

证券研究报告

行业研究

行业点评

传媒行业

投资评级 看好

上次评级 看好

冯翠婷 传媒互联网及海外
首席分析师

执业编号: S1500522010001

邮箱: fengcuiting@cindasc.com

信达证券股份有限公司

CINDA SECURITIES CO., LTD

北京市西城区闹市口大街9号院1号楼

邮编: 100031

阶跃星辰发布万亿参数 MoE 语言大模型， Kimi 长文本处理提升至 200 万字

2024 年 3 月 24 日

近期 AI 相关热点更新:

- **阶跃星辰发布 Step 系列通用大模型，万亿参数 MoE 大语言模型亮眼:**3 月 23 日，阶跃星辰发布 Step 系列通用大模型，包括 Step-1 千亿参数语言大模型、Step-1V 千亿参数多模态大模型，以及 Step-2 万亿参数 MoE 语言大模型的预览版，提供 API 接口给部分合作伙伴试用。目前，基于 Step-1 和 Step-1V 千亿参数大模型的产品效率工具跃问和 AI 开放世界平台冒泡鸭已经全面开放。
- **Kimi 再次突破上下文窗口技术，无损上下文长度提升至 200 万字:** 3 月 18 日，月之暗面宣布 Kimi 智能助手在长上下文窗口技术上再次取得突破，无损上下文长度提升了一个数量级到 200 万字。并从即日起启动“内测”，对大模型超长无损上下文能力有需求的用户，可到 Kimi 智能助手网页版 kimi.ai 首页申请抢先体验。本次无损上下文长度上量级的提升，有望帮助用户探索更多的 AI 应用方式，给用户带来更大的价值。
- **OpenAI 携 Sora 模型有望与好莱坞进行合作。**人工智能公司 OpenAI 计划将其突破性的 AI 视频生成技术带入电影产业。该公司即将与洛杉矶的制片厂和媒体高管展开会谈，探讨合作机会。OpenAI 的 Sora 服务能根据文本提示生成真实感视频，尽管尚未公开发布，但已在好莱坞引起广泛关注。这一技术的发展和应用，预示着电影制作领域的未来可能将迎来较为明显的变化。
- **谷歌发布“Vlogger”模型，利用单张图片生成 10 秒视频:** 3 月 20 日，谷歌发布了一个新的视频框架：只需要一张头像图片、一段讲话录音，就能得到一个栩栩如生的本人演讲视频。视频时长可变，目前看到的示例最高为 10s，无论是口型还是面部表情，视频表现都非常自然。如果输入图像囊括整个上半身，视频也能配合产出丰富的手势。
- **Suno AI 推出音乐生成模型-V3，面向所有用户免费开发:**AI 初创公司 Suno AI 重磅推出了第一款可制作「广播级」的音乐生成模型——V3，仅用几秒的时间，V3 便可以创作出 2 分钟的完整歌曲。用户只用几个简短的词，用户就可以用任何语言创作一首歌曲。
- **英伟达 GTC 大会开幕，Blackwell 架构 GPU-B200 问世:** 时隔两年，英伟达官宣新一代 Blackwell 架构，定位为“新工业革命的引擎”。本次大会亮点频出：宣布全新 GPU B200、超级芯片 GB200、新一代计算单元 GB200 NVL72，与相同数量的 72 个 H100 相比，GB200 NVL72 对于大模型推理性能提升高达 30 倍，成本和能耗降低高达 25 倍；官宣 ai.nvidia.com 页面，目标是让任何人都可以通过易于使用的用户界面

体验各种 AI 模型和应用，同时，企业使用这些服务在自己的平台上创建和部署自定义应用，同时保留对其知识产权的完全所有权和控制权。

投资建议:

- 受益于全球生成式 AI 持续发展，文生图/视频/音乐/代码等多方面应用的迭代更新有望长期给诸多行业带来降本增效，重点关注：1) **AI+影视**：万达电影、华策影视、慈文传媒、上海电影、光线传媒、芒果超媒、超讯通信、因赛集团、视觉中国、华数传媒；2) **AI+IP**（涉及版权、算料等）：中广天择、中文在线、上海电影、华策影视、奥飞娱乐、果麦文化、捷成股份、南方传媒；3) **AI+3D**：大丰实业、丝路视觉、凡拓数创、恒信东方、风语筑、锋尚文化、岭南股份；4) **AI+营销**：分众传媒、三人行、因赛集团、易点天下、值得买、蓝色光标、兆讯传媒、浙文互联、每日互动；5) **AI+教育/阅读**：南方传媒、盛通股份、世纪天鸿、掌阅科技等；6) **AI+出版**：果麦文化、南方传媒、皖新传媒、时代出版、城市传媒、中南传媒、凤凰传媒、新华传媒等；7) **AI+游戏**：巨人网络、恺英网络、完美世界、紫天科技、盛天网络、昆仑万维、神州泰岳、三七互娱、掌趣科技、宝通科技、吉比特、名臣健康、姚记科技、顺网科技、汤姆猫、星辉娱乐；8) **其他 AI 应用**：昆仑万维、万兴科技、人民网、新华网、福昕软件、金山办公等；9) **新增关注**：新华传媒、松炆资源等。
- **风险因素**：生成式 AI 公司资本开支不及预期、AI 应用迭代不及预期。

表 1：海外科技企业近期动态汇总

日期	公司	事件
3.20	Google	谷歌发布了一个新的视频框架：只需要一张头像、一段讲话录音，就能得到一个栩栩如生的本人演讲视频。视频时长可变，目前看到的示例最高为 10s。无论是口型还是面部表情，视频表现都非常自然。如果输入图像囊括整个上半身，视频也能配合产生丰富的手势。这个框架名叫 VLOGGER，主要基于扩散模型，并包含两部分：一个是随机的人体到 3D 运动 (human-to-3d-motion) 扩散模型。另一个是用于增强文本到图像模型的新扩散架构。其中，前者负责将音频波形作为输入，生成人物的身体控制动作，包括眼神、表情和手势、身体整体姿势等等。后者则是一个时间维度的图像到图像模型，用于扩展大型图像扩散模型，使用刚刚预测的动作来生成相应的帧。VLOGGER 的训练是在一个 (名叫 MENTOR) 的大型数据集上完成的，全长 2200 小时，共包含 80 万个人物视频。
3.19	Stability AI	3月19日，Stable Diffusion 背后公司 Stability AI 再次上新图生 3D 模型 SV3D (Stable Video 3D)，模型基于 Stable Video Diffusion，只用一张图片就能生成高质量 3D 网格。SV3D 首次将视频扩散模型应用到了 3D 生成领域，基于此，SV3D 大大提高了 3D 生成的质量和视图一致性。引入潜在视频扩散模型，SV3D 的核心目的是利用视频模型的时间一致性来提高 3D 生成的一致性，并且视频数据本身也比 3D 数据更容易获得。Stability AI 这次提供两个版本的 SV3D：一个是 SV3D_u：基于单张图像生成轨道视频。另一个是 SV3D_p：扩展了 SV3D_u 的功能，可以根据

		指定的相机路径创建 3D 模型视频。从与其他大模型的横向对比上, 在新视角合成 (NVS) 和 3D 重建方面, SV3D 超过了现有其他方法, 达到 SOTA。从定性比较的结果来看, SV3D 生成的多视角试图, 细节更丰富, 更接近与原始输入图像, 在理解和重构物体的 3D 结构方面, 能够更准确地捕捉到细节, 并保持视角变换时的一致性。
3.19	英伟达	时隔两年, 英伟达官宣新一代 Blackwell 架构, 定位为“新工业革命的引擎”。本次大会亮点频出: 宣布全新 GPU B200: 两个芯片合在一起, 以 10TB 每秒的带宽互联, 组成 B200 GPU, 总计包含 2080 亿晶体管; 超级芯片 GB200: 两个 B200 GPU 与 Grace CPU 结合, 就成为 GB200 超级芯片; 新一代计算单元 GB200 NVL72: 与相同数量的 72 个 H100 相比, GB200 NVL72 对于大模型推理性能提升高达 30 倍, 成本和能耗降低高达 25 倍; 官宣 ai.nvidia.com 页面, 目标是让任何人都可以通过易于使用的用户界面体验各种 AI 模型和应用, 同时, 企业使用这些服务在自己的平台上创建和部署自定义应用, 同时保留对其知识产权的完全所有权和控制权。此外, 在这次大会上, 还宣布了与苹果在 Vision Pro 方面的合作, 让开发者可以在工业元宇宙里进行空间计算。此前推出的新光刻技术 cuLitho 软件库也有了新进展, 合作方包括台积电和新思科技。
3.21	腾讯	在游戏开发者大会 (GDC) 上, 腾讯发布自研游戏 AI 引擎 GiiNEX。基于生成式 AI 和决策 AI 技术, GiiNEX 可以支持游戏从研发到运营的全生命周期需求。具体来说, GiiNEX 能高效覆盖包括 AI NPC 的对话生成, 场景制作中的 3D 城市建造, 以及剧情、关卡、音乐等内容生成的多个领域。GiiNEX 的整体架构包含前沿算法模型、高效训练平台、在线推理引擎三大核心: 其中, 基于强化学习、自然语言处理等 AI 基础研究能力构建的统一算法底层模型, 能够支持 MOBA、FPS、派对游戏等十余种游戏类型。高性能训练平台方面, 面向智能体和大模型, GiiNEX 可支持万卡规模资源调度。在线推理引擎适配主流设备, 可实现移动端与云端的混合部署, 保障多端协同。在 GDC 现场, 腾讯展示了 GiiNEX 在游戏 3D 城市生成和 UGC 关卡设计场景中的具体应用。
3.18	月之暗面	3 月 18 日, 月之暗面宣布 Kimi 智能助手在长上下文窗口技术上再次取得突破, 无损上下文长度提升了一个数量级到 200 万字。并从即日起启动“内测”, 对大模型超长无损上下文能力有需求的用户, 可到 Kimi 智能助手网页版 kimi.ai 首页申请抢先体验。Kimi 可以根据用户的问题, 主动去互联网上搜索、分析和总结最相关的多个页面, 生成更直接、更准确的答案。Kimi 智能助手的智能搜索, 离不开大模型的无损长上下文能力。Kimi 主动搜索得到的多篇资料, 会作为上下文的一部分交给模型去推理。正是因为 Kimi 大模型支持的上下文窗口足够长, 窗口内的信息损失足够低, Kimi 智能助手才能输出高质量的结果, 为用户带来截然不同的搜索体验。因此, 本次无损上下文长度上量级的提升, 有望帮助用户探索更多的 AI 应用方式, 给用户带来更大的价值。
3.23	阶越星辰	3 月 23 日, 阶越星辰发布 Step 系列通用大模型, 包括 Step-1 千亿参数语言大模型、Step-1V 千亿参数多模态大模型, 以及 Step-2 万亿参数 MoE 语言大模型的预览版, 提供 API 接口给部分合作伙伴试用。阶越星辰的团队的数据团队核心骨干来自必应搜索引擎, 曾支持全球 100 多种语言, 为 200 多个国家和地区提供服务, 对全球互联网高质量语料的分布有深入了解, 并建立起强大的数据处理和知识图谱流水

		线。团队自建机房+租用算力，积极进行算力储备，实践单集群万卡以上的系统建设与管理，训练千亿模型的 MFU（有效算力输出）达 57%。目前，基于 Step-1 和 Step-1V 千亿参数大模型的产品效率工具跃问和 AI 开放世界平台冒泡鸭已经全面开放。
3.23	OpenAI	人工智能公司 OpenAI 计划将其突破性的 AI 视频生成技术带入电影产业。该公司即将与洛杉矶的制片厂和媒体高管展开会谈，探讨合作机会。OpenAI 的 Sora 服务能根据文本提示生成真实感视频，尽管尚未公开发布，但已在好莱坞引起广泛关注。这一技术的发展和运用，预示着电影制作领域的未来可能将迎来较为明显的变化。
3.18	xAI	xAI 开源史上最大参数量大语言模型 Grok-1。Grok-1 是马斯克 xAI 自 2023 年 7 月 12 日成立以来发布的首个自研大模型，Grok-1 是一个混合专家 (Mixture-of-Experts, MOE) 大模型，是迄今参数量最大的开源大语言模型，遵照 Apache 2.0 协议开放模型权重和架构。这种 MOE 架构重点在于提高大模型的训练和推理效率，其基础模型基于大量文本数据训练，未针对特定任务进行微调。xAI 从 2023 年 10 月开始使用自定义训练堆栈在 JAX 和 Rust 之上从头开始训练，模型拥有 3140 亿参数，其对于每个 token，活跃权重比例为 25%。此次 xAI 并未放出 Grok-1 的具体测试成绩，其与 OpenAI 即将发布的 GPT-5 的大模型进行比较，值得持续关注。
3.24	Suno AI	AI 初创公司 Suno AI 重磅推出了第一款可制作「广播级」的音乐生成模型——V3，仅用几秒的时间，V3 便可以创作出 2 分钟的完整歌曲。Suno v3 包含丰富的音乐风格和流派选项，比如古典音乐、爵士乐、Hiphop、电子等新潮曲风。用户只用几个简短的词，用户就可以用任何语言创作一首歌曲。相比与之前的版本，v3 生成的音乐质量更高，而且能制作各种各样的风格和流派的音乐和歌曲。伴随着 v3 版本的推出，Suno AI 发布了 AI 音乐水印系统，每段由平台生成的音乐都添加了人声无法识别的水印，从而在未来能够保护用户在 Suno 的创作，打击抄袭，防止将 Suno 产生的音乐进行滥用。目前，Suno AI V3 已经向所有用户免费开放。

资料来源：智东西，量子位，月之暗面官方公众号，阶越星辰官方公众号，AI 范儿，新智元，信达证券研发中心

研究团队简介

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

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

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

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

白云汉，信达证券传媒互联网及海外团队成员。美国康涅狄格大学金融学硕士，曾任职于腾讯系创业公司投资部，一级市场从业2年。后任职于私募基金担任研究员，二级市场从业3年，覆盖传媒互联网赛道，拥有游戏领域产业链资源。2023年加入信达证券研究所，目前主要专注于微软、网易等美股研究以及结合海外映射对A股港股的覆盖。

分析师声明

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

免责声明

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

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

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

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

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

本报告版权仅为信达证券所有。未经信达证券书面同意，任何机构和个人不得以任何形式翻版、复制、发布、转发或引用本报告的任何部分。若信达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，信达证券对此等行为不承担任何责任。本报告同时不构成信达证券向发送本报告的机构之客户提供的投资建议。

如未经信达证券授权，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。信达证券将保留随时追究其法律责任的权利。

评级说明

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

风险提示

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

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