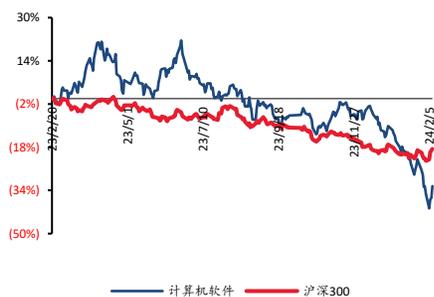


计算机

Sora 和 Gemini 1.5 发布，多模态大模型更进一步

■ 走势比较



■ 子行业评级

基础软件及管理办公软件	无评级
行业应用软件	无评级
新兴计算机软件	无评级

■ 推荐公司及评级

相关研究报告

<<2023 年计算机行业业绩预告及基金持仓总结>>—2024-02-05

证券分析师：曹佩

电话：

E-MAIL: caopeisz@tpyzq.com

执业资格证书编号：S1190520080001

证券分析师：王景宜

电话：

E-MAIL: wangjy@tpyzq.com

执业资格证书编号：S1190523090002

报告摘要

OpenAI 发布视频生成模型 Sora, AI 视频生成迎来里程碑。2月16日, OpenAI 宣布推出全新的生成式人工智能模型“Sora”。Sora 是一种通用的视觉数据模型, 能够生成各种持续时间、宽高比和分辨率的视频和图片。与此前的视频生成模型相比, Sora 主要实现了以下突破: 1) 视频时长达到 60 秒: Sora 可以通过文本指令直接输出长达 60 秒的视频, 并保持视频主体与背景的高度流畅性与稳定性。2) 文本的深度理解能力: Sora 可以准确理解用户的文本指令, 无论是复杂的动作场景还是细腻的情感表达, Sora 都能够精确捕捉并展现。3) 对真实世界的理解: Sora 对物理规律的遵循程度较高, 对于光影反射、运动方式、镜头移动等细节的呈现效果较为逼真。4) 长序列连贯性和目标持久性: Sora 能在单个视频中生成同一角色的多个镜头, 并在整个视频中保持其外观。即使人、动物和物体被遮挡或离开画面, Sora 模型也能使其保持不变。

谷歌发布 Gemini 1.5 Pro, 上下文长度突破百万。2月15日, 谷歌宣布了下一代大模型 Gemini 1.5 Pro。Gemini 1.5 Pro 是一种中型多模态模型, 针对多种任务的扩展进行了优化, 其性能水平与谷歌迄今为止最大的模型 1.0 Ultra 类似, 并引入了长上下文理解方面的突破性实验特征。Gemini 1.5 Pro 基于 Transformer 和 MoE 架构建立。MoE 模型分为更小的“专家”神经网络, 根据给定输入的类型, MoE 模型学会选择性地仅激活其神经网络中最相关的专家路径, 这种专业化极大地提高了模型的效率。目前, 少数开发人员和企业客户可以通过 AI Studio 和 Vertex AI 的私人预览版在最多 100 万个 token 的上下文窗口中进行尝试, 这意味着使用 1.5 Pro 能够一次性处理海量信息, 比如 1 小时的视频、11 小时的音频、超过 30,000 行的代码库, 或是超过 700,000 个单词。

投资建议: Sora 以及 Gemini 大模型的发布进一步推动了多模态大模型的进步, 对于计算机板块而言, 仍然从三个角度筛选公司:

1) 算法层面, 重点关注科大讯飞、云从科技;

2) 应用层面, 重点关注金山办公、新国都、万兴科技、虹软科技、熵基科技等;

3) 算力层面, 重点关注海光信息、农尚环境、景嘉微、神州数码、中科曙光、寒武纪、浪潮信息等。

风险提示: 技术研发不及预期; 下游景气度不及预期; 生态建设不及预期; 市场竞争加剧。

一、 OpenAI 发布视频生成模型 Sora，AI 视频生成迎来里程碑

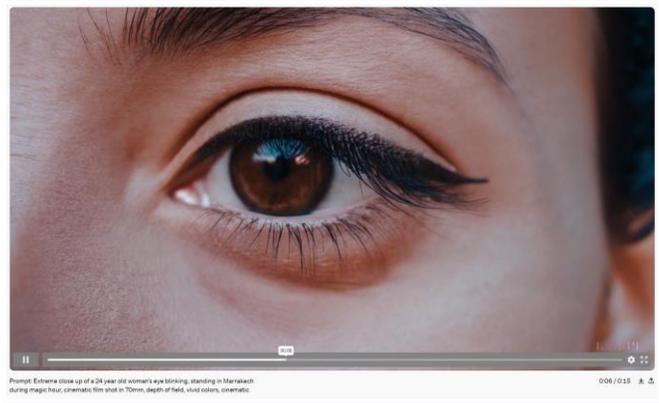
OpenAI 发布视频生成模型 Sora。2 月 16 日，OpenAI 宣布推出全新的生成式人工智能模型“Sora”。Sora 是一种通用的视觉数据模型，能够生成各种持续时间、宽高比和分辨率的视频和图片。OpenAI 声称，如果给定一段简短或详细的描述或一张静态图片，Sora 就能生成类似电影的 1080p 场景，其中包含多个角色、不同类型的动作和背景细节。OpenAI 官方公布了 48 个用 Sora 生成的视频示例，充分展示了 Sora 模型的强大能力。

图表1: Sora 生成视频示例 1



资料来源: OpenAI 官网, 太平洋证券整理

图表2: Sora 生成视频示例 2



资料来源: OpenAI 官网, 太平洋证券整理

与此前的视频生成模型相比，Sora 主要实现了以下突破：

- **视频时长达到 60 秒：**不同于此前发布的模型只能生成几秒到十几秒长度的视频，Sora 可以通过文本指令直接输出长达 60 秒的视频，并保持视频主体与背景的高度流畅性与稳定性。
- **对文本的深度理解能力：**Sora 可以准确理解用户的文本指令，无论是复杂的动作场景还是细腻的情感表达，Sora 都能够精确捕捉并展现。
- **对真实世界的理解：**Sora 不仅了解用户在提示中要求的内容，还了解这些东西在物理世界中的存在方式，对于光影反射、运动方式、镜头移动等细节的呈现效果较为逼真。
- **长序列连贯性和目标持久性：**Sora 能在单个视频中生成同一角色的多个镜头，并在整个视频中保持其外观。即使人、动物和物体被遮挡或离开画面，Sora 模型也能使其保持不变。

图表3: Sora 与其他视频生成模型效果对比



资料来源：机器之心，太平洋证券整理

- **多模态输入处理能力**:除了文本提示外, Sora 还能够接受静态图像或已有视频作为输入, 进行内容的延伸、填充缺失帧或进行风格转换等操作。如可以在两个输入视频之间逐渐进行插值, 从而在具有完全不同主题和场景构成的视频之间创建无缝过渡。

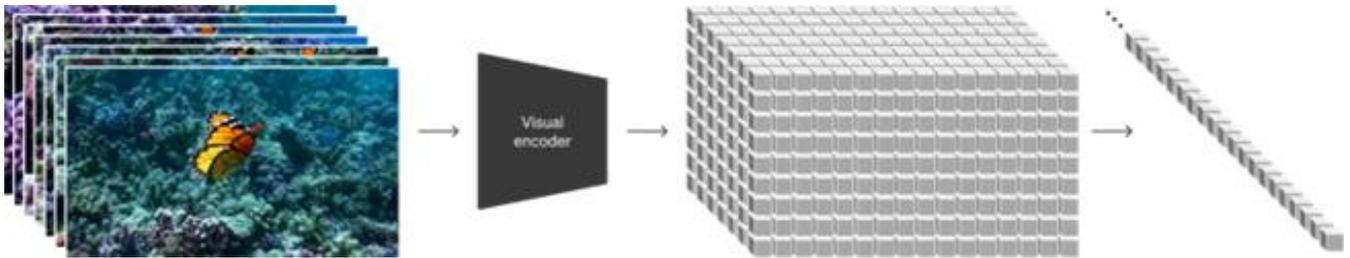
图表4: Sora 在两个输入视频之间逐渐进行插值合成新的视频



资料来源：OpenAI 官网，太平洋证券整理

Sora 把视频和图像分解为统一的数据单元 Patch。Sora 首先将不同类型的视觉数据转换成统一的视觉数据表示（视觉 patch），然后将原始视频压缩到一个低维潜在空间，并将视觉表示分解成时空 patch（相当于 Transformer token），让 Sora 在这个潜在空间里进行训练并生成视频。这种统一的数据表示方法能够在更广泛的视觉数据上训练模型，覆盖了不同的持续时间、分辨率和纵横比。

图表5: Sora 将视频数据转换为视觉块 (patch)



资料来源: OpenAI 官网, 太平洋证券整理

Sora 将 Transformer 和扩散模型相结合。Sora 是一种扩散模型，输入一个噪声 Patch，经过多步骤的噪声去除过程，来预测下一个 Patch。Sora 采用了 DALL-E 3 的重标注技术，通过为视觉训练数据生成详细描述的标题。同时，利用 GPT 将简短的用户提示转换为更长的详细字幕，然后发送到视频模型。这使 Sora 能够准确遵循用户提示生成高质量的视频。

图表6: Sora 将 Transformer 和扩散模型相结合

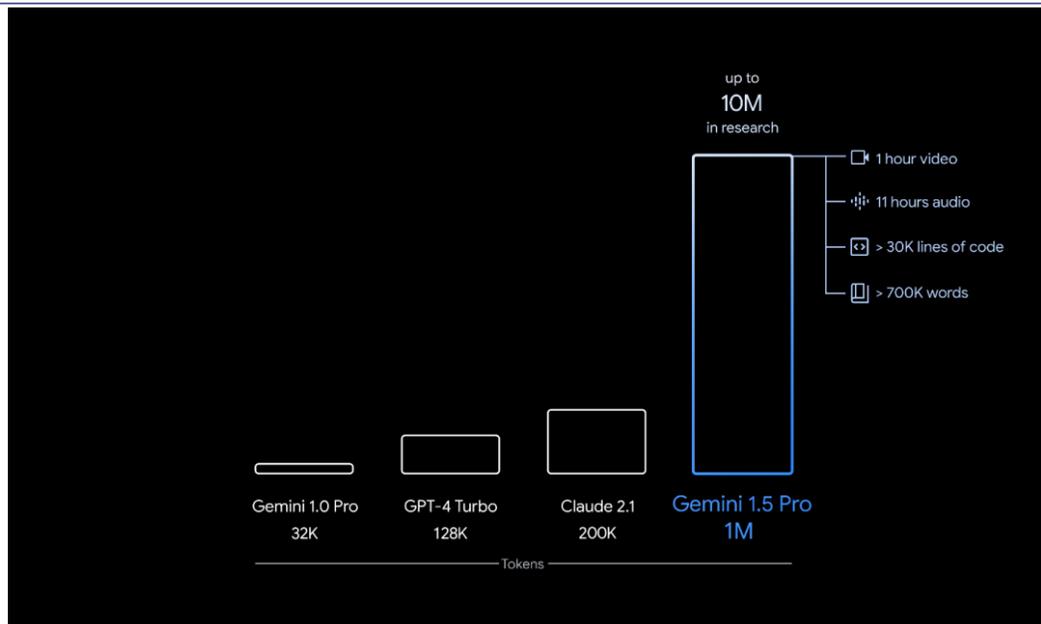


资料来源: OpenAI 官网, 太平洋证券整理

二、谷歌发布 Gemini 1.5 Pro，上下文长度突破百万

2月15日，谷歌宣布了下一代大模型 Gemini 1.5 Pro。Gemini 1.5 Pro 是一种中型多模态模型，针对多种任务的扩展进行了优化，其性能水平与谷歌迄今为止最大的模型 1.0 Ultra 类似，并引入了长上下文理解方面的突破性实验特征。目前，少数开发人员和企业客户可以通过 AI Studio 和 Vertex AI 的私人预览版在最多 100 万个 token 的上下文窗口中进行尝试，这意味着使用 1.5 Pro 能够一次性处理海量信息，比如 1 小时的视频、11 小时的音频、超过 30,000 行的代码库，或是超过 700,000 个单词。

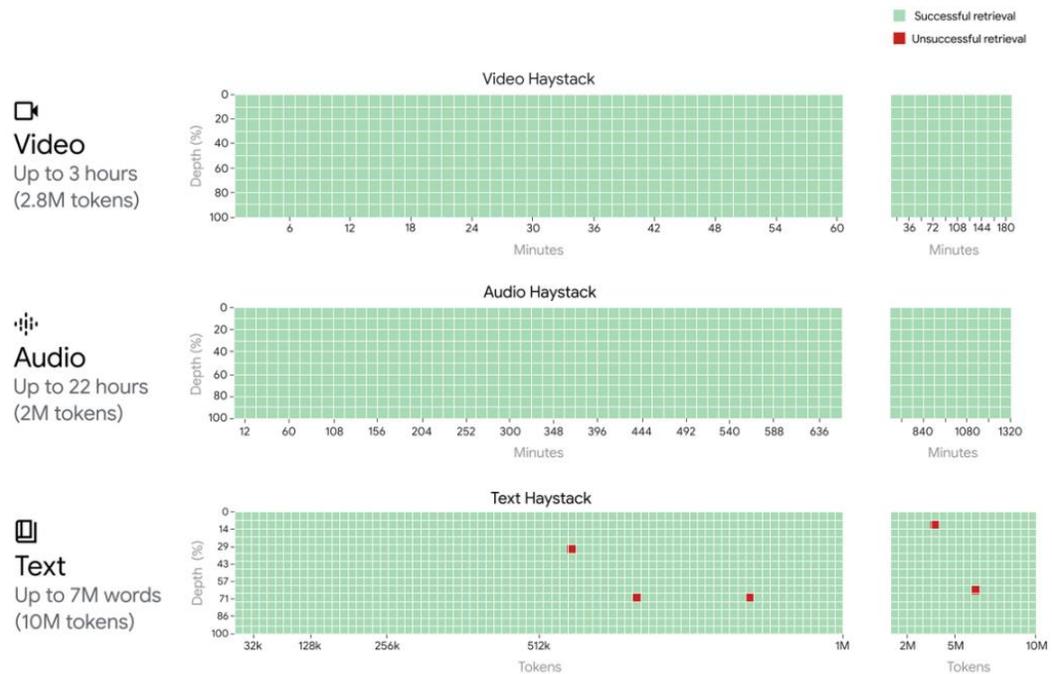
图表7: Gemini 1.5 Pro 上下文窗口长度远超现有其他大模型



资料来源: Google 官网, 太平洋证券整理

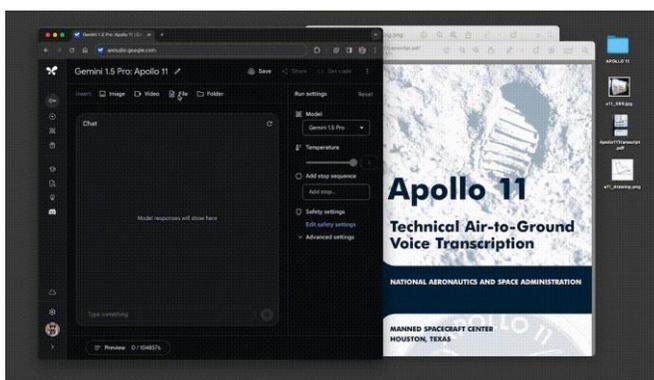
Gemini 1.5 Pro 基于 Transformer 和 MoE 架构建立。传统 Transformer 充当一个大型神经网络，而 MoE 模型则分为更小的“专家”神经网络。根据给定输入的类型，MoE 模型学会选择性地仅激活其神经网络中最相关的专家路径，这种专业化极大地提高了模型的效率。Google 是深度学习 MoE 技术的先驱，其对该技术的深入理解保障了 Gemini 系列模型的快速迭代。

Gemini 1.5 Pro 具备精确检索海量信息的能力。在多模态海底捞针测试中，对于文本处理，Gemini 1.5 Pro 在处理高达 530,000 token 的文本时，能够实现 100% 的检索完整性，在处理 1,000,000 token 的文本时达到 99.7% 的检索完整性。甚至在处理高达 10,000,000 token 的文本时，检索准确性仍然高达 99.2%。在音频处理方面，Gemini 1.5 Pro 能够在大约 11 小时的音频资料中，100% 成功检索到各种隐藏的音频片段。在视频处理方面，Gemini 1.5 Pro 能够在大约 3 小时的视频内容中，100% 成功检索到各种隐藏的视觉元素。

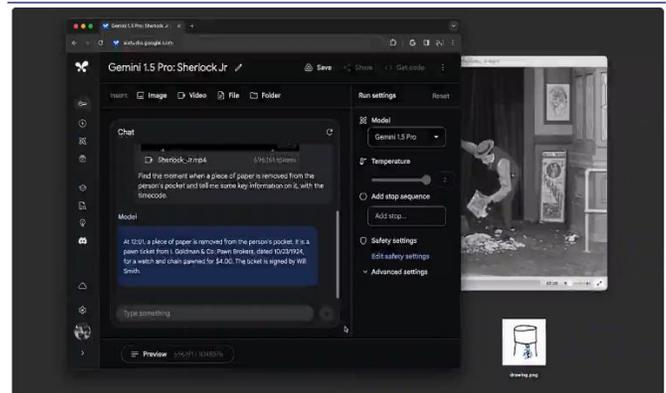
图表8: Gemini 1.5 Pro 多模态海底捞针测试表现


资料来源: Google 官网, 太平洋证券整理

Gemini 1.5 Pro 能够深入理解多模态海量信息。 1.5 Pro 可以针对包括文本、图片、视频在内的不同模式执行高度复杂的理解和推理任务。例如，输入一份阿波罗 11 号任务到月球的 402 页飞行记录，可以对文档中的复杂信息进行提问，模型会给出精确的回答。当给定一部 44 分钟的无声电影时，该模型可以准确分析各种情节点和事件，甚至推理出电影中容易被忽略的小细节。

图表9: Gemini 1.5 Pro 能够读取 402 页 PDF


资料来源: Google 官网, 太平洋证券整理

图表10: Gemini 1.5 Pro 能够理解电影中的细节


资料来源: Google 官网, 太平洋证券整理

三、投资建议

Sora 以及 Gemini 大模型的发布进一步推动了多模态大模型的进步，对于计算机板块而言，仍然从三个角度筛选公司：

- 1) 算法层面，重点关注科大讯飞、云从科技；
- 2) 应用层面，重点关注金山办公、新国都、万兴科技、虹软科技、熵基科技等；
- 3) 算力层面，重点关注海光信息、农尚环境、景嘉微、神州数码、中科曙光、寒武纪、浪潮信息等。

四、风险提示

技术研发不及预期；下游景气度不及预期；生态建设不及预期；市场竞争加剧。

投资评级说明

1、行业评级

看好：预计未来 6 个月内，行业整体回报高于沪深 300 指数 5%以上；

中性：预计未来 6 个月内，行业整体回报介于沪深 300 指数-5%与 5%之间；

看淡：预计未来 6 个月内，行业整体回报低于沪深 300 指数 5%以下。

2、公司评级

买入：预计未来 6 个月内，个股相对沪深 300 指数涨幅在 15%以上；

增持：预计未来 6 个月内，个股相对沪深 300 指数涨幅介于 5%与 15%之间；

持有：预计未来 6 个月内，个股相对沪深 300 指数涨幅介于-5%与 5%之间；

减持：预计未来 6 个月内，个股相对沪深 300 指数涨幅介于-5%与-15%之间；

卖出：预计未来 6 个月内，个股相对沪深 300 指数涨幅低于-15%以下。

公司地址

北京市西城区北展北街 9 号华远企业号 D 座二单元七号

上海市浦东南路 500 号国开行大厦 17 楼 太平洋证券

深圳市福田区商报东路与莲花路新世界文博中心 19 层 1904

广州大道中圣丰广场 988 号 102 太平洋证券



研究院

中国北京 100044

北京市西城区北展北街九号

华远·企业号 D 座

投诉电话： 95397

投诉邮箱： kefu@tpyzq.com

重要声明

太平洋证券股份有限公司具有经营证券期货业务许可证，公司统一社会信用代码为：91530000757165982D。

本报告信息均来源于公开资料，我公司对这些信息的准确性和完整性不作任何保证。负责准备本报告以及撰写本报告的所有研究分析师或工作人员在此保证，本研究报告中关于任何发行商或证券所发表的观点均如实反映分析人员的个人观点。报告中的内容和意见仅供参考，并不构成对所述证券买卖的出价或询价。我公司及其雇员对使用本报告及其内容所引发的任何直接或间接损失概不负责。我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业务服务。本报告版权归太平洋证券股份有限公司所有，未经书面许可任何机构和个人不得以任何形式翻版、复制、刊登。任何人使用本报告，视为同意以上声明。