

2025 年 12 月 31 日



华鑫证券
CHINA FORTUNE SECURITIES

MiniMax 发布 MiniMax M2.1 大模型，清华大学发布 TurboDiffusion

—计算机行业周报

推荐(维持)

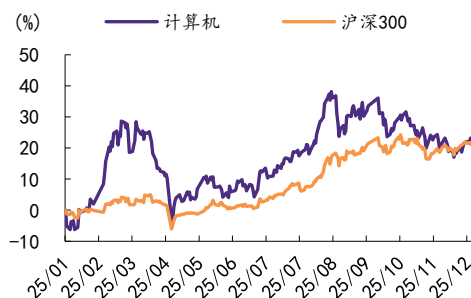
分析师：任春阳 S1050521110006

rency@cfsc.com.cn

行业相对表现

表现	1M	3M	12M
计算机(申万)	-0.2	-7.6	18.2
沪深 300	2.3	-0.2	17.7

市场表现



资料来源：Wind，华鑫证券研究

相关研究

- 1、《Monday.com(MNDY.O)：AI 多产品矩阵超预期贡献，获客渠道转型筑基长期增长》2025-12-31
- 2、《ServiceNow(NOW.N)：智能体 AI 驱动平台价值重塑，业绩指引上修》2025-12-31
- 3、《HubSpot(HUBS.N)：智能代理平台驱动产品化转型，多产品协同带动大客户渗透》2025-12-31

投资要点

算力：算力租赁价格平稳，MiniMax 发布 MiniMax M2.1 大模型

12.23 日，MiniMax 发布 MiniMax M2.1 大模型。模型实现多项关键技术突破，在 SWE-bench Multilingual 多语言评测中斩获 72.5% 的 SOTA 成绩，性能超越多款主流竞品。为支撑其全栈开发能力，MiniMax 同步开源 VIBE 基准测试体系，模型在该测试中以 88.6 分展现卓越性能。作为率先引入 Interleaved Thinking 的开源模型系列，M2.1 系统性问题解决能力升级，同时实现响应速度提升与 Token 消耗下降。此次突破补齐了 MiniMax 在逻辑推理与代码生成领域的短板，为 MiniMax 技术产业化及市场拓展奠定关键基础。

AI 应用：Discord 周访问量环比+9.44%，清华大学发布 TurboDiffusion

2025 年 12 月 23 日，由清华大学 TSAIL 实验室与生数科技联合研发并开源发布了视频生成加速框架 TurboDiffusion，标志着这一领域取得了突破性进展。该框架的推出在全球人工智能社区内引发了强烈反响，获得了包括 OpenAI、Meta 以及 vLLM 在内的众多顶尖机构与开源社区研究者及工程师的关注与认可。它成功地将 AI 视频生成从传统的“分钟级”耗时，一举推进到了“秒级”乃至“实时生成”的新时代。该框架能够在几乎不影响最终视频质量的前提下，实现高达 100 至 200 倍的加速。

AI 融资动向：LemonSlice 获千万融资，押注 AI 实时交互头像素赛道

AI 初创企业 LemonSlice 宣布完成 1050 万美元种子轮融资，投资方涵盖知名机构与行业大咖，融资将用于迭代实时交互式头像素技术及推动商业落地。该公司 2024 年成立且为 YCombinator 毕业企业，核心产品 LemonSlice-2 模型（200 亿参数）可将单张静态图转化为实时交互的视频角色，支持全身形象、多风格生成及文本操控，凭借因果注意力、量化等技术实现 2.8 秒平均响应，视频生成耗时仅占 26%。其核心优势在于无需海量数据即可实时创建角色，技术由大规模视频扩散变换器驱动，交互性与自然度更优。此次融资印证 AI 正从文本向多模态交互演进，也为中国创业者提示了本地化、情感化 AI 交互界面的差异化竞争方向，尤其在教育等垂直领

域潜力显著。

投资建议

2025 年 12 月 25 日消息，英伟达宣布以约 200 亿美元现金通过非排他性授权协议收购 AI 芯片新创公司 Groq 的核心资产，这笔交易成为英伟达史上最大收购案，其核心内容包括获得 Groq 全部资产与技术授权、吸纳 Groq 创始人及核心团队成员，而 Groq 旗下 GroqCloud 云端业务将保持独立运营。公开资料显示，Groq 成立于 2016 年，核心团队源自谷歌 TPU 工程团队，具备深厚的 AI 芯片研发积淀。该公司推出的基于 TSA 架构的 LPU 芯片，在推理速度与能耗性价比方面表现亮眼，性能显著优于常规 GPU 及 TPU 产品。截至目前，Groq 已为超 200 万名开发者及多家财富 500 强企业提供服务，业务覆盖范围与行业认可度持续提升。2025 年 9 月，Groq 完成 7.5 亿美元新一轮融资，投后估值达 69 亿美元，为此次收购奠定了重要基础。此次收购得益于英伟达丰厚的现金储备，也是其巩固 AI 芯片领域领导地位的重要举措，近期英伟达还开展了对 OpenAI、英特尔等的投资及与新思科技的战略合作等多项生态布局。英伟达通过此次战略举措，成功将 Groq 的 LPU 技术纳入麾下，进一步卡位高效推理这一关键赛道。Groq 的 LPU 在特定推理任务中具备显著优势，与英伟达生态形成完美互补，强化了其全栈护城河。这一动作不仅印证了算力赛道的高景气，更凸显了推理端的战略价值。龙头通过整合不断抬高行业门槛，我们维持对 AI 算力板块的长期看好态度。

中长期，建议关注新能源业务高增并供货科尔摩根等全球电机巨头的唯科科技（301196.SZ），AI 智能文字识别与商业大数据领域巨头的合合信息（688615.SH）、深耕工业 AI 与软件并长期服务高端装备等领域头部客户的能科科技（603859.SH）。

风险提示

- 1) AI 底层技术迭代速度不及预期。2) 政策监管及版权风险。3) AI 应用落地效果不及预期。4) 推荐公司业绩不及预期风险。

重点关注公司及盈利预测

公司代码	名称	2025-12-31	EPS			PE			投资评级
		股价	2024	2025E	2026E	2024	2025E	2026E	
301196.SZ	唯科科技	80.57	1.76	2.53	3.34	45.78	31.85	24.12	买入
603859.SH	能科科技	42.05	0.78	0.96	1.18	53.91	43.80	35.64	买入
688615.SH	合合信息	227.58	4.01	3.37	4.11	56.75	67.53	55.37	买入

资料来源：Wind，华鑫证券研究

正文目录

1、 算力动态：算力租赁价格平稳，MINIMAX 发布 MINIMAX M2.1 大模型 4

 1.1、 Tokens 跟踪..... 4

 1.2、数据跟踪：算力租赁价格平稳 5

 1.3、产业动态：MiniMax 拓展核心板块，发布 MiniMax M2.1 大模型..... 5

2、 AI 应用动态：DISCORD 周访问量环比+9.44%，清华大学发布 TURBODIFFUSION 8

 2.1、 周流量跟踪：Discord 周访问量环比+9.44%..... 8

 2.2、 产业动态：清华大学发布 TurboDiffusion，引领 AI 视频生成进入秒级实时时代 8

3、 AI 融资动向：LEMONSLICE 获千万融资，押注 AI 实时交互头像素赛道 12

4、 行情复盘 14

5、 投资建议 16

6、 风险提示 16

图表目录

图表 1: TOKENS 规模 LEADERBOARD 示意..... 4

图表 2: 市场份额占据示意 4

图表 3: 上周算力租赁情况 5

图表 4: SOTA 对比示意图..... 5

图表 5: VIBE 测试结果示意..... 6

图表 6: SWE-BENCH VERIFIED 测试示意图..... 6

图表 7: 2025.12.19-2025.12.25AI 相关网站流量..... 8

图表 9: VLLM 上对 TURBODIFFUSION 的评价..... 9

图表 10: RTX5090 显卡上 TURBODIFFUSION 视频生成速度提升的表现..... 9

图表 11: TURBODIFFUSION 以 1.3B 大小的模型生成 5 秒的视频的表现..... 10

图表 12: TURBODIFFUSION 以 14B 大小的模型生成 5 秒 720P 的视频的表现..... 10

图表 12: 上周 AI 初创公司融资动态 13

图表 13: 上周（12.22-12.26 日）指数日涨跌幅..... 14

图表 14: 上周（12.22-12.26 日）AI 算力指数内部涨跌幅度排名 14

图表 15: 上周（12.22-12.26 日）AI 应用指数内部涨跌幅度排名 15

图表 16: 重点关注公司及盈利预测 16

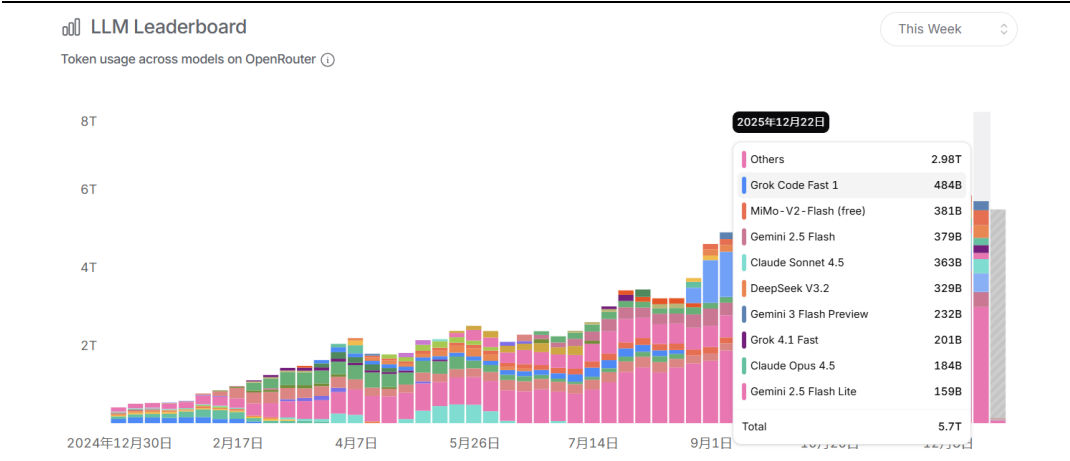
1、算力动态：算力租赁价格平稳，MiniMax 发布 MiniMax M2.1 大模型

1.1、Tokens 跟踪

根据 OpenRouter 公开数据，2025 年 12 月 22 日至 12 月 28 日，周度 token 消耗量有所上升，调用量为 5.7T，环比上周-2.56%。在 tokens 规模 leaderboard 前五名中，xAI 旗下 GrokCodeFast1 以 484Btokens 位居榜首；xiaomi 的 MiMo-V2-Flash(free) 以 381B 位列第二；Google 的 Gemini2.5Flash 以 379Btokens 排第三；Anthropic 的 ClaudeSonnet4.5 和 deepseek 的 DeepSeekV3.2 分别以 363B、329Btokens 位列第四、第五。

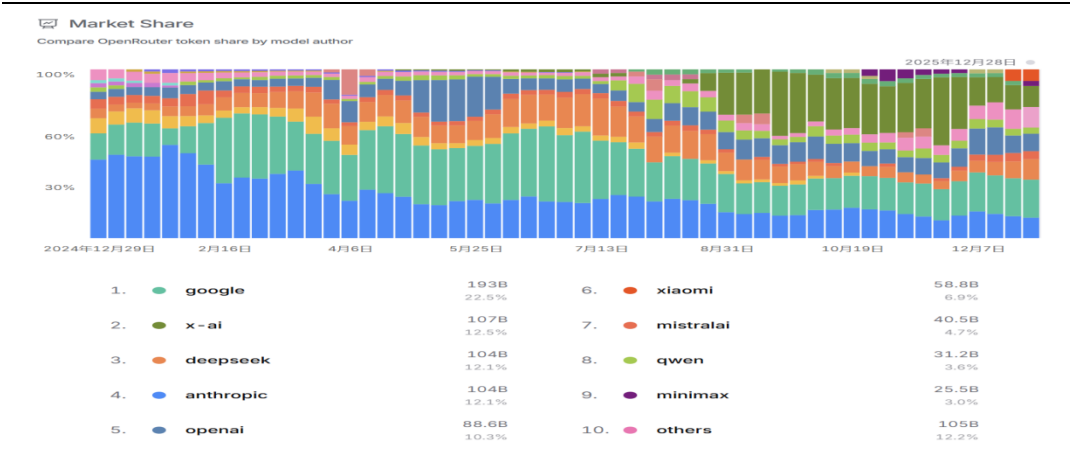
从市场份额维度来看，Google 以 193Btokens 占据 22.5% 的份额，稳居首位；xAI 以 107Btokens 占比 12.5%，位列第二；Deepseek、Anthropic、OpenAI 则分别以 104B、104B、88.6Btokens，对应占据 12.1%、12.1%、10.3% 的市场份额。

图表 1：tokens 规模 leaderboard 示意



资料来源：OpenRouter，华鑫证券研究

图表 2：市场份额占据示意



资料来源：OpenRouter，华鑫证券研究

1.2、数据跟踪：算力租赁价格平稳

上周算力租赁价格平稳。具体来看，显卡配置为 A100-40G 中，腾讯云 16 核+96G 价格为 28.64 元/时, 阿里云 12 核+94GiB 价格为 31.58 元/时；显卡配置为 A800-80G 中，恒源云 16 核+256G 价格为 7.50 元/时。

图表 3：上周算力租赁情况

显卡配置	CPU	内存	磁盘大小 (G)	平台名称	价格（每小时）	价格环比上周
A100-40G	16	96	可自定，额外收费	腾讯云	28.64/元	6.19%
	12 核	94G	可自定，额外收费	阿里云	31.58/元	0.00%
A100-80G	13	128	系统盘：20G 数据盘：50GB	恒源云	-	0.00%
A800-80G	16	256	系统盘：20G 数据盘：50GB	恒源云	7.50/元	0.00%

资料来源：腾讯云，阿里云，恒源云，华鑫证券研究

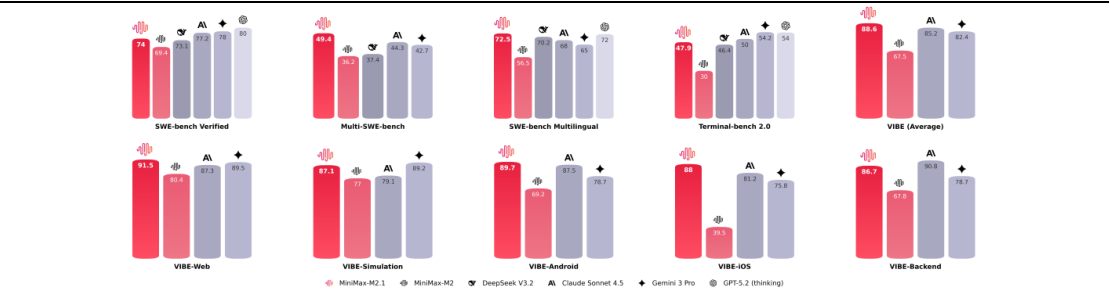
1.3、产业动态：MiniMax 拓展核心板块，发布 MiniMax M2.1 大模型

12 月 23 日，MiniMax 发布了其最新的 MiniMax M2.1 大模型。该模型实现多项核心技术突破，不仅在多语言代码评测中取得最优性能表现，更补齐逻辑推理与代码生成两大核心模块短板；其全栈开发能力、系统问题解决能力及工具调用能力的显著升级，为公司技术产业化落地及市场拓展奠定关键基础。

代码评测：SWE-bench Multilingual 斩获 SOTA

据官方披露信息，M2.1 模型在 SWE-bench Multilingual 多语言评测中以 72.5% 的成绩斩获当前最优结果（SOTA），性能超越 Gemini 3 Pro 与 Claude Sonnet 4.5。模型突破了此前局限于 Python 或前端代码生成的能力边界，全面覆盖 Rust、Java、C++ 等后端编程语言，针对性解决了过往大模型代码可运行性差、工程实用性不足等问题。

图表 4：SOTA 对比示意图



资料来源：机器之心，华鑫证券研究

开发能力：强化移动应用，推动 vibe coding 落地

与此同时，M2.1 大幅强化了原生 Android 与 iOS 应用开发能力。MiniMax 系统性提升了模型在 Web 与 App 场景中的设计理解，能够出色地构建复杂交互、3D 科学场景模拟与高质量可视化表达，推动 vibe coding 成为可持续、可交付的生产实践。

测试体系：开源 VIBE 基准，全栈性能卓越

为支撑上述提到的全栈开发能力，MiniMax 同步构建并开源全新基准测试体系 VIBE。与传统基准测试不同，VIBE 覆盖 Web 开发、仿真测试、Android 开发、iOS 开发及后端开发五大核心领域，并创新引入 Agent-as-a-Verifier 范式，可自动评估生成应用在真实运行环境中的交互逻辑与视觉呈现效果。在该全栈构建能力终极测试中，M2.1 模型以平均 88.6 分的成绩展现出卓越性能，不仅在几乎所有测试子集上显著优于 Claude Sonnet 4.5，更接近 Claude Opus 4.5 的性能水平。

图表 5：VIBE 测试结果示意

Benchmark	MiniMax-M2.1	MiniMax-M2	Claude Sonnet 4.5	Claude Opus 4.5	Gemini 3 Pro
VIBE (Average)	88.6	67.5	85.2	90.7	82.4
VIBE-Web	91.5	80.4	87.3	89.1	89.5
VIBE-Simulation	87.1	77	79.1	84	89.2
VIBE-Android	89.7	69.2	87.5	92.2	78.7
VIBE-iOS	88	39.5	81.2	90	75.8
VIBE-Backend	86.7	67.8	90.8	98	78.7

资料来源：机器之心，华鑫证券研究

问题解决：升级系统能力，提升实操效率

M2.1 模型在实际问题解决层面亦有突出表现。作为开源模型中率先系统性引入 Interleaved Thinking 的模型系列，M2.1 systematic problem-solving 能力再次升级。模型不仅关注代码执行是否正确，同时关注模型对“复合指令约束”的整合执行能力，在真实办公场景具备更高的可用性。并且，MiniMax M2.1 的模型回复以及思维链更加简洁，在实际编程与交互体验中，响应速度显著提升，Token 消耗明显下降，在 AI Coding 与 Agent 驱动连续工作流中更加流畅和高效。

图表 6：SWE-bench Verified 测试示意图

Benchmark	MiniMax-M2.1	MiniMax-M2	Claude Sonnet 4.5	Claude Opus 4.5	Gemini 3 Pro	GPT-5.2 (thinking)	DeepSeek V3.2
SWE-bench Verified	74	69.4	77.2	80.9	78	80	73.1
Multi-SWE-bench	49.4	36.2	44.3	50	42.7	x	37.4
SWE-bench Multilingual	72.5	56.5	68 ± 0.5	77.5 ± 1.5	65	72	70.2
Terminal-bench 2.0	47.9	30	50	57.8	54.2	54	46.4

资料来源：MINIMAX，华鑫证券研究

M2.1 模型补齐了公司在逻辑推理与代码生成模块的短板，核心目标在于渗透 B 端生产力场景。Toolathon 作为行业内对智能体能力评估标准的典型代表，涵盖 32 款专业软件及 600 余项工具，要求模型在平均 20 轮交互中完成复杂长周期任务。M2.1 模型强化代码解释器与工具调用能力，正是为了应对真实商业场景的复杂度。测试中，模型能够熟练操作 Docker 容器、管理日程安排并自动处理电商订单，其正逐步从 C 端娱乐工具向 B 端生产力工具升级。目前，MiniMax 形成 C 端产品提供数据与收入支持、底层模型通过混合专家系统控制推理成本、开放平台依托 M2.1 能力切入企业级市场的业务结构。此次 M2.1 模型发布，为其技术产业化及市场拓展奠定基础。

2、AI 应用动态：Discord 周访问量环比 +9.44%，清华大学发布 TurboDiffusion

2.1、周流量跟踪：Discord 周访问量环比+9.44%

本期（2025.12.19–2025.12.25）AI 相关网站流量数据：访问量前三位分别为 ChatGPT（1129.0M）、Bing（739.2M）和 Gemini（383.2M），访问量环比增速第一为 Discord（9.44%）；平均停留时长前三位分别为 Character.AI（00:18:06）、Discord（00:10:41）和 Kimi（00:08:49）；平均停留时长环比增速第一为文心一言（2.98%）。

图表 7：2025.12.19–2025.12.25AI 相关网站流量

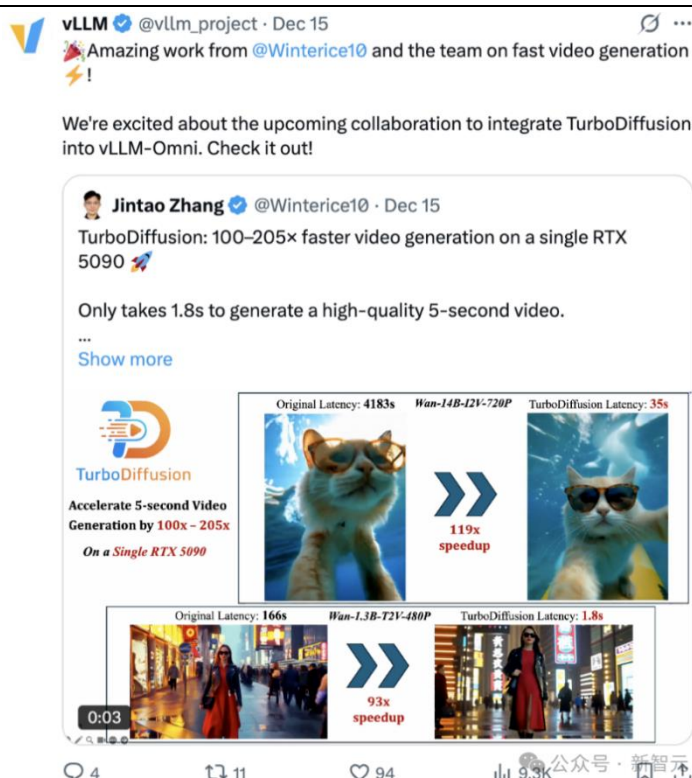
应用	应用类型	归属公司	周平均访问量 (M)	访问量环比	平均停留时长	时长环比
ChatGPT	聊天机器人	OpenAI	1129.0	-12.95%	6:31	0.77%
Bing	搜索	微软	739.2	-3.88%	7:56	-0.42%
Gemini	聊天机器人	谷歌	383.2	-2.77%	7:12	0.93%
Canva	在线设计	Canva	164.8	-22.04%	6:06	0.83%
Discord	游戏社区	微软	147.2	9.44%	10:41	0.00%
Github	代码托管	微软	109.8	-3.94%	6:21	-1.55%
Character.AI	聊天机器人	Character.AI	43.04	0.84%	18:06	-0.55%
Perplexity	AI 搜索	Perplexity	36.62	-14.98%	4:30	-0.37%
NotionAI	文本/笔记	Notion	32.13	-13.21%	8:04	0.41%
DeepL	翻译工具	DeepL	24.27	-14.78%	2:34	1.32%
Kimi	聊天机器人	MoonshotAI	8.268	-5.50%	8:49	1.34%
QuillBot	释义工具	QuillBot	7.73	-33.95%	3:11	1.06%
文心一言	聊天机器人	百度	0.997004	6.13%	2:53	2.98%

资料来源：similarweb, 华鑫证券研究

2.2、产业动态：清华大学发布 TurboDiffusion，引领 AI 视频生成进入秒级实时时代

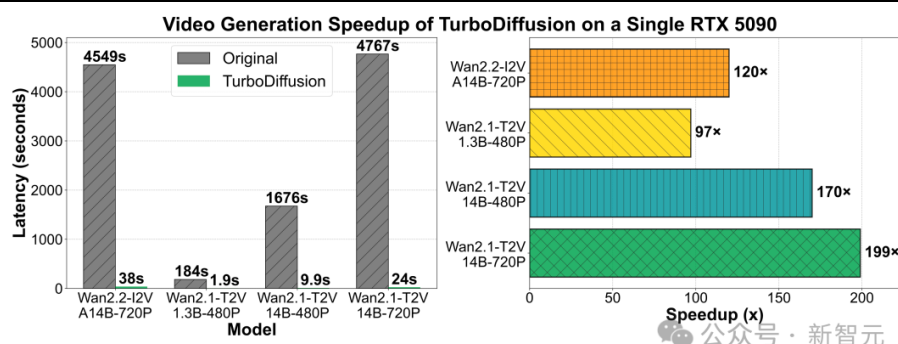
在 AI 视频生成因速度瓶颈而严重影响应用体验的背景下，2025 年 12 月 23 日，由清华大学 TSAIL 实验室与生数科技联合研发并开源发布了视频生成加速框架 TurboDiffusion。它成功地将 AI 视频生成从传统的“分钟级”耗时，一举推进到了“秒级”乃至“实时生成”的新时代。该框架能够在几乎不影响最终视频质量的前提下，实现高达 100 至 200 倍的加速。

图表 9: vLLM 上对 TurboDiffusion 的评价



资料来源：新智元，华鑫证券研究

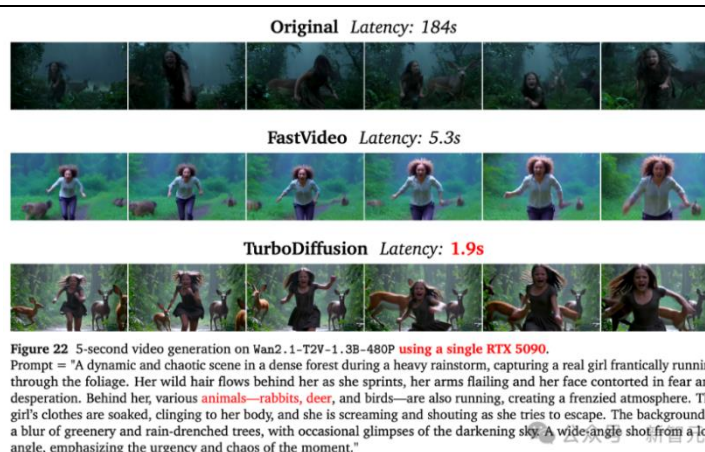
图表 10: RTX5090 显卡上 TurboDiffusion 视频生成速度提升的表现



资料来源：新智元，华鑫证券研究

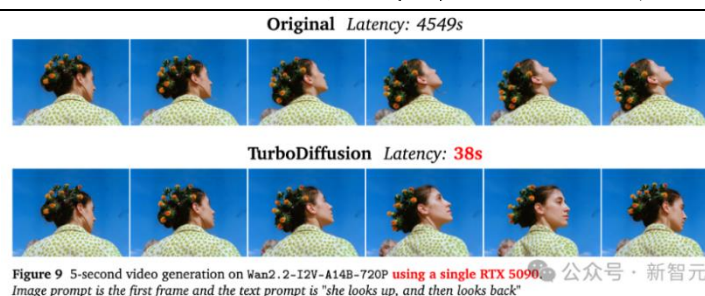
TurboDiffusion 是一个专为 Diffusion 模型加速而生的工具，尤其擅长处理视频生成场景。它能够高效处理从图像生成视频与从文本生成视频等多种任务，即便在面对高分辨率、长时间视频内容的生成需求时，依然能保持卓越的加速性能。其实测结果充分证明了其有效性：例如，对于一个参数规模为 1.3B 的模型，生成一段 5 秒的视频，标准实现需要 184 秒，而使用 TurboDiffusion 仅需 1.9 秒，加速比接近 97 倍，且生成内容在视觉上无明显差异。在更具挑战性的场景中，如使用 14B 参数规模的图生视频模型生成 720P 分辨率的 5 秒视频，标准实现耗时 4549 秒，而 TurboDiffusion 将其缩短至 38 秒，实现了约 120 倍的加速，同时确保了水下自拍、戴墨镜猫咪冲浪等复杂画面元素的完整保留与高质量呈现。同样，对于 14B 规模的文生视频模型，在相同条件下也能实现约 200 倍的加速。

图表 11: TurboDiffusion 以 1.3B 大小的模型生成 5 秒的视频的表现



资料来源：新智元，华鑫证券研究

图表 12: TurboDiffusion 以 14B 大小的模型生成 5 秒 720P 的视频的表现



资料来源：新智元，华鑫证券研究

TurboDiffusion 所包含的加速技术具备良好的通用性。例如，将其技术应用于 Vidu 这类视频生成模型时，同样能带来速度提升。原本需要 900 秒生成的一段 8 秒、1080P 高清视频，在应用加速技术后，生成时间锐减至 8 秒，真正达到了“所见即所得”的实时体验。

TurboDiffusion 之所以能实现如此惊人的性能突破，依赖于其集成的四项核心技术创新。首先，是名为 SageAttention 的低比特量化注意力加速技术。该技术由清华大学自主研发，旨在解决传统 Transformer 注意力机制在处理高分辨率视频数据时产生的巨大计算开销。通过对注意力计算进行低比特量化，该技术能够充分挖掘显卡硬件的计算潜力，实现极致的速度优化。此项技术不仅是全球首个实现注意力计算量化加速的方案，而且已经成功集成到 NVIDIA 的 TensorRT 推理引擎中，并在华为昇腾、摩尔线程等主流计算平台完成了部署，被腾讯混元、字节豆包、阿里 Tora、生数 Vidu、智谱清影、百度飞桨、昆仑万维、GoogleVeo3、商汤、vLLM 等国内外头部科技企业及团队所采用，创造了显著的经济价值。

其次，框架引入了稀疏线性注意力技术。该技术与低比特量化加速形成正交互补，通过构建在 SageAttention 之上，利用稀疏计算特性显著减少了全连接矩阵乘法中的冗余运算，从而在推理过程中能够额外获得数倍的加速效果。第三项关键技术是来自 NVIDIA 开源实验室的 rCM 步数蒸馏方法。该方法通过先进的训练策略，使得模型能够用远少于原始所需的采样步数，例如从 50-100 步压缩到 4-8 步，生成质量相当的视频内容，从而大幅降低生成过程的延迟。最后，框架在线性层应用了 W8A8INT8 量化策略。该策略将模型权重和激活值映射到 8 位整数空间，并在 128×128 的块粒度上进行分块量化，在保证精度的同时，显著降低了推理过程中的内存占用与功耗。

对于广大开发者与研究者而言，TurboDiffusion 的使用门槛极低。用户仅需简单的几步操作：安装 TurboDiffusion 仓库中的 Python 包、下载对应的 Checkpoints、并运行仓库中附带的推理脚本，即可快速生成高质量视频。综上所述，TurboDiffusion 将视频生成从需要漫长等待的专业化过程，转变为近乎实时的平民化工具，极大地释放了 AI 在视频内容创作方面的潜力。

3、AI 融资动向：LemonSlice 获千万融资，押注 AI 实时交互头像素赛道

12 月 24 日，LemonSlice 宣布完成 1050 万美元种子轮融资，投资方包括 MatrixPartners、YCombinator、Dropbox 首席技术官 ArashFerdowsi、Twitch 首席执行官 EmmettShear 以及音乐组合 TheChainsmokers。该公司致力于为 AI 智能体引入交互式视频头像素，将聊天体验从单纯文本拓展到视觉互动层面。这笔融资将用于拓展其实时交互式头像技术，并推动其 API 和嵌入式头像素产品在商业领域的更广泛应用。

LemonSlice 由 LinaColucci、SidneyPrimas 和 AndrewWeitz 于 2024 年创立（YCombinator2024 年冬季批次毕业企业），拥有 8 名员工，正在构建自己的通用视频扩散模型。公司的 AI 视频头像素技术，能够将单一静态图像转化为完全交互式的会说话视频角色。公司的最新达到模型 LemonSlice-2（一款视频扩散转换器模型和推理框架）允许用户上传照片（无论是公司头像、卡通形象还是绘画作品），然后立即对该角色进行实时视频通话。LemonSlice 联合创始人 LinaColucci 表示：“在生成式 AI 早期，我的联合创始人开始小试不同的视频模型，我们意识到视频将变得交互式。ChatGPT 等工具的吸引力在于它们的交互性，我们希望视频也具备这一层面。”

该 LemonSlice-2 模型拥有 200 亿参数，以每秒 20 帧的速度流式传输视频，并通过 API 或可嵌入小部件提供服务，只需一行代码即可添加；采用高效的注意力机制和缓存策略，使其在交互环境中实现超快速的响应时间，并支持无限长度视频生成且无误差累积。LemonSlice-2 支持全身虚拟形象生成，能够展现富有表现力和语义感知的手势动作。为了突破实时生成的瓶颈，LemonSlice-2 采用了因果注意力、新型分布匹配采样可回退的训练范式、高效缓存、CUDA 图加速和量化等技术策略。用户使用 LemonSlice-2 时，平均响应时间仅为 2.8 秒，其中视频生成仅占用 26% 的时间。该模型的核心能力包括：可以基于单一图像和音频样本在一个 GPU 上实时生成任何角色的视频；支持多种风格的视频实时生成；能够表现富有表现力的手势和场景感知；作为自回归模型，不受固定长度限制且不存在误差累积问题；支持通过文本提示对视频内容进行实时操控；实现任何角色的实时互动，并提供可嵌入的小部件。用户可以随时更新头像素的外貌、背景和风格，并能生成人类和拟人角色。创建头像素后，用户可随时更改角色的背景、风格和外貌。公司还使用 ElevenLabs 的 AI 语音技术，为平台生产的 AI 头像素视频，配备语音内容。公司表示已实施保障措施，防止未经授权的面部或声音克隆，并使用 AI 进行内容审核。

LemonSlice 区别于竞争对手的关键在于，其角色是实时创建的，无需大量训练数据或视频片段，也不会施加严格的风格限制。该公司的技术由大规模视频扩散变换器驱动，能够即时生成动画的每个像素，包括面部表情、手势和全身动作。LemonSlice 表示，这项技术使其能够提供富有表现力和情感吸引力的头像素，比典型的静态或有限头像素感觉更自然、更具交互性。公司将与 D-ID、HeyGen 和 Synthesia 等视频生成公司以及其他数字头像素制作商 Genies、SoulMachine、Praktika 和 AvatarOS，在 AI 视频生成赛道，展开激烈竞争。

实时交互 AI 头像素技术代表了数字人交互的下一代范式，LemonSlice 的成功融资

案例表明 AI 正从纯文本交互向多模态体验演进。对中国创业者而言，这预示着构建本地化、情感化的 AI 交互界面将成为未来差异化竞争的关键，特别是在教育、客服和医疗等对情感交流要求较高的垂直领域。

图表 12：上周 AI 初创公司融资动态

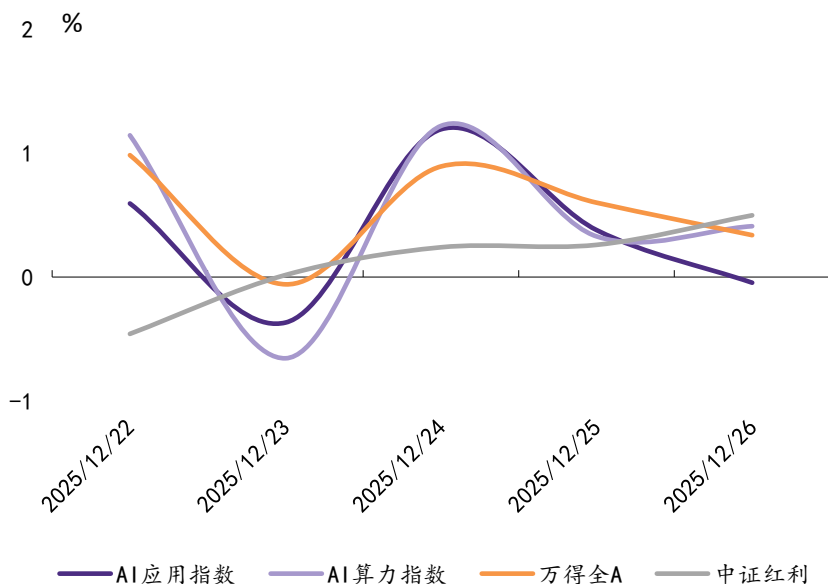
应用	应用类型	领投方	融资轮	融资额	目前累计 融资额	目前估值
LemonSlice	AI 实时交互	MatrixPartners	种子轮	1050 万美 元	1050 万美 元	-
NeuralConcept	企业级 AI 工程设计平 台	高盛另类投资部门 (GrowthEquity)	C 轮	1 亿美元	13800 万 美元	-
Lovable	AI 驱动的低 代码/无代 码软件开发 平台	CapitalG	B 轮	3.3 亿美元	2025 年累 计超 50000 万 美元	66 亿美元

资料来源：Saasverse, wind, 华鑫证券研究

4、行情复盘

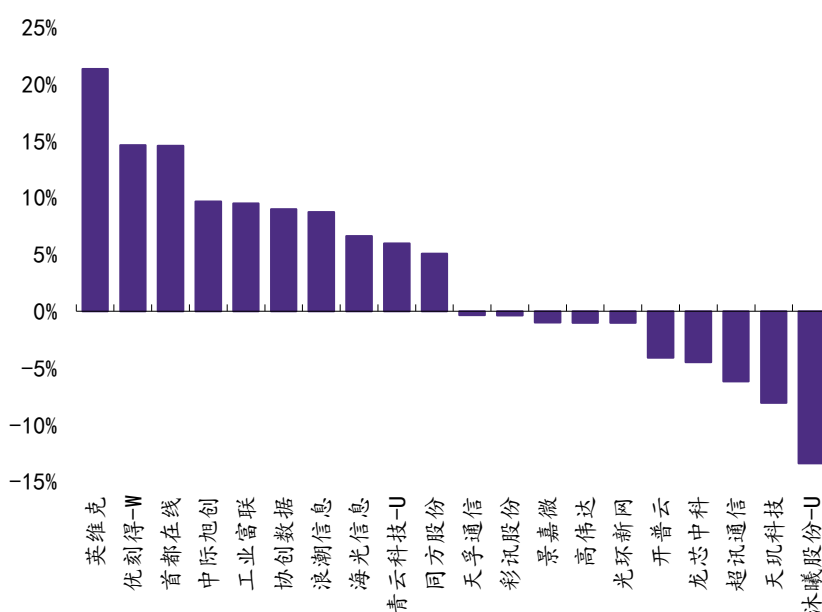
上周（12.22-12.26日），AI应用指数/AI算力指数/万得全A/中证红利日涨幅最大值分别为1.19%/1.22%/0.98%/0.50%，AI应用指数/AI算力指数/万得全A/中证红利日跌幅最大值分别为-0.37%/-0.65%/-0.06%/-0.46%。AI算力指数内部，英维克以+21.35%录得上周最大涨幅，沐曦股份-U以-13.40%录得上周最大跌幅。AI应用指数内部，生益科技以+23.28%录得上周最大涨幅，美年健康以-12.85%录得上周最大跌幅。

图表 13：上周（12.22-12.26日）指数日涨跌幅



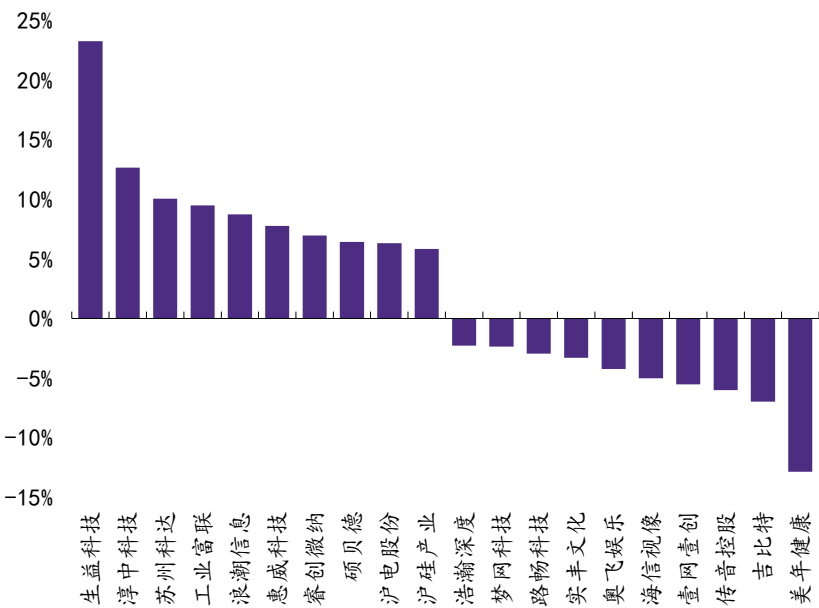
资料来源：wind, 华鑫证券研究

图表 14：上周（12.22-12.26日）AI算力指数内部涨跌幅度排名



资料来源：wind, 华鑫证券研究

图表 15：上周（12. 22-12. 26 日）AI 应用指数内部涨跌幅度排名



资料来源：wind, 华鑫证券研究

5、投资建议

2025 年 12 月 25 日消息，英伟达宣布以约 200 亿美元现金通过非排他性授权协议收购 AI 芯片新创公司 Groq 的核心资产，这笔交易成为英伟达史上最大收购案，其核心内容包括获得 Groq 全部资产与技术授权、吸纳 Groq 创始人及核心团队成员，而 Groq 旗下 GroqCloud 云端业务将保持独立运营。公开资料显示，Groq 成立于 2016 年，核心团队源自谷歌 TPU 工程团队，具备深厚的 AI 芯片研发积淀。该公司推出的基于 TSA 架构的 LPU 芯片，在推理速度与能耗性价比方面表现亮眼，性能显著优于常规 GPU 及 TPU 产品。截至目前，Groq 已为超 200 万名开发者及多家财富 500 强企业提供服务，业务覆盖范围与行业认可度持续提升。2025 年 9 月，Groq 完成 7.5 亿美元新一轮融资，投后估值达 69 亿美元，为此次收购奠定了重要基础。此次收购得益于英伟达丰厚的现金储备，也是其巩固 AI 芯片领域领导地位的重要举措，近期英伟达还开展了对 OpenAI、英特尔等的投资及与新思科技的战略合作等多项生态布局。英伟达通过此次战略举措，成功将 Groq 的 LPU 技术纳入麾下，进一步卡位高效推理这一关键赛道。Groq 的 LPU 在特定推理任务中具备显著优势，与英伟达生态形成完美互补，强化了其全栈护城河。这一动作不仅印证了算力赛道的高景气，更凸显了推理端的战略价值。龙头通过整合不断抬高行业门槛，我们维持对 AI 算力板块的长期看好态度。

中长期，建议关注新能源业务高增并供货科尔摩根等全球电机巨头的唯科科技（301196.SZ），AI 智能文字识别与商业大数据领域巨头的合合信息（688615.SH）、深耕工业 AI 与软件并长期服务高端装备等领域头部客户的能科科技（603859.SH）。

图表 16：重点关注公司及盈利预测

公司代码	名称	2025-12-31	EPS			PE			投资评级
		股价	2024	2025E	2026E	2024	2025E	2026E	
301196.SZ	唯科科技	80.57	1.76	2.53	3.34	45.78	31.85	24.12	买入
603859.SH	能科科技	42.05	0.78	0.96	1.18	53.91	43.80	35.64	买入
688615.SH	合合信息	227.58	4.01	3.37	4.11	56.75	67.53	55.37	买入

资料来源：Wind，华鑫证券研究

6、风险提示

1) AI 底层技术迭代速度不及预期。2) 政策监管及版权风险。3) AI 应用落地效果不及预期。4) 推荐公司业绩不及预期风险。

■ 中小盘&主题&北交所组介绍

任春阳：华东师范大学经济学硕士，6 年证券行业经验，2021 年 11 月加盟华鑫证券研究所，从事计算机与中小盘行业上市公司研究

周文龙：澳大利亚莫纳什大学金融硕士

陶欣怡：毕业于上海交通大学，于 2023 年 10 月加入团队。

倪汇康：金融学士，2025 年 8 月加盟华鑫证券研究所。

■ 证券分析师承诺

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并注册为证券分析师，以勤勉的职业态度，独立、客观地出具本报告。本报告清晰准确地反映了本人的研究观点。本人不曾因，不因，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收到任何形式的补偿。

■ 证券投资评级说明

股票投资评级说明：

	投资建议	预测个股相对同期证券市场代表性指数涨幅
1	买入	>20%
2	增持	10%—20%
3	中性	-10%—10%
4	卖出	<-10%

行业投资评级说明：

	投资建议	行业指数相对同期证券市场代表性指数涨幅
1	推荐	>10%
2	中性	-10%—10%
3	回避	<-10%

以报告日后的 12 个月内，预测个股或行业指数相对于相关证券市场主要指数的涨跌幅为标准。

相关证券市场代表性指数说明：A股市场以沪深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；美国市场以道琼斯指数为基准。

■ 免责声明

华鑫证券有限责任公司（以下简称“华鑫证券”）具有中国证监会核准的证券投资咨询业务资格。本报告由华鑫证券制作，仅供华鑫证券的客户使用。本公司不会因接收人收到本报告而视其为客户。

本报告中的信息均来源于公开资料，华鑫证券研究部门及相关研究人员力求准确可靠，但对这些信息的准确性及完整性不作任何保证。我们已力求报告内容客观、公正，但报告中的信息与所表达的观点不构成所述证券买卖的出价或询价的依据，该等信息、意见并未考虑到获取本报告人员的具体投资目的、财务状况以及特定需求，在任何时候均不构成对任何人的个人推荐。投资者应当对本报告中的信息和意见进行独立评估，并应同时结合各自的投资目的、财务状况和特定需求，必要时就财务、法律、商业、税收等方面咨询专业顾问的意见。对依据或者使用本报告所造成的一切后果，华鑫证券及/或其关联人员均不承担任何法律责任。本公司或关联机构可能会持有报告中所提到的公司所发行的证券头寸并进行交易，还可能为这些公司提供或争取提供投资银行、财务顾问或者金融产品等服务。本公司在知晓范围内依法合规地履行披露。

本报告中的资料、意见、预测均只反映报告初次发布时的判断，可能会随时调整。该等意见、评估及预测无需通知即可随时更改。在不同时期，华鑫证券可能会发出与本报告所载意见、评估及预测不一致的研究报告。华鑫证券没有将此意见及建议向报告所有接收者进行更新的义务。

本报告版权仅为华鑫证券所有，未经华鑫证券书面授权，任何机构和个人不得以任何形式刊载、翻版、复制、发布、转发或引用本报告的任何部分。若华鑫证券以外的机构向其客户发放本报告，则由该机构独自为此发送行为负责，华鑫证券对此等行为不承担任何责任。本报告同时不构成华鑫证券向发送本报告的机构之客户提供的投资建议。如未经华鑫证券授权，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。华鑫证券将保留随时追究其法律责任的权利。请投资者慎重使用未经授权刊载或者转发的华鑫证券研究报告。