中泰证券 ZHONGTAI SECURITIES

中泰证券研究所 专业 | 领先 | 深度 | 诚信

证券研究报告

Sora是如何成功的？

——技术复盘与产业分析

2024.02.20

分析师：闻学臣
执业证书编号：S0740519090007

分析师：苏仪
执业证书编号：S0740520060001

联系人：王雪晴
Email：wangxq03@zts.com.cn

核心观点

■ **本篇报告中我们深入分析了Sora的各项能力、采用的技术路线和创新性的工作。我们认为Sora是将之前的研究工作上进行了很好的综合，并在强大的算力、工程能力以及GPT和DALL·E模型技术积累下诞生的。随着OpenAI将这种具有开创性的技术路径走通，国内模型和应用厂商有望快速迭代出类Sora的视频生成模型和应用产品。**

■ **Sora的突破可以概括为以下几点：**

- 从生成效果看，突破此前视频生成模型的时长限制，能够生成60s时长、分辨率1080p的视频，可用性极高。
- 从技术路线看，依旧遵从LLM范式“大力出奇迹”，通过patches向量化与transformer架构结合，使得训练数据能够使用大小、尺寸、分辨率不同的视频，能够让模型学习到视频的规律乃至世界的规律；使用GPT生成prompt，在训练和推理过程中解决了模态之间的对齐问题，大大提升了生成效果。
- 从产业发展看，Sora通用性极强，有望统一视频生成生态；能够进一步赋能与促进下游应用发展，未来有望成为真正的“世界模拟器”。

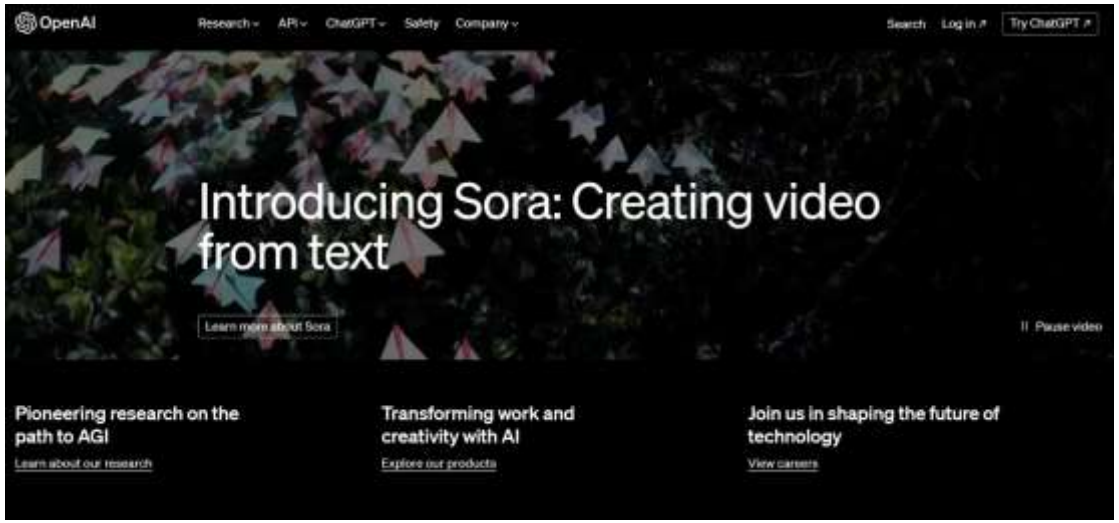
■ **推荐关注的方向：**

- 算力：依旧是确定性最强的方向，“大力出奇迹”的路线会持续带动算力的需求，建议关注浪潮信息、中科曙光、神州数码、紫光股份、首都在线等；
- 应用：强大工具产品和视频模态相关标的有望受益，建议关注万兴科技、焦点科技、金山办公、科大讯飞、萤石网络、中科创达、格灵深瞳、彩讯股份、鼎捷软件、金现代、商汤、云天励飞、云从科技、神思电子、百融云等；
- 安全：安全是软件应用的基础，公司需要针对更严峻的Deepfakes问题进行安全布局，建议关注启明星辰、三未信安、深信服、美亚柏科、天融信、安恒信息、迪普科技、盛邦安全、永信至诚、安博通、信安世纪、绿盟科技、亚信安全、奇安信、麒麟信安、中孚信息等。

1.1 Sora：大规模训练的视频生成模型，支持60s 1080p视频生成

- **2024年2月15日，OpenAI推出了视频生成模型Sora**。Sora进行了大规模的训练，并使用了不同尺寸、分辨率和时长的视频进行训练，并沿用了扩散模型（Diffusion Model）的思路在Transformer架构上进行训练。
- **Sora能够生成长达60s的1080p清晰度视频**。OpenAI认为，构建Sora这样的缩放视频生成模型（Scaling Video Generation Model）是构建物理世界通用模拟器的可行方法。
- **目前Sora内测开放给OpenAI的红队成员，正在对其危害或风险进行评估。**

图表：OpenAI官网对Sora的宣传页



资料来源：OpenAI、中泰证券研究所

图表：Sora可生成1080p清晰度视频

资料来源：OpenAI、中泰证券研究所

1.2 能够理解和生成复杂场景，但对客观物理规律理解不足

- **Sora能够生成具有多个角色、特定类型的动作以及主体和背景的准确细节的复杂场景。**它不仅理解用户在提示中提出的要求，还理解这些事物在现实世界中的存在方式。
- **目前Sora的不足在于难以模拟现实世界中的物理规律，且对于事物发生的因果、时序和空间关系理解不足。**例如模型能够生成一个人咬一口饼干，但饼干可能没有咬痕；模型可能在某些视频中混淆左右；而且可能难以精确描述随着时间推移而发生的事件，例如跟踪特定的相机轨迹等。

图表：Sora生成视频示例

Sora能够生成具有复杂场景的视频

资料来源：OpenAI、中泰证券研究所

图表：Sora生成视频示例

Sora难以理解杯子倾斜后，液体才流到桌面上的时序逻辑

资料来源：OpenAI、中泰证券研究所

1.3 支持多模态结合输入，可编辑、延伸或生成自定义尺寸视频

- **Sora支持图片和视频的多模态输入，以及多模态的混合输入。**除能够根据用户的文本输入生成视频之外，还能够基于DALL·E 2和DALL·E 3生成的图片再进行视频生成。通过输入原视频和文本提示，Sora能够对目标视频风格进行编辑。Sora还可以在输入的多个视频之间生成转场镜头，将不同视频丝滑地连接起来。
- **Sora同样支持（时序）向前或向后延续生成视频，以及可直接以原始尺寸为不同设备生成视频。**它还允许用户先以较低的分辨率快速生成内容，再提升分辨率，以提高生成效率。除此之外Sora还能直接生成高达2048*2048分辨率的图片。

图表：Sora基于DALL·E生成结果（左图）生成的视频（右图）

提示：一只戴贝雷帽穿黑色高领毛衣的柴犬

资料来源：OpenAI、中泰证券研究所

图表：Sora根据视频+文本的输入完成对视频风格的转换编辑

资料来源：OpenAI、中泰证券研究所

1.4 结合GPT、DALL·E的能力与方法，语言理解能力强

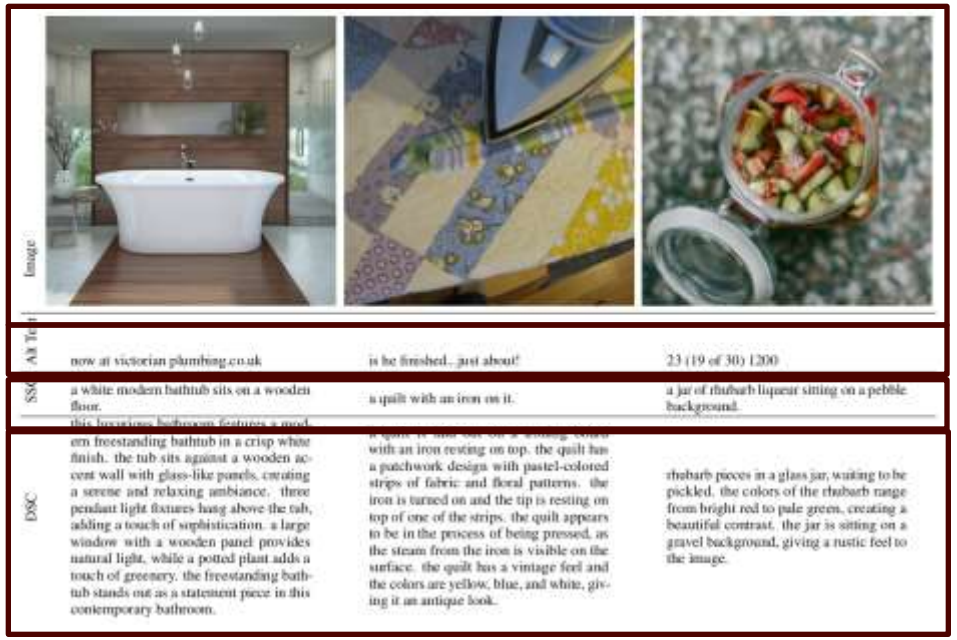
- **基于强大的GPT和DALL·E模型，Sora在训练过程中完成了文本和视频模态的“对齐”，从而能够理解提示中的词汇含义，并能够将其对应到生成视频中的事物上，大大提升了生成视频的准确性。**Sora采用的推理方法与DALL·E 3类似，同样是利用GPT将简短的用户提示转化为更详细的描述，并将其发送给视频模型，用以提高视频生成质量。同时Sora的训练思想也与DALL·E 3类似，也是通过训练文本生成器caption重述文本，再使用生成的文本和视频对模型进行训练。
- **我们认为，OpenAI采用的合成数据方式在大幅提升模型效果的同时，相似风格的文本prompt也能够提升Sora与GPT、DALL·E的联动效果，使得Sora更接近一个“全能”的多模态模型。**

图表：Sora能够理解用户提示中的名词

只改变提示中的部分名词、其他不变，Sora能够准确生成出对应的视频

资料来源：OpenAI、中泰证券研究所

图表：DALL·E 3使用的合成数据方法示意



图片数据

随图抓取的文本（Alt-Text）

合成的短文本（SSC）

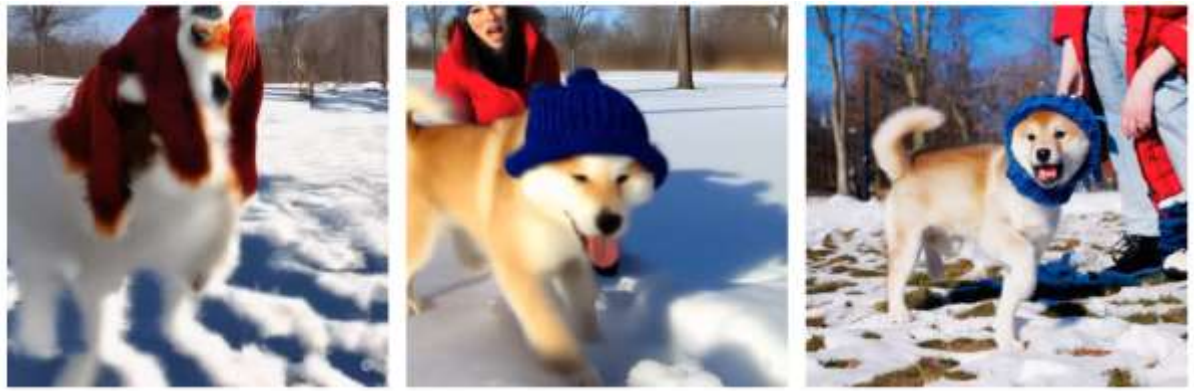
合成的描述性长文本（DSC）

资料来源：Improving Image Generation with Better Captions、中泰证券研究所

1.5 依旧“大力出奇迹”，在大算力下的效果有显著优势

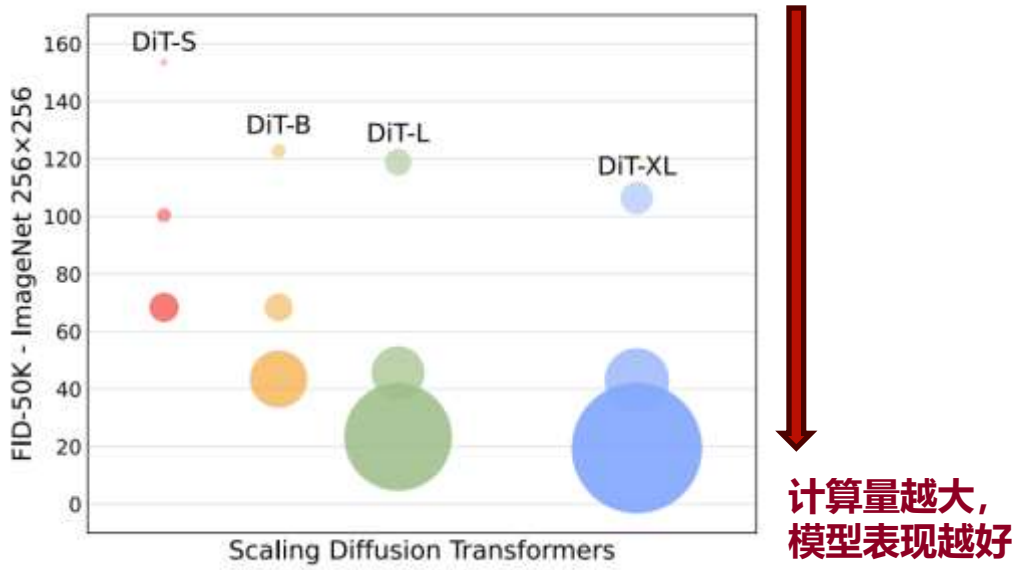
- **Sora作为一种DiT（Diffusion Transformer），也在训练过程中表现出了类似的缩放效应，即在更大算力水平下表现更好。**如图所示在训练过程中，随着计算量的增加，对于固定的输入和种子生成出的样本，质量也也有显著提高。
- 对这个效应DiT论文中进行过更直观的描述：对于同样大小的一组DiT模型（图示中同色），拥有更大气泡面积（即更大运算量）的模型，在纵轴参数FID-50K上的表现越低（该指标越低越好）。
- **我们认为，Sora表现出了类似LLM的Scaling law的效应，“大力出奇迹”的训练方法在视频模态依旧有效。**

图表：大算力等级下的生成效果优势显著



资料来源：OpenAI、中泰证券研究所

图表：DiT模型中的缩放效应



资料来源：Scalable Diffusion Models with Transformers、中泰证券研究所

1.6.1 涌现出更多模拟能力，未来的“世界模拟器”

- **OpenAI发现大规模训练下Sora产生了很多有趣的规模涌现能力，这些能力在构建未来的世界模拟器（Simulators of the Physical World）可能发发挥更大的作用：**
- 1）**3D一致性：**Sora可以生成动态视角下的视频。随着相机的移动和旋转，人物和场景元素在三维空间中始终保持一致。
- 2）**长期一致性和物体持久性：**Sora经常（并不总是）能够有效地模拟短期和长期依赖关系。例如即使在人、动物和物体被遮挡或离开画面时，它们的外观也能保持一致。

图表：Sora生成的视频元素在动态视角下保持一致

资料来源：OpenAI、中泰证券研究所

图表：镜头切换后元素保持外观一致

资料来源：OpenAI、中泰证券研究所

1.6.2 涌现出更多模拟能力，未来的“世界模拟器”

- 3）**与世界交互**：Sora有时可以以简单的方式模拟影响世界状态的动作。例如，画家可以在画布上留下新的笔迹，这些笔迹会随着时间的推移而持续存在；表现一个人吃汉堡时，汉堡上可以留下咬痕。
- 4）**模拟数字世界**：Sora还能够模拟人工过程，例如电子游戏中，Sora可以根据基本策略控制Minecraft中的玩家，同时也可以高保真地渲染世界及其动态。
- **我们认为随着模型持续迭代，这些能力有望得到加强，在3D建模、数字孪生、元宇宙等领域均能得到更好的应用。**

图表：Sora生成画家绘画的视频

资料来源：OpenAI、中泰证券研究所

图表：Sora在Minecraft中操作人物角色

资料来源：OpenAI、中泰证券研究所

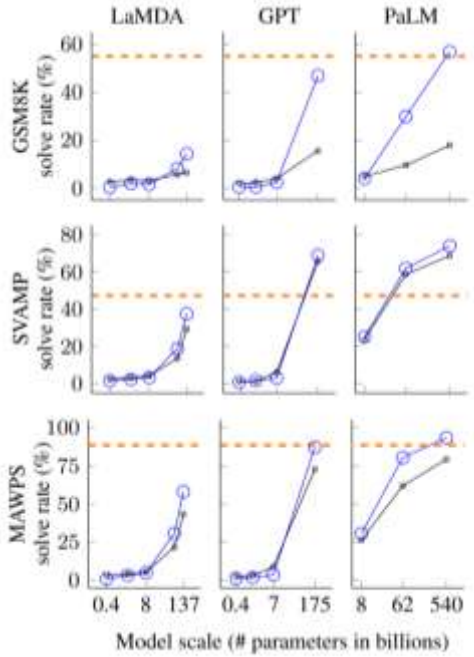
2.1 训练思路：Diffusion Model思想与LLM范式的成功结合

- Sora的训练思想是一次Diffusion Model思想与LLM范式的成功结合：
- Diffusion Model思想指先将图片加入噪声，使模型学习过程，再反向去噪生成“干净”图片的思路。
- LLM范式是指通过大规模无监督训练使模型得到通用涌现能力，Sora的涌现的模拟能力正是这一范式的成功延续。在GPT系列模型中，OpenAI不断增加训练集和模型规模，在更大数量级算力的模型上得到了强大的能力。

图表：Diffusion Model训练思想

资料来源：OpenAI、中泰证券研究所

图表：LLM的能力涌现

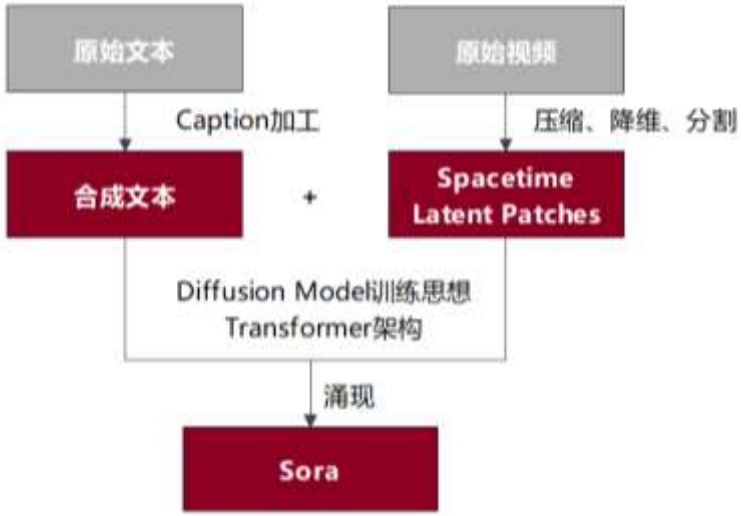


资料来源：Chain-of-Thought Prompting Elicits Reasoning in Large Language Models、中泰证券研究所

2.2 技术创新：采用Patches嵌入，关键在时空信息的处理

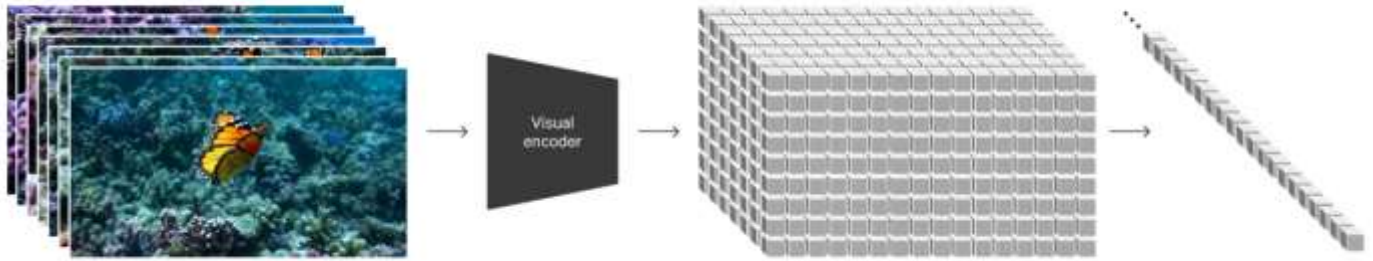
- 在LLM的训练过程中，使用token统一嵌入数字、代码等数据并输入给大模型，是涌现能力能够出现的关键。类似地，在Sora的训练过程中，OpenAI采用了spacetime latent patches作为视频数据的基本嵌入方法。通过这种数据的标准化处理，使Sora能够同时使用不同尺寸的原视频进行训练，从而学到原视频中的构图和取景方法，能够生成更完整的镜头。
- **使用Patches嵌入的优势：**1）Patches能够使用Transformer架构进行处理（而非传统扩散模型的U-Net架构），因此能够支持大规模、多层的训练，构成LLM范式复现的基本条件；2）这种包含了原视频的时空信息，而非单纯按帧进行图像分割，因此能够更加完整地保留原视频的信息和规律。

图表：Sora训练过程示意



资料来源：OpenAI、中泰证券研究所

图表：Sora的嵌入方式



资料来源：OpenAI、中泰证券研究所

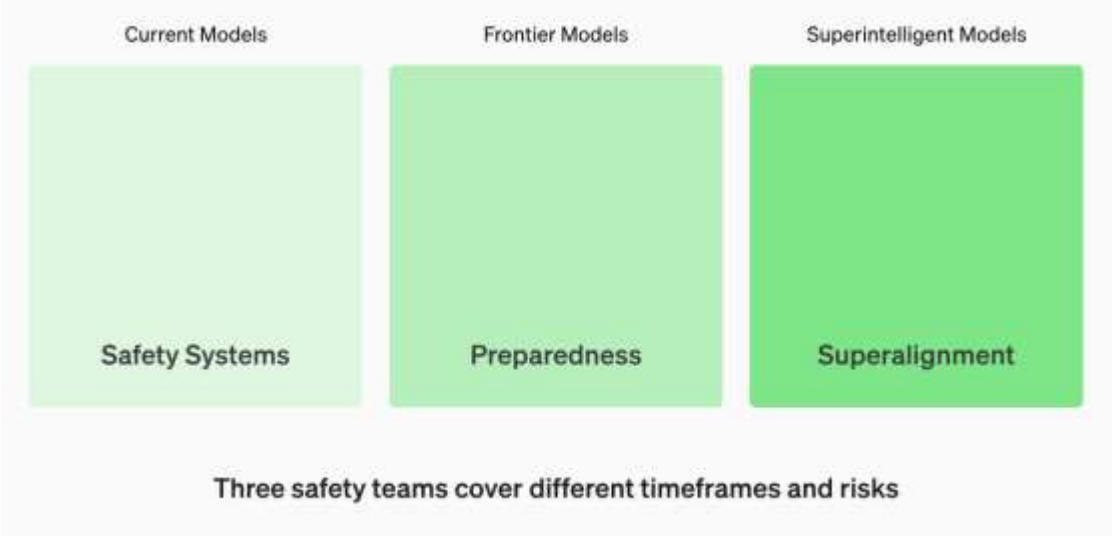
3 Sora的成功对产业意味着什么？

- **Sora的诞生无异是产业的里程碑，以其为代表的“多模型协同”方式是接近AGI的可行道路。**与Gemini这样的多模态模型不同，Sora的核心能力依旧在视频生成领域，且在推理时需要调用GPT的能力重述prompt。这种方式可能不如Gemini符合直觉，但效果非常显著，大大加速了产业走向AGI的过程。
- **Sora代表LLM的通用和涌现范式在视频领域的成功复现，因此算力依旧是模型与应用厂商布局的关键。**通过巧妙的patches嵌入方法，Sora能够运用高效的Transformer架构在海量的视频上进行训练，因此也涌现了模拟现实世界的能力。在其他技术路径的模型获得更好的效果之前，这种“大力出奇迹”的训练方式将依旧是产业的主流，算力需求将持续迎来更大的爆发。
- **Sora可能成为视频生成领域的Base Model，模型层的竞争格局可能走向收敛。**相比其他轻应用，Sora的生成时长更长、质量更高，能够完全替代这些轻应用。因此在多数场景下，Sora都能取代其他的生成模型和应用，最终使视频模型格局走向收敛。

4.1 安全问题：类Sora的视频生成模型可能加大Deepfakes威胁

- **AI技术的崛起已经让大众认识到了深度伪造（Deepfakes）问题。**这种技术能将视频中的脸孔替换成别人的脸孔，甚至创造出虚假的场景。当前有专家强调，实施有针对性的防御措施至关重要，这可能包括为人工智能生成的内容打上独特的标识符或“水印”，以便准确追踪信息源头，及时遏制虚假信息的传播。
- OpenAI的产品中提供Sora之前，也采取了一些安全措施，这些措施包括运用自动化流程，以防止其商业AI模型生成极端暴力、色情内容、仇恨图像以及涉及真实政治领导人或名人的描述。

图表：OpenAI现有的安全团队



资料来源：OpenAI、中泰证券研究所

图表：OpenAI招募红队成员进行安全测试

资料来源：OpenAI、中泰证券研究所

风险提示

- **AI技术落地不及预期：**AI技术更新迅速，如果公司无法跟上技术应用的步伐，可能会被竞争对手超越。同时AI技术的使用会改变用户的工作方式，如果用户不愿意接受这些改变，可能会影响公司的潜在业务增长速度。即使AI技术在实验室环境中表现优秀，但在实际应用中可能遇到许多未预见的问题和挑战。

- **数据更新不及时：**AI领域变化较快，报告中引用的部分图表和数据存在一定的时效性，因此可能面临数据更新不及时的风险

- **安全风险：**AI生成作为新技术，可能带来潜在的安全问题。如果不能妥善解决，会对公司的产业进展、业务研发等产生负面影响。

中泰证券 ZHONGTAI SECURITIES

重要声明

- 中泰证券股份有限公司（以下简称“本公司”）具有中国证券监督管理委员会许可的证券投资咨询业务资格。。本公司不会因接收人收到本报告而视其为客户。
- 本报告基于本公司及其研究人员认为可信的公开资料或实地调研资料，反映了作者的研究观点，力求独立、客观和公正，结论不受任何第三方的授意或影响。本公司力求但不保证这些信息的准确性和完整性，且本报告中的资料、意见、预测均反映报告初次公开发布时的判断，可能会随时调整。本公司对本报告所含信息可在不发出通知的情形下做出修改，投资者应当自行关注相应的更新或修改。本报告所载的资料、工具、意见、信息及推测只提供给客户作参考之用，不构成任何投资、法律、会计或税务的最终操作建议，本公司不就报告中的内容对最终操作建议做出任何担保。本报告中所指的投资及服务可能不适合个别客户，不构成客户私人咨询建议。
- 市场有风险，投资需谨慎。在任何情况下，本公司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。
- 投资者应注意，在法律允许的情况下，本公司及其本公司的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，并可能为这些公司正在提供或争取提供投资银行、财务顾问和金融产品等各种金融服务。本公司及其本公司的关联机构或个人可能在本报告公开发布之前已经使用或了解其中的信息。
-