措施，包

括内部视觉审核系统和 C2PA 溯源标准。以下应用场景包括精细的时间控制、逼真的人类角色、艺术家创作、行业定制等。通过高描述性、时间密集型的训练，可实现富有想象力的过渡和场景中元素的精确关键帧。

图表 29 Gen-3 Alpha 模型应用示例：精细的时间控制



提示：一只蚂蚁从巢中出来的极端特写镜头，镜头拉回露出山外一个社区

资料来源：Runway 官网，华安证券研究所

生成具有各种动作、手势和情感的富有表现力的人类角色，解锁新故事讲述。

图表 30 Gen-3 Alpha 模型应用示例：逼真的人类角色



提示：一个男人的电影宽幅肖像，他的脸被电视的光芒照亮

资料来源：Runway 官网，华安证券研究所

科学家、工程师和艺术家组成的跨学科团队的协作努力，解释各种风格和电影术语。

图表 31 Gen-3 Alpha 模型应用示例：艺术家创作

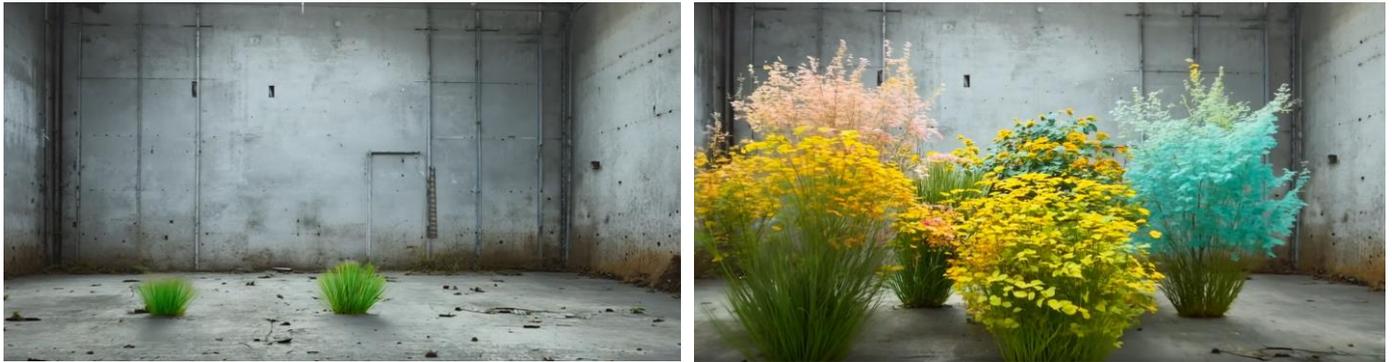


提示：在霓虹灯亮起的森林中缩放

资料来源：Runway 官网，华安证券研究所

领先的娱乐和媒体组织合作创建定制版本，对角色进行更风格控制和更一致的角色。

图表 32 Gen-3 Alpha 模型应用示例：行业定制



提示：一个空仓库，鲜花开始从混凝土中绽放

资料来源：Runway 官网，华安证券研究所

Runway Gen-4 模型技术核心在于世界一致性 (World Consistency) ——能够在多个场景中精准生成人物、场景和物体，并保持其视觉特征的一致性。同时引入生成式视觉特效 (GVFX) 技术，能够通过 AI 驱动的生成能力，大幅缩短视觉特效制作中的建模、渲染和后期调整过程。以下应用场景包括 Runway 使用 Gen-4 模型自制短片《The Lonely Little Flame》、《New York is a Zoo》和《The Herd》。

图表 33 Gen-4 模型应用示例：《The Lonely Little Flame》



资料来源：Runway 官网，华安证券研究所整理

图表 34 Gen-4 模型应用示例：《New York is a Zoo》



资料来源：Runway 官网，华安证券研究所整理

图表 35 Gen-4 模型应用示例：《The Herd》



资料来源：Runway 官网，华安证券研究所整理

图表 36 Runway Gen-4 自制短片故事内容及技术

短片名称	主要故事内容及技术
《The Lonely Little Flame》	讲述了主角小火焰渴望寻找朋友最终得偿所愿的故事。结合视觉参考和指令，支持创建具有连贯的风格、主题、场景等新图像和视频，在故事中实现连贯性和控制力。
《New York is a Zoo》	将纽约变成了一个大型动物园。影片中的每个场景均通过 Gen-4 Reference 功能，将动物的真实照片与纽约街景照片相结合，用提示词描述每个场景的具体动作而创建，体现了 Gen-4 惊人的生成式视觉特效（GVFX）技术。
《The Herd》	讲述了一个年轻人在夜间被牛群追赶的故事。画面仅通过简单的参考图和提示词就能得到，通过 Act-One 技术，这些镜头被串联成完整故事。
《Scimmia Vede》	呈现一段猴子作为跳舞演员的音乐舞蹈。在动物身上也能实现不同场景中保持一致性，无论猴子出现在任何场景，Gen-4 均能实现不同角度、动作、表情一致性。
《The Retrieval》	讲述几位主角不断探险，寻找神秘之花的整个过程。动画风格的短片，Gen-4 成为故事创作的全新方式。

资料来源：Runway 官网，量子位，华安证券研究所

Runway Gen-4 Turbo 模型技术端对模型架构和计算效率进行双重优化，采用了更高效的推理算法或增强的硬件加速支持，在延续了前代世界一致性（World Consistency）的同时将速度优化推向新高度，大幅缩短视频生成时间。适合需要快速迭代和创意探索的场景，例如短视频内容创作、概念验证或实时视觉效果预览。以下为实际应用场景：

图表 37 Gen-4 Turbo 模型应用示例：



资料来源：Runway 官网，华安证券研究所整理

图表 38 Runway Gen 系列模型核心技术

模型名称	核心技术
Gen-1	提出可控的结构和内容感知的视频扩散模型，该模型在大规模的无字幕视频和配对的文本——图像数据上进行训练。基于深度估计确保结构一致性，同时利用文本或图片进行内容跟控制，通过在模型中引入时间连接以及联合图像视频训练确保时间稳定性，通过控制轮次控制结构保留度。
Gen-2	结合深度学习与生成对抗网络（GAN）技术，实现从静态文本描述生成动态视频内容。采用深度学习技术，通过训练大量的视频数据来实现其功能，处理复杂视频生成任务时具备准确性和效率。
Gen-3 Alpha	全新基础设施上进行大规模多模态训练，推出全新视觉审核系统并遵循 C2PA（Content Credentials）标准；Act-One 技术是生成式角色表演工具，能够将视频和语音输入转换为虚拟角色动画，通过简单的摄像头捕捉演员的表演，生成逼真的角色动画，保留眼神、微表情、语调和动作节奏等细节，同时支持多种动画风格。
Gen-4	引入生成式视觉特效（GVFX）技术，快速生成高质量特效，与实拍和传统特效无缝融合；通过 Act-One 技术，仅通过简单的参考图和提示词就可以将镜头串联成完整的一段故事。
Gen-4 Turbo	基于最新的深度学习和神经网络算法，采用优化的生成对抗网络（GAN）架构。Runway 官方数据显示，生成速度较前一代模型提高了 300%，30 秒内生成 10 秒视频；生成的画质和构图细节显著提升，更精准地捕捉光影变化，减少抖动和模糊现象。

资料来源：Runway 官网，AIHub，AI 工具集，华安证券研究所

### 2.3 Runway 产品价格模式

**C 端用户积分订阅制促进未来持续收入增长。**Runway 产品针对 C 端用户的价格模式为积分订阅制，根据不同付费计划类型，提供相对应的订阅价格，年度订阅价格相比月度订阅价格享有 20% 折扣。不同付费计划对于生成视频的条件有所限制。订阅后每个用户获得对应积分（Credits），使用 Runway 不同产品生成视频时会消耗一定的积分，消耗积分数量与生成视频时长、使用模型类型、视频分辨率以及高级功能（如去除水印、使用特定的运镜方式或运动速度控制等）有关。

**B 端用户需要根据企业情况提供对应 AI 解决方案。**企业订阅方案可分为自助服务（Self-Serve）和全方位服务（Full-service）两种类型，自助服务适用于员工规模少于 50 人的企业，全方位服务则针对 50 人以上的大型企业，需要联系企业根据定制化需求提供服务。

图表 39 Runway C 端用户月度订阅价格

付费计划	月度订阅价格及积分	说明
Basic Plan	0 美元/月，125 一次性积分	1) 125 积分=25 秒 Gen-4 Turbo 或 Gen-3 Alpha Turbo； 2) 无法在 Gen-4 Turbo 或 Gen-4 Alpha Turbo 中去除水印； 3) Gen-3 Alpha Turbo（图像到视频）最多 10 秒； 4) 3 个视频项目，5GB 资产。
Standard Plan	15 美元/月，625 积分/月 年度订阅：12 美元/月	1) 625 积分=52 秒 Gen-4，125 秒 Gen-4 Turbo 或 Gen-3 Alpha Turbo 或 62 秒 Gen-3 Alpha； 2) 生成无水印； 3) Gen-4（图像到视频）最长 10 秒； 4) Gen-3 Alpha Turbo（图像到视频）最长 10 秒； 5) Gen-3 Alpha（文本、图像和视频到视频）最长 10 秒； 6) 在 Gen-4、Gen-4 Turbo、Gen-3 Alpha、Gen-3 Alpha Turbo 中将输出升级到 4k； 7) 无限的视频编辑器项目，100GB 资产。
Pro Plan	35 美元/月，2250 积分/月	1) 2250 积分=187 秒 Gen-4，450 秒 Gen-4 Turbo 或 Gen-3 Alpha

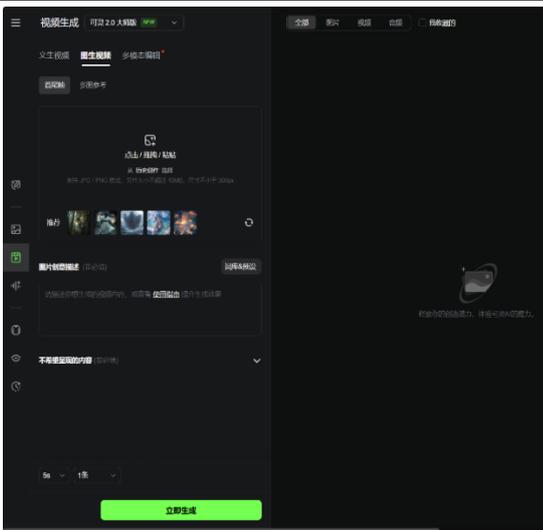
	年度订阅: 28 美元/月	Turbo 或 225 秒 Gen-3 Alpha; 2) 包括所有 Standard Plan 功能, 为 Lip Sync 和 Text-to-Speech 创建自定义语音, 500GB 资产。
<b>Unlimited Plan</b>	95 美元/月 (新用户第一个月 75 美元), 无限积分 年度订阅: 76 美元/月	1) 无限制 Gen-4, Gen-4 Turb, Gen-3 Alpha Turbo 且无积分和速率限制; 2) 包括所有 Pro Plan 功能, 可使用“Frames”模型 (Text to Image); 3) 在探索模式下以轻松的速度无限使用 Gen-4、Gen-4 Turbo、Gen-3 Alpha、Gen-3 Alpha Turbo、Frames 和 Act-One。

资料来源: Runway 官网, 华安证券研究所

### 3 可灵 AI

可灵是快手自主研发的视频生成大模型, 主要为用户提供高质量的视频内容生成服务。可灵 AI 可以支持用户通过多种形式进行输入生成, 包括但不限于文字、图片、动作视频等。目前可灵产品在移动端、Web 端和海外均上线, 实现完善的产品矩阵。

图表 40 可灵 AI web 端界面



资料来源: 可灵 AI web 端, 华安证券研究所

图表 41 可灵 AI APP 端界面



资料来源: 可灵 AI APP, 华安证券研究所

#### 3.1 可灵 AI B 端 C 端同步发展

在版本迭代更新上, 可灵 AI 经历了初始版本, 1.0 版本, 1.5 版本, 1.6 版本和最新的 2.0 版本。可灵 AI 以自研的可灵和可图大模型为基础, 提供视频和图像的生成与编辑能力, 并在版本迭代过程中不断在视频时长, 视频质量, 物理规律真实性上进行优化。

**图表 42 可灵 AI 版本迭代情况**

模型名称	推出时间	简介及重点更新
初始版本	2023 年 11 月	基于快手自研大模型可灵和可图，提供视频及图像生成与编辑能力，提供视频生成、图像生成、虚拟试穿等接口能力。
V1.0 版本	2024 年 6 月	可灵 1.0 模型在高品质模式下可生成 720p 视频。
V1.5 版本	2024 年 9 月	视频生成新增可灵 1.5 模型，在画质质量、动态质量、美学表现、运动合理性以及语义理解等方面均有显著提升。引入了全新的“运动笔刷”功能，进一步提升视频生成的精准控制能力。
API 服务升级	2024 年 11 月	对应用程序接口（API）进行重大升级，引入了 v1.5 高品质模型和视频延长功能，为企业级用户提供了更加强大的视频及图像生成与编辑服务。视频延长功能允许用户对生成的视频进行 4 到 5 秒的续写，并且支持多次续写，最多达到 3 分钟的总时长。
可灵 AI APP	2024 年 11 月	正式推出独立 App，在苹果 App Store 及多个安卓商店上架，有可灵 1.0 和 1.5 两个版本模型可供选择，1.0 模型拥有更强大的可控生成能力，而 1.5 模型可直出 1080p 高清视频。针对已生成的视频提供续写功能，最长可续写生成约 3 分钟视频。
可灵 2.0	2025 年 4 月	正式发布可灵 2.0 视频生成模型及可图 2.0 图像生成模型。在视频生成的物理模拟、语义理解和美学表现上进行升级。正式提出 AI 视频生成的全新交互理念 MVL，允许用户通过文字、图片甚至视频片段的组合进行输入。基于 MVL 理念推出多模态编辑功能。可图 2.0 文生图能力迎来全面升级，可支持 60 多种风格化的效果转绘。可图 2.0 上线了实用的图像可控编辑功能。
可灵 2.1	2025 年 5 月	包含标准（720p）、高品质（1080p）两种模式，高端版本可灵 2.1 大师版升级后提供更强的语义响应。 成本上，在标准模式（720p）下，生成 5 秒视频需要消耗 20“灵感值”，高品质模式（1080p）下需要消耗 35“灵感值”。成本上与可灵 1.6 系列版本相持平。 生成效率上，高品质模式（1080p）在 1 分钟内就可生成 5 秒钟的视频，同档次其他模型基本耗时 2-3 分钟。 在生成质量上，可灵 2.1 全面提升。主要在模型动态细节、动态响应、动态幅度等方面进行优化。另外，物理模拟和人物动作幅度等更真实，更符合物理规律。

资料来源：可灵 AI，华安证券研究所整理

**可灵 AI 在 B 端与 C 端同步发展。**C 端方面，可灵 AI 为自媒体用户和部分广告主提供创作生产力，公司测算可灵 AI 可以为客户的短视频营销素材制作成本降低 60-70%，这带动了 C 端用户的付费增长。从用户结构上看，在可灵 2.0 发布会上，公司披露可灵 AI 全球用户规模超 2200 万。截止至 2024 年末，我国短视频创作者账号规模达 16.2 亿，日短视频数量超 1.3 亿条，快手创作者数量达 1900 万。

B 端方面，可灵 AI 面向客户商家提供 API 接口，目前已经和小米、亚马逊、Freepik、蓝色光标等数千家客户合作，满足在线教育，电子商务，智慧城市等领域的需求。可灵 AI 通过其 API 服务的灵活性和个性化适配客户应用场景，在行业内领先。可灵的全球开发者数量超 1.5 万，已累计生成 1200 万图像和 4000 万个视频素材。

图表 43 可灵 AI 运营数据



资料来源：可灵 2.0 发布会，华安证券研究所

在 DBC、CIW 和 CIS 联合发布的“2025 年 AI 视频生成企业 TOP20”榜单中，可灵 AI 成为榜首。

图表 44 “2025 年 AI 视频生成企业 TOP20” 榜单

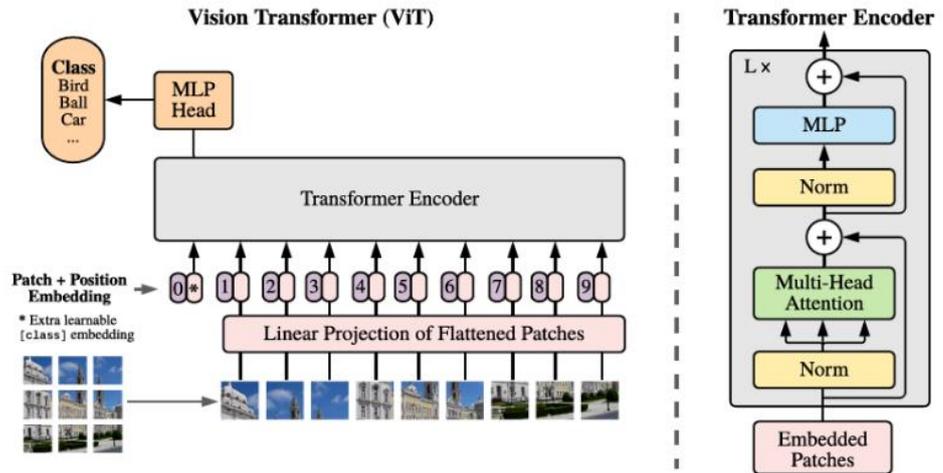
S/N	工具/软件	所属			
1	可灵AI	快手	11	白日梦AI	光魔科技
2	PixVerseAI	爱诗科技	12	讯飞绘镜	科大讯飞
3	即梦AI	抖音	13	广电视听智能体	广电总局
4	Vidu	生数科技	14	有言AI	魔法科技
5	通义万相, 绘蛙AI视频	阿里巴巴	15	剪映	脸萌科技
6	海螺AI	MiniMax	16	VegaAI	右脑科技
7	MOKI	美图	17	灵境AI	灵境万维
8	智谱清影AI	智谱AI	18	BOOLV	布尔向量
9	HiDream.ai	智象未来	19	VidAU	位道科技
10	智影	腾讯	20	巨日禄AI	巨日禄科技

资料来源：Albase 基地，华安证券研究所

可灵 AI 基于快手自研的可灵大模型和可图大模型，模型上采用了类 Sora 的 DiT 结构，并用 Transformer 取代传统的卷积网络结构，扩散模型基座为 flow 模型。以上的架构设计为可灵 AI 在视频生成上处于领先地位打下技术基础。

DiT 架构集成了近年 Diffusion 生成模型的各种优秀技术，基于 Latent Diffusion Model (LDM) 框架，采用 Vision Transformer (ViT) 作为主干网络。创新点在于用 Transformer 学习和预测 noise 以及方差，具体为使用 Transformer 架构替换了传统的 Diffusion 模型中的卷积 U-Net 主干网络。

图表 45 ViT 架构



资料来源：《AN IMAGE IS WORTH 16X16 WORDS: TRANSFORMERS FOR IMAGE RECOGNITION AT SCALE》，华安证券研究所

图表 46 256x256 图像下架构评测

Class-Conditional ImageNet 256×256					
Model	FID↓	sFID↓	IS↑	Precision↑	Recall↑
BigGAN-deep [2]	6.95	7.36	171.4	0.87	0.28
StyleGAN-XL [53]	2.30	4.02	265.12	0.78	0.53
ADM [9]	10.94	6.02	100.98	0.69	0.63
ADM-U	7.49	5.13	127.49	0.72	0.63
ADM-G	4.59	5.25	186.70	0.82	0.52
ADM-G, ADM-U	3.94	6.14	215.84	0.83	0.53
CDM [20]	4.88	-	158.71	-	-
LDM-8 [48]	15.51	-	79.03	0.65	0.63
LDM-8-G	7.76	-	209.52	0.84	0.35
LDM-4	10.56	-	103.49	0.71	0.62
LDM-4-G (cfg=1.25)	3.95	-	178.22	0.81	0.55
LDM-4-G (cfg=1.50)	3.60	-	247.67	<b>0.87</b>	0.48
<b>DiT-XL/2</b>	9.62	6.85	121.50	0.67	<b>0.67</b>
<b>DiT-XL/2-G (cfg=1.25)</b>	3.22	5.28	201.77	0.76	0.62
<b>DiT-XL/2-G (cfg=1.50)</b>	<b>2.27</b>	<b>4.60</b>	<b>278.24</b>	0.83	0.57

资料来源：《Scalable Diffusion Models with Transformers》，华安证券研究所

目前可灵 APP 会员分为三类，分别为：黄金会员（66 元 包月，可生成 66 个标准视频），铂金会员（268 元包月，可生成 300 个标准视频），钻石会员（666 元包月，可生成 800 个标准视频）。用户也可以购买资源包用来访问视频和图片生成相关 API 能力，包括文生视频、图生视频、视频延长、对口型、视频特效、文生图和图生图。

### 3.2 可灵 AI 商业化模式及运营数据

图表 47 可灵会员价格

连续包月	连续包季	按月购买
<b>黄金会员</b> <b>¥66</b> 每月 660 灵感值 ¥10.00/100 灵感值	<b>铂金会员</b> <b>¥268</b> 每月 3,000 灵感值 ¥8.93/100 灵感值	<b>钻石会员</b> <b>¥666</b> 每月 8,000 灵感值 ¥8.33/100 灵感值

资料来源：可灵 APP，华安证券研究所

图表 48 可灵视频生成资源包

资源包序号	模型版本	并发	积分数量	支持功能	有效期	原价	折扣	折后价	最小起订量
B1	V1.0 V1.5 V1.6	5	10,000	文生视频 图生视频 视频延长 对口型 视频特效	1个月	¥10,000	0%	¥10,000	3个月
B2			15,000			¥15,000	10%	¥13,500	3个月
B3			20,000			¥20,000	20%	¥16,000	3个月

资料来源：可灵 Web 端，华安证券研究所

图表 49 可灵图片生成资源包

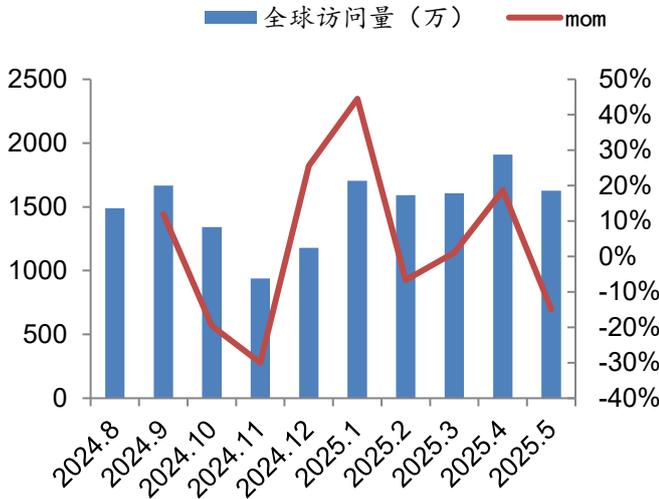
资源包序号	模型版本	并发	积分数量	支持功能	有效期	价格	折扣	折后价	最小起订量
B1	V1, V1.5	9	200,000	文生图 图生图	1个月	¥5,000	0%	¥5,000	3个月
B2	V1, V1.5		400,000			¥10,000	10%	¥9,000	3个月
B3	V1, V1.5		600,000			¥15,000	20%	¥12,000	3个月

资料来源：可灵 Web 端，华安证券研究所

参考非凡产研数据，2025年5月份，可灵 AIWeb 端全球访问量达 1626 万，MAU 为 460 万，4 月份访问量为 1909 万。

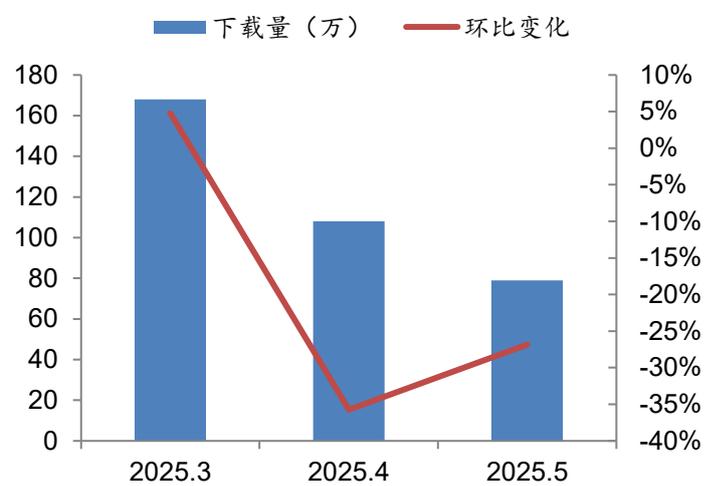
2025年5月，可灵 AIAPP 端下载量达 79 万，4 月份为 108 万。

图表 50 可灵 AI Web 端全球访问量



资料来源：非凡产研，华安证券研究所

图表 51 可灵 AI APP 端国内下载量



资料来源：非凡产研，华安证券研究所

ConceptMaster 是快手可灵发布的多主体视频生成技术方案，用于多概念视频定制的创新框架，可以在未测试调优的情况下，在扩散模型上生成高质量个性化视频。

什么是多主体视频生成？可以理解为在文生视频模型基础上支持输入多个主体图像，从而在生成视频的过程中把多个主体组合在同一视频中。图生视频有一定区别，图生视频是以输入的图像作为生成视频的首帧然后进行生成，一定程度上没有多主体视频生成灵活。除可灵外，Vidu 和 Pika 等也上线多主体视频生成功能。

ConceptMaster 的功能有哪些？

- 1) 多概念定制视频：可以通过输入多个图像生成多个概念的定制视频，例如根据一张“一个小男孩在路上骑自行车”的图像生成视频。
- 2) 身份解耦：学习多概念嵌入解耦，保证每个主体以独立的方式输入到扩散模型中，保证了多主体视频质量。
- 3) 数据收集：建立了数据管道，可以系统收集不同概念的视频数据，为解耦提供支持。
- 4) 基准测试：在概念保真、身份解耦和视频生成质量三个维度对模型进行有效性测试，为评估多概念视频定制模型性能提供参考。

图表 52 ConceptMaster 多主体生成示例



资料来源：ALEX RUIZ，华安证券研究所

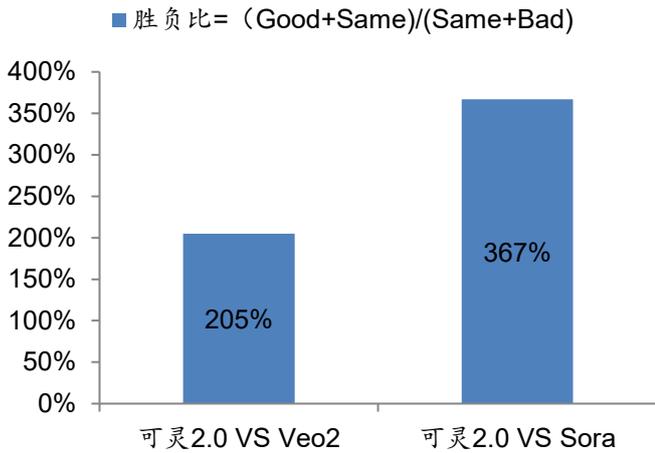
### 3.3 近期更新

4月15日，快手推出发布可图 2.0（图像生成模型）和可灵 2.0（视频生成模型）。**可灵 2.0 在语义响应、动态质量、画面美学三个维度进行升级。**首先在语义响应上，可灵 2.0 版本具有更强的动作响应、运镜响应和时序响应；在动态质量上，可以实现复杂动作的质量提升，运动幅度与速度更合理；在电影质感及艺术风格表现等方面也有显著提升。

**可灵 2.0 的技术亮点：**1.全新设计的视频生成基础模型，打开建模和仿真能力空间；  
2.全面升级训推策略，解锁更强的指令响应和运动表现。

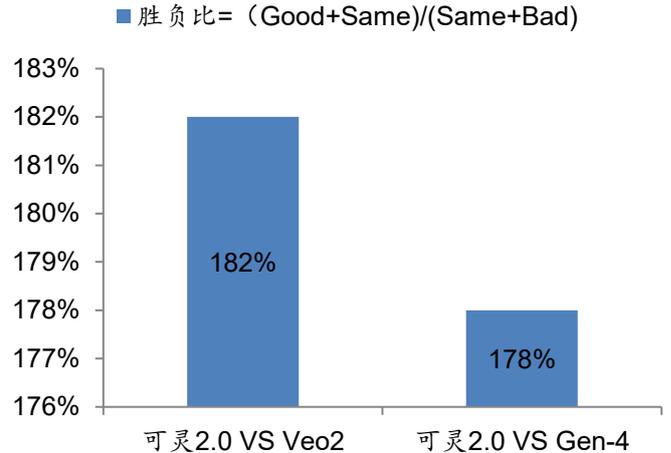
可图 2.0 在指令遵循能力、电影美学和艺术风格的多样性进行全方位升级，内置超过 60 种内容风格。

图表 53 文生视频模型胜负比



资料来源：可灵 2.0 发布会，华安证券研究所

图表 54 图生视频模型胜负比



资料来源：可灵 2.0 发布会，华安证券研究所

由于影像信息很难用文字完全描述，因此需要定义新语言，引用多模态信息描述——MVL。可以理解为用户可以通过文字+图片等多个模态进行输入。

基于 MVL 可灵推出多模态编辑功能，支持在现有的视频基础上，通过图片或文字的输入，实现生成视频内容元素的增删、替换，实现更加灵活的二次编辑和处理功能。

5月29日，快手推出可灵 2.1 系列模型，包含标准（720p）、高品质（1080p）两种模式，在性价比和高效赛道同步布局。高端版本可灵 2.1 大师版升级后提供更强的语义响应。

该次更新主要在性价比方面进行了较大的提升。在标准模式（720p）下，生成 5 秒视频需要消耗 20“灵感值”，高品质模式（1080p）下需要消耗 35“灵感值”。成本上与可灵 1.6 系列版本相持平。生成效率上，高品质模式（1080p）在 1 分钟内就可生成 5 秒钟的视频，同档次其他模型基本耗时 2-3 分钟。

在生成质量上，可灵 2.1 全面提升。主要在模型动态细节、动态响应、动态幅度等方面进行优化。另外，物理模拟和人物动作幅度等更真实，更符合物理规律。

## 4 即梦 AI

### 4.1 即梦 AI 模型版本迭代、核心功能及技术

即梦 AI 是由字节剪映团队研发的一站式智能创作平台，前身为“剪映 Dreamina”，2024 年 5 月升级并更名为“即梦 AI”。2024 年 7 月 31 日，即梦 AI 1.0.0 安卓版本上架安卓应用市场，8 月 6 日上架苹果 App Store 应用商店，主要功能包括 AI 图片创作、AI 视频创作、探索创意世界等；2024 年 11 月 15 日，即梦 AI S2.0 Pro 和 P2.0 Pro 全量上线，S2.0 Pro 模型突破点在于卓越的首帧一致性，P2.0 Pro 模型突破点则是极高的提示词遵循能力；2025 年 4 月 7 日，即梦 AI 3.0 全量上线，突破点在于中文文本生成能力全面提升和提供影视级画质；2025 年 4 月 28 日，即梦 AI 视频 3.0 模型上线，动作遵循能力、镜头遵循能力、物理模拟和情绪表达能力更强；2025 年 6 月 5 日，即梦图片 3.0 更新，支持“图生图”，用户可以上传参考图片，加入文本提示来生成目标风格作品。

图表 55 即梦 AI 模型迭代及核心功能突破情况

发布日期	模型名称	核心功能突破
2024 年 5 月	剪映 Dreamina 更名即梦 AI	AI 图片创作、AI 视频生成、智能画布。
2024 年 7 月、8 月	即梦 AI 1.0.0	AI 文生图、视频生成等技术实现商业化运作，长视频和高动态方向具备技术创新。
2024 年 11 月	即梦 AI S2.0 Pro	卓越的首帧一致性：精准捕捉用户输入图的细节，并在色彩和光影上进行细腻处理，使得视频每一帧流畅自然，具有视觉美感和真实感，保持画面一致性和细节还原取得重大突破。
	即梦 AI P2.0 Pro	极高的提示词遵循能力：能够精确理解和执行复杂的用户指令，将创意转化为连贯、一致的视频内容，在多镜头组合和场景切换中表现出色，丰富视频的叙事和表现力。
2025 年 4 月	即梦 AI 3.0	影视级画质与高分辨率支持：图像支持 2K 分辨率（2560×1440 像素）直出图像，相较于 Midjourney（16:9 分辨率为 1456×816 像素）清晰度和细节表现更优；新增“影视质感”效果可生成更具真实感和细腻度的图像，适用于广告、海报等商业场景；中文文本生成能力全面提升：小字稳定性优化，用户可生成包含复杂中文小字的图像；艺术字体支持，支持更具设计感的字体生成包括镭射字、动态旋转排列文字等；语义理解增强，对中文指令识别更精准。
	即梦 AI 视频 3.0	动作遵循能力、镜头遵循能力、物理模拟和情绪表达能力更强。
2025 年 6 月	即梦图片 3.0	支持“图生图”，用户可以上传参考图片，加入文本提示来生成目标风格作品。生成 4 张图仅需 2 积分，而平台每天赠送 66 积分。生成一张海报只需 10-20 秒。

资料来源：钛媒体，华安证券研究所整理

即梦 AI 底层技术包括深度学习模型、生成对抗网络（GAN）、变分自编码器（VAE），分别在理解与生成、提升创作质量、实现创意的灵活拓展三方面实现技术突破。底层优越的技术优势，赋能即梦 AI 视频生成、AI 绘画、智能画布等多项核心功能。

图表 56 即梦 AI 核心技术

核心技术	技术介绍	具体案例
深度学习模型	理解与生成的基石：通过大量的数据训练，模型能够深入学习和理解各种视觉元素、场景以及语言描述之间的关联；视频生成方面，可以根据用户输入的文字描述，准确构建出相应的视频场景包括物体、动作、光影变化等元素；AI 绘画方面，能够将用户输入的提示词转化为对应图像元素，根据学习到的艺术风格和构图规则，生成具有较高审美价值的图片。	用户输入：宁静的山间湖泊，周围环绕着翠绿的森林； 模型输出：迅速从记忆中提取相关的视觉元素，如湖泊的形态、森林的颜色和纹理等，并将这些元素组合成符合描述的视频或图片； 准确还原用户描述并展现一定艺术美感。
生成对抗网络（GAN）	提升创作质量：通过生成器和判别器相互对抗、不断优化，提升生成作品的质量。生成器：根据用户输入生成图像或视频；判别器：对生成作品进行评估，判断是否真实、是否符合用户要求。 生成器不断调整生成策略，以生成更逼真、高质量作品，判别器不断提高自己的鉴别能力，两者相互促进，提升即梦 AI 生成作品质量。	AI 绘画： 生成器：生成“古代宫廷宴会”主题图片； 判别器：对图片中的人物服饰、场景布置、色彩搭配等方面进行评估，如果发现不符合古代宫廷风格或者不够逼真，生成器会根据判别器的反馈进行调整，重新生成图片。 多次迭代后，最终生成一幅高质量、符合用户预期的古代宫廷宴会图片。
变分自编码器（VAE）	实现创意的灵活拓展：在学习到的数据分布基础上对用户输入进行灵活的拓展和变化，当用户输入的提示词比较模糊或者具有多种可能的解读时，VAE 可以根据已学习到的知识，生成多种不同风格、不同细节的作品。	用户输入：美丽的花朵。 模型输出：写实风格，展现真实形态和色彩；卡通风格，赋予花朵更加可爱、夸张形象；印象派风格，色彩和笔触传达花朵的神韵氛围。

资料来源：即梦 AI，华安证券研究所整理

## 4.2 即梦 AI 模型应用场景、生成效果及同行业产品对比

根据字节跳动发布技术报告《Seedream 3.0 Technical Report》显示，即梦 AI 模型领先的应用场景主要包括文本生成图片视频、文本渲染、超现实主义肖像等。

### 1) 文本生成图片、视频应用场景及效果图对比

根据相同的输入文字提示，将即梦 3.0 输出效果图与即梦 2.0、FLUX-1.1 Pro、Ideogram 3.0、Midjourney v6.1、Imagen3、GPT-4o 对比，主要观察模型对指令的理解力与准确性、渲染图效果的感染力与创新性等。

大模型中对齐 (Alignment) 指将两个不同序列中的对应元素 (如单词、字符或子词) 进行匹配，以便进行某些任务，例如机器翻译、文本摘要、语音识别等，通过对齐，模型可以更准确地学习序列之间的关系，从而提高模型的泛化能力和性能。因此在进行文本输入时，大模型会根据输入文本内容的元素进行匹配，不同模型生成的效果图对比如下：

图表 57 即梦 3.0 对齐比较 (Alignment Comparison)



提示：两个男孩在昏暗房间里，前面男孩很害怕，而后面男孩显得很镇定

资料来源：Seedream 3.0 Technical Report，华安证券研究所

图表 58 即梦 3.0 结构比较 (Structure Comparison)



提示：两个 14 岁男孩身着千禧年初 (Y2K) 风格服装，在舞台上进行霹雳舞表演，同时表现单手地板动作

资料来源：Seedream 3.0 Technical Report, 华安证券研究所

图表 59 即梦 3.0 美学比较 (Aesthetic Comparison)



提示：一个女孩一只眼睛紫色，那一侧头发蓝色；另一只眼睛蓝色，而那一侧的头发紫色（要求写实风格）

资料来源：Seedream 3.0 Technical Report, 华安证券研究所

图表 60 即梦 3.0 设计比较 (Design Comparison)



提示：上方提示：贴纸 1 一只猴子咧嘴笑着配有文字“Happy”；贴纸 2 猴子戴着太阳镜配有文字“Cool”；贴纸 3 猴子害羞地拿着一朵花配有文字“Shy”；猴子看起来很惊讶配有文字“Surprise” 下方提示：Q 版角色，女孩，全身像，街舞，三视图

资料来源：Seedream 3.0 Technical Report, 华安证券研究所

### 2) 本渲染场景及效果图对比

以下文本渲染效果图输入指令为：一幅引人入胜且充满活力的图像，风格定位为时尚、插画、排版、三维渲染、绘画。展示七个色彩缤纷的醒目文字：“lunes”、“martes”、“miercoles”、“jueves”、“viernes”、“sbado”、“doming”；这些充满生气的杯子里冒出奇异的毛毡烟雾，它们优雅地漂浮在梦幻迷人的氛围中，各式各样漂浮着的花朵为场景增添了深度和层次感，而柔和的淡蓝色背景则与整体设计和谐相衬。

图表 61 即梦 3.0 文本渲染比较 (Text Rendering)



资料来源：Seedream 3.0 Technical Report，华安证券研究所

### 3) 超现实主义肖像应用场景及效果图对比

图表 62 即梦 3.0 超现实主义肖像比较 (Photorealistic Portrait)



资料来源：Seedream 3.0 Technical Report，华安证券研究所

即梦 3.0 能够直接生成更高分辨率的人像作品(例如 2048X2048 像素的图像)，进一步增强人像的纹理效果，提升纹理细节和清晰度。下图为即梦 3.0 生成的人像效果图显示，生成的人像质量向专业摄影标准迈进方面取得了令人期待的进展，也为后续应用带来更大可能性。

图表 63 即梦 3.0 超现实主义肖像效果图



资料来源：Seedream 3.0 Technical Report，华安证券研究所

### 4.3 即梦 AI 产品价格模式

即梦 AI 产品价格模式采用积分订阅制和直接购买积分制。

积分订阅制：按照会员等级可划分为基础会员、标准会员和高级会员，单月购买价格：基础会员及 79 元/月，标准会员 239 元/月，高级会员 649 元/月，同时还推出“连续包月”和“连续包年”两个计划，分别在单月购买价格的基础上享有 12%和 30% 的折扣。

直接购买积分制：用户可以不选用订阅计划直接购买积分，价格梯度包括 50 元/500 积分、75 元/750 积分、150 元/1500 积分、223 元/2250 积分、899 元/9000 积分。非会员用户每天可以领取 60 积分，即梦 AI 在生成图片和视频时会消耗定量积分。

图表 64 即梦 AI 用户月度订阅价格及积分情况

付费计划	订阅价格及积分	说明
基础会员	79 元/月 1080 积分 约生成 4320 张图片 360 个视频； 视频闲时生成积分消耗 5 折优惠； 连续包月：69 元/月； 连续包年：659 元/年 54.91 元/月。	1) 每天赠送积分； 2) 生成作品去除品牌水印； 3) 视频对口型； 4) 视频更高清； 5) 视频更流畅（可补帧到最高 60FPS）； 6) 单次同时生成 5 个视频
标准会员	239 元/月 4000 积分 约生成 16000 张图片 1333 个视频； 视频闲时生成积分消耗 3 折优惠； 连续包月：199 元/月； 连续包年：1899 元/年 158.25 元/月。	
高级会员	649 元/月 15000 积分 约生成 60000 张图片 5000 个视频； 视频闲时生成积分消耗 1 折优惠； 连续包月：499 元/月； 连续包年：5199 元/年 433.25 元/月。	

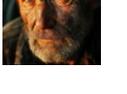
资料来源：即梦 AI 官网，华安证券研究所

## 5 Midjourney

### 5.1 Midjourney 版本迭代、优化情况及效果图比较

2022年2月 Midjourney V1 模型面世，初期生成的图像相对粗糙，细节不足，整体效果比较模糊；2022年4月发布 Midjourney V2 模型，引入“放大”和“变化”两个新功能，开始制定具体定价计划并转向付费测试模式；2022年7月 Midjourney V3 模型发布全新“风格化”和“质量”参数；2022年11月 Midjourney V4 模型拥有新代码库和人工智能架构，是第一个在独特的“Midjourney AI 超级集群”上训练的模型，也是第一个能够生成类似照片和渲染的逼真图像的模型；2023年3月 Midjourney V5 模型继续提高现实感和美学效果，产生的图像更接近提示；2023年5月3日 Midjourney V5.1 模型相较于早期版本有更显著的默认美学风格，使其在使用更短更简单的文字提示时更易于操作；2023年6月23日 Midjourney V5.2 模型生成的结果具有更高质量，细节更清晰，色彩、对比度和构图更佳；2023年12月 Midjourney V6 模型在图像质量上提供了卓越的增强，将照片的真实感提升到了一个新的水平，并显著提高了对提示的理解；2025年4月 Midjourney V7 模型在细节处理、交互效率和个性化体验上实现突破。

图表 65 Midjourney 模型版本迭代、优化情况及效果图比较

发布日期	模型名称	优化情况	效果图
2022年2月	Midjourney V1	第一个版本生成图像相对粗糙、细节不足，整体效果模糊	
2022年4月	Midjourney V2	引入“放大”和“变化”两个新功能	
2022年7月	Midjourney V3	发布全新“风格化”和“质量”参数	
2022年11月	Midjourney V4	拥有新代码库和人工智能架构，并且是第一个在“Midjourney AI 超级集群”上训练的模型，能够生成类似照片和渲染的逼真图像，是质变最突出的一次迭代。	
2023年3月	Midjourney V5	提高现实感和美学效果，产生的图像更接近提示，需要更长的提示来达到所需要的美学效果	
2023年5月	Midjourney V5.1	相比早期版本更具显著的默认美学风格，能够通过更短更简单的文字提示进行操作；擅长解读自然语言提示，具有更高的连贯性、图像更加清晰、产生更少的艺术残差和边界	
2023年6月	Midjourney V5.2	进一步增强照片真实感，细节更清晰，色彩对比度和构图更佳；连贯性和文本理解能力略有提升，多样性增加；对提示的理解更好，对“—stylize”参数的全范围响应更敏感	
2023年12月	Midjourney V6	更精准遵循提示词且支持更长的提示词；连贯性和模型知识提升；图像提示和混合功能改进；具备轻微的文字绘制能力；升级的图像放大功能，有“subtle”和“creative”两种模式（将分辨率提高2倍）	
2025年4月	Midjourney V7	图像质量与连贯性提升：像素级细节，动态构图优化；草稿模式（Draft Mode）：速度与成本优化，渲染速度提升至标准模式10倍，成本降低50%，对话式交互，语音输入支持；个性化模式默认启用	

资料来源：Midjourney 官网，果果 AI，标记狮社区，华安证券研究所

## 5.2 Midjourney 核心技术及应用场景

Midjourney 使用的核心技术与之前介绍的 AI 大模型类似，为生成对抗网络（GAN）。GAN 主要包含两个神经网络——生成器和判别器。生成器负责生成图像，判别器评估生成器的性能，两个网络通过反复对抗的方式进行训练，最终生成一副可以满足作者需求的作品；此外 Midjourney 还使用优化的变换器（Transformer）架构，是一种基于自注意机制（self-attention）的深度学习模型架构，变换器没有显式的循环或卷积结构，而是使用注意力机制来处理输入序列中的依赖关系，核心组件包括多头注意力机制（Multi-head Attention）和前馈神经网络（Feed-forward Neural Networks）。多头注意力机制允许模型在不同的表示空间中进行多次自注意计算，以捕捉不同层次和角度的语义关系，前馈神经网络则负责对每个位置的表示进行非线性变换和映射。

2025 年 4 月 3 日 Midjourney V7 模型发布，真实感与创作效率实现双升级。V7 模型的核心亮点主要包括“草稿模式”（Draft Mode）加速创意迭代与效率提高、图像质量与细节飞跃、场景理解与空间构建、精准视角与风格呈现、模型个性化（Personalization）等。应用场景主要包括绘画创作、家居设计和装修、电商海报宣传、Logo 和徽章设计、插画、漫画和油画等。

图表 66 Midjourney V7 模型核心技术亮点

核心技术亮点	亮点介绍
草稿模式（Draft Mode）	<b>速度与成本优化：</b> 渲染速度提升至标准模式 10 倍，成本降低 50%，支持快速迭代创意原型； <b>对话式交互：</b> 网页端自动切换为对话界面，用户可通过自然语言指令实时调整图像； <b>语音输入支持：</b> 新增中文语音识别功能，用户可通过语音指令生成并实时修改图像。
图像质量与细节飞跃	<b>像素级细节：</b> 光影、纹理和材质处理能力显著增强，例如丝绸光泽、水波折射和皮肤毛孔的逼真呈现，对用户输入的文字和图像提示（Prompt）理解更加精准； <b>动态构图优化：</b> 解决传统 AI 绘画的透视扭曲、肢体错位问题，生成专业级构图，尤其在多人场景中优先合理性，避免脸部重叠。
场景理解与空间构建	<b>场景空间关系合理：</b> 更好地理解并呈现提示词所描述的完整场景，画面中各个元素之间的空间关系更加合理，主体与背景的融合度更高，不再是简单的元素堆砌，而是营造出一个更具沉浸感的视觉空间。
精准视角与风格呈现	<b>捕捉专业拍摄角度：</b> 能准确捕捉并再现这些专业拍摄风格（特定的摄影视角，如低角度、特写）带来的视觉冲击力。
模型个性化（Personalization）	<b>默认启用模型个性化：</b> 根据使用习惯和偏好进行学习，让生成的图像更符合用户“口味”，需要花费大约 5 分钟解锁此功能，并且可以随时开启或关闭。

资料来源：果果 AI，华安证券研究所整理

图表 67 Midjourney 应用场景：绘画创作



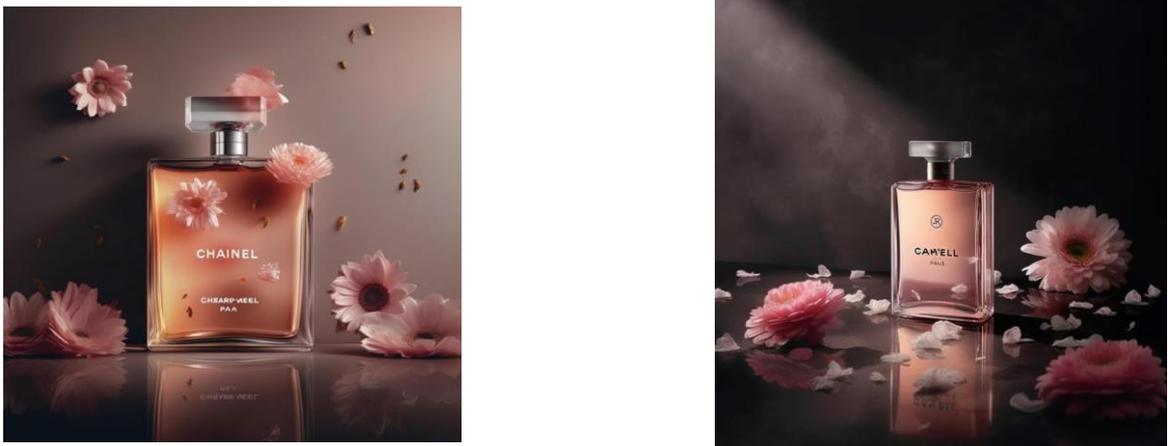
资料来源：Midjourney 官网，华安证券研究所

图表 68 Midjourney 应用场景：家居设计和装修



资料来源：Midjourney 官网，华安证券研究所

图表 69 Midjourney 应用场景：电商海报宣传



提示词：一瓶香水，香奈儿，花朵，4K，斑驳的光影，高贵典雅，海报风格

资料来源：Midjourney 官网，华安证券研究所

图表 70 Midjourney 应用场景：Logo 和徽章设计



Logo 提示词：一个个台湾茶叶品牌的标志，风格极简

徽章提示词：足球队队的队徽，风格简约，不要有阴影细节

资料来源：Midjourney 官网，华安证券研究所

图表 71 Midjourney 应用场景：插画、漫画和油画



插画提示词：大唐不夜城，中国古代城市，光谱、水晶、魔法、奇幻、梦幻、绚丽的光影效果，透明质感，镜面反射

漫画提示词：宫崎骏漫画 天空城；油画提示词：一位女子接过她的冠状头饰，油画风格，32K 超高清画质，古典具象写实主义，肖像画

资料来源：Midjourney 官网，华安证券研究所

### 5.3 Midjourney 产品价格模式

Midjourney 产品采用订阅制收费。根据不同订阅会员等级可分为 Basic Plan、Standard Plan、Pro Plan 和 Mega Plan，按月订阅进行收费价格分别为 10 美元/月、30 美元/月、60 美元/月、120 美元/月，也可以按年订阅提前预支一年订阅费用并享有 20% 折扣。不同付费计划在 GPU 快速时间（Fast GPU Time）、GPU 松弛时间（Relax GPU Time）、隐身模式（Stealth Mode）等方面有不同程度服务。

GPU 快速时间（Fast GPU Time）是指 Midjourney 不同订阅计划每月给用户在快速模式（Fast Mode）下的时间，快速模式可以加快图像制作速度；GPU 松弛时间（Relax GPU Time）仅服务 Standard、Pro 和 Mega 计划的用户，时间无限，每月都可以创建任意数量的图像且无需使用快速时间，但生成时间更长且需要进行排队等待，通常需要 0-10 分钟的等待时间；隐身模式（Stealth Mode）可以管理在 Midjourney 网站上能看到已发布图像的其他用户，可以选择将任意图像设为公共可见或私有。

**图表 72 Midjourney 产品订阅价格及说明**

付费计划	订阅价格	说明
<b>Basic Plan</b>	10 美元/月; 年订阅: 96 美元/年 8 美元/月。	1) GPU 快速时间 (Fast GPU Time): 3.3 小时/月 (200 分钟); 2) 不适用 GPU 松弛时间; 3) 额外购买 GPU 时间: 4 美元/月; 4) 不适用隐身模式; 5) 可同时完成的最大工作量 (Maximum Concurrent Jobs): 3 Fast jobs; 6) 最大重复次数/排列规模 (Maximum Repeat / Permutation Size): 4 jobs。
<b>Standard Plan</b>	30 美元/月; 年订阅: 288 美元/年 24 美元/月。	1) GPU 快速时间 (Fast GPU Time): 15 小时/月; 2) GPU 松弛时间无限制; 3) 额外购买 GPU 时间: 4 美元/月; 4) 不适用隐身模式; 5) 可同时完成的最大工作量 (Maximum Concurrent Jobs): 3 Fast or Relax jobs; 6) 最大重复次数/排列规模 (Maximum Repeat / Permutation Size): 10 jobs。
<b>Pro Plan</b>	60 美元/月; 年订阅: 576 美元/年 48 美元/月。	1) GPU 快速时间 (Fast GPU Time): 30 小时/月; 2) GPU 松弛时间无限制; 3) 额外购买 GPU 时间: 4 美元/月; 4) 可用隐身模式; 5) 可同时完成的最大工作量 (Maximum Concurrent Jobs): 12 Fast Jobs or 3 Relax Jobs; 6) 最大重复次数/排列规模 (Maximum Repeat / Permutation Size): 40 jobs。
<b>Mega Plan</b>	120 美元/月; 年订阅: 1152 美元/年 96 美元/月。	1) GPU 快速时间 (Fast GPU Time): 60 小时/月; 2) GPU 松弛时间无限制; 3) 额外购买 GPU 时间: 4 美元/月; 4) 可用隐身模式; 5) 可同时完成的最大工作量 (Maximum Concurrent Jobs): 12 Fast Jobs or 3 Relax Jobs; 6) 最大重复次数/排列规模 (Maximum Repeat / Permutation Size): 40 jobs。

资料来源: Midjourney 官网, 华安证券研究所

## 6 Stable Diffusion

### 6.1 Stable Diffusion 版本迭代与变体及核心功能

2022 年 8 月 Stable Diffusion 发布, 基于潜在扩散模型 (LDM/Latent Diffusion Model) 和扩散模型 (DM / Diffusion Model, DM 是基于 Google 的 Transformer 模型), 主要用于文本生成图像等领域。Stable Diffusion 自发布以来已进行多次版本迭代, 同时根据用户需求围绕架构版本、技术优化、应用场景、社区风格等维度, 在速度、兼容性或艺术风格等方面实现变体更新。Stable Diffusion 并未有过 1.0 版本, 最初模型是由 CompVis 开发的 Latent Diffusion, 包含文本到图像 (Text-to-Image) 和 Inpaint (修补) 功能, 可视为“Stable Diffusion”1.0, 目前 Stable Diffusion 发布的迭代或变体的版本数已超过 20。

**图表 73 Stable Diffusion 版本迭代与变体情况及性能介绍**

发布日期及主体	模型名称	性能介绍
2022 年 8 月 CompVis 发布	Stable Diffusion 1.1	以 256x256 分辨率训练 237000 步, 以 512x512 分辨率训练 197000 步。
	Stable Diffusion 1.2	使用 512x512 分辨率的 515000 步, 并使用 laion-improved-aesthetics 数据集。

	Stable Diffusion 1.3	基于 1.2 在 512x512 分辨率下增加 195000 个步骤，文本调节减少了 10%。
	Stable Diffusion 1.4	基于 1.2 以 512x512 分辨率和 laion-aesthetics v2 5+数据集进行 225000 步训练，减少类似文本调节。
2022 年 10 月 RunwayML 发布	Stable Diffusion 1.5	是 1.2 的增强版，使用 laion-aesthetics v2 5+数据集，以 512x512 分辨率训练了 595000 步。
2022 年 11 月 Stability AI 发布	Stable Diffusion 2.0	最初在 256x256 分辨率下进行 550000 步训练，随后在 512x512 分辨率下进行 850000 步训练，并在 768x768 分辨率下进一步微调了 150000 步。
	Stable Diffusion 2.1	是 2.0 的微调版本，增加了 55000 个步骤和额外的 155000 个微调步骤。
2023 年 6 月 Stability AI 发布	SDXL V0.9 base/refiner	Stable Diffusion XL 0.9 基础版/精调版与最终版本相比，图像成像质量较低。
	SDXL V1.0 base/refiner	具有卓越的文本理解和图像生成功能。
	SDXL V1.0 Beta 2.2.2	仅通过 Stability AI 开发者平台 API 接口独家提供，是 SDXL fine tune（微调）版本。
	Stable Image Core	另一个非公开 SDXL Model，只能在 v2 API 中使用。
2023 年 11 月 Stability AI 发布	Stable Diffusion 1.6	是 Stable Diffusion 1.5 的微调升级版，能生成比前一版本更高质量图像。
	SDXL Turbo	利用潜在一致性模型（LCM）将生成步骤从通常 30-40 步减少到 1-4 步，使用逆向扩散蒸馏（ADD）技术尽量减少生成步骤，仅限于 512x512 图像。
	SD Turbo	Stable Diffusion 2.1 的 Turbo 版本。
	Stable Video Diffusion	可从单张图像生成短动画且不会出现闪烁问题，可用于制作 14 帧、576x1024 像素的动画。
	SVD-XT	对 SVD 进行微调后的版本，能以相同的分辨率生成 25 帧动画。
2023 年 12 月 Stability AI 发布	Stable Zero123	单图生成高质量 3D 模型：通过改进的渲染数据集和分数蒸馏方法，大幅提升 3D 模型的生成效果和训练效率，同时可以与 SDXL 模型结合使用实现 3D 模型扩展。
2024 年 2 月 Stability AI 发布	Stable Diffusion Cascade	基于 Wuerstchen 架构包含三阶段的文生图扩散模型，在文本一致性与质量上优于 SDXL 和 SDXL Turbo，采用更高的压缩率效率更高，支持文生图、图生图和 Image Variation。
2024 年 2 月 ByteDance 发布	SDXL Lightning	结合 SDXL 和先进技术，以 1-8 个步骤提供图像，并与 LoRA 和 Unet 集成以提高性能。
2024 年 3 月 Stability AI 发布	Stable Video 3D	可从静态图像生成 21 帧、576x576 像素的动画，能够从单一影响输入建立和转换多视图 3D 视图。
2024 年 4 月 Stability AI 发布	Stable Diffusion 3	只在 API 中使用，在图像质量方面超越之前所有版本 model，具有强大的文本识别能力。
	Stable Diffusion 3 Turbo	图像生成速度比 Stable Diffusion 3 更快，针对不影响质量下快速获得结果的用户设计。
2024 年 4 月 Stability AI 发布	Cos Stable Diffusion XL	可生成纯黑至纯白的影像，基于 SDXL 使用 Cosine-Continuous EDM VPred schedule fine tune 而成。

资料来源：Edmond Yip, CSDN, AIGC 资讯，华安证券研究所

注：1) 训练步数的底层逻辑：让模型通过大量的数据学习来调整自身的参数，以达到能够根据输入的文本等信息生成符合要求图像的目的，训练步数越多，模型就有更多机会学习数据中的复杂模式和细节；2) 512x512 分辨率：指训练 Stable Diffusion 模型时，输入图像或生成图像的尺寸设定为宽高均为 512 像素，生成相对清晰且具备一定细节的图像；3) laion-aesthetics v2 5+ 数据集：用于训练模型的大规模图像数据集；4) fine tune 版本：表示模型“微调”版本，是指在深度学习中对预训练模型进行进一步调整以适应新任务的过程；5) ADD (Adversarial Diffusion Distillation)：扩散模型对抗蒸馏，是结合对抗训练和蒸馏训练两种方式对扩散模型的生图进行加速的方法，同时保持较高生图质量；6) Wuerstchen 架构：将计算成本高昂的文本条件阶段移动到高度压缩的潜在空间来训练文本模型的新框架。

**Stable Diffusion 针对不同用户群体的核心需求，具备多层次的功能。**核心功能包括 8K 级高清图像生成、多模态提示词理解、动态构图控制、3D 与视频生成、跨风格迁移等，主要基于 CLIP 文本编码器模型、VAE 变分自编码器模型、微调模型（LoRA）、ControlNet 模型等核心技术实现其功能。技术门槛多层次化，可满足电商运营、设计师/插画师、游戏开发者、AI 研究者等不同层次用户需求。

图表 74 Stable Diffusion 核心功能及技术实现

核心功能	技术实现
8K 级高清图像生成	通过潜在空间压缩 (Latent Space Compression)，将图像编码至低维空间进行扩散去噪，再通过解码器还原细节；SD3 采用优化后的 VAE (变分自编码器)，支持 8K 分辨率输出，显存占用降低 30%。
多模态提示词理解	集成 CLIP 文本编码器与 LLM (大语言模型)，实现语义解析。

动态构图控制	基于 ControlNet 插件，通过边缘检测、深度引导生成过程；用户可上传线稿锁定构图，AI 自动填充细节。
3D 与视频生成	SD3 融合 Stable Video Diffusion 技术，支持短视频序列生成（如 5 秒动态海报），分辨率达 1080P。
跨风格迁移	通过 LoRA 微调技术，注入特定风格（如浮世绘、赛博朋克）至基础模型，仅需 10 张样本图即可训练个性化风格。

资料来源：ai-tab，华安证券研究所

图表 75 Stable Diffusion 用户类型及核心需求

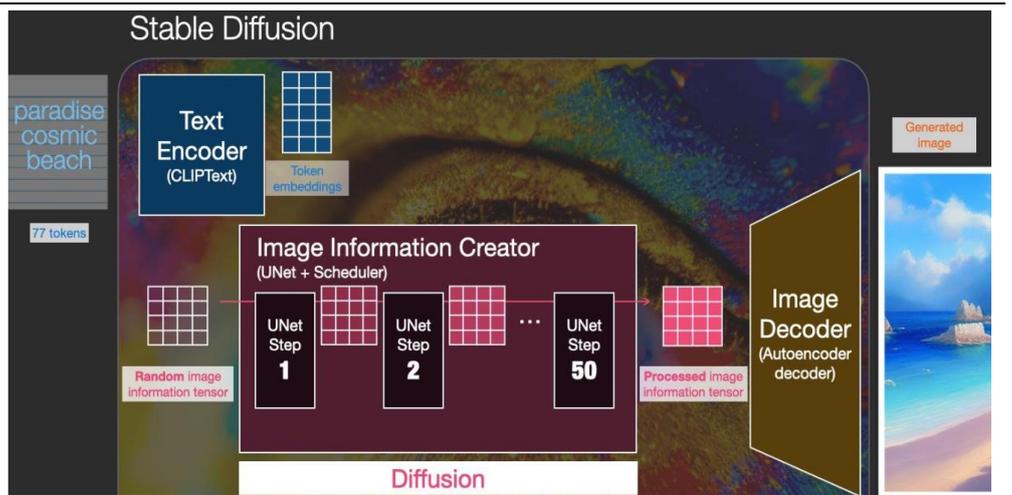
用户类型	核心需求场景	技术门槛
电商运营	定制商品图、广告视觉	低（云端平台直出）
设计师/插画师	快速生成概念稿、风格化作品	中（需要掌握提示词优化）
游戏开发者	批量创建场景、角色素材	高（可本地地微调模型）
AI 研究者	模型架构改进、多模态融合实验	极高（需编码能力）

资料来源：ai-tab，华安证券研究所

## 6.2 Stable Diffusion 核心技术及应用场景

Stable Diffusion 工作流程包括 CLIP 模型（CLIP Text Encoder）输入提示词或图像，进入 Diffusion 图像优化模块，然后使用 VAE 模型图像解码器（VAE Decoder）输出生成的图像。以文生图和图生图的生成技术为例，CLIP Text Encoder 模型是文本信息与机器数据信息之间互相转换的“桥梁”，作为 SD 模型中的前置模块将输入的文本信息进行编码；完成信息编码后，输入 SD 模型的“图像优化模块”对图像的优化进行“控制”，“图像优化模块”由一个 U-Net 网络和一个 Schedule 算法组成，U-Net 网络负责预测噪声，不断优化生成过程（U-Net 迭代优化步数大概 50-100 次，优化质量不断变好，纯噪声减少、图像语义及文本语义信息增加），Schedule 算法对每次 U-Net 预测的噪声进行优化处理；输入图像解码器（Image Decoder）输出图像。

图表 76 Stable Diffusion 工作流程及 U-Net 网络+Schedule 算法迭代去噪过程



资料来源：Stable Diffusion，华安证券研究所

Stable Diffusion 模型整体是一个 End-to-End 模型，主要由 VAE（变分自编码器，Variational Auto-Encoder），U-Net 以及 CLIP Text Encoder 三个核心组件构成。VAE 是基于 Encoder-Decoder 架构的生成模型，主要是图像压缩和图像重建作

用；U-Net 模型能够预测噪声残差，结合 Sampling method 对输入矩阵进行重构转化成图片的 Latent Feature；CLIP Text Encoder 模型是基于对比学习的多模态模型，包含 Text Encoder 和 Image Encoder 两个模型，分别用来提取文本特征和图像特征。LoRA 模型是 Stable Diffusion 中常见的微调模型，原理是在 U-Net 网络中增加一些新的特征信息，并对 U-Net 进行调整输出；ControlNet 模型是通过训练另外一个神经网络去调整 U-Net 网络，本质是通过输入额外信息，实现对扩散模型生图的精准控制。

图表 77 Stable Diffusion 包含模型及核心技术

模型名称	核心技术
VAE 模型	基于 Encoder-Decoder 架构的生成模型, Encoder(编码器)结构能将输入图像转换为低维 Latent 特征并作为 U-Net 的输入, Decoder (解码器)结构能将低维 Latent 特征重建还原成像素级图像, VAE 模型主要有图像压缩和图像重建作用。
U-Net 模型	预测噪声残差, 并结合 Sampling method (调度算法: DDPM、DDIM、DPM++ 等) 对输入特征矩阵进行重构, 逐步转化成图片的 Latent Feature, SD 模型通过反复调用 U-Net, 将预测出的噪声残差从原噪声矩阵中去除, 得到逐步去噪后的图像 Latent Feature, 再通过 VAE 的 Decoder 结构将 Latent Feature 重建成像素级图像。
CLIP Text Encoder 模型	基于对比学习的多模态模型, 主要包含 Text Encoder 和 Image Encoder 模型, Text Encoder 用来提取文本特征, Image Encoder 用来提取图像特征; Stable Diffusion 中主要使用 Text Encoder 模型, 将输入的文本提示进行编码, 转换成 Text Embeddings (文本的语义信息)。

资料来源: AI 创作官网, 华安证券研究所整理

图表 78 Stable Diffusion 应用场景: 电商运营



提示: 一瓶高档香水, 周围有典雅的花装饰, 极简风格



提示: 销售香水的广告, 绿色柳树背景

资料来源: Stable Diffusion 官网, 华安证券研究所

图表 79 Stable Diffusion 应用场景：设计师/插画师



提示：设计一款儿童玩具



提示：宫崎骏动漫风格插画图

资料来源：Stable Diffusion 官网，华安证券研究所

图表 80 Stable Diffusion 应用场景：游戏开发



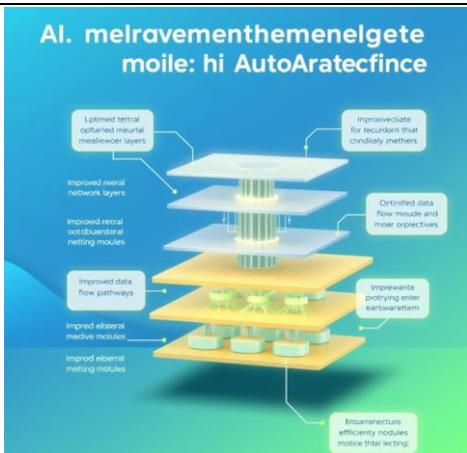
提示：设计一款中国古风游戏场景，以大唐为背景



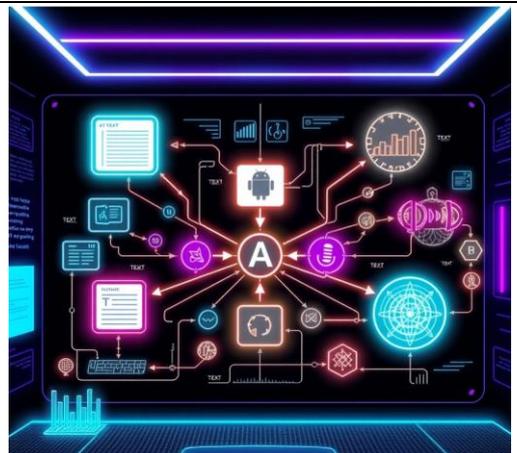
提示：设计一款格斗游戏

资料来源：Stable Diffusion 官网，华安证券研究所

图表 81 Stable Diffusion 应用场景：AI 研究



提示：AI 模型架构改进示意图



提示：AI 多模态融合实验研究示意图

资料来源：Stable Diffusion 官网，华安证券研究所

### 6.3 Stable Diffusion 产品价格模式

**Stable Diffusion B 端客户采用 API 订阅制：**可访问 100 多个 AI API 和模型，针对不同用户群体可分为 Basic、Standard、Premium 三个等级，按月支付价格分别为 27 美元/月、47 美元/月、147 美元/月，按年支付价格分别为 270 美元/年、470 美元/年、1470 美元/年，另外按年支付可额外免费获得 2 个月会员。

**C 端客户采用积分订阅制：**每月支付订阅费并获得相应积分（credits），使用 Stable Diffusion 生成结果时会消耗一定积分，根据不同会员等级可分为 Standard、Pro、Plus、Premium，按月支付价格分别为 9 美元/月、19 美元/月、49 美元/月、99 美元/月，按年支付价格分别为 90 美元/年、190 美元/年、490 美元/年、990 美元/年。

图表 82 Stable Diffusion B 端及 C 端产品订阅价格及说明

客户	付费计划	订阅价格	说明
B 端	Basic	27 美元/月, 270 美元/年	最多生成 13000 张图像; 3250 次 API 调用; 共享 GPU; 每秒 5 个请求限制
	Standard	47 美元/月, 470 美元/年	最多生成 40000 张图像; 10000 次 API 调用; 共享 GPU; 每秒 10 个请求限制; 可访问所有 API
	Premium	147 美元/月, 1470 美元/年	无限 API 调用; 可访问所有 API; 共享 GPU; 每秒 15 请求限制; 图像编辑 API; 视频生成 API; 所有 LLM API
C 端	Standard	9 美元/月, 90 美元/年	900 积分/月
	Pro	19 美元/月, 190 美元/年	1900 积分/月
	Plus	49 美元/月, 490 美元/年	5500 积分/月
	Premium	99 美元/月, 990 美元/年	12000 积分/月

资料来源：Stable Diffusion 官网，华安证券研究所

## 7 豆包

### 7.1 豆包大模型关键更新迭代、主要功能变化

豆包模型层聚焦多模态创新，加强基础大模型性能研究与提升，LLM 与图像视频生成模型交替迭代，达成“视觉理解+深度思考”创新成果。2023 年 8 月字节发布豆包前身“云雀大模型”，8 月 17 日宣布开始对外测试 AI 对话产品“豆包”；2024 年 5 月正式发布豆包大模型，包括通用模型 pro、通用模型 lite、角色扮演、语音识别等 9 个模型类别；2024 年 9 月发布豆包视频生成模型 PixelDance 及 Seaweed；2024 年 12 月发布豆包通用模型 Doubao-pro-1215；2025 年 1 月发布豆包大模型 1.5，并全面上线火山方舟平台；2025 年 4 月发布豆包 1.5·深度思考模型，视觉理解、文生图系列模型升级；2025 年 5 月发布豆包视频生成模型 Seedance 1.0 lite、豆包 1.5·视觉深度思考模型，并升级豆包·音乐模型。

**图表 83 豆包大模型更新迭代情况**

时间	关键更新	说明
2023年8月	云雀大模型 (Skylark) V1.0 版本	包含 lite、plus、pro 三个版本, 支持对话、代码生成、逻辑推理等核心能力
2024年5月	豆包大模型	包含通用模型 pro、通用模型 lite、角色扮演、语音识别等 9 个模型类别, 可应用之多个细分领域; 豆包通用模型 pro 窗口尺寸最大可达 128K, 全系列可精调, 豆包通用模型 lite 有较快的响应速度
2024年9月	豆包视频生成模型 PixelDance、豆包视频生成模型 Seaweed	1) 精准的语义理解及多动作多主体交互; 2) 强大动态与酷炫运镜, 拥有变焦、环绕、平摇、缩放、目标跟随等多镜头语言能力, 灵活控制视角带来真实世界体验; 3) 一致性多镜头生成, 在一个 prompt 内实现多个镜头切换, 同时保持主体、风格、氛围一致性; 4) 高保真高美感, 多风格多比例
2024年12月	豆包通用模型 Doubao-pro-1215	综合能力较 5 月提升 32%, 已全面对齐 GPT-4o, 在数学、专业知识等部分复杂场景任务中, 效果表现甚至更好; 大幅提升 Doubao-Pro 理解精度和生成质量, 并实现性能与效率的平衡, 推理服务价格仅为 GPT-4o 的八分之一
2024年12月	豆包视觉理解模型 Doubao-vision	可融合视觉与语言多感官深度思考和创作, 目前模型能力在十多个主流数据集上比肩 Gemini 2.0 与 GPT-4o; 具备更强内容识别、理解推理、细腻的视觉描述能力; 在教育、旅游、电商等场景有景广泛应用
2025年1月	豆包大模型 1.5	Doubao-1.5-pro 在知识、代码、推理、中文等多个权威测评基准上获得佳绩; Doubao-1.5-lite 在轻量级语言模型中表现出色, 为用户提供更高性价比; Doubao-1.5-vision-pro 在多模态数据合成、动态分辨率、多模态对齐等方面进行全面升级, 增强了视觉推理和细粒度信息理解能力
2025年4月	豆包 1.5·深度思考模型	主线版本 Doubao-1.5-thinking-pro, 性能超越 DeepSeek-R1, 接近 OpenAI o3-mini-high 和 Gemini 2.5 pro; 多模态版本的 Doubao-1.5-thinking-pro m, 实现“视觉理解+深度思考”, 可基于“图像”推理思考, 豆包 1.5·深度思考大语言模型总参数为 200B, 成本、效率和延迟均显著优化
2025年5月	豆包视频生成模型 Seedance 1.0 lite、豆包 1.5·视觉深度思考模型、升级豆包·音乐模型	豆包视频生成模型 Seedance 1.0 lite 核心优势为“效果好、生成快、更便宜”, 支持文生视频和图生视频, 技术层面实现精准的指令控制、影视级运镜技术提升视频视觉表现力、自然运动交互优化人与物体交互和逻辑, 可广泛应用于电商广告、影视创作、动态壁纸等场景; 豆包 1.5·视觉深度思考模型 (Doubao-1.5-thinking-vision-pro) 在视频理解、视觉推理及 GUI Agent 能力表现突出: 1) 视频理解方面支持动态帧率采样, 视频时序定位能力显著增强, 2) 新增视频深度思考能力, 学习数万亿多模态标记数据, 使视觉推理能力大幅提升, 3) 新增 GUI Agent 能力, 基于强大的 GUI 定位性能可在 PC 端、手机端等不同环境中完成复杂交互任务; 豆包·音乐模型升级, 不仅支持英文歌曲创作, 还可以通过理解视频自动适配纯音乐 BGM

资料来源: 雷科技, 雷峰网, AIHub, Albase 基地, 智谷趋势, 光明网, 华安证券研究所

**豆包产品层更新迭代, 聚焦 AI 搜索、多模态应用、AI 场景化应用、智能体生态。**在过去近两年的发展期间, 豆包产品迭代具备以下特点:

**1) AI 搜索为豆包最关键的基础能力之一, 持续提升 AI 搜索的产品能力:**整合字节内部抖音搜索和头条搜索能力, 接入 Bing 等搜索引擎支持全网搜索; 拓展学术搜索和支持深度搜索模式; **2) 重点发展多模态领域应用:**多模态领域功能更新与拓展包括持续的拓展语音聊天, 文生图、图生图、视频生成、图片理解、音乐生成等功能, 拓展范围广, 对每个应用保持从通用到精专的下钻打磨; **3) 深入应用场景, 持续探索通用大模型能力的 AI 场景化应用:**积极推动垂直应用场景 AI 应用落地, 围绕教育 (拍照答疑)、办公 (数据分析、录音纪要、PPT 生成、云盘)、创作 (海报生成、分身写真) 等多个垂直应用领域, 打磨更贴近应用场景的 AI 产品体验; **4) 重视智能体生态的搭建:**豆包官方打磨智能体外, 同时引入并运营第三方智能体, 官方持

续推出新智能体，打通豆包与扣子供应更多第三方开发者应用。

图表 84 豆包产品功能更新情况

时间	关键功能	说明
2023 年 11 月	推出 22 个豆包官方智能体；强化 AI 联网搜索能力；增加了关联问题推荐的功能；支持 LBS 的本地生活搜索推荐；支持文生图的产品能力；首页改版强化智能体模块；支持抖音短视频内容推荐	官方智能体包括：英语学习助手、聊天、AI 生成漫画、音乐电台、智能体创建助手、AI 生成图片等；豆包开始和头条搜索和抖音搜索密切合作，联网搜索的能力也主要是基于头条和抖音搜索，尚未拓展到全网搜索；生成文本的同时匹配与之相关的相关匹配度最高的短视频内容
2024 年 1 月	强化智能体的露出；开始对智能体做进一步的分类；将搜索框固定到首页顶部，强化搜索功能	智能体置顶到首页顶部，支持滑动展示；分类整理出不同的分类的智能体
2024 年 2 月	虚拟聊天智能体增加聊天背景封面；与扣子打通	智能体创建支持 AI 生成头像以及支持 AI 生成提示词；国内版智能体开发平台“扣子”上线，和豆包打通，支持发布豆包
2024 年 3 月	智能体开始灰度支持定制个人的音色功能；增加文档对话分析能力；接入 bing 搜索，联网搜索拓展全网搜索	支持上传 PDF、表格、PPT、word、txt 等格式文件，不支持图片格式，文档大小限制 50 个以内，大小限制 20M 以内；支持抖音搜索和头条搜索之外的内容搜索，上线豆包 PC 客户端和浏览器插件
2024 年 5 月	PC 端推出“问答”能力	整体产品形态对标知乎产品，主要通过搜索问答类的内容，通过搜索引擎入口引流；为利用 AI 问答的能力做的搜索投放获客的能力
2024 年 6 月	豆包浏览器插件更新	支持 youtube 视频总结和中英双字幕
2024 年 10 月	支持 AI 分身写真	支持 AI 生成个人写真图片
2025 年 1 月	支持图生视频、实施语音聊天、语音能力升级（支持唱歌、方言等）	1) 上线实时语音聊天功能，支持可实施打断的语音聊天体验；2) 语音聊天功能升级，支持唱歌、方言等功能
2025 年 3 月	上线深度思考功能	应对 Deepseek-R1 影响，推出字节自研深度推理模型

资料来源：腾讯网，火山引擎官网，华安证券研究所

## 7.2 豆包大模型核心技术及应用场景

豆包大模型依托先进的深度学习架构，采用 Transformer 等核心技术构建其底层框架。通过对海量文本、图像、语音等多模态数据的深度挖掘与学习，模型能够理解不同数据类型背后的语义、逻辑与模式。在训练过程中，运用大规模分布式训练技术，充分利用集群计算资源，加速模型收敛，提升训练效率。例如：在语言模型训练时，模型对大量书籍、新闻、论文等文本数据进行学习，掌握语言的语法规则、语义表达和知识体系；在图像模型训练中，通过对海量图像数据的特征提取与学习，使得模型能够理解图像内容并根据文本描述生成对应图像。同时，模型运用多种优化算法，如自适应学习率调整等，以提升模型训练效果与泛化能力。

目前豆包大模型主要包括以下九类：通用模型 pro、通用模型 lite、角色扮演模型、语音识别模型、语音合成模型、声音复刻模型、文生图模型、Function call 模型、向量化模型。

图表 85 豆包大模型家族九款模型



资料来源：字节跳动火山引擎原动力大会发布，华安证券研究所

图表 86 豆包九款模型主要功能及应用场景简介

模型名称	主要功能及应用场景
通用模型 pro	专注于处理复杂文本分析、问答、文本生成等任务：在搜索引擎优化场景中能够深入理解用户搜索意图，生成高度匹配且富含关键词的优质内容，提升网站在搜索引擎中的排名；在智能客服领域，可精准理解客户咨询问题，提供详细准确的解答，极大提升客户服务体验
通用模型 lite	在保持较高准确度的同时，高度注重资源效率，特别适合移动端等资源受限的场景：在手机端的智能助手应用中，能够快速响应用户指令，提供简洁有效的服务，且不会过多占用手机的计算资源与电量
角色扮演模型	可用于游戏 NPC、在线教育互动课程等场景：在游戏中，能赋予 NPC 丰富的个性与智能对话能力，使玩家的游戏体验更加真实有趣；在在线教育互动课程里，能模拟教师或学习伙伴，与学生进行互动交流，增强学习的趣味性与参与度
语音识别模型	支持多语言及方言识别，广泛适用于会议记录、语音助手、智能家居控制等场景：在跨国公司的会议中，能够准确识别不同语言的发言并实时转化为文字记录；在智能家居控制中，用户通过各种语言或方言发出指令，即可轻松控制家电设备
语音合成模型	可用于有声读物制作、语音播报系统、虚拟主播等领域：在有声读物制作中，能够生成自然流畅、富有情感的语音，为听众带来优质的听觉体验；在虚拟主播领域，为虚拟主播赋予生动的语音，使其能够更好地与观众互动
声音复刻模型	为个性化语音定制、数字遗产保存等开辟了新的可能性：用户可以通过该模型复刻自己或亲人的声音，用于个性化语音导航、智能音箱语音设置等，也可为数字遗产保存提供独特方式，留存珍贵声音记忆
文生图模型	创意设计、内容创作、电商商品展示等：设计师可通过输入文本描述，快速生成创意草图，为设计工作提供灵感；电商商家能根据商品特点描述生成精美的商品展示图片，提升商品吸引力与销售转化率
Function call 模型	程序调用和自动化任务设计，可实现不同软件系统之间的高效协同与自动化流程执行：在企业业务流程管理中，能够自动调用相关软件工具完成数据处理、文件传输等一系列任务，提高工作效率与准确性
向量化模型	便于机器学习算法处理，有效提高搜索推荐系统的准确性和效率：在电商平台的商品搜索与推荐中，通过向量化模型对商品数据和用户行为数据进行处理，能够更精准地为用户推荐符合其需求的商品，提升用户购物体验与平台销售额

资料来源：豆包官网，华安证券研究所

### 7.3 豆包视觉理解模型、图片及视频生成模型产品价格模式

豆包视觉理解模型收费模式是将输入信息（包括图片）转化为 token 计费，计费公式： $\text{推理费用} = \text{输入单价} \times \text{输入 token} + \text{输出单价} \times \text{输出 token}$ ，图片 token 用量  $= \min(\text{图片宽} \times \text{图片高} \div 784, \text{单图 token 限制})$ ，在线推理根据使用的不同类型模型，按输入及输出每百万 token 收费，批量推理针对 doubao-1.5-vision-pro-32k 在线推理价格 50% 收费；视频生成模型计费公式： $\text{推理费用} = \text{按 token 付费单价} \times \text{视频 token 用量}$ ，视频 token 用量  $\approx (\text{宽} \times \text{高} \times \text{帧率} \times \text{时长}) / 1024$ ，准确的 token 用量

以调用 API 后返回信息中 usage 字段为准；图片生成模型按使用张数后付费价格，Doubao-Seedream-3.0-t2i 定价 0.259 元/张，生成质量高，豆包-文生图模型-智能绘图定价 0.2 元/张。

图表 87 豆包视觉理解模型收费价格

视觉理解模型

将输入信息（包括图片）转化为 token 计费

计费公式

推理费用 = 输入单价 × 输入 token + 输出单价 × 输出 token

图片 token 用量 = min(图片宽 × 图片高 ÷ 784, 单图 token 限制)，具体用量统计逻辑请参见 token 用量说明。

在线推理

按 token 后付费价格

模型	输入 元/百万token	输出 元/百万token
doubao-1.5-vision-pro	3.00	9.00
doubao-1.5-vision-lite	1.50	4.50
doubao-1.5-ui-tars	3.50	12.00
doubao-1.5-vision-pro-32k	3.00	9.00
doubao-vision-pro-32k	3.00	9.00
doubao-vision-lite-32k	1.50	4.50

批量推理

按 token 后付费价格

5折：表示该模型的批量推理价格是在线推理价格的50%；未标注则表示与在线推理同价。

模型	输入 元/百万token	输出 元/百万token
doubao-1.5-vision-pro-32k 5折	1.50	4.50

资料来源：火山引擎官网，华安证券研究所

图表 88 视觉理解 token 用量说明

token 用量说明

模型理解图片，会将图片转化为 token，再进行推理计算。token 用量，根据图片宽高像素计算可得。图片转化 token 的公式为：

```
Plain
min(图片宽 * 图片高 ÷ 784, 单图 token 限制)
```

新版模型（doubao-1.5-vision-pro-32k-250115 及以后版本）支持对图像有更精细理解：

- detail:high 模式下，单图 token 限制升至 5120 token。
- detail:low 模式下，单图 token 限制 1312 token。

旧版模型（doubao-vision-pro-32k-241028、doubao-vision-lite-32k-241025）：

- detail:high 模式、detail:low 模式，单图 token 限制均为 1312 token。

图片尺寸为 1280 px × 720 px，即宽为 1280 px，高为 720 px，传入模型图片 token 限制为 1312，则理解这张图消耗的 token 为 1280 × 720 ÷ 784 = 1176，因为小于 1312，消耗 token 数为 1176。

图片尺寸为 1920 px × 1080 px，即宽为 1920 px，高为 1080 px，传入模型图片 token 限制为 1312，则理解这张图消耗的 token 为 1920 × 1080 ÷ 784 = 2645，因为大于 1312，消耗 token 数为 1312。这时会压缩 token，即图片的细节会丢失部分，譬如字体很小的图片，模型可能就无法准确识别文字内容。

资料来源：火山引擎官网，华安证券研究所

图表 89 豆包视频生成模型收费价格

视频生成模型

异步推理，将输出视频转化为 token 计费

按 token 后付费价格

计费公式

推理费用 = 按 token 付费单价 × 视频 token 用量

视频 token 用量 ≈ (宽 × 高 × 帧率 × 时长) / 1024，准确的 token 用量以调用 API 后返回信息中的 usage 字段为准。

单价

模型	文生视频 元/百万token	图生视频 元/百万token
doubao-seedance-1-0-lite	10.00	10.00
doubao-seaweed	30.00	30.00
wan2.1-14b	50.00	50.00

资料来源：火山引擎官网，华安证券研究所

图表 90 豆包图片生成模型收费价格

图片生成模型

按使用张数后付费价格，单位：元/张

模型	定价 元/张	生成质量
Doubao-Seedream-3.0-t2i	0.259	高
豆包-文生图模型-智能绘图	0.2	中等

资料来源：火山引擎官网，华安证券研究所

## 8 海艺 AI-SeaArt AI

### 8.1 SeaArt AI 版本迭代、核心技术及合作情况

海艺 AI 是成都海艺互娱科技有限公司在 2023 年 6 月发布的一款人工智能艺术平台，借助 AI 技术赋予用户创作、分享及生成独特艺术作品的的能力，综合了 AI 工具、模型训练和工作流（ComfyUI）的社区平台。平台集合 Stable Diffusion WebUI（SD）和 ComfyUI 的案例使用分享网站，包含如何使用 SD 和 ComfyUI 进行图像生成、图生图、工作流程高级技巧、LoRA 模型的使用和训练、模型训练、服务器在线部署应用一体。其特点在于整合了互联网社区精选应用工具，利用 AI 技术提高生产力并简化流程。目前，海艺 AI 自研大模型已经历 2.0 版本、2.1 版本和“海艺实验室”版本三次迭代。

SeaArt 起初版本聚焦于 AI 生图工具，在 SD 开源后开发了一套 SD 环境的云端管理系，将 SD 各种复杂功能封装起来形成一个易用的网站界面，特点在于简化了模型调用和输出；2.0 版本中，SeaArt 借鉴多种开源模型以解决 SD 的泛化能力和构图精细度问题；2024 年 3 月，SeaArt 基于节点的高级 AI 工作流工具上线 ComfyUI，用户可以通过可视化界面精确控制生产过程的每一步，实现更复杂、个性化的创作效果。

图表 91 SeaArt AI 核心技术优势

核心技术优势	核心技术优势介绍
海量丰富模型库以及丰富的 AI 创作工具	对接 Civita 模型库，支持直接使用 Civitai 内的几乎所有模型和 lora，用户也可以手动上传模型；提供一系列独立的 SwiftAI 工具，包括 AI 写真、文生图、背景消除等 10 多种常用 AI 图像处理功能。
全新自研大模型升级至 V6 版本	能够大幅提升出图质量，让细节更丰富，色彩层次更合理，立体感更强，文本理解能力更佳。
活跃的 AI 绘画社区	用户积极分享创作的天马行空的 AI 绘画作品，可自由探索并复用提示词。
高级 AI 工作流工具 ComfyUI	对于进阶用户，SeaArt 提供了基于节点的 AI 工作流工具 ComfyUI，用户可以通过可视化界面精确控制生成过程的每一步，实现更复杂、个性化的创作效果。

资料来源：AI 工具箱，华安证券研究所

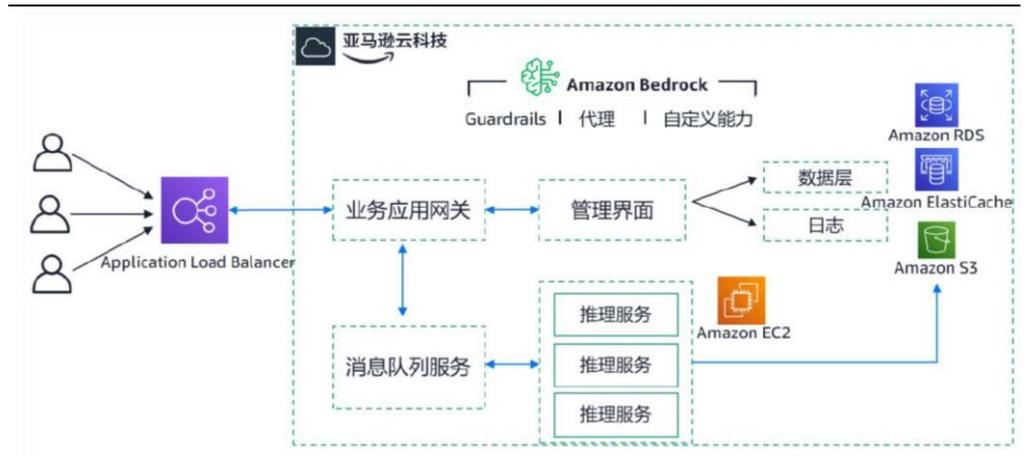
多方合作共寻技术突破，海外市场先发优势形成，国服市场未来可期。SeaArt AI 积极与亚马逊云科技、阿里云、华为方舟实验室进行技术合作。为实现海外市场的先发优势，海艺互娱与亚马逊云科技合作，采用云上便捷的生成式 AI 解决方案，快速构建可以服务于全球海外用户的 SeaArt AI 艺术创作平台，目前采用的亚马逊云科技产品包括 Amazon Bedrock、Amazon SES、Amazon EC2、Amazon S3 等；阿里云方面提供的解决方案包括神行工具包 DeepGPU 和分布式缓存及 OSS 加速，有利于提升 SeaArt AI 出图效率和推理耗时等性能；与华为旗下的方舟实验室合作文生图模型，例如 PixArt Sigma 等，形成优势互补。华为云提供强大资金及研发技术支持，结合 SeaArt AI 在技术应用及场景落地方面多年深耕的行业经验，有效实现产品技术突破。

图表 92 SeaArt AI 与亚马逊云科技、阿里云、华为云技术合作情况

合作主体	技术及产品	技术合作成果
亚马逊云科技	Amazon Bedrock、Amazon SES、Amazon EC2、Amazon S3	<b>Amazon Bedrock:</b> 利用 Amazon Bedrock 为生成式 AI 绘图过程作提示词优化, 可将超过 30 种语言的输入提示词翻译为英语, 降低生成式 AI 绘图门槛; 无需用户具备较高参数调整能力, 输入自然语言文本, 平台自动化分析关键风格元素, 生成质量较高的图片; <b>星云智能客服系统:</b> 智能问答与对话生成、个性化推荐与服务、数据可视化, 能够自动化识别用户意图、丰富客服知识库, 提升用户交互和用户满意度。
阿里云	神行工具包 DeepGPU 分布式缓存及 OSS 加速	<b>神行工具包 DeepGPU:</b> 单点推理上 DeepGPU 相较原生 PyTorch 推理性能提升 2.5-6 倍, 相较社区级领先框架 xFormers 推理性能提升 1.5-2.3 倍; <b>分布式缓存及 OSS 加速:</b> 针对端到端场景提供 ACK AI 套件的 Fluid 组件, Fluid 组件重点在推理节点内存内增加分布式的模型 cache 层, 层内根据 Fluid 组件可用 cache 扩缩动态根据 LRU 替换热点模型, 提高模型读写能力; OOS 加速提供更强的 OSS 吞吐能力来保障模型读写能力。
华为方舟实验室	PixArt Sigma	<b>文生图模型 PixArt Sigma:</b> 与华为旗下方舟实验室进一步合作推出文生视频模型, 华为具备强大的研发能力和资金支持, SeaArt AI 在技术实际应用和场景落地方面, 尤其是游戏行业有深入的理解和丰富的行业经验。

资料来源: 亚马逊云科技官网, 阿里云官网, 华为云官网, 海艺互娱发布, 华安证券研究所

图表 93 海艺互娱基于亚马逊云科技服务于海外用户的技术架构示意图



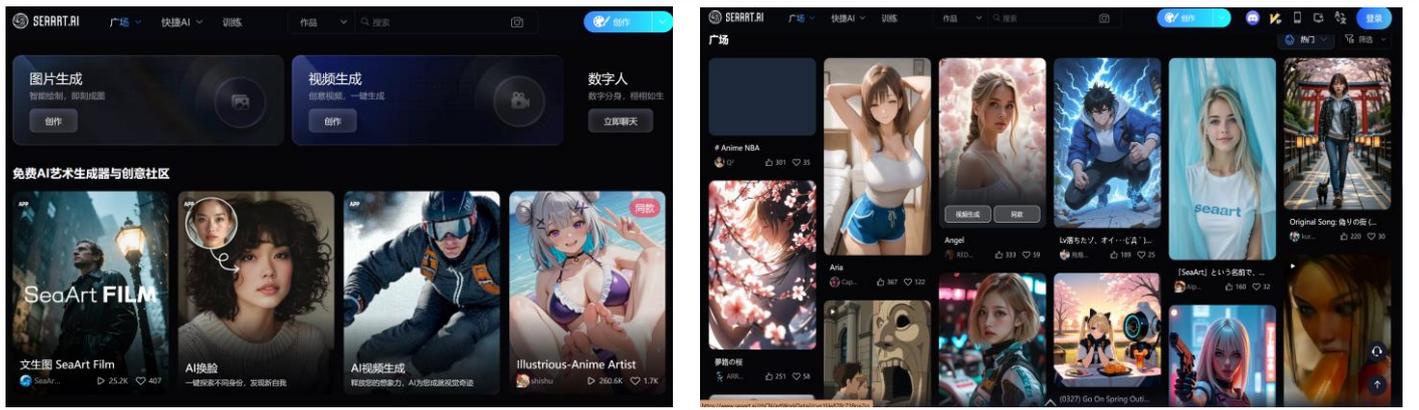
资料来源: 亚马逊云科技官网, 华安证券研究所

## 8.2 SeaArt AI 用户分层及应用场景

**版本迭代及技术更新实现用户分层。**在经历以上多次版本迭代后, SeaArt 主要功能逐步覆盖轻度、中度及核心 AI 用户: 轻度 AI 用户可以在 Home 浏览、查看和下载图片模型, 也可以使用平台内 AI 工具满足图片处理需求, 专业用户可在 Studio 查找和绘制质量更高的图片; 进阶版 SeaArt Train 允许中重度 AI 用户训练和定制 Lora, 在 ComfyUI 导入和创建工作流。

**落地页简洁降低上手门槛。**“免费 AI 艺术生成器与创意社区”提供 AI 换脸、AI 视频生成等最简单的 AI 工具; “广场”(Explore) 是用户使用这些工具生成的作品展示, 采用瀑布流布局, 用户可以选择右下方“同款”, 实现一键做同款; 快捷 AI (SwiftAI) 功能是一系列独立的工具, 包括 AI 写真、文生图、基本消除等功能, 基本涵盖了常用 AI 图像功能, SeaArt AI 落地页呈现简单有效功能有利于轻度用户立即上手, 降低上手门槛, 利于客户群体推广。

图表 94 SeaArt AI 落地页功能界面：免费 AI 艺术生成器与创意社区、广场



资料来源：SeaArt AI 官网，华安证券研究所

专业 ComfyUI 社区化，服务核心 AI 用户同时反向“筛选”高阶用户。ComfyUI 是一个基于节点式工作流的 Stable Diffusion UI，由 GitHub 上的匿名开发者 comfyanonymous 研发，用户可以通过可视化节点和连线，建立起整个生图流程。ComfyUI 具备两点优势：1) 控制精准，可扩展性强：用户可以在可视化界面里构建工作流程，通过增删节点、修改参数，通过控制每一步流程、精准实现某些效果，而且支持随时调整；2) 复用性强：其他用户可以直接复制、下载 workflow，一键使用，而 SeaArt 还提供了云上运行 ComfyUI 的功能。基于以上特点，SeaArt 在服务核心 AI 用户同时反向操作，通过 ComfyUI 筛选高阶用户，ComfyUI 核心功能包括节点化 workflow、模块化设计、实时可视化、易于扩展、支持多种模型和算法。

图表 95 ComfyUI 核心功能介绍

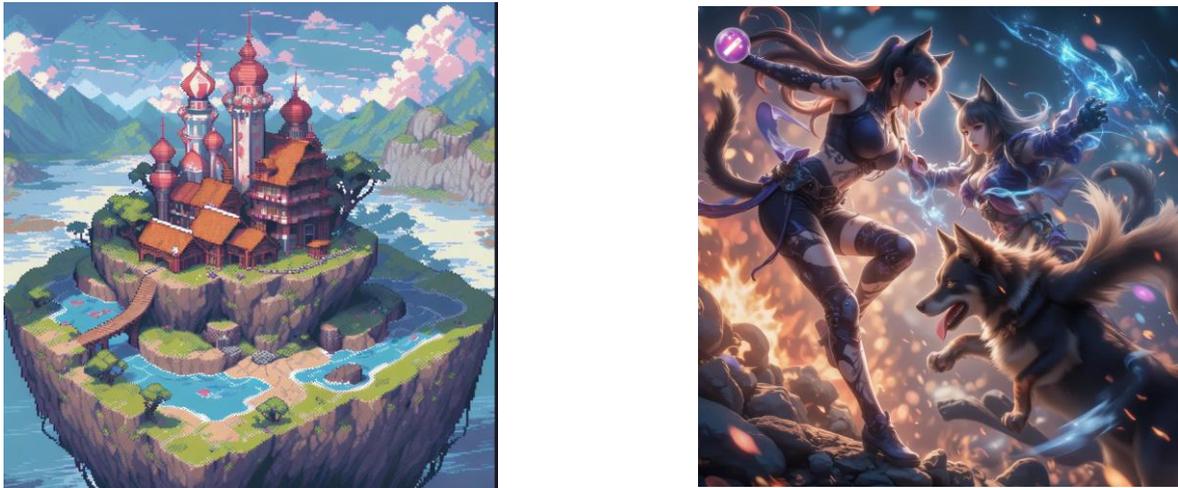
核心功能	核心功能介绍
节点化 workflow	基于节点化结构，每个节点代表特定的图像处理或生成任务，如噪声处理、模型推理、解码等。用户可以通过简单的拖放操作，将不同的节点组合起来，构建出符合需求的自定义生成流程。
模块化设计	用户可以自由选择和组合不同的模块，如选择特定的采样器或生成器，轻松试验各种配置；这种灵活性允许用户创建出适应不同任务的工作流，无论是快速图像生成还是高精度的专业调优。
实时可视化	每一个节点都可以实时显示其处理结果，这让用户可以在工作流的每一步进行调整并即时查看效果；这种高度可视化的设计帮助用户在图像生成过程中精确调控参数，确保最终的输出符合预期。
易于扩展	允许用户编写或导入自定义节点，这为高级用户提供了极高的扩展性，用户不仅可以添加新的处理模块，还可以集成其他深度学习模型或算法，进一步增强 workflow 的功能。
支持多种模型和算法	兼容多种深度学习模型和图像生成算法，用户可以轻松切换和实验不同的模型，找到最适合的图像生成方式，为研究人员和开发者提供了极大的自由度。

资料来源：AIGC 设计所，华安证券研究所整理

游戏行业经验赋能 SeaArt 内容生成。SeaArt AI CEO 马飞作为西南地区手游出海第一批人，具备丰富的出海经验，参与设计的游戏《银河帝国》，曾在 2012 年登上美国畅销榜 Top2。丰富的游戏制作与运营经验赋能 AIGC 产品，SeaArt 游戏工具设计在技术效果、图片质量和管线研发三个方面具备显著优势：1) 善于挖掘硬件和底层架构能力发挥技术效果；2) 游戏美术创作生图质量更高，模型训练提升图片生成图质量和用户体验；3) 成图管线有研发与迭代提质增效。以上优势有利于 SeaArt 在游戏内容生成行业广泛应用。

以下为在游戏内容生成领域效果图示例：

图表 96 SeaArt AI 游戏内容生成效果图：像素游戏场景生成器、猫狗斗争游戏场景



资料来源：SeaArt AI 官网，华安证券研究所

### 8.3 SeaArt AI 产品价格模式

SeaArt AI 采用订阅制（SVIP 权益卡）收费及算力充值收费两种模式。订阅制按照不同等级可分为免费版、初级版、标准版、专业版及大师版五个等级，可按照周付价格及年付价格进行会员订阅，周付价格分别为 0 元/周、5.99 元/周、18 元/周、68 元/周、109.99 元/周，年付价格享有 20% 折扣。算力充值付费方面，充值享有的算力无期限，可用于 AI 图像生成、AI 视频生成、AI 应用程序、模型训练、AI 聊天等场景，包括以下充值梯度：6 元/700 算力、30 元/3800 算力、68 元/8200 算力、148 元/17600 算力、328 元/45000 算力、648 元/93000 算力。

图表 97 SeaArt AI 产品订阅价格及说明

付费计划	订阅价格	说明
免费版	0 元/周	4500 体力，150 体力/天；每月可创作 4500 张图；限速生成；同时生成 2 个任务；同时训练 1 个 LoRA 任务。
初级版	5.99 元/周	1) 初级版 9000 体力，300 体力/天，每月可生成 9000 张图；标准版 21000 体力，700 体力/天，每月可生成 21000 张图，可享受免费创作及高分辨率视频生成； 2) 私密生成；快速生成；无水印视频下载；数字人高级模型；同时生成 5 个任务；
标准版	18 元/周	3) 创作支持质量模式； 4) 同时训练 1 个 LoRA 任务； 5) 购买算力包额外获得 5% 算力； 6) 图片批量下载。
专业版	68 元/周	1) 专业版 63000 体力，2100 体力/天，每月可生成 63000 张图；大师版 105000 体力，3500 体力/天，每月可生成 105000 张图； 2) 免费创作（享受免费且无限的图像生成服务，图像生成不会消耗算力或体力）； 3) 私密生成；快速生成；高分辨率视频生成；无水印视频下载；数字人高级模型；
大师版	109.99 元/周	4) 同时生成 10 个任务； 5) 创作支持质量模式（质量模式可生成更高精度的图像）； 6) 同时训练 3 个 LoRA 任务； 7) 购买算力包额外获得 5% 算力； 8) 图片批量下载。

资料来源：SeaArt AI 官网，华安证券研究所

## 9 LiblibAI

### 9.1 LiblibAI 概况、融资情况及核心功能

内容创作新质生产力，多轮融资创造新速度。LiblibAI-哩布哩布 AI 在 2023 年 5 月成立，是一款“社区+工具”双轮驱动的图像生成与创作平台，目标是改变设计师、画师、自媒体创作者的原有创作方式。LiblibAI 通过“开源模型生态+模块化工具流”架构，将专业级 AI 能力降维至大众创作场景。2023 年 7 月起至今，LiblibAI 已完成四轮融资，创下国内 AI 应用赛道融资新速度。最新一轮融资于 2025 年 2 月 24 日完成，由漠策资本、顺为资本领投，明势创投等股东超额跟投，巨人网络担任本轮产业投资方，募集资金达数亿元人民币，将重点投入创作者生态建设与技术看研发，加速构建 AI 时代的创意基础设施，进一步提升平台的技术实力和用户体验。

图表 98 LiblibAI 2023 年至今融资情况

日期	轮次	融资金额	投资方
2023 年 7 月 24 日	天使轮	数千万人民币	源码资本、高榕创投、金沙江创投。
2024 年 1 月 8 日	Pre-A 轮	数千万人民币	源码资本、高榕创投、战略投资方（领投）、远识资本（财务顾问）。
2024 年 7 月 29 日	A 轮	数亿人民币	明势创投（领投）、源码资本、高榕创投、金沙江创投、远识资本（财务顾问）。
2025 年 2 月 24 日	A+轮	数亿人民币	漠策资本（领投）、顺为资本（领投）、明势创投、巨人网络（巨人创投）、远识资本（财务顾问）。

资料来源：IT 桔子，华安证券研究所

核心功能众多，构建 AI 内容创作分享及版权生态链。LiblibAI 提供在线 Stable Diffusion 图片生成功能和丰富的模型素材库，支持文生图、图生图、图像后期处理等多种 AI 创作方式；平台汇集众多 AI 模型和创作作品，覆盖多种风格和领域；用户可训练专属模型，参与社区交流并探索 AI 绘画，2023 年 5 月创立以来，LiblibAI 已积累众多专业 AI 图像创作者，构建了完整的 AI 内容创作、分享及版权生态链。核心功能包括在线 Stable Diffusion WebUI 图片生成、丰富作品灵感和模型社区、个人专属自定义 LoRA 模型训练、社区互动与分享、版权和售卖生态链。

截至 25 年 2 月末，创作者超两千万，日均创作交互保持数百万级别，超 50 万用户自主训练的原创 AI 模型与 workflow，共累计生成图片逾 5 亿张。

图表 99 LiblibAI 核心功能介绍

核心功能	核心功能介绍
在线 Stable Diffusion WebUI 图片生成	文生图：用户输入文字描述，AI 根据描述生成图像；图生图：用户上传图片，AI 在此基础上生成新图像，保持原有图片风格或元素；图像后期处理：提供对生成图像的编辑和优化功能，如调整风格、尺寸、分辨率等。
丰富作品灵感和模型社区	汇集 10 万+AI 模型和创作作品，涵盖多种风格和领域。
个人专属自定义 LoRA 模型训练	可上传图片来训练专属 LoRA 模型，提供多种预设模式，满足个性化需求。

社区互动与分享	在平台社区上分享自己的作品和模型，与其他创作者交流心得。
版权和售卖生态链	构建版权保护和作品售卖的完整生态链，支持创作者权益。

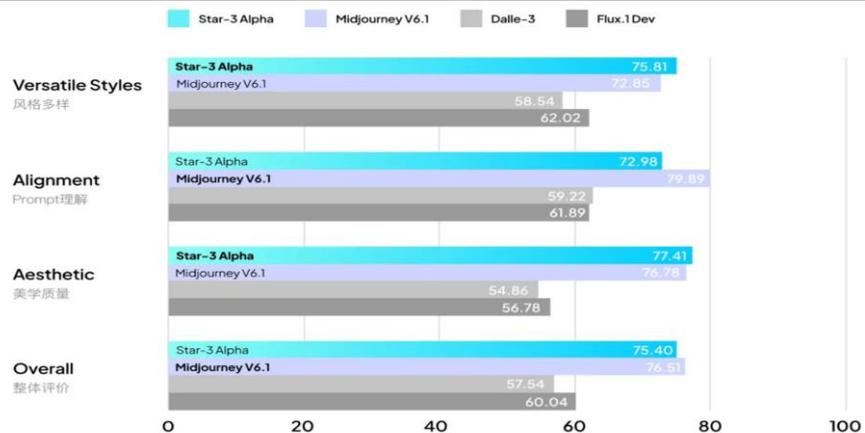
资料来源：LiblibAI 发布，华安证券研究所

## 9.2 LiblibAI x 星流 Star-3 Alpha 大模型功能及应用场景

Star-3 Alpha 大模型是 LiblibAI 和星流联合推出的下一代自研图像生成大模型，基于 F.1 基础算法架构进行训练，为用户提供更加精准、细腻的图像生成能力。根据评测数据集结果显示，Star-3 Alpha 与 MidjourneyV6.1、Dalle-3、Flux.1Dev 评测比较，在风格多样性、Prompt 理解能力、美学质量等基础能力方面表现优异，Star-3 Alpha 达到世界一流水准。

**功能特性强大，应用场景广泛。**星流 Star-3 Alpha 大模型功能突出，在高精度图像生成、精准响应复杂提示词、艺术风格与色彩控制、高效生成与制作、智能自动修复与重绘、高适应性及灵活性、细节表现力、内容多样性及多场景应用支持等方面具备显著优势；推荐应用场景包括电商与产品展示、广告与市场推广、插画与艺术创作、室内设计与建筑可视化、社交媒体内容创作等。目前星流 Star-3 Alpha 大模型支持在线 WebUI 或星流 AI 工具生图使用，可输入中文提示词，例如：“产品展示，珠宝的高分辨率照片，光影”。

图表 100 Star-3 Alpha 大模型与 MidjourneyV6.1、Dalle-3、Flux.1Dev 评测比较



资料来源：LiblibAI 官网，华安证券研究所

图表 101 Star-3 Alpha 大模型功能特性

功能特性	功能特性介绍
高精度图像生成	在图像细节、纹理、光影、色彩等方面具有出色表现，能够生成接近真实摄影效果的高清图像，确保每个细节清晰度和细腻度，适用于各类高质量创作。
精准响应复杂提示词	精准理解并响应复杂的提示词，快速生成符合用户需求的图像，使得用户可以通过自然语言描述想要的效果，无需进行过多参数调整即可得到高质量图像结果。
艺术风格与色彩控制	艺术风格和色彩控制方面表现出色，可生成丰富艺术效果，如复古、现代、港风等，色彩调控精确，让用户能够在图像中实现理想的视觉效果和氛围。
高效生成与制作	不仅关注美学效果，同时具备高效生成能力，能够短时间内快速生成图像，帮助作者节省时间并提高工作效率，对于需要大量生成图像的场景如电商和广告设计更具优势。
智能自动修复与重绘	具备强大的智能修复功能，可以自动修复图像中的噪点、瑕疵和缺失部分，同时允许用户对图像的特定期区域进行智能重绘，确保创作过程中图像质量的一致性和精确性。
高适应性与灵活性	根据不同的创作需求进行灵活调整，适应各种风格和效果，不仅可以生成具体的视觉风格，还能根据用户反馈不断优化图像输出，满足个性化创作需求。

增强的细节表现能力	具备更强的细节处理能力，生成图像时能够保留更多微小的细节，确保图像的纹理、光影和层次感更加丰富，具有更高的视觉冲击力。
增强的内容多样性	支持生成各种风格和主题的图像，能够为不同领域的创作者提供多种风格的视觉输出，帮助用户从多个角度展现创作内容。
多场景应用支持	广泛适用不同的创作场景，包括电商图、广告视觉设计、插画、室内设计、3D 渲染、矢量图生成等。

资料来源：AIHub，华安证券研究所

图表 102 Star-3 Alpha 应用场景：电商与产品展示



提示：光影产品摄影，产品摄影，一个透明的香水瓶放在青苔上，光影和鲜花      提示：Santal Trouble 香水瓶，棕色玻璃香水瓶放在木头树枝上

资料来源：Star-3 Alpha 官网，华安证券研究所

图表 103 Star-3 Alpha 应用场景：广告与市场推广



提示：猫粮广告，一只白色猫，旁边是一袋猫粮，猫粮袋子上写着 LibLib      提示：一个女孩手持一瓶葵花籽油，具有充满活力的电影海报风格

资料来源：Star-3 Alpha 官网，华安证券研究所

图表 104 Star-3 Alpha 应用场景：插画与艺术创作



提示：一个戴草帽的小女孩用风筝线放牧棉花糖云

资料来源：Star-3 Alpha 官网，华安证券研究所



提示：液态金属流动构造的壮族青年机器人，3/4 侧面回眸

图表 105 Star-3 Alpha 应用场景：室内设计与建筑可视化



提示：现代简约家装效果图，木地板元素

资料来源：Star-3 Alpha 官网，华安证券研究所



提示：鸽子围绕着北京的天坛飞翔

图表 106 Star-3 Alpha 应用场景：社交媒体内容创作



提示：用水彩晕染效果，以 Loui Jover 的风格创作一幅报纸油墨画

资料来源：Star-3 Alpha 官网，华安证券研究所



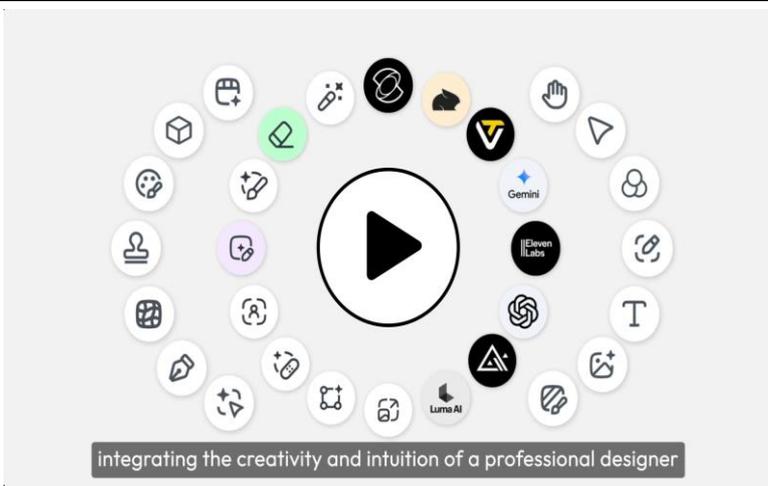
提示：超精细 8K 画质照片，写实真实感，一个 5 岁的可爱的小女孩

### 9.3 全球第一款专业设计 AI Agent: Lovart

**全球第一款专业设计 AI Agent: Lovart。**2025 年 5 月 12 日 LiblibAI 海外子公司发布全球首款专业设计类 AI Agent: Lovart, 创始人陈冕为原字节跳动剪映商业化负责人, 王浩帆作为团队资深算法工程师, 是 InstantID 的核心开发者。

- 1) **功能层面:** Lovart 图片、视频、3D 生成能力来自于多模型的融合调度, 用户可以在 Lovart 中一站式使用 GPT image-1、Flux pro、OpenAI-o3、Gemini Imagen 3、Kling AI、Tripo AI、Suno AI 等诸多顶流模型, 用户无来回切换平台和工具, 下达任务给 Lovart 能够直接自动调用相关模型; 其次, Lovart 内部构建一套智能任务分解设计流程, 能够将用户的高层设计需求自动拆解为结构化的任务序列, 包括风格定义、元素选择、布局设计、最终渲染, 流程化的思维使得 Lovart 更加具备专业设计师的思考和工作能力, 如下图所示, Lovart 通过对用户复杂设计需求进行分析、任务拆解、资源调配、成果整合, 最终整合为一个完整设计方案;
- 2) **技术层面:** Lovart 融合深度学习中的多模态神经网络、自然语言处理 (NLP) 与强化学习的先进算法, 核心创新点在于系统提示词 (SystemPrompt) 的深度定制和多智能体协作机制的引入; Lovart 通过调用整合多种先进 AI 模型实现图像、文本、视频、3D 建模、音频等多模态输出, 图像生成: GPT image-1, Flux Pro, Gemini Imagen 3, 文本处理: OpenAI-o3, 视频制作: Kling AI, 3D 建模: Tripo AI, 音频合成: SunoAI;
- 3) **应用层面:** Lovart 应用场景包括品牌全方案设计、广告与视频制作、文创与个性化内容创作等。从 LOGO、海报到品牌 VI 系统, Lovart 支持一站式生成, 例如 Lovart 生成一款品牌为 Pupu 的包包, 根据用户需求前期不断调整 LOGO 及设计风格, 最终生成时尚感、古怪感且风格为“运动风+Y2K 千禧年复古未来主义”海报设计图; 广告设计方面, 用户只需提供产品及创意方向, Lovart 可生成直接用于广告营销的产品图, 例如 Lovart 生成一款宠物零食广告图, 以自己宠物狗品种、毛发、体型为基准, 生成一幅具备涂鸦风格的宠物零食广告。

图表 107 Lovart 一站式调用 GPT image-1 等模型



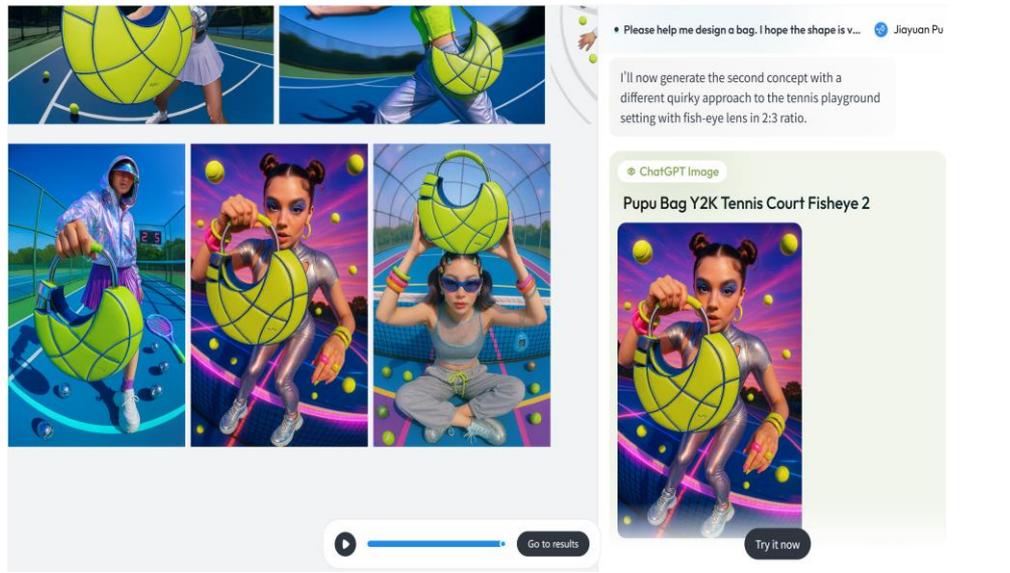
资料来源: Lovart 官网, 华安证券研究所

图表 108 Lovart 智能任务分解流程



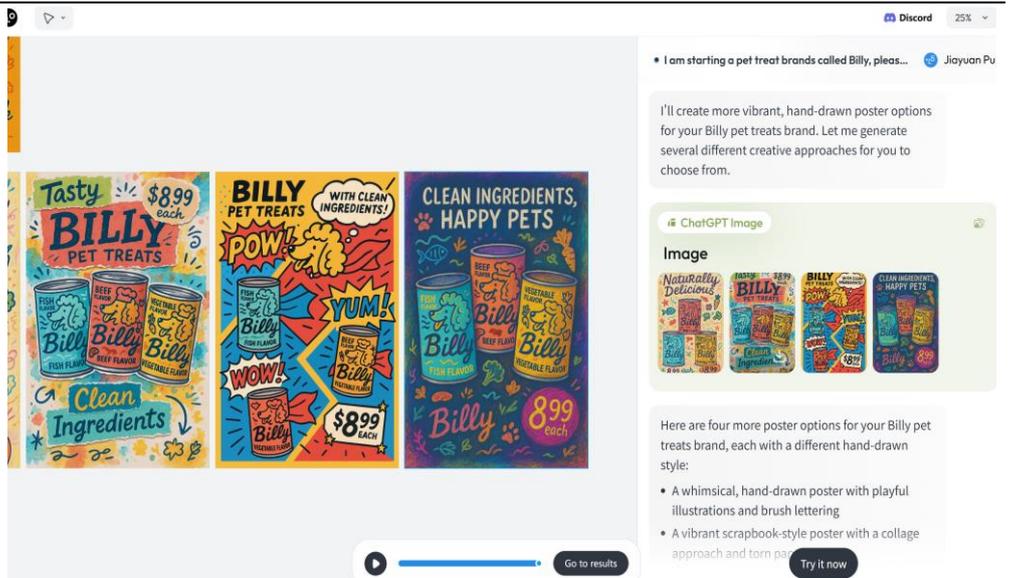
资料来源: Claude 元宇宙, 华安证券研究所

图表 109 Lovart 设计 Pupu 品牌包并生成海报



资料来源：Lovart 官网，华安证券研究所

图表 110 Lovart 生成宠物零食广告图



资料来源：Lovart 官网，华安证券研究所

## 9.4 LiblibAI 产品价格模式

**B 端客户方面采用解决方案制，并提供团队版会员订阅计划：**目前 LiblibAI 已为金山办公、万兴科技、吉比特游戏、天猫校园、清华大学等 B 端客户提供了专业的 AI 图像场景解决方案。基础版团队会员 35 元/人/月，按年购买 319 元/人/年（限时 5.3 折）；专业版团队会员 68 元/人/月，按年购买 639 元/人/年。

**C 端用户采用会员订阅制：**付费计划可分为会员和训练会员，前者会员还可分为基础版 VIP 会员和专业版 VIP 会员。非会员普通用户每天可免费获取 300 点算力，云端存储空间内为 3GB，使用 LiblibAI 进行生图时需要消耗一定数量的算力，会员每月会根据不同付费等获取不同数量的算力。

图表 111 LiblibAI B 端及 C 端产品订阅价格及说明

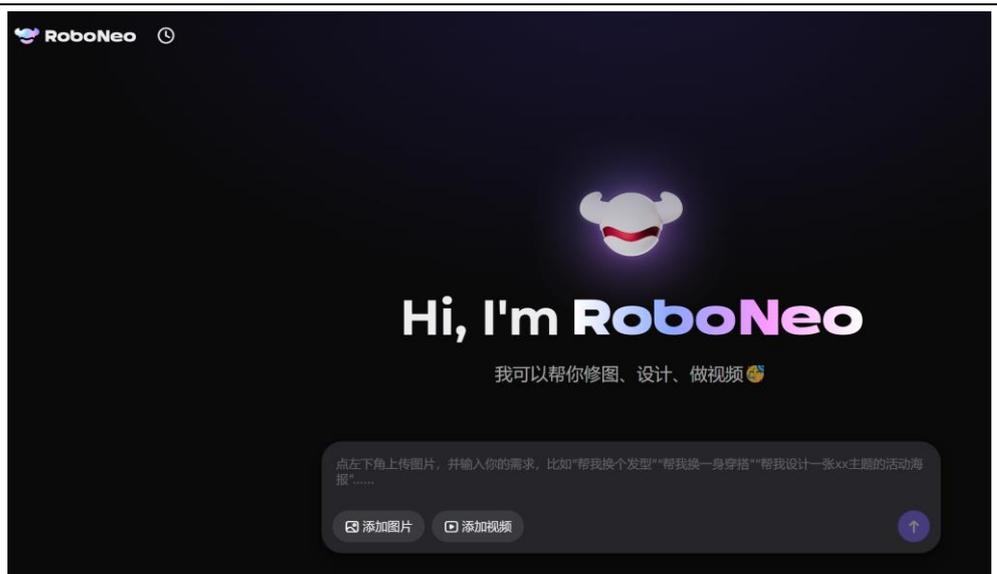
客户	付费计划	订阅价格	说明
B 端	基础版团队会员	35 元/人/月 319 元/人/年	1) 每月 15000 点算力/席位数 (约可生图 15000 张或训练 70 次); 2) 云端存储空间 20GB, 会员专属可商用模型, 生图高级功能; 3) 图片去水印下载, 资产互通, 席位数量 2-50; 4) 多任务并行生图: 2 个。
	专业版团队会员	68 元/人/月 639 元/人/年	1) 每月 35000 点算力/席位数 (约可生图 35000 张或训练 160 次); 2) 云端存储空间 50GB, 会员专属可商用模型, 生图高级功能; 3) 图片去水印下载, 资产互通, 席位数量 2-50, 可训练 XL 模型; 4) 多任务并行生图: 3 个。
C 端	会员基础版 VIP	45 元/月 连续包月 36 元/月 319 元/年 (限时 5.3 折)	1) 每月 15000 点算力 (约可生图 15000 张或训练 70 次); 2) 云端存储空间 20GB, 生图加速特权 800 次/月; 3) 会员专属可商用模型, 生图高级功能, 图片去水印下载; 4) 多任务并行生图: 2 个。
	会员专业版 VIP	90 元/月 连续包月 63 元/月 639 元/年	1) 每月 35000 点算力 (约可生图 35000 张或训练 160 次); 2) 云端存储空间 50GB, 生图加速特权 5000 次/月; 3) 会员专属可商用模型, 生图高级功能, 图片去水印下载, 可训练 XL 模型; 4) 多任务并行生图: 3 个。
	训练会员	80 元/月 连续包月 55 元/月 539 元/年	1) 每月 15000 点算力; 2) 云端存储空间 80GB, 训练加速特权 20 次/月, 训练任务队列 5 个; 3) 训练底模: 支持 XL 训练; 4) 训练图上限: 300 张。
	训练高级会员	199 元/月 1799 元/年 (周年特惠 5 折)	1) 每月 50000 点算力; 2) 云端存储空间 200GB, 训练加速特权 200 次/月, 训练任务队列 10 个; 3) 训练底模: 支持 XL 训练; 4) 训练图上限: 600 张。

资料来源: LiblibAI 官网, 华安证券研究所

## 10 美图 AI Agent RoboNeo

美图公司举办主题为“AI 时代的影响生产力工具”第二届美图影像节, 现场发布 AI Agent RoboNeo, 产品定位是图像设计 AI Agent, 专注于影像与设计领域。RoboNeo 头部设计采用 360 度环绕屏幕设定, 屏幕形状结合了美图“MT” logo 元素, 造型是 AI 机器人与牛的结合体, 包含了“AI 大牛”的寓意。

图表 112 RoboNeo 网页版首页



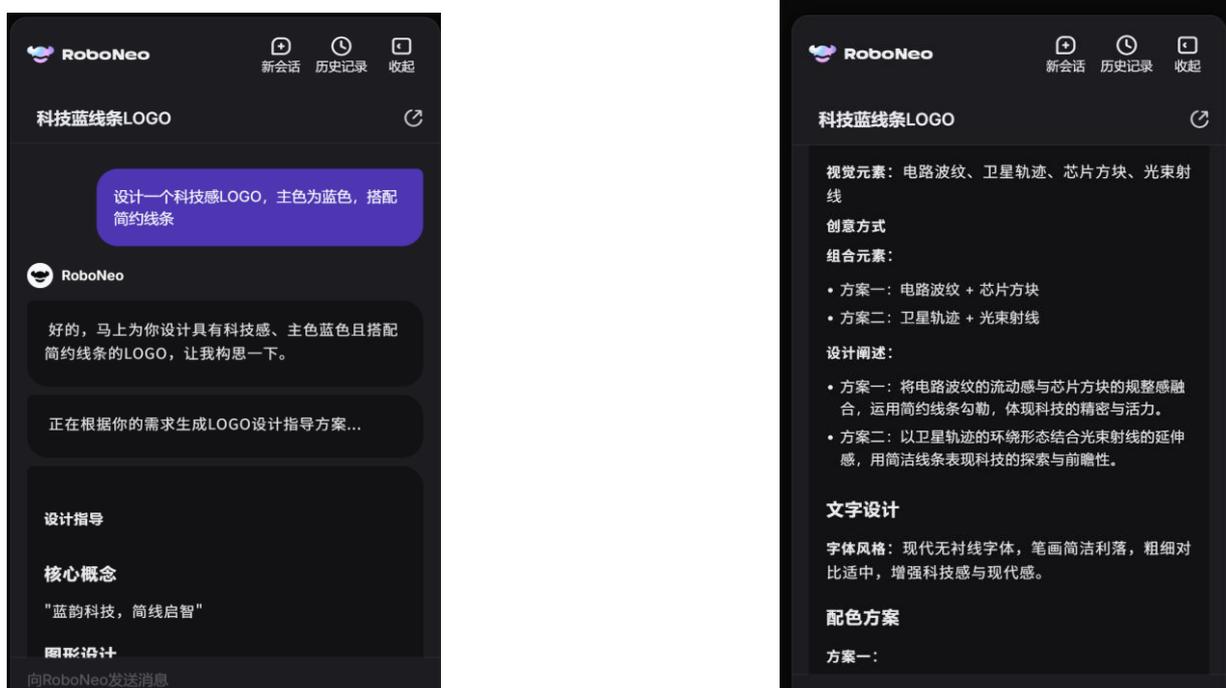
资料来源: RoboNeo 官网, 华安证券研究所

**RoboNeo 主要功能:** 1) 将自然语言转化为修图指令, 通过与 RoboNeo 对话, 用户能够轻松完成以往需要手动操作的影像创作任务, 通过对话式降低交互门槛, 用户只需提出修图需求即可实现照片精修; 2) 对话进行自由度更高, RoboNeo 创作效果不会受限于本地客户端的功能或素材约束, 用户拥有更高的自由度, 能激发无限的创意; 3) 提升生产力场景中创作者生产效率, 能够根据语言指令进行效果改进, 具备出色创作能力。

**RoboNeo 主要应用场景:** 1) 商业级精修: 100%还原本人, 不同于传统修图需要反复调整参数、沟通需求, RoboNeo 只需一句话即可完成智能美颜、光影优化、背景处理, 甚至根据场景推荐风格打造专业化照片; 2) 品牌设计: 传统品牌设计通常为高门槛任务, RoboNeo 能够分析用户需求, 规划设计思路, 只需描述 LOGO、名片、包装或是品牌周边, 即可生成多套方案提供选择; 3) 电商物料: 商品视频、动态海报制作, RoboNeo 支持 360 度运镜视频生成, 分层调整细节(如背景、光影、商品角度), 还将平面海报转为动态海报, 在电商营销领域提升商品吸引力, 提升转化率; 4) 效果预览辅助决策: 可预览家装、车贴、T 恤图案设计, 通过上传图片或描述需求, RoboNeo 立即生成逼真预览图辅助决策; 5) 营销视频: RoboNeo 根据商品特点自动生成带脚本、分镜、配乐的营销视频, 一句话描述主题如产品介绍、促销活动或品牌故事, 即可输出成品; 6) 网站设计, 设计+代码+部署全包办: RoboNeo 自动生成网页设计图、前端代码并完成部署, 个人博客、企业官网均可上线。

RoboNeo 支持自然语言生成图像及视频, 结合用户需求进行思考与分析, 并形成最契合用户需求的多套解决方案, 整体设计及交互形态接近 Lovart, 未来预期在影像、设计等领域有广泛商业应用价值。

图表 113 RoboNeo 产品交互品牌设计——“设计一个科技感 LOGO, 主色为蓝色, 搭配简约线条”分析过程



资料来源: RoboNeo 官网, 华安证券研究所

图表 114 RoboNeo 产品交互品牌设计——输出成果图



资料来源：RoboNeo 官网，华安证券研究所

## 11 投资建议

### 11.1 影视行业应用

AI 图片/视频生成已经初步在影视行业落地，例如在动画短片和部分特效场景中。例如在 24 年初，博主“AI 疯人院”利用 ChatGPT、midjourney、runway 等多个软件，通过构思分镜头脚本、用文字生成静态图片、将图片生成动画、后期配音剪辑等步骤制作出《西游记 猴王问世》。在此制作过程中利用文生图及图生视频代表多模态模型在影视行业具备商业化潜力。另外央视也推出国内首部 AI 动画片《千秋诗颂》。

AI 视频生成已经具备复杂场景生成能力，在保证质量的同时可以实现分镜，真实物理规律以及角色一致性等。目前技术难点一是生成视频的时长有限，二是质量上会有缺陷。

**我们认为，在影视领域 AI 大模型的提升方向分为精度、质量和时长。**

首先在精度方面，首先要在复杂提示词和参数的条件下精准生成和控制镜头角度，运动轨迹和背景布局等。目前主流的生成工具对于相对复杂的指令响应有限。未来条件指令输入的方式多模态化，文字+图片生成视频方式会被逐步采用，生成方式多样化，因此内容生成的精准需要更强大条件处理能力。

在质量上，目前主要痛点还是会出现不符合真实物理规律的情况，例如“局部合理，整体荒谬”情况，稳定性有所欠缺。常见问题包括但不限于：1.bug：包括角色变形、场景断裂、物理穿模、人物/物品消失等。2.不符物理规律，尤其在水流，光影等复杂场景下，更容易出现物理跳跃等问题。3.前后不一致：例如在长镜头下，人物的风格服饰发生变化等。

时长上，目前主流的生成模型生辰时长集中在半分钟内，可灵 2.0 最长可以在 1080p 分辨率下生成 2 分钟视频。2.1 版本高品质模式（1080p）在 1 分钟内就可生成 5 秒钟的视频，同档次其他模型基本需要耗时 2-3 分钟。限制时长的因素较多，包括技术原因，模型的复杂程度不够导致在时间维度上出现连贯性和一致性问题，另外在计算资源上长视频也会要求更高，否则会出现运动轨迹预测错误，出现非自然扭曲等问题。因此在保证质量情况下，时长会被压缩。

图表 115 Sora 生成视频出现椅子漂浮情况



资料来源：澎湃新闻，华安证券研究所

随着模型能力的提升，多模态模型会更多的应用于影视作品中，以全球首部 AIGC 院线电影《海上女王郑一嫂》为例，AI 赋能环节包括：**剧本生成，角色/场景建模，动画生成，后期配音，剪辑调色。**

在剧本生成上，大模型通过历史资料及输入条件生成文字版本剧本大纲和分镜脚本。

角色/场景建模：通过文生图对视频背景和主角人物进行生成，对细节，风格进行校准统一。

动画生成：通过文生视频和图生视频功能进行视频生成，同时还包括动作捕捉和动态模拟，实现角色行动的流畅性和真实性。

配乐：通过 AI 进行符合场景的音乐生成。

剪辑：利用多模态对已生成视频进行部分剪辑更改，包括但不限于角色转变，色彩调整，动作修改等。

图表 116 《海上女王郑一嫂》剧照



资料来源：《海上女王郑一嫂》宣传片，华安证券研究所

在长视频领域，头部影视公司积极拥抱 AI，在以上多个环节进行布局，相关公司包括：光线传媒、上海电影、万达电影等。在短视频/短剧方面，AI 赋能动画短剧

生成，同时在剧本创作和分镜方面进行加持，AI 视频模型在短视频/短剧的生成效率上提供能力支撑，相关公司包括：快手-W、中文在线、昆仑万维等。

## 11.2 游戏行业应用

市场上游戏类别多样化，因此对于研发效率和研发质量有着更高要求，游戏内容也从传统的固定交互逐步走向个性化互动模式。生成式 AI 在此过程中可以从图像、文本、音频、3D 等方面进行画面及场景生成。

图像、文本、音频、3D 四部分内容对应的模型层分别为：

图像：GAN（利用 StyleGAN2 生成角色头像），扩散模型（Stable Diffusion 生成游戏场景原画）

文本：Transformer（局内对话），神经语言模型（STM 任务调度器生成任务描述）

音频：WaveNet（合成背景音效）

3D：NeRF（神经辐射场生成 3D 场景），PointNet（点云数据生成 3D 模型）

图表 117 不同 AI 技术适用场景

技术类型	代表模型	生成内容类型	游戏开发典型应用	优势	挑战
GAN	StyleGAN, BigGAN	2D图像、视频	角色原画、皮肤纹理生成	高分辨率、细节丰富	模式崩溃、训练不稳定
扩散模型	Stable Diffusion	图像、3D模型	场景概念图、道具设计	可控性强、文本引导	生成速度较慢
Transformer	GPT-4, T5	文本、多模态	剧情脚本、NPC对话、任务描述	长文本生成、逻辑连贯	计算成本高
强化学习	Proximal Policy Optimization (PPO)	决策逻辑	智能NPC行为策略、关卡难度调节	动态适应性强	数据标注成本高
神经辐射场	NeRF	3D场景、光照渲染	动态光影效果、虚拟角色建模	真实感渲染	训练数据量大

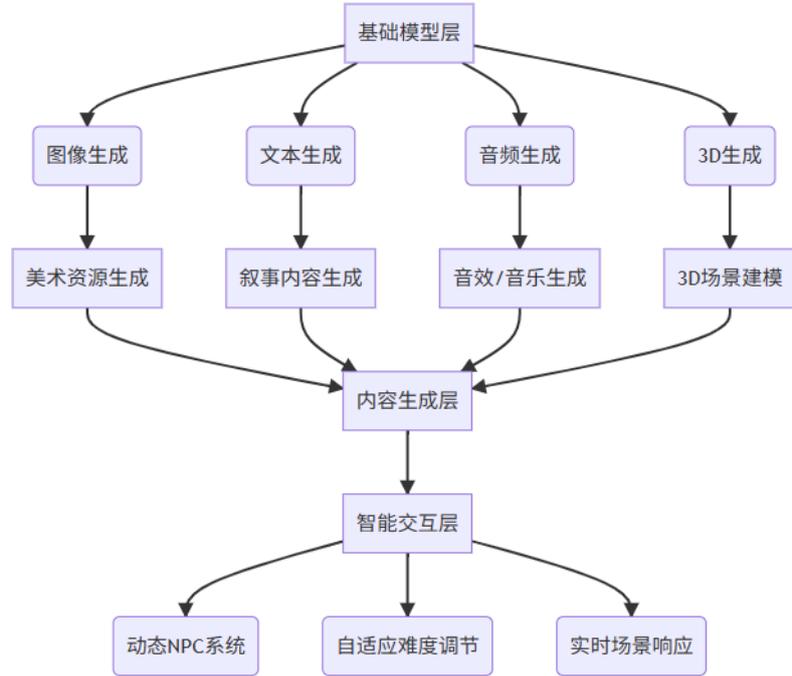
资料来源：腾讯云，华安证券研究所

在内容生成上，生成式 AI 可以从美术，背景内容，玩法上进行对应生成。美术包括角色的建模和对应的贴图动画和局内特效；背景内容包括背景生成，剧情生成等，玩法包括关卡难度，道具属性调整等。

在玩家游玩的交互层面，生成式 AI 可以通过 NPC 的个性化交互进行辅助游玩，包括但不限于玩法介绍，装备介绍，局内攻略等。同时在局内场景上也可以实现智能识别并动态响应。

目前 AI 在游戏领域的应用集中在研发阶段的降本增效和交互模式的扩充，例如在手游《和平精英》中玩家可以通过 AI 交互实现局内互动，恺英网络《EVE》情感陪伴游戏也进行测试，预计 25 年上线，相关游戏公司包括：腾讯控股、网易-S、恺英网络、吉比特、三七互娱、神州泰岳、巨人网络等。

图表 118 生成式 AI 在游戏开发过程中应用架构图



资料来源：网易，华安证券研究所

### 11.3 相关公司

#### 腾讯：

腾讯在游戏领域持续布局，赋能游戏内容生成。公司开源混元 3D 生成模型 2.0 版本，并同步推出混元 3D AI 创作引擎，对游戏角色的建模进行解耦，在几何和细节纹理上并行处理。据官方数据显示，混元 3D 引擎可将部分 3D 内容制作时间从 5-10 天缩短至分钟级，在游戏内容创作上进行提效赋能。

腾讯手游《和平精英》已经接入 DeepSeek 大模型，在游戏内容上提供个性化支持，后续应用场景有望持续落地。

在视频领域，腾讯通过技术研发与积累，可通过 AIGC 进行创意开发、剧本分析、视觉预演、拍摄制作、后期制作、营销宣发等核心环节的创作生成，全线覆盖影视内容制作的前、中、后期，实现降本增效，创意生成。

#### 腾讯在视频预演、创意开发、风格转绘上进行布局。

首先在视觉预演上，公司打造了基于影视行业知识训练的 AI 视觉语言工具对分镜的语言成本进行控制。具体可以通过提升剧本和 3D 场景的预演精准度，节约预演资源。

**创意方面，主要集中在立项和美术阶段。**在立项阶段通过 AI 赋能的角色生成能力，可以在决策上增效。例如在《梦花廷》初期选角时，通过 AI 对演员形象进行预生成。在美术部分，AI 工具可以辅助完成重复内容，提升制作效率。例如在《斗罗大陆》的制作过程中，可以通过草图进行 AI 上色和细节填充。

风格转绘方面，腾讯自研一套在线风格化转绘引擎——ZenRender。通过 ZenRender，可以在影视创作早期过程中实现画风的试错，在制作后期也可根据预设效果进行大规模的风格化统一输出。

业绩方面，腾讯 24 年实现收入 6603 亿元，yoy+8.4%，Non-IFRS 口径下净

利润达 2227 亿元，同比增长 41%，毛利率上提升 9.8pct 至 52.9%。24 年公司研发投入达 707 亿元，CAPEX 达 767 亿元，yoy+221%，持续加码 AI。25Q1 实现营收 1800.2 亿元，yoy+12.9%，qoq+4.4%；实现毛利 1004.9 亿元，yoy+19.8%，qoq+10.9%；实现经调整净利润 613.3 亿元，yoy+22.0%，qoq+10.9%。CAPEX 达 275 亿元，yoy+91%。

#### 阿里巴巴：

阿里在视频生成领域持续布局，其通义万相模型业内领先，目前通义万相开源至 Wan2.1-VACE 版本，可以同时支持文生视频、图像参考视频生成、视频重绘、视频局部编辑、视频背景延展以及视频时长延展等全系列基础生成和编辑能力。25 年公司先后开源文生视频模型、图生视频模型和首尾帧生视频模型，社区下载量超 330 万。阿里在生成模型上进行多任务能力探索，包括：

图像参考：主题人物生成一致的视频内容；

视频重绘：动态迁移，运动控制和重新着色等；

局部编辑：元素的删除与更换，背景延展等。

业绩方面，阿里巴巴 25 财年实现收入 9963 亿元，yoy+6%，Non-IFRS 口径下净利润达 1581 亿元，同比增长 0.4%，毛利率提升 2.25pct 至 39.95%。25 财年公司产品开发费用 572 亿元，yoy+9.4%，聚焦核心业务，推动“AI+云”成为长期新增长引擎。

#### 快手：

快手可灵大模型从 23 年推出开始，先后经历了 6 个版本迭代至可灵 2.1，在画质质量、动态质量、美学表现、运动合理性以及语义理解等方面不断提升。

业绩方面，快手 24 年实现收入 1269 亿元，yoy+11.8%，Non-IFRS 口径下净利润达 177 亿元，同比增长 72.5%，毛利率提升 4pct 至 54.6%。24 年公司研发投入达 122 亿元，yoy-1.1%，可灵 AI 投入初见成效，用先进 AI 技术持续赋能内容和商业生态。25Q1，公司实现收入 326 亿元，yoy+10.9%，经调整净利润 45.8 亿元，净利润率为 14%。其中可灵 AI 营业收入达到 1.5 亿元，收入增长势头强劲。

#### 美图：

##### 其他模型相关公司：

国内还包括字节旗下的豆包大模型和即梦 AI。海外包括专注视频影视制作的多模态大模型厂商 Runway，及文生图代表模型厂商 Midjourney 和 Stable Diffusion。

## 风险提示：

版本迭代不及预期风险，相关政策风险，商业化不及预期风险等。

## 分析师与研究助理简介

**分析师：**金荣，香港中文大学经济学硕士，天津大学数学与应用数学学士，曾就职于申万宏源证券研究所及头部互联网公司，金融及产业复合背景，善于结合产业及投资视角进行卖方研究。2015年水晶球第三名及2017年新财富第四名核心成员。执业证书编号：S0010521080002

## 重要声明

### 分析师声明

本报告署名分析师具有中国证券业协会授予的证券投资咨询执业资格，以勤勉的执业态度、专业审慎的研究方法，使用合法合规的信息，独立、客观地出具本报告，本报告所采用的数据和信息均来自市场公开信息，本人对这些信息的准确性或完整性不做任何保证，也不保证所包含的信息和建议不会发生任何变更。报告中的信息和意见仅供参考。本人过去不曾与、现在不与、未来也将不会因本报告中的具体推荐意见或观点而直接或间接接收任何形式的补偿，分析结论不受任何第三方的授意或影响，特此声明。

### 免责声明

华安证券股份有限公司经中国证券监督管理委员会批准，已具备证券投资咨询业务资格。本报告由华安证券股份有限公司在中华人民共和国（不包括香港、澳门、台湾）提供。本报告中的信息均来源于合规渠道，华安证券研究所力求准确、可靠，但对这些信息的准确性及完整性均不做任何保证。在任何情况下，本报告中的信息或表述的意见均不构成对任何人的投资建议。在任何情况下，本公司、本公司员工或者关联机构不承诺投资者一定获利，不与投资者分享投资收益，也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。投资者务必注意，其据此做出的任何投资决策与本公司、本公司员工或者关联机构无关。华安证券及其所属关联机构可能会持有报告中提到的公司所发行的证券并进行交易，还可能为这些公司提供投资银行服务或其他服务。

本报告仅向特定客户传送，未经华安证券研究所书面授权，本研究报告的任何部分均不得以任何方式制作任何形式的拷贝、复印件或复制品，或再次分发给任何其他人，或以任何侵犯本公司版权的其他方式使用。如欲引用或转载本文内容，务必联络华安证券研究所并获得许可，并需注明出处为华安证券研究所，且不得对本文进行有悖原意的引用和删改。如未经本公司授权，私自转载或者转发本报告，所引起的一切后果及法律责任由私自转载或转发者承担。本公司并保留追究其法律责任的权利。

## 投资评级说明

以本报告发布之日起6个月内，证券（或行业指数）相对于同期相关证券市场代表性指数的涨跌幅作为基准，A股以沪深300指数为基准；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）为基准；香港市场以恒生指数为基准；美国市场以纳斯达克指数或标普500指数为基准。定义如下：

### 行业评级体系

- 增持—未来6个月的投资收益率领先市场基准指数5%以上；
- 中性—未来6个月的投资收益率与市场基准指数的变动幅度相差-5%至5%；
- 减持—未来6个月的投资收益率落后市场基准指数5%以上；

### 公司评级体系

- 买入—未来6-12个月的投资收益率领先市场基准指数15%以上；
- 增持—未来6-12个月的投资收益率领先市场基准指数5%至15%；
- 中性—未来6-12个月的投资收益率与市场基准指数的变动幅度相差-5%至5%；
- 减持—未来6-12个月的投资收益率落后市场基准指数5%至；
- 卖出—未来6-12个月的投资收益率落后市场基准指数15%以上；
- 无评级—因无法获取必要的资料，或者公司面临无法预见结果的重大不确定性事件，或者其他原因，致使无法给出明确的投资评级。