

## 计算机

### Sora 实现文生视频跨越式突破，AIGC 持续正反馈

#### ► 行业事件：

美国时间 2024 年 2 月 15 日，OpenAI 发布文生视频模型 Sora；谷歌推出最新 Gemini 1.5 Pro 模型。

#### ► 视频生成模型 Sora 实现众多突破

OpenAI 发布了文生视频大模型 Sora，可根据简短提示词，生成长达 60s 高清短视频，将此前行业视频生成长度大幅提升。该模型拥有文本到视频，长时生成能力；复杂场景和角色生成能力；语言理解能力；多镜头生成能力；物理世界模拟能力等。Sora 的诞生预示着用户可以省去繁琐冗杂的制作过程，将自我想象力转换为视频作品，有望提升短视频创作效率，加速短视频平台发展。

#### ► Gemini 1.5 Pro 长上下文能力提升

谷歌推出 Gemini 1.5，其建立在对于 Transformer 和 MoE 架构的领先研究之上，模型效率极大提升，上下文窗口容量可以在生产环境中运行多达 100 万个 token，明显超过现有其他模型。基于长上下文能力提升，Gemini 1.5 可对大量信息进行复杂推理，实现更好地多模态理解和推理，高效处理更长代码相关问题。此外在综合性测试中，Gemini 1.5 Pro 仍能保持高水平的性能，与 1.0 Ultra 表现相差无几。长上下文能力的提升，有望进一步扩展大模型应用场景。

#### ► 投资建议

Sora 的发布标志着文生视频领域的跨越式突破，Gemini 1.5 pro 的推出大大提升了长上下文能力。随着技术的进步和应用场景的不断拓展，AIGC 的四个核心要素：数据、算力、模型和应用，正在不断实现正反馈，推动整个行业的创新和发展。

**建议关注：**(1) **国产算力基础设施：**海光信息、寒武纪-U、中科曙光、浪潮信息、紫光股份、锐捷网络等；(2) **视频应用：**万兴科技、海康威视、大华股份等；(3) **网络安全：**国投智能、启明星辰、电科网安等；(4) **游戏：**巨人网络（传媒互联网行业）、昆仑万维（传媒互联网行业）等；(5) **图像应用：**美图公司（传媒互联网行业）等。

**风险提示：**AI 技术发展演进不及预期；商业化进程不及预期；法律政策监管风险；行业竞争加剧等。

投资建议： 强于大市（维持）  
上次建议： 强于大市

#### 相对大盘走势



#### 作者

分析师：姜青山  
执业证书编号：S0590523050001  
邮箱：jiangqs@glsc.com.cn  
分析师：黄楷  
执业证书编号：S0590522090001  
邮箱：huangk@glsc.com.cn

联系人：周砾灵  
邮箱：zhoull@glsc.com.cn

#### 相关报告

1、《计算机：计算机迎 10 年估值底，中长期投资价值显现》2024. 02. 08  
2、《计算机：重仓持股比例环比下降，AI 个股持仓分化》2024. 02. 03

## 1. 视频生成模型 Sora 实现众多突破

### 1.1 Sora 模型实现众多突破

2月16日, OpenAI 发布了文生视频大模型 Sora, 可根据简短提示词, 生成长达 60s 高清短视频, 将此前行业视频生成长度大幅提升。该模型拥有文本到视频, 长时生成能力; 复杂场景和角色生成能力; 语言理解能力; 多镜头生成能力; 物理世界模拟能力等。Sora 的诞生预示着用户可以省去繁琐冗杂的制作过程, 将自我想象力转换为视频作品, 有望提升短视频创作效率, 加速短视频平台发展。

图表1: Sora 可根据文本描述生成长达 60s 的视频



资料来源: OpenAI 官网, 国联证券研究所

#### ➤ 文本到视频, 长时生成能力

Sora 能够根据用户提供的文本描述生成长达 60s 的视频, 这些视频不仅保持了视觉品质, 而且完整准确还原了用户的提示语。

#### ➤ 复杂场景和角色生成能力

Sora 可以生成具有多个角色、特定的运动类型以及细节精准的主题与背景所构建的复杂场景。通过对生动角色以及复杂运镜的应用, 视频具有高度真实性和叙事效果。我们认为对于短视频创作行业, 制作者可以快速将创意进行视频化, 并带来逼真的视觉效果。

图表2: Sora 发布视频展现复杂场景和角色生成能力



资料来源: OpenAI 官网, 国联证券研究所

### ➤ 深刻语言理解能力

Sora 对语言拥有深刻理解能力, 可以准确解释提示并生成可以表达丰富情感、引人注目的角色。我们认为 Sora 对语言的理解能力可以准确的反应用户的指令, 从而创作出更加优质的作品。

图表3: Sora 发布视频展现语言理解能力



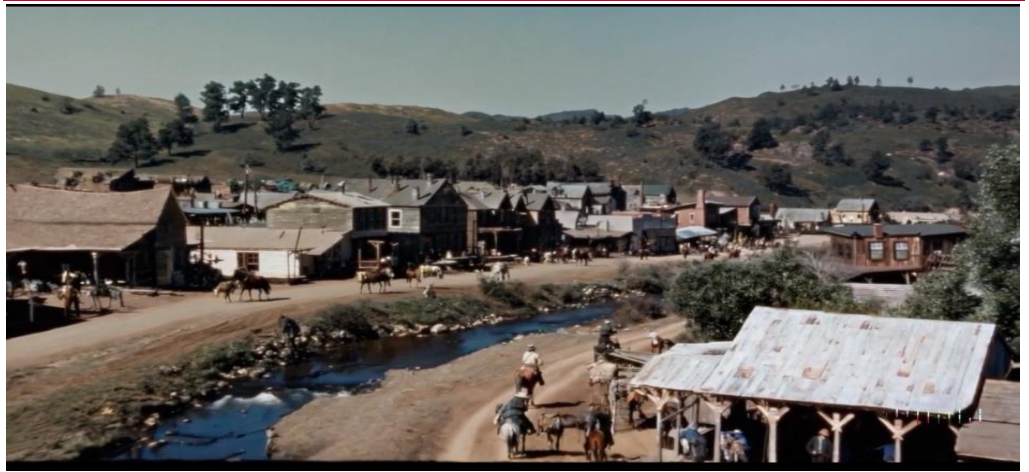
资料来源: OpenAI 官网, 国联证券研究所

### ➤ 多镜头生成能力

Sora 可以在单个生成的视频中创建多个镜头, 同时可保持角色与视觉风格的一致性。多镜头生成功能对于影视制作及其他多视角展示作品可起到极大辅助作用。



图表4: Sora 发布视频展现多镜头生成能力

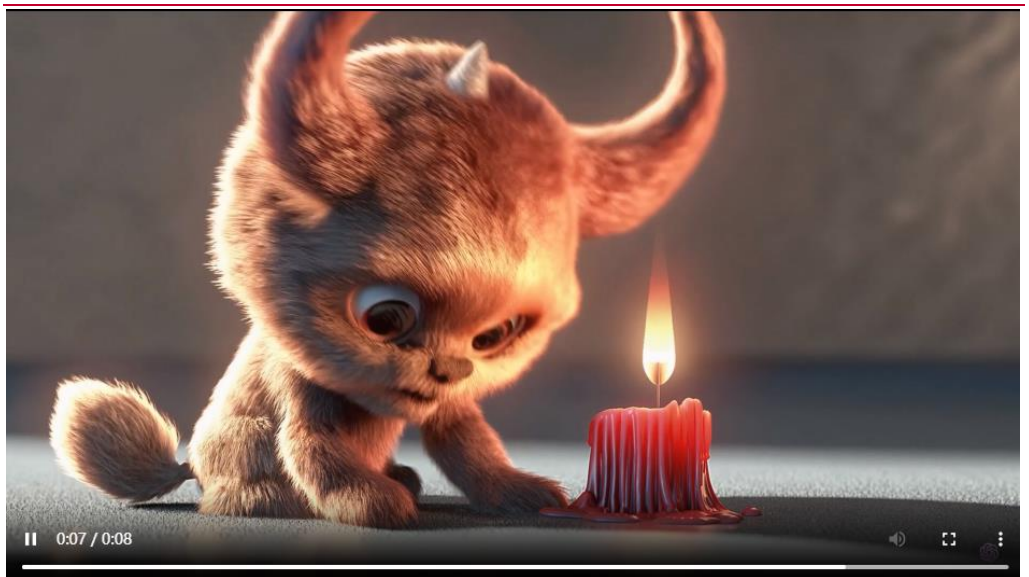


资料来源: OpenAI 官网, 国联证券研究所

### ➤ 物理世界模拟能力

Sora 不仅可以理解用户在提示中输出的指令, 还可以理解所生成的事物在物理世界存在与运行的方式, 从而展现了人工智能在理解现实世界并与之互动的能力。Sora 可以模拟真实物理世界的运动, 例如物体的移动、习性以及相互作用。

图表5: Sora 发布视频展现物理世界模拟能力



资料来源: OpenAI 官网, 国联证券研究所

### ➤ 模型存在一定限制, 仍可逐步提升

Sora 的诞生代表着 AI 行业在视频生成领域产生较大突破, 但其仍存在一定弱点需逐步改善。Sora 可能难以准确模拟在某些复杂场景中物体的物理特性, 并且可能无法理解某些特定的场景物体的因果关系。例如, 一个人咬一口饼干, 但饼干可能没有咬痕。该模型或可能混淆用户生成指令的空间细节, 例如, 左右混淆, 或可能难以精确描述随时间推移发生的事件, 例如遵循特定的相机轨迹。

图表6: Sora 可能无法理解场景物体的因果关系



资料来源: OpenAI 官网, 国联证券研究所

图表7: Sora 可能无法精确描述事件推移发生的事件



资料来源: OpenAI 官网, 国联证券研究所

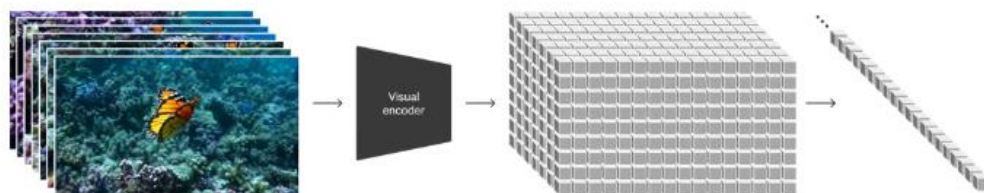
## 1.2 Sora 采用了扩散模型结合 transformer

过去的研究中, 许多团队聚焦于使用递归网络、生成对抗网络、自回归 Transformer 和扩散模型等各种方法, 对大模型进行学习。这些工作局限于较小类别视觉数据、较短或固定大小的视频上。Sora 作为通用视觉模型, 可以生成不同时长、纵横比和分辨率的视频和图像, 最多可达一分钟的高清视频。

### ➤ 将可视数据转换成数据包 (patches)

大语言模型通过 token 将各种形式的文本代码、数学和自然语言统一起来, OpenAI 从中获得灵感, Sora 通过视觉包 (patches) 实现了类似的效果。对于不同类型的视频和图像, 包是一种高度可扩展且有效的表示方式, 对于训练生成模型具有重要意义。

图表8: OpenAI 将输入将视频转换成数据包 patches



资料来源: OpenAI 官网, 国联证券研究所

### ➤ 视频压缩网络

OpenAI 训练一个降低视觉数据维度的网络。这个网络将原始视频作为输入, 并输出经过压缩的潜在表示。Sora 在这个压缩的潜在空间中接受训练并生成视频。

OpenAI 还设计了相应的解码器模型，将生成的潜在表示映射回像素空间，从而生成可视的视频或图像。

### ➤ 时空包

当给定一个压缩后的输入视频时，OpenAI 会从中提取出一系列的时空包，这些包被用作转换 token。这一方案不仅适用于视频，也适用于图片，因为视频由连续帧构成的，所以图像可以视为单帧的视频。通过这种时空包的表示方式，Sora 可以对不同分辨率、持续时间和纵横比的视频和图像进行训练。在推理阶段，OpenAI 可以在适当大小的网格中排列随机初始化的时空包，从而控制生成视频的大小。

### ➤ 用于视频生成的缩放 Transformers

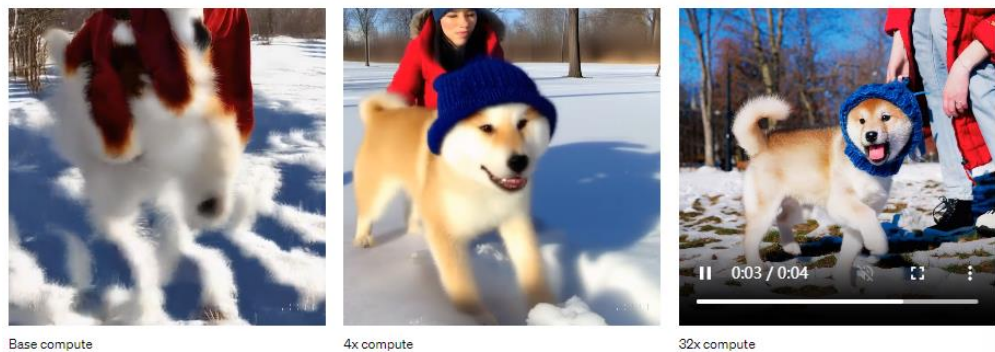
Sora 是一个扩散模型，可输入噪声包（以及如文本提示等条件性输入信息），被训练去预测原始的“干净”包。Sora 是一个基于扩散的转换器模型，这种模型已经在语言建模、计算机视觉以及图像生成等领域，展现了显著的扩展性。

图表9：随着训练量的增加，扩散转换器生成的样本质量有了明显提高



资料来源：OpenAI 官网，国联证券研究所

图表10：随着训练量的增加，扩散转换器生成的样本质量有了明显提高



资料来源：OpenAI 官网，国联证券研究所

### 1.2.1 丰富的持续时间、分辨率与纵横比

过去，图像和视频生成方法需要将视频调整大小、裁剪或修剪至标准尺寸，如 4 秒、256x256 分辨率的视频。OpenAI 发现，对原始大小的数据进行训练有几个优势。



**采样更灵活：**Sora 可以采样宽屏 1920x1080p 视频、垂直 1080x1920 视频以及介于两者之间的所有视频。这使得 Sora 可以直接为各种设备生成与其原始纵横比完美匹配的内容。它还使 OpenAI 能够在以全分辨率生成之前快速以较小尺寸制作内容原型。这些都得益于使用相同的模型。

图表11: Sora 丰富的持续时间、分辨率与纵横比



资料来源：OpenAI 官网，国联证券研究所

**改进构图与框架：**OpenAI 根据经验发现，以原始纵横比对视频进行训练可以改善构图和取景。OpenAI 将 Sora 与将所有训练视频与裁剪为正方形的模型版本进行比较，可发现在方形裁剪（左）上训练的模型有时会生成仅部分可见主体的视频。相比之下，Sora（右）的视频取景有所改善。

图表12: Sora 改进构图与框架



资料来源：OpenAI 官网，国联证券研究所

### 1.2.2 语言理解深化

训练文本到视频生成系统需要大量带有相应文本字幕的视频。OpenAI 将 DALL·E3 中引入的 re-captioning 技术应用于视频。OpenAI 首先训练一个高度描述性的字幕生成器模型，然后使用它为训练集中的所有视频生成文本字幕。OpenAI 发

现，对高度描述性视频字幕进行训练可以提高文本保真度以及视频的整体质量。与 DALL·E 3 类似，OpenAI 还利用 GPT 将简短的用户提示转换为较长的详细字幕，然后发送到视频模型，这使得 Sora 能够生成高质量视频。

### 1.2.3 图片与视频提示

Sora 还能接受其他类型的输入提示，如预先存在的图像或视频。这种多样化的提示方式使 Sora 能够执行广泛的图像和视频编辑任务，如创建完美的循环视频、将静态图像转化为动画、向前或向后扩展视频等。

**将 DALL·E 图片变成动画：**Sora 还能在提供图像和提示作为输入的情况下生成视频。下面展示的示例视频截图基于 DALL·E 2 和 DALL·E 3 的图像生成的。这些示例展示了它在图像和视频编辑领域的无限潜力。

图表13: Sora 将图片拓展为视频



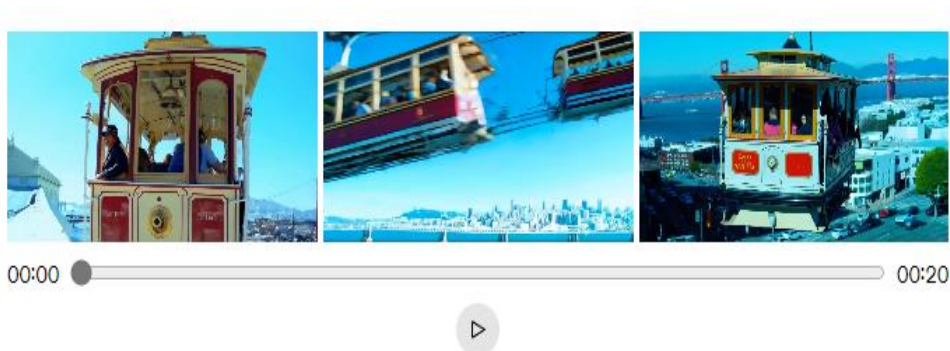
A Shiba Inu dog wearing a beret and black turtleneck.

资料来源：OpenAI 官网，国联证券研究所

**扩展生成视频：**Sora 不仅具备生成视频的能力，更能在时间维度上实现向前或向后的无限扩展。以下三个视频截图便是从同一生成视频片段出发，逐步向后扩展的示例。尽管它们的起始部分各异，但结局却一致。



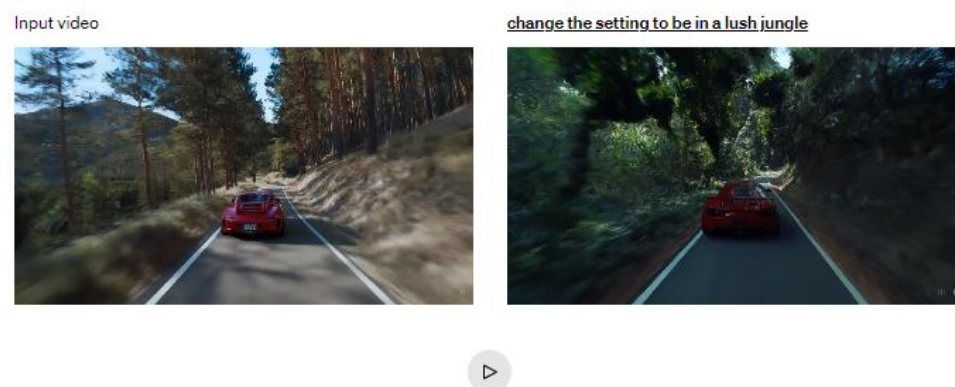
图表14: Sora 对视频扩展生成能力



资料来源: OpenAI 官网, 国联证券研究所

**视频到视频编辑:** 扩散模型已经实现了从文本提示编辑图像和视频的多种方法。下面 OpenAI 应用了其中一种方法, 将 SDEdit 的技术到 Sora。这种技术使得 Sora 能够零样本 (zero-shot) 地转换输入视频的风格和环境。

图表15: Sora 视频从到视频编辑能力



资料来源: OpenAI 官网, 国联证券研究所

**视频的无缝连接:** Sora 能在两个截然不同的输入视频之间实现无缝过渡。通过逐渐插入技术, OpenAI 能够在具有完全不同主题和场景构图的视频之间创建出流畅自然的过渡效果。

#### 1.2.4 图片生成能力

Sora 还可以生成图像。这得益于一种独特的算法, 该算法可在精确时间范围内, 巧妙地在空间网格中排列高斯噪声补丁。它可以根据用户需求, 生成不同大小的图像, 最高可达惊人的  $2048 \times 2048$  分辨率。

图表16: Sora 图片生成能力



Close-up portrait shot of a woman in autumn, extreme detail, shallow depth of field



Vibrant coral reef teeming with colorful fish and sea creatures

资料来源: OpenAI 官网, 国联证券研究所

### 1.2.5 新的模拟能力

OpenAI 发现, 视频模型在展现时表现出许多新兴能力。这些能力使得 Sora 能够模拟物理世界中的一些人、动物和环境的某些方面。这些属性在没有通过对 3D 建模、物体等的任何显性归纳偏差的情况下出现——它们纯粹是模型规模现象。

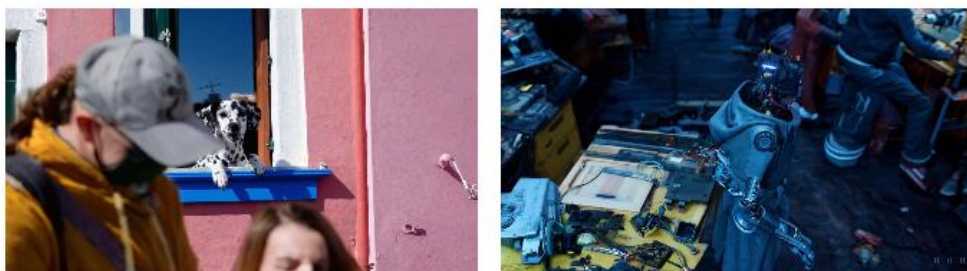
图表17: Sora 展示物理世界模拟能力



资料来源: OpenAI 官网, 国联证券研究所

**较长视频相关性和物体持久性:** 视频生成系统面临的一个重大挑战是在生成视频时保持时间一致性。OpenAI 发现 Sora 通常 (尽管并非总是) 能够有效地对短期和长期物体间的依赖关系进行建模。例如, OpenAI 的模型可以保留人、动物和物体, 即使它们被遮挡或离开画面后。同样, 它可以在单个样本中生成同一角色的多个镜头, 并在整个视频中保持其外观的一致性。

图表18: Sora 展现较长视频相关性和物体持久性



资料来源: OpenAI 官网, 国联证券研究所

**与世界互动:** SORA 可以用简单的方式模拟影响世界状况的动作。例如, 画家可以在画布上留下新的笔触, 并随着时间的推移而存在, 或者一个人吃汉堡并留下咬痕。

图表19: Sora 展示与世界互动特点



资料来源: OpenAI 官网, 国联证券研究所

**模拟数字世界:** Sora 还能够模拟人工过程——一个例子是视频游戏。Sora 可以同时通过基本策略控制《我的世界》中的玩家, 同时以高保真度渲染世界及其动态。这些能力可以通过用提及“我的世界”的标题提示生成视频, 无需额外训练数据或模型参数调整。这些功能表明, 视频模型的持续扩展是开发物理和数字世界以及生活在其中的物体、动物和人的高性能模拟器的一条高效途径。

图表20: Sora 操控 Minecraft 玩家



资料来源: OpenAI 官网, 国联证券研究所



## 2. 谷歌推出 Gemini 1.5 Pro，长上下文能力提升

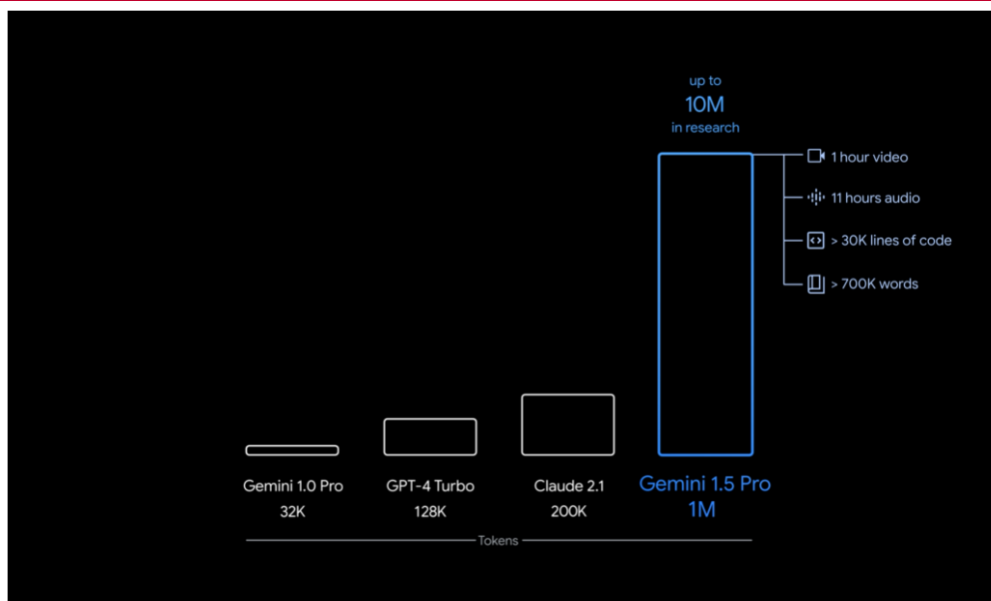
2月16日，谷歌宣布推出其最新的AI模型——Gemini 1.5，该模型在信息处理能力上实现了重大突破。与前代产品 Gemini 1.0 Ultra 相比，Gemini 1.5 在多个方面均有所改进，可更高效、更长时间地处理信息。

**上下文窗口越大，输出效果越好。**AI模型的“上下文窗口”由 token 组成，而 token 是用于处理信息的构建块，可以是文字、图像、视频、音频或代码的整个部分或子部分。模型的上下文窗口越大，它在给定提示中可以接收和处理的信息就越多，从而使其输出更加一致、相关和有用。

**MoE 架构提升模型效率。**Gemini 1.5 建立在谷歌对于 Transformer 和 MoE 架构的领先研究之上。不同于传统的作为一个庞大的神经网络运行的 Transformer，MoE 模型由众多小型的“专家”神经网络组成。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。

**Gemini 1.5 Pro 上下文容量远超 1.0。**通过一系列机器学习创新，Gemini 1.5 Pro 可以在生产环境中运行多达 100 万个 token，远远超出了 Gemini 1.0 最初的 32,000 个 token 的上下文窗口容量。这意味着 1.5 Pro 可以一次性处理大量信息，如 1 小时的视频、11 小时的音频、超过 30,000 行代码或超过 700,000 个单词的代码库。谷歌还成功测试了多达 1000 万个 token。

图表21: Gemini 1.5 可以在生产环境中运行多达 100 万个 token



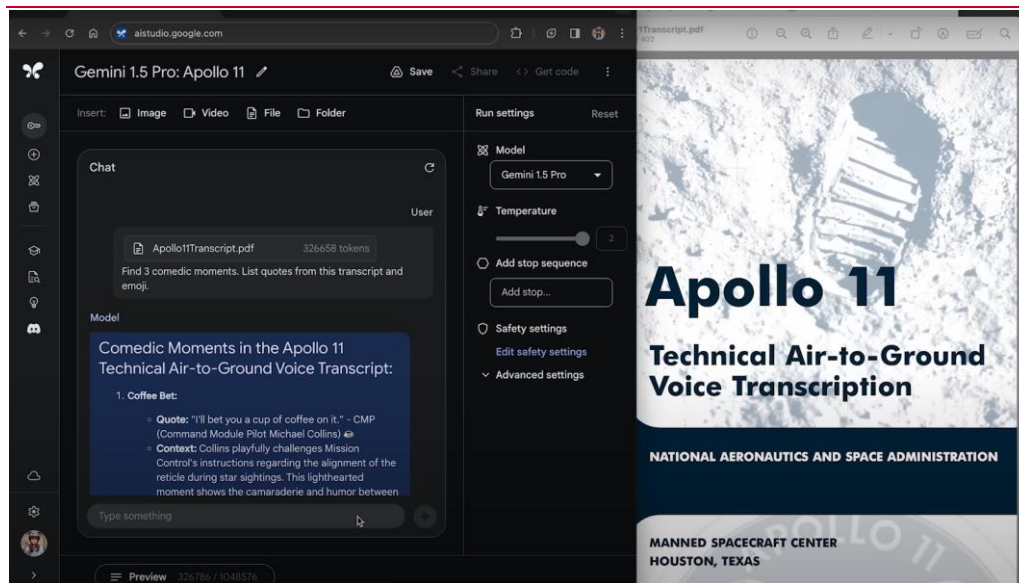
资料来源：谷歌官网，国联证券研究所

### ➤ 对大量信息进行复杂推理

Gemini 1.5 Pro 可以在给定提示内无缝分析、分类和总结大量内容。例如，当给出阿波罗 11 号登月任务的 402 页记录时，它可以推理整个文档中的对话、事件

和细节。

图表22: Gemini 1.5 Pro 可以对大量信息进行复杂推理

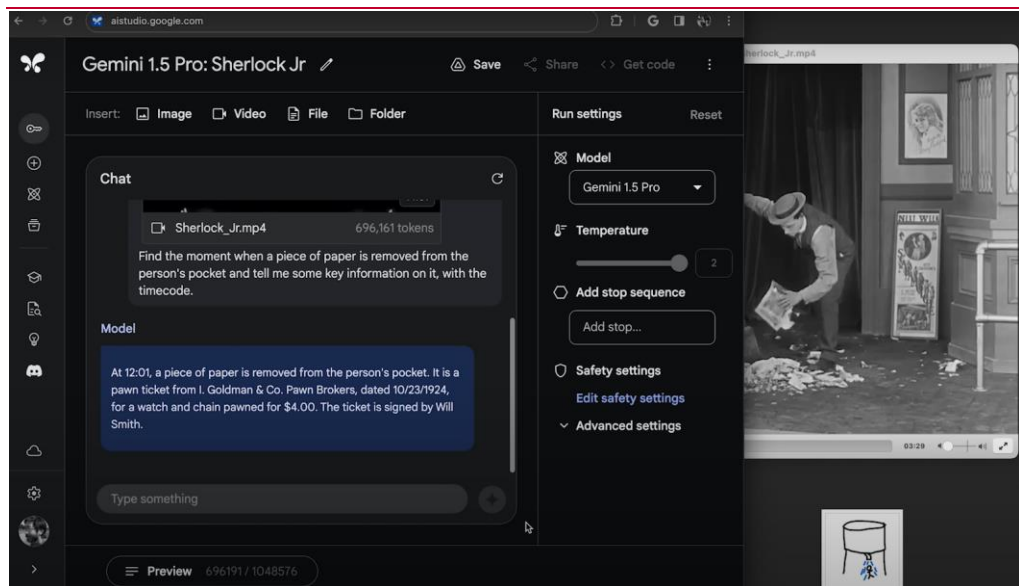


资料来源：谷歌官网，国联证券研究所

### ➤ 更好地多模态理解和推理

1.5 Pro 可以针对包括视频在内的不同模态执行高度复杂的理解和推理任务。例如，当给定一部 44 分钟的巴斯特·基顿无声电影时，该模型可以准确分析各种情节节点和事件，甚至推理出电影中容易被忽略的小细节。

图表23: Gemini 1.5 Pro 可以更好地多模态理解和推理

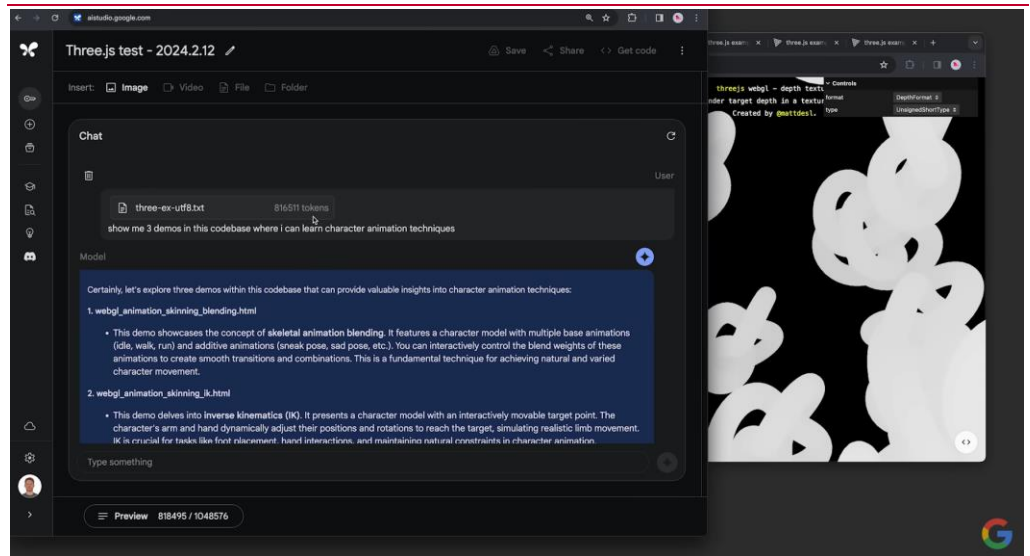


资料来源：谷歌官网，国联证券研究所

### ➤ 高效处理更长代码相关问题

1.5 Pro 可以跨较长的代码块执行相关的问题解决任务。当给出超过 100,000 行代码的提示时，它可以更好地推理示例、建议有用的修改并解释代码不同部分的工作原理。

图表24: Gemini 1.5 Pro 可以高效处理更长代码相关问题



资料来源：谷歌官网，国联证券研究所

➤ Gemini 1.5 Pro 在扩大上下文窗口后，保持了高性能

即使上下文窗口增加，Gemini 1.5 Pro 仍能保持高水平的性能。在涵盖文本、代码、图像、音频和视频的综合性测试中，1.5 Pro 在 87% 的基准测试上超越了 1.0 Pro。与 1.0 Ultra 在相同基准测试的比较中，1.5 Pro 的表现也相差无几。

图表25: Gemini 1.5 Pro 在扩大上下文窗口后，保持了高性能

Core Capability		Relative to	
		1.0 Pro	1.0 Ultra
Text	Math, Science & Reasoning	+28.9%	+5.2%
	Multilinguality	+22.3%	+6.7%
	Coding	+8.9%	+0.2%
	Instruction following	+9.2%	+2.5%
Vision	Image understanding	+6.5%	-4.1%
	Video understanding	+16.9%	+3.8%
Audio	Speech recognition	+1.2%	-5.0%
	Speech translation	+0.3%	-2.2%

资料来源：《Gemini 1.5: Unlocking multimodal understanding across millions of tokens of context》，国联证券研究所

### 3. 投资建议

Sora 的发布标志着文生视频领域的跨越式突破，Gemini 1.5 pro 的推出大大提升了长上下文能力。随着技术的进步和应用场景的不断拓展，AIGC 的四个核心要素：数据、算力、模型和应用，正在不断实现正反馈，推动整个行业的创新和发展。

**建议关注：**(1) **国产算力基础设施：**海光信息、寒武纪-U、中科曙光、浪潮信息、紫光股份、锐捷网络等；(2) **视频应用：**万兴科技、海康威视、大华股份等；(3) **网络安全：**国投智能、启明星辰、电科网安等；(4) **游戏：**巨人网络（传媒互联网行业）、昆仑万维（传媒互联网行业）等；(5) **图像应用：**美图公司（传媒互联网行业）等。



#### 4. 风险提示

AI 技术发展演进不及预期；商业化进程不及预期；法律政策监管风险；行业竞争加剧等。

## 分析师声明

本报告署名分析师在此声明：我们具有中国证券业协会授予的证券投资咨询执业资格或相当的专业胜任能力，本报告所表述的所有观点均准确地反映了我们对标的证券和发行人的个人看法。我们所得报酬的任何部分不曾与，不与，也将不会与本报告中的具体投资建议或观点有直接或间接联系。

## 评级说明

投资建议的评级标准		评级	说明
报告中投资建议所涉及的评级分为股票评级和行业评级（另有说明的除外）。评级标准为报告发布日后6到12个月内的相对市场表现，也即：以报告发布日后的6到12个月内的公司股价（或行业指数）相对同期相关证券市场代表性指数的涨跌幅作为基准。其中：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；香港市场以摩根士丹利中国指数为基准；美国市场以纳斯达克综合指数或标普500指数为基准；韩国市场以柯斯达克指数或韩国综合股价指数为基准。	股票评级	买入	相对同期相关证券市场代表指数涨幅20%以上
		增持	相对同期相关证券市场代表指数涨幅介于5%~20%之间
		持有	相对同期相关证券市场代表指数涨幅介于-10%~5%之间
		卖出	相对同期相关证券市场代表指数跌幅10%以上
	行业评级	强于大市	相对同期相关证券市场代表指数涨幅10%以上
		中性	相对同期相关证券市场代表指数涨幅介于-10%~10%之间
		弱于大市	相对同期相关证券市场代表指数跌幅10%以上

## 一般声明

除非另有规定，本报告中的所有材料版权均属国联证券股份有限公司（已获中国证监会许可的证券投资咨询业务资格）及其附属机构（以下统称“国联证券”）。未经国联证券事先书面授权，不得以任何方式修改、发送或者复制本报告及其所包含的材料、内容。所有本报告中使用的商标、服务标识及标记均为国联证券的商标、服务标识及标记。

本报告是机密的，仅供我们的客户使用，国联证券不因收件人收到本报告而视其为国联证券的客户。本报告中的信息均来源于我们认为可靠的已公开资料，但国联证券对这些信息的准确性及完整性不作任何保证。本报告中的信息、意见等均仅供客户参考，不构成所述证券买卖的出价或征价邀请或要约。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。客户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、税收等方面咨询专家的意见。对依据或者使用本报告所造成的一切后果，国联证券及/或其关联人员均不承担任何法律责任。

本报告所载的意见、评估及预测仅为本报告出具日的观点和判断。该等意见、评估及预测无需通知即可随时更改。过往的表现亦不应作为日后表现的预示和担保。在不同时期，国联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。

国联证券的销售人员、交易人员以及其他专业人士可能会依据不同假设和标准、采用不同的分析方法而口头或书面发表与本报告意见及建议不一致的市场评论和/或交易观点。国联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。国联证券的资产管理部门、自营部门以及其他投资业务部门可能独立做出与本报告中的意见或建议不一致的投资决策。

## 特别声明

在法律许可的情况下，国联证券可能会持有本报告中提及公司所发行的证券并进行交易，也可能为这些公司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。因此，投资者应当考虑到国联证券及/或其相关人员可能存在影响本报告观点客观性的潜在利益冲突，投资者请勿将本报告视为投资或其他决定的唯一参考依据。

## 版权声明

未经国联证券事先书面许可，任何机构或个人不得以任何形式翻版、复制、转载、刊登和引用。否则由此造成的一切不良后果及法律责任有私自翻版、复制、转载、刊登和引用者承担。

## 联系我们

**北京：**北京市东城区安定门外大街208号中粮置地广场A塔4楼

**无锡：**江苏省无锡市金融一街8号国联金融大厦12楼

电话：0510-85187583

**上海：**上海浦东新区世纪大道1198号世纪汇一座37楼

**深圳：**广东省深圳市福田区益田路4068号卓越时代广场1期13楼