2024 年 05 月 20 日

# Sora 和世界模型殊途同归，自动驾驶行业有望加速

## ——中小盘策略专题

中小盘研究团队

任浪（分析师）
renlang@kysec.cn
证书编号：S0790519100001

赵旭杨（分析师）
zhaoxuyang@kysec.cn
证书编号：S0790523090002

相关研究报告

《世界模型——自动驾驶的下一站—中小盘周报》-2024.5.18

《需求驱动，智驾渗透加速可期，新增推荐三夫户外—中小盘周报》-2024.5.12

《深耕户外历浮沉，坚定转型焕新颜—中小盘首次覆盖报告》-2024.5.6

● **Sora 横空出世，践行规模法则叠加强大工程化能力构筑精品**

**2024 年初，Sora 横空出世，凭借惊艳的视频生成效果和分钟级的时长引领市场。**Sora 生成长达 60 秒的视频，并且可以通过自然语言、视频、图片作为提示词实现视频生成，相比此前的其他文生视频工具性能优势显著。此外 Sora 生成的视频还呈现出时间一致性、空间一致性和因果一致性，被 OpenAI 称为世界模拟器。Sora 在数据、算法、算力上均大胆创新，数据方面，采用了特殊的视频编码模式将视频模块化和压缩，构建适用于视频生成模型的时空模块，通过原本具备的大语言模型能力构建高质量的视频文本数据集合文本生成提示词等。算法层面，引入 DiT 算法增强可扩展性，同时加入某些自回归任务加强模型的帧间信息处理能力。最后 OpenAI 的强大算力也是 Sora 诞生的必要因素。

● **世界模型——自动驾驶的下一站**

**世界模型是预测未来的梦境，将全面赋能自动驾驶。**世界模型即通过对世界基础运行规律的理解来实现对未来的预测。在自动驾驶领域，预测未来可以被用于：生成逼真、稀缺的驾驶场景助力模型的训练以及仿真验证，同时模型也可以直接生成驾驶策略指导自动驾驶运行。而在端到端算法时代，产业对合成数据、闭环验证的需求进一步增强，世界模型的重要性凸显。目前在自动驾驶领域，特斯拉开发了 World Model、Wayve 开发了 GAIA-1、英伟达亦推出自身的基础模型，诸多玩家推出相应产品来实现驾驶场景的视频生成等任务。而在学术界，多种世界模型亦层出不穷，以 DriveDreamer 为例，模型不仅可以实现驾驶场景的生成，更能生成驾驶场景下所应该实现的驾驶行为，为世界模型应用打开想象空间。

● **世界模型、视频生成殊途同归，自动驾驶有望迎加速**

**面向相似的目标，采用相近的方法，多种任务殊途同归，自动驾驶未来已来。**视频生成领域，Sora、Runway 等均表达了希望进军世界模型的想法，而“预测未来”对自动驾驶乃至具身智能都存在不可替代的意义，长时间、稳定的对未来的场景进行预测是诸多行业面临的难点。而在算法架构方面，我们看到视频生成和自动驾驶的世界模型均有诸多相似之处，均将复杂外部世界获取的数据进行编码和压缩、抽象成为低维度的向量，并采用 Transformer 或者其他模型在时空维度学习这些知识进而实现预测，再通过不同类型的解码器将之前生成的潜在空间的向量解码成为我们所需要的信息形式，如视频、点云、甚至执行器的控制信息等。而我们也看到在 Sora 的启发下，OpenSora、Vidu 等视频生成工具迭出，效果不俗。大模型开发和自动驾驶汇集 AI 领域诸多优秀人才和资源，相似的开发方向有望让产业互相借鉴，加速产业发展，推动自动驾驶加速实现。

● **推荐及受益标的：**推荐标的：长安汽车、比亚迪、长城汽车、德赛西威、经纬恒润-W、均胜电子、华阳集团、美格智能、华测导航。受益标的：小鹏汽车-W、理想汽车-W、蔚来-SW、中科创达等。

● **风险提示：**技术进步不及预期、市场需求不及预期、重大事故致行业受挫等。

# 目 录

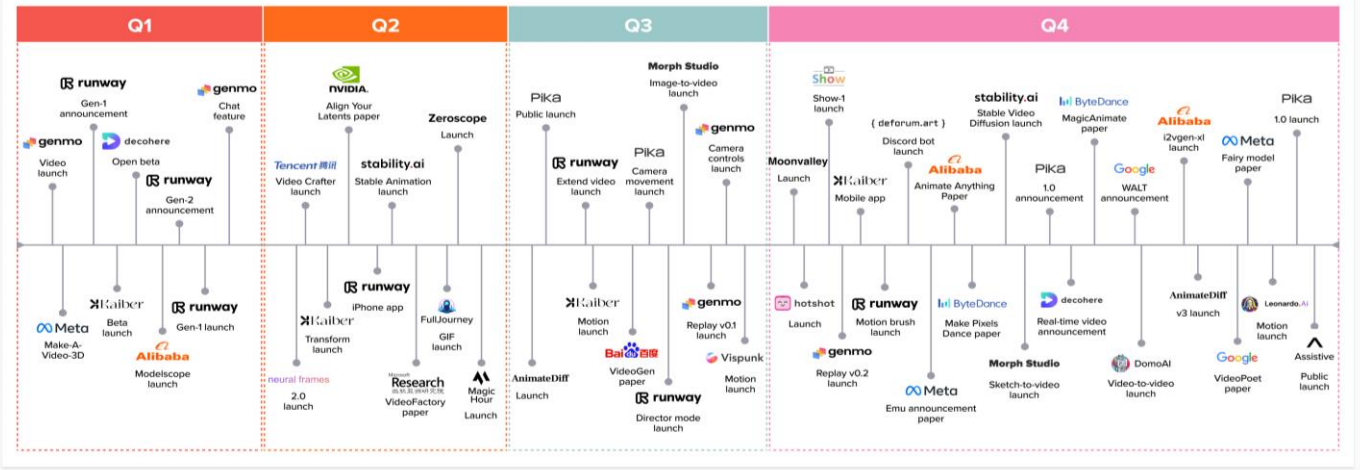


# 图表目录

## 1、 Sora 横空出世，世界模拟器惊艳世人

**AI 生成视频从 2023 年以来呈现快速增长态势，但模型性能一度遇到瓶颈。**根据 A16Z 的统计，AI 视频生成模型在 2023 年四季度呈现井喷式增长。然而在如火如荼的模型发布热潮中，模型本身的进步却难言迅速，大多视频生成模型都遇到了类似的瓶颈：实现较好控制性难度高——即如何让模型精准按照语言的描述控制视频中发生的场景。实现时间一致性难度大——如何让角色、对象和背景在帧之间保持一致，而不会变成其他的东西或者扭曲不易实现，这也直接决定模型生成视频的时长。因此我们通常看到的生成式视频，通常会快速切换画面，并且内容天马行空，这正是为了规避模型弊端采取的举措。

**图1：AI 视频生成模型加速发展 2023 年四季度迎井喷**




资料来源：a16z 官网

### 1.1、 Sora 横空出世，引燃市场热情

**Sora 凭借惊艳的视频生成效果和分钟级的时长引领市场。**前述视频生成模型所遇到的问题在 Sora 诞生后出现根本改变。2023 年 2 月 16 日凌晨，OpenAI 发布了文生视频大模型 Sora，能够根据用户提供的文本描述生成长达 60 秒的视频，同时视频精准反应提示词内容，复杂且逼真，效果惊艳，引燃市场热情。

**图2：Sora 生成东京街头女士，场景复杂**

资料来源：OpenAI 官网

**图3：特写镜头细节饱满效果逼真**

资料来源：OpenAI 官网

### 1.1.1、 功能强大，可完成多种视频图片生成任务

**Sora 功能强大，可实现多种视频和图像生成任务。**（1）Sora 可通过提示词生成视频并灵活改变视频持续时间、分辨率和宽高比，即可以为各类不同的设备生成内容相同或相似的视频；（2）通过图片提示生成视频，如基于 DALL·E2 和 E3 生成的静态图片生成具有动态效果的视频。（3）通过视频提示生成视频，如将不同开头的视频最终生成相同的结局，或者生成无限循环的视频，以及对视频进行编辑，改变其中的某些元素和环境，同时也可以将不同的视频进行拼接。（4）通过提示词生成高分辨率的精美图片。

**图4：Sora 功能强大，可按需完成各类视频生成任务**


资料来源：OpenAI 官网

### 1.1.2、 性能优异，对比其他产品形成显著优势

对比其他的视频生成工具，Sora 的性能优异呈现出碾压式的优势：（1）视频时长：可生成时长长达 1 分钟的视频，并且品质优异，内容稳定；（2）场景复杂内容逼真：Sora 可生成包含多个角色、特定运动类型以及主题精确背景细节复杂的场景，视频效果逼真。（3）语言理解能力优异：Sora 能够深入理解提示词并且精准、忠实的表达。（4）灵活度高：Sora 可随意生成不同时长、长宽比、分辨率的视频。

**表1：Sora 的性能显著优于其他竞品**

| | Open AI sora | 其他 |
|---|---|---|
| 视频时长 | 60 秒 | 至多十几秒 |
| 视频长宽比 | 1920×1080 与 1080×1920 之间任意尺寸 | 固定尺寸，如 16:9/9:16/1:1 等 |
| 清晰度 | 1089P | Upscale 之后达到 4K |
| 扩展视频 | 向前/向后扩展 | 仅支持向后扩展 |
| 视频连接 | 支持 | 不支持 |
| 运动相机模拟 | 强 | 弱 |
| 依赖关系建模 | 强 | 弱 |
| 影响世界状态（世界交互） | 强 | 弱 |

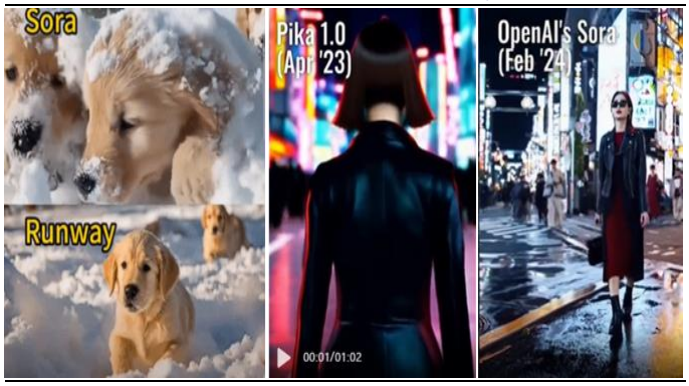
资料来源：元宇宙头条官方公众号、开源证券研究所

以最热门的 Pika 和 RunwayML 以及 Stable Video 和 Sora 做比较，可发现相同的提示词下，Sora 生成的视频不仅时长远超其他，效果也优于同时期其他产品。

**图5：Sora 的视频生成时长远超其他竞品**


资料来源：创业邦公众号

**图6：Sora 的视频生成效果优于 Runway 和 Pika**




资料来源：机器之心公众号

### 1.1.3、 起于视频生成，迈向世界模拟器

**Sora 在进行视频生成任务时，生成的视频一定程度上能够遵循现实世界的物理规律，这使得其模拟现实世界中的人物、动物、环境等，拥有了更广阔的想象空间。（1）空间一致性：**Sora 能够生成带有动态摄像头的运动视频，随着摄像头的移动和旋转，人物和场景元素在三维空间中始终保持一致的运动规律。**（2）时间一致性：**在 Sora 生成的长视频中，元素之间通常能够保持较好的时空一致性，如即使动物被遮挡，或离开画面，在后续的视频中仍然能被较好的呈现。**（3）因果一致性：**Sora 生成的视频可呈现一定的因果关系。比如画家可在画布上留下笔触，人吃汉堡也能在汉堡上留下痕迹。Sora 还能够模拟人工过程，如视频游戏，可用基本策略控制《我的世界》，无需特殊的微调，在 Sora 中提示“我的世界”即可实现。

**图7：Sora 生成的视频在动态变化的过程中，视频元素的 3D 形状和位置保持一致**


资料来源：OpenAI 官网

**Sora 也呈现出一定的局限性，对物理规律的遵循没有那么严格。**在某些场景下无法准确还原物理交互过程，如无法完美的模拟水杯打碎液体飞溅的场景，有些视频中物体会凭空起飞等，表明 Sora 仍然具有较大的提升空间。

**图8：Sora 仍有瑕疵，呈现打碎被子液体飞溅的场景模拟并不准确**

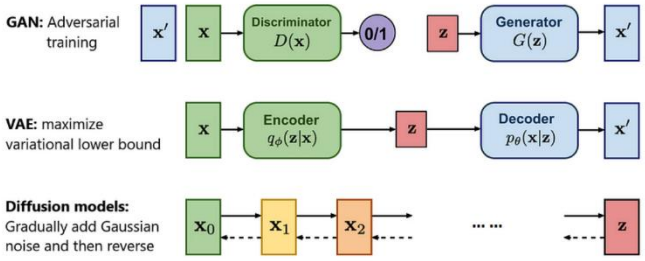

资料来源：OpenAI 官网

## 1.2、 Diffusion 构成 Sora 基座，不断进步羽翼渐丰

### 1.2.1、 扩散模型逐渐成为 AI 视觉生成的主流方案

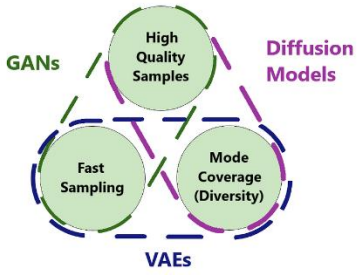
**生成式模型在人工智能领域由来已久，近年随着大模型的兴起，生成式模型逐步占到了舞台中央。**生成式模型类型丰富，常见的有生成对抗网络（GAN，Generative Adversarial Networks）、变分自编码器（VAE，Variational Autoencoders）、扩散模型（Diffusion）、Transformer 等。早年，GAN 和 VAE 模型占据生成式模型市场的主流，GAN 的生成效果尚可但收敛难训练困难，而 VAE 虽然易于训练，但生成效果一般，常常出现样本失真等问题，并不具备大规模使用的基础。Diffusion 生成效果优异样本多样性好，相对更容易收敛，逐步引发市场关注，当然方案本身也存在样本生成速度慢、对算力消耗大等问题，近年亦涌现出基于掩码的自回归视频生成算法，总体而言，在生成式模型领域，算法不断演进，性能亦不断提升。

**图9：生成式模型主要有 GAN、VAE、Diffusion 等**




资料来源：TowardsAI 官网

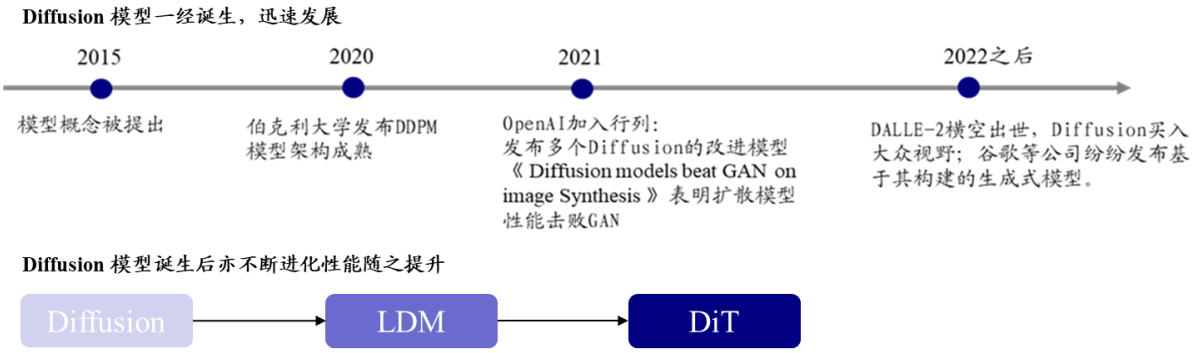
**图10：三大生成式模型性能各有优劣**




资料来源：TowardsAI 官网

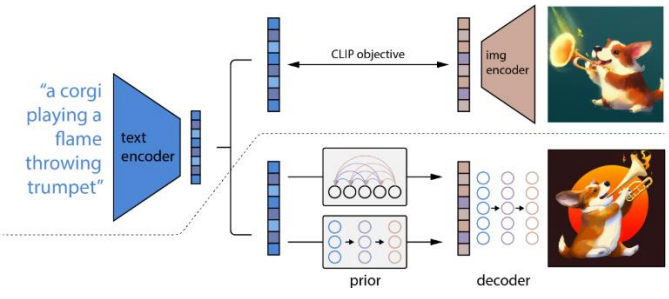
**Diffusion 模型历经发展逐步确立地位。**扩散模型最初在 2015 年被提出，2020 年伯克利大学发布 DDPM 的论文，标志着架构上扩散模型逐步迈向成熟，其后不断有新的机构将扩散模型不断完善，OpenAI 也加入行列之中，发表了“Improved Diffusion”、“Classifier Guidance”、“Classifier Free Guidance”等模型，2021 年 OpenAI 发表文章《Diffusion models beat GAN on image Synthesis》表明扩散模型的性能已经超越其他模型方案。2022 年 DALLE-2 横空出世，通过利用扩散模型和海量数据，该模型呈现出前所未有的理解和创造能力，将扩散模型彻底引入公众视野。此后不到一个月时间谷歌发布文生图模型 Imagen、Stability AI 公司发布 Stable Diffusion 的基石模型 Laion-5B、系列的基于扩散模型的生成式模型不断出现，持续掀起市场热潮，扩散模型逐步衍生出潜在扩散模型（LDM）、Diffusion Transformer 等架构，后期包括 Sora 等文生视频的模型以及部分文生 3D 的模型均以扩散模型作为基础，确立了 Diffusion 模型在视觉生成领域的地位。

**图11：Diffusion 模型近年逐步成为生成式模型的主流方案之一，架构上也不断演进性能持续提升**




资料来源：量子位公众号、《Diffusion models beat GAN on image Synthesis》（Prafulla Dhariwal 等）、开源证券研究所

**图12：DALLE2 采用 Diffusion 模型构建**




资料来源：《Hierarchical Text-Conditional Image Generation with CLIP Latents》（Aditya Ramesh 等）

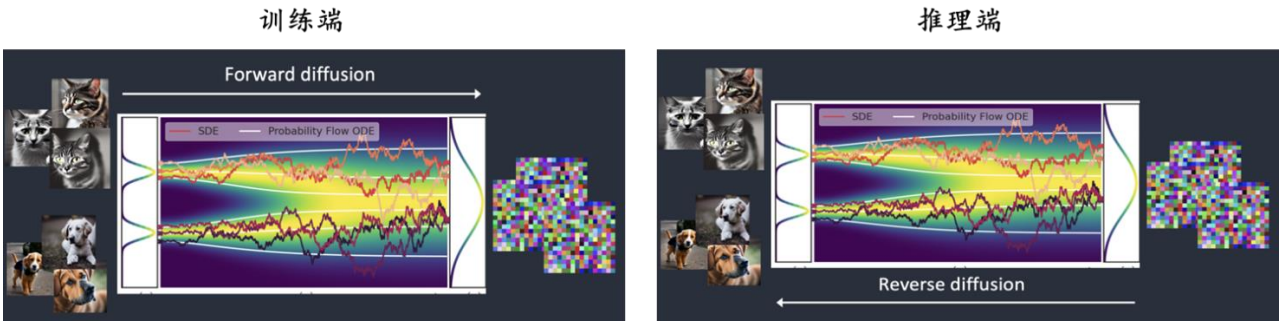
**图13：GAN（左）生成的图片效果不及 Diffusion（中）**

资料来源：《Diffusion models beat GAN on image Synthesis》（Prafulla Dhariwal 等）

### 1.2.2、 扩散模型依靠噪声的添加和祛除实现图像生成

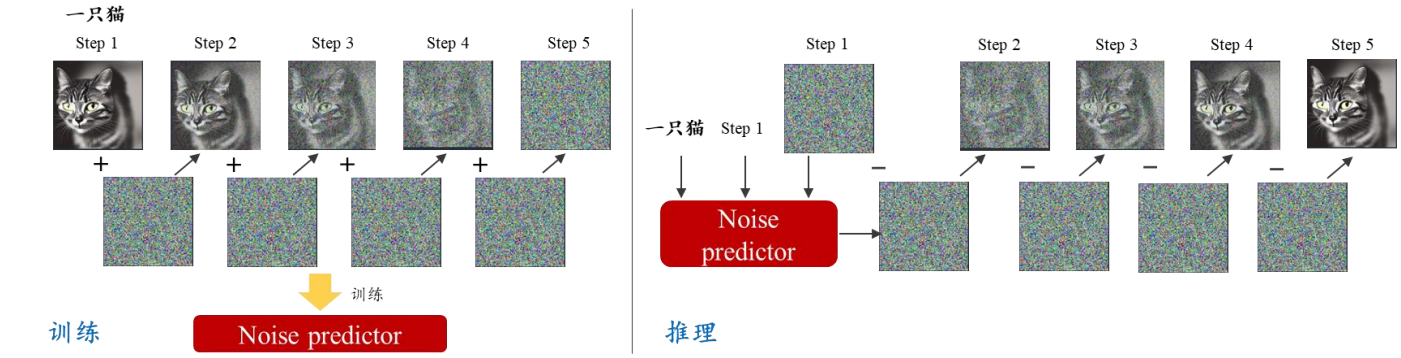
**“加噪声”，“祛噪声”形成扩散模型基本原理。**扩散模型最初受到了热力学扩散定理模型的启发，像墨水滴入清水中一样，通过前向加噪声训练，反向去噪声推理，经过多个步骤渐进式实现视觉内容的生成。具体而言，为了让扩散过程可以逆转，会训练一个神经网络称为噪声预测器（Noise Predictor）。在训练过程中，建立一个噪声预测器神经网络，选择一张照片，加入文字条件，并逐步骤加入噪声使图像变得嘈杂，最终生成纯噪声图片。这一过程中噪声预测器将学习到中间加入了多少次噪声以及每次加入的是何种噪声。在推理过程中，将训练步骤反向操作，让噪声预测器预测并生成当前步骤下图片中的噪声，从前一步噪声图片中减去该步骤下噪声预测器预测的噪声，图像即变得更加清晰，经过多次迭代即可还原出对应的图片。

**图14：“加噪声、“祛噪声”构成 Diffusion 模型的基本原理**




资料来源：太平洋科技搜狐号、开源证券研究所

**图15：Diffusion 模型的核心在于“噪声预测器”**



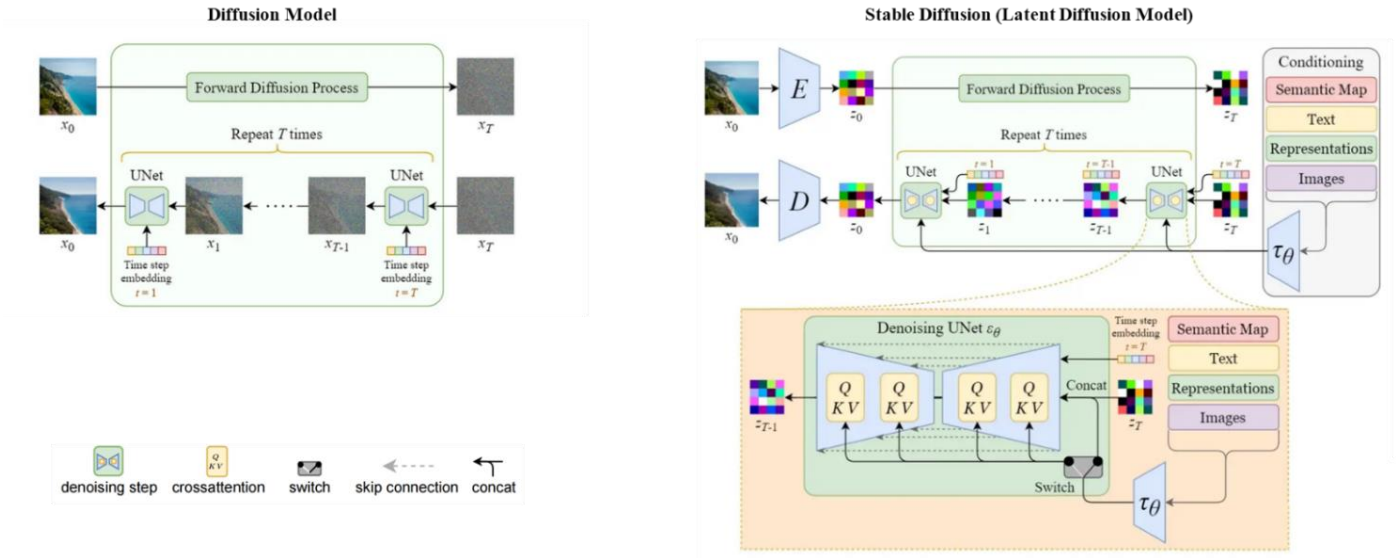

资料来源：太平洋科技搜狐号、开源证券研究所

### 1.2.3、 Stable Diffusion 推动模型迈向更广泛受众

**潜在扩散模型提升计算效率，增强算法能力，助力扩散模型更广泛推开。**前述提到的扩散模型，是在像素空间运行，模型对于算力的消耗巨大，为了解决这一问题，诞生了潜在扩散模型（LDM、稳定扩散模型，Stable Diffusion）。其先通过编码器将图像压缩到一个称作潜在空间的区域中，这时扩散模型将面向潜在空间中的张量来进行添加噪声和祛除噪声的过程，进而大幅减少计算量，之后再将生成的张量通过解码器还原成为图像即可。这样的算法帮助 Stable Diffusion 能够在个人电脑上运行，同时这样的方式也被诸多后续的文生图乃至文生视频的算法所采用包括 OpenAI 的 DALLE-3、甚至 Sora 等。潜在空间（Latent Space）即为抽象的多维空间，能够展示出数据在抽象层面的一些有意义特征和共性，模型通过这些共性的特征可以实现对数据的识别、归类、处理等任务。以人感知世界为例，识别“椅子”时通

常会观察其是否包含四只腿和靠背，而颜色、材质则会被忽略，近似的我们将椅子的概念在大脑中压缩成为“带有四个腿、靠背的物体”。

**图16：相比扩散模型，潜在扩散模型将扩撒步骤压缩到潜在空间进行，大幅提升计算效率**




资料来源：数据派 THU 公众号

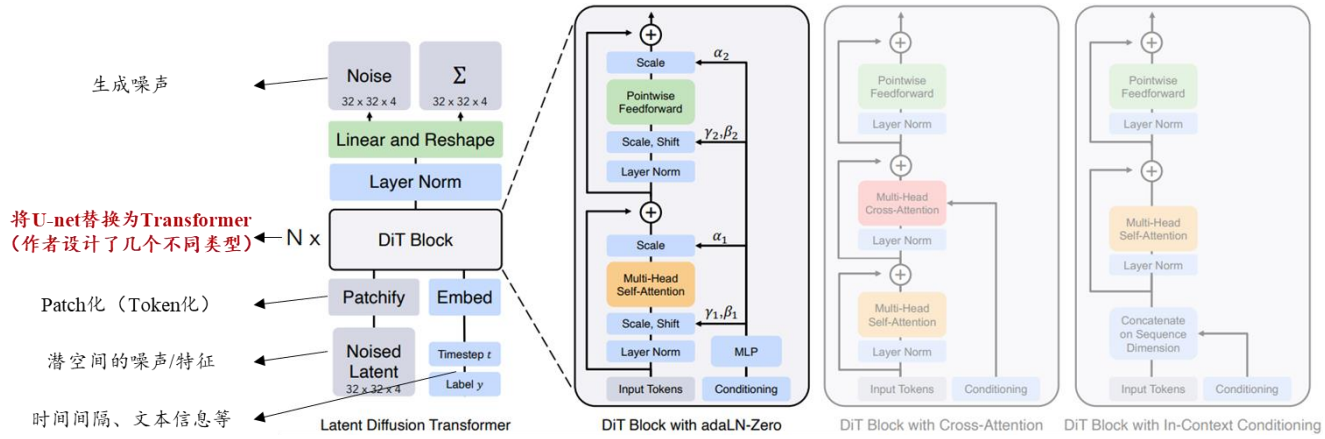
### 1.2.4、 Transformer 作为主干的扩散模型 DiT，规模优势凸显

**Diffusion 进一步进化，与 Transformer 结合，Diffusion Transformer 横空出世。**扩散模型中的噪声预测器是决定模型生成质量的关键，在扩散模型的奠基性文章 DDPM 中，作者采用 U-net 作为噪声预测器的基础网络，U-net 为卷积神经网络（CNN）的一种，具有简洁、语义连贯性强等特点，输入和输出的维度相同，天然适合扩散模型，但在和文本融合的过程中 U-net 会遇到一定的问题。在 2022 年，伯克利大学的 William Peebles 和纽约大学的谢赛宁，发表了论文《Scalable Diffusion Models with Transformers》，在扩散模型中采用 Transformer 替代了传统的 U-net，通过实验证明这样的架构体现出明显的规模效应，其运算速度更快、并且生成的图像效果更佳。

具体而言，模型首先采用类似 Stable Diffusion 的架构，将图像通过解码器（Encoder）压缩至潜在空间，之后参考 ViT（Vision Transformer，视觉 Transformer，用 Transformer 来实现图像分类等任务），将图像压缩并分割（Patchify）成为小的序列（Tokens）。之后送入基于 Transformer 构建的扩散模型中，这里作者设计了四种不同类型架构。最后将生成的序列编码进行解码，输出相应噪声，实现图像生成。

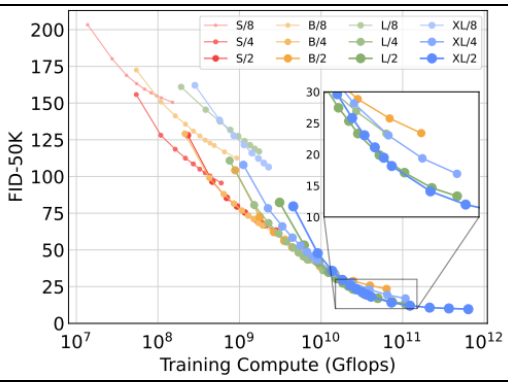
**DiT 的多模态能力提升，视频效果优异，规模效应增强。**相比 CNN，Transformer 在图像处理领域拥有更好的性能表现，同时更加擅长处理多模态的任务，而这一特点也被 Transformer 带到了 Diffusion Transformer 中， DiT 拥有更好的多模态信息处理能力和视频生成效果。除此之外，模型显示出显著的规模效应：更大的计算量会显示出明显更优的计算效果。更小的 Patch Size（意味着更大的计算量）和更大的参数量都带来更好的图像生成效果。这表明 DiT 是一个非常适合于通过规模来提升显示效果的模型。

**图17：谢赛宁等人提出 Vision Transformer，进一步提升扩散模型的性能和规模性**



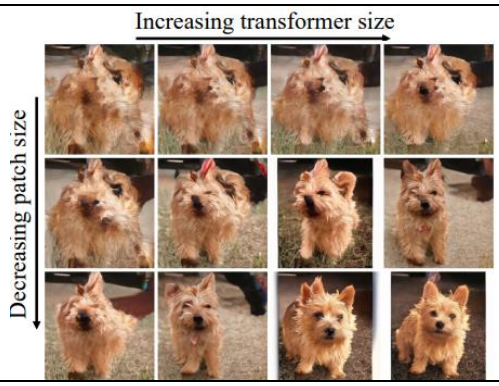

资料来源：《Scalable Diffusion Models with Transformers》（William Peebles 等）、开源证券研究所

**图18：随着算力的提升模型的误差显著降低**




资料来源：《Scalable Diffusion Models with Transformers》（William Peebles 等）

**图19：更大计算量和更多模型参数都将带来性能提升**

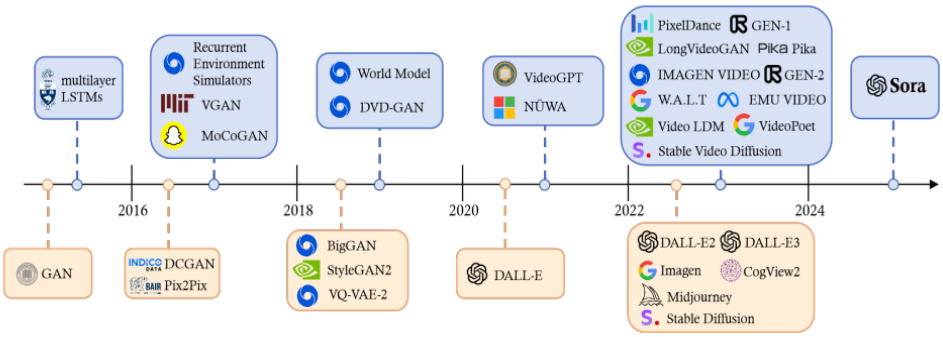



资料来源：《Scalable Diffusion Models with Transformers》（William Peebles 等）

### 1.2.5、 视频生成历经发展，Diffusion 模型逐步占据主要市场

**Diffusion 和 Transformer 结合在视频生成领域崭露头角。**文生视频和文生图拥有着千丝万缕的联系，按照传统，视频本身可以拆分为不同帧的图像，因此文生视频算法的发展通常会伴随文生图算法的演绎路径。早期文生视频领域多采用 GAN 或 VAE 架构，如 VGAN、VQGAN、DVDGAN 等，这一时期生成的视频分辨率低、效果差。随后受到文本 GPT3 和大规模预训练 Transformer 架构的启发，很多玩家开始开发基于 Transformer 架构的文生视频工具如 VideoGPT、NUWA 等。最后伴随扩散模型的广泛应用，人们开始逐步将其应用在视频生成领域，伴随着文生图工具的井喷，文生视频行业在 2023 年也迎来了蓬勃发展的状态。而在 2024 年，Sora 的横空出世更是将文生视频的水平推升到新的高度。后续推出的 Open Sora 采用了类似 DiT 架构，清华大学和生数科技推出的 Vidu 采用了自研的 Diffusion 和 Transformer 融合的 U-ViT 架构，均实现了惊艳的视频生成效果。

**图20：在视频生成领域 Diffusion 模型亦逐步占据主流**



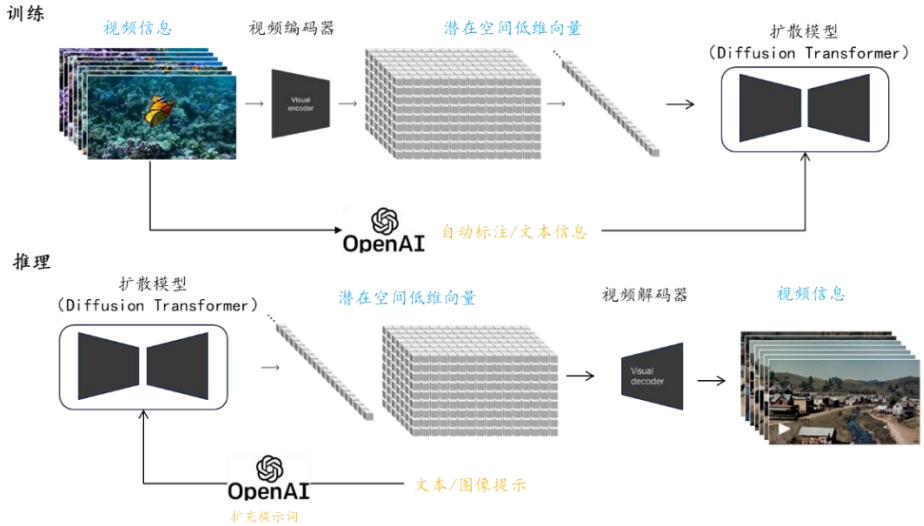

资料来源：《Sora: A Review on Background, Technology, Limitations, and Opportunities of Large Vision Models》（Yixin Liu）

## 1.3、 Sora——践行 Scaling law+强大工程化能力下的产物

### 1.3.1、 Sora 以扩散模型为基是多种技术的结合体

**Sora 是扩散模型和 Transformer 以及视频压缩网络的综合体。**我们可以大致推断 Sora 模型的技术架构。Sora 的主干网络是一个 Diffusion Transformer 模型，在训练过程中采用了特殊设计的编码器将图像和视频信息进行编码，之后将视频数据压缩为隐变量，输入 Diffusion Transformer 模型中对模型进行训练。推理的过程中，将自然语言（文字）或者图像乃至视频作为提示词输入到模型中，通过扩散模型输出相应的去噪之后的隐变量并通过解码器将信息解码成为视频，即可输出品质优越的视频结果。

**图21：Sora 采用了 Diffusion 和 Transformer 结合的架构**




资料来源：OpenAI 官网、开源证券研究所

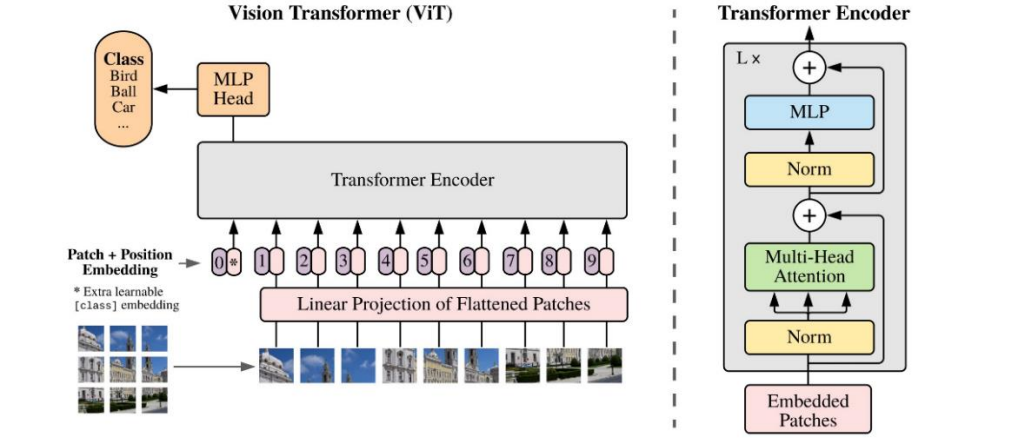
### 1.3.2、 坚定 Scaling law+强大工程化构筑最强视频生成模型

对于大模型来说，除了足够的算力之外，算法结构、数据处理亦是非常重要的环节。相比传统的视频生成模型，Sora 模型在数据、算法等几个方面呈现出明显的特点：

- **数据处理：视频分割和压缩方式、丰富的数据集、强大的自动标注很关键**

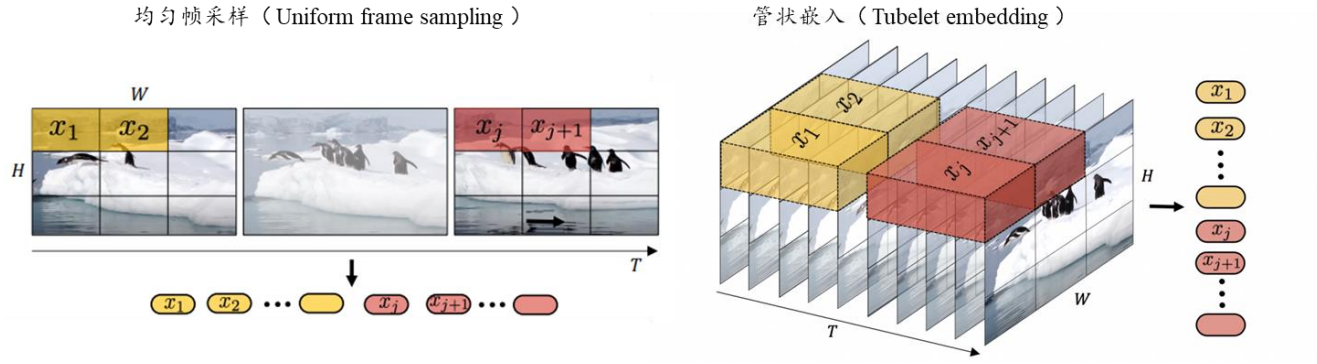
**（1）采用特殊的编码方式对视频进行模块化，构建适合于视频生成模型训练的时空模块（SpaceTime Patchs）。**大语言模型通常将文本信息转变成 Token 输入模型进行训练，而在训练视频生成模型时，如何将整段视频经过压缩、转换之后，恰当的分解成小的片段（Patchs）交给扩散模型训练很关键，这一定程度上将决定模型能从输入的视频信息中学到什么。Patch 早年在 ViT 模型中即有体现，在采用 Transformer 处理图像问题时，会将图像进行分块，进而提升模型处理效率。在视频领域，Sora 技术文档中引用的论文《ViViT: A Video Vision Transformer》介绍了几类视频切分方式：第一种是均匀帧采样，即将每一帧图片进行切分，最后将这些不同帧切分出来的模块一起送入 Transformer；另一种方法则将 T 时间内视频分块，形成管状的模块，模块中同时包含时序和空间信息，这样有效捕捉视频的动态性。我们推断 Sora 采用了第二种视频数据切分方式。而这样的训练方式或许能够让模型获得更好的帧间关系处理能力。

**图22：ViT 中为了更高效的处理图像信息，引入了“Patch”**




资料来源：《An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale》（Alexey Dosovitskiy 等）

**图23：管状嵌入的视频分割模式有望帮助模型获得更好的帧间关系处理能力**




资料来源：《ViViT: A Video Vision Transformer》（Anurag Arnab 等）、开源证券研究所

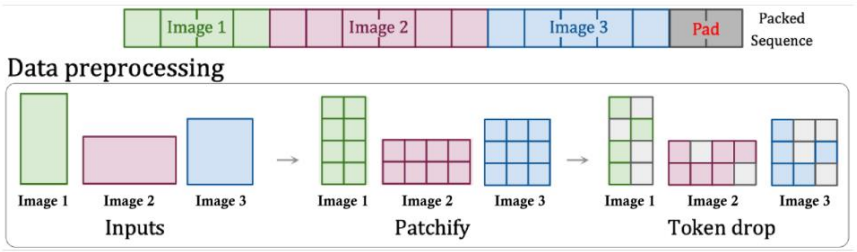
**（2）通过原有文生图等能力，构建高质量的视频文本数据集和文本生成提示词。**在构建 Sora 的过程中，OpenAI 训练了专门的模型对视频进行描述，实现对模型内容的标注。在将文本和视频关联起来的过程中，模型是否能够从训练数据中习得“准确的文字描述”对模型性能会产生关键影响。OpenAI 采用类似 DALLE 的技术训练自动标注模型来对所有视频生成文字字幕，这有助于提升视频质量。同时在推理过程中，Sora 还利用 GPT 将剪短的用户提示扩充为较长且非常详细的提示词输入模型，进而让所生成的视频能够忠实的反应客户提示词。

**（3）Sora 采用了特殊的数据处理形式，能够保证视频以原有尺寸进行训练。**通常情况下的图像和视频生成方法在训练模型时会将视频调整大小、剪裁到标准尺寸。OpenAI 则通过特殊的数据处理方法，可以允许视频数据以原始尺寸进行训练。这使得 Sora 能够以灵活的分辨率生成视频，同时能够改善生成视频的构图。研究人员根据 Sora 技术报告引用的文献推测，OpenAI 采用了论文《Patch n' Pack: NaViT, a Vision Transformer for any Aspect Ratio and Resolution》中的“patch n' pack”的数据处理方式，来实现对多种分辨率/持续时间/宽高比的视频的适应。

**图24：Sora 可以灵活的生成各种尺寸的视频**

资料来源：OpenAI 官网

**图25：OpenAI 或采用了 Pack n'Pack 的处理方式适应不同类型的视频**




资料来源：《Patch n' Pack: NaViT, a Vision Transformer for any Aspect Ratio and Resolution》（Mostafa Dehghani 等）

**（4）特殊的数据集。**Sora 的训练采用了独特的数据集，有研究人员提出，Sora 模型可能采用了游戏引擎的数据来进行训练，到底 OpenAI 采用了什么样的数据集来训练模型仍然是未知数。以往的经验来看，比如此前的研究发现对大语言模型进行代码训练，会显著提高模型的逻辑性。不同的数据集对模型的性能也会有较大影响。

- **算法层面：引入 DiT 算法大幅增强可扩展性**

（1）Sora 采用了 DiT（Diffusion Transformer）算法，将传统 Diffusion 模型中的类卷积神经网络 U-net 替换成为 Transformer 模型，这一方案在 2022 年被伯克利大学的 William Peebles 和纽约大学的谢赛宁提出，并发表在论文《Scalable Diffusion Models with Transformers》中，而 William Peebles 也是 Sora 的主要作者之一。主干网络替换为 Transformer 拥有明显优势，其一，Transformer 目前已经被应用于各类多模态数据的处理，因此本身 Transformer 更适合处理多样化的视频生成任务；其二，Transformer 天然具备捕捉长程或者不规则时间依赖性的能力，在处理长时间维度之间的信息时具有性能有时；最后 Transformer 的规模效应远远好于其他模型，能够充分发挥规模化的优势。（2）OpenAI 的开发者可能在模型训练中增添了某些自回归任务，以让模型能够更好地学习帧与帧之间的关系。

➢ **算力层面：强大算力是实现模型优异效果的基础**

OpenAI 强大的算力基础亦给与模型强力支持。在其技术报告中显示，当算力提升时，模型的推理效果也显著提高。

**图26：算力的提高可显著提升模型的视频生成效果**


资料来源：OpenAI 官网

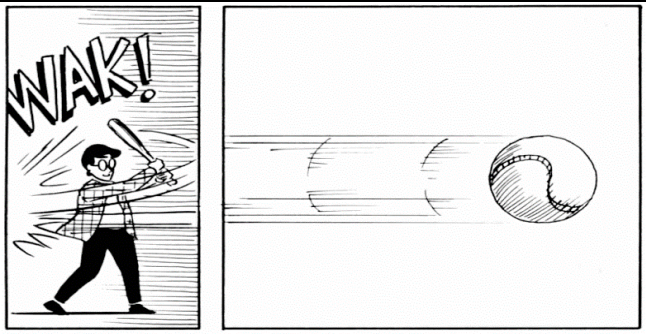
Sora 的出现是 OpenAI 强大工程化能力的综合体现，尽管 OpenAI 并非所有算法的原创，但无论数据集、数据预处理、算法架构上 OpenAI 都进行了诸多探索，寻找出一套行之有效的方案，结合自身强大的算力基础，将 Scaling law 推升到极致，最终诞生现象级的产品。

## 2、 世界模型——自动驾驶的下一站

### 2.1、 世界模型——理解世界，预测未来

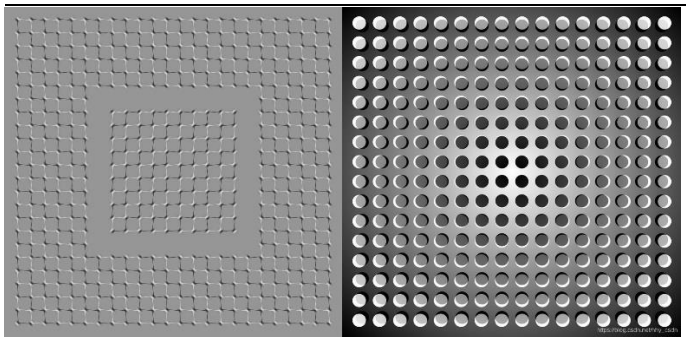
**世界模型——预测未来的“梦境”。**对于世界模型，最早可以追溯到 20 世纪四五十年代的心理学研究，认为任何动物可以依靠世界模型预测世界的下一个状态。谷歌在 2018 年发表了影响深远的论文《World Models》，对世界模型做出了如下描述：人类通常会以有限的感官所能感知到的事物为基础，在内心建立一个世界模型，我们所有的行为都基于这个内部的模型来展开。这样的模型不仅能够预测未来，而且能够根据我们当前的运动行为来预测未来的感官数据，我们能够基于这种预测迅速采取行动。以棒球为列，棒球运动员只有毫秒级的击球时间，甚至比视觉信号从眼球传到大脑还短，因此运动员根本无法在挥棒过程中调整和规划路线，之所以能够提前控制肌肉以正确的方式挥出球棒并击中棒球，得益于他们大脑中的“预测模型”，这个能够预测未来世界状态，在我们大脑中凭空演示一遍的“梦境”，就被称为世界模型。之后这篇论文里还构建了一套世界模型的体系，并通过游戏实验发现，如果让模型在“梦境”中预测未来会发生的事情，那么模型的游戏技能将明显提升。

**图27：运动员击中的实际为人类“世界模型”中的棒球**




资料来源：《World Models》（David Ha 等）

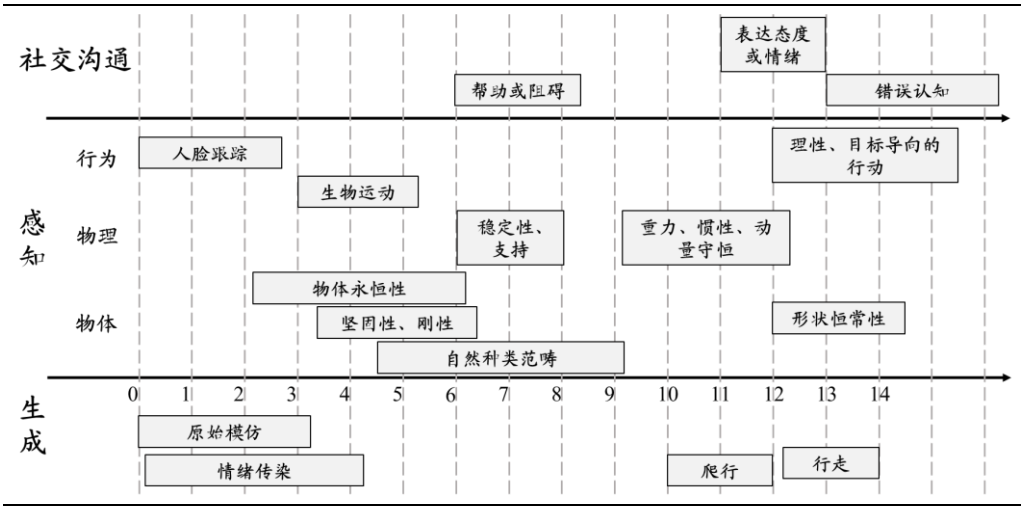
**图28：人类会预测感官未来感知到的事物**



资料来源：《World Models》（David Ha 等）

**世界模型也是对物理世界“常识”的理解。**另外一个被经常提到的是人工智能三巨头之一的 Yann Lecun 的“World Model”。在 Yann Lecun 的著名论文《A Path Towards Autonomous Machine Intelligence》中提到，青少年可以在 20 小时练习中学会开车，而人工智在大量的训练中仍然无法实现良好的自动驾驶，因此动物和人对世界的理解能力远超当前人工智能和机器学习系统。人类和动物快速学习能力是基于对于世界的基础认识和常识。在人类最初出生的几天，几周，几个月就会学习了大量关于世界如何运转的基本知识，如左右眼有视差，物体不会凭空产生、消失、变形或传送，有些物体的轨迹可预测（如无生命的物体）、有些物体的行为方式有些不好预测（如风沙、水、风中的树叶等），在此基础之上，又会形成一些如稳定性、重力、惯性等概念。后续抽象的概念正式基于这样简单的概念建立。有了这些世界的知识（常识），动物或者人类就可以快速学习新事物，对合理与否进行判断。当然“具备常识”可以被认为是一种实现的路径，基于对世界更深入的理解，最终更好的“预测未来”。

**图29：人类在出生后很短的时间内迅速学习到了世界运行的规律**

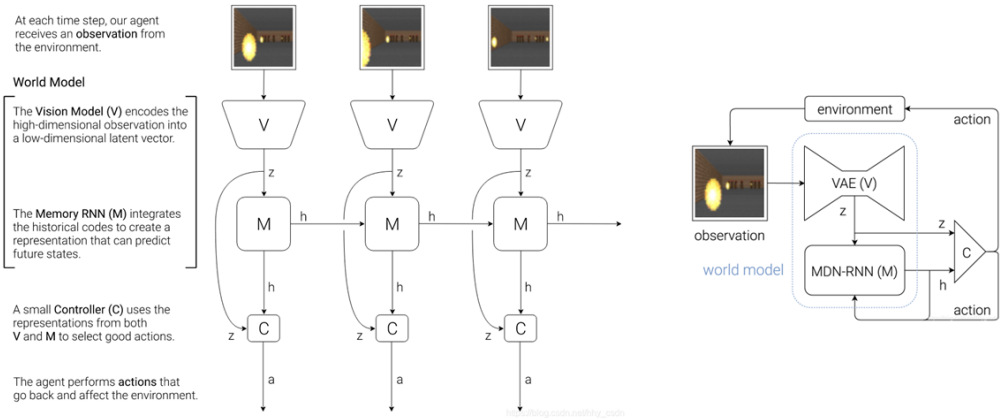



资料来源：《A Path Towards Autonomous Machine Intelligence》（Yann LeCun）、开源证券研究所

## 2.2、 世界模型——感知/记忆/控制的综合体

**世界模型的构建——感知、记忆、控制模块齐备。**在谷歌和 DeepMind 的论文中，提出了一个简单的世界模型构建的方法，包含视觉感知组件（Vision Model，V）、记忆组件（Memory RNN，R）、和控制组件（Controller，C），几个部分。其中重点在于视觉感知和记忆组件。视觉感知模块采用变分自编码器（VAE）学习一个抽象的、压缩的表示来描述每一帧的图像输入，这一点与人类相似，人类在观察事物的时候也会对其进行处理，将抽象的信息如物体之间的相对关系、位置等总结出来，而不是观察其绝对尺寸形状等；记忆组件将历史信息进行关联和压缩，并对下一个状态的信息进行预测，当然也会受到自身行为带来的外部环境变化的反馈所影响；世界模型主要由“感知”和“记忆”模块构成；控制组件负责确定所要采取的行动。经过试验发现，世界模型能够以远超其他方法的分数完成赛车游戏。

**图30：谷歌的论文中提出了一种感知、记忆、控制模块齐备的“世界模型”**

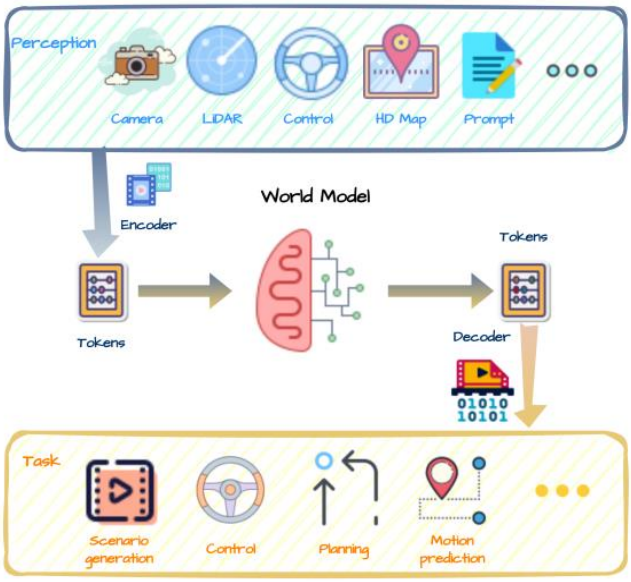



资料来源：《World Models》（David Ha 等）

## 2.3、 世界模型是自动驾驶助推器，助力模型训练、验证测试、甚至推理

**世界模型可以有效赋能智驾。**在自动驾驶领域，能够准确预测驾驶场景未来的演变至关重要，通过对场景即将发生的事件进行预判，汽车可以自如的进行规划和控制做出更明智的决策。怎样构建这样一个可以通过理解世界运行规律进而预测未来的模型，学界和工业界都进行了不懈的探索。与大语言模型类似，玩家采用自回归的模型，将数据压缩和提炼，在潜在空间通过无监督的训练构建模型对未来进行预测，之后通过不同的解码器将预测好的信息解码成为需要的表达方式进而构建世界模型。从结果来看模型可以通过海量的数据一定程度上总结世界运行的物理规律并对未来进行一定程度的预测。在自动驾驶领域，世界模型可以用来生成场景，也可以直接用来做决策规划。具体而言：（1）可以生成诸多逼真的场景，生成稀缺、难以采集的场景，为模型训练提供足量的数据；（2）同样生成的场景亦可以作为仿真测试工具对算法进行闭环验证；（3）多模态的世界模型亦可以直接生成驾驶策略来指导自动驾驶行为。

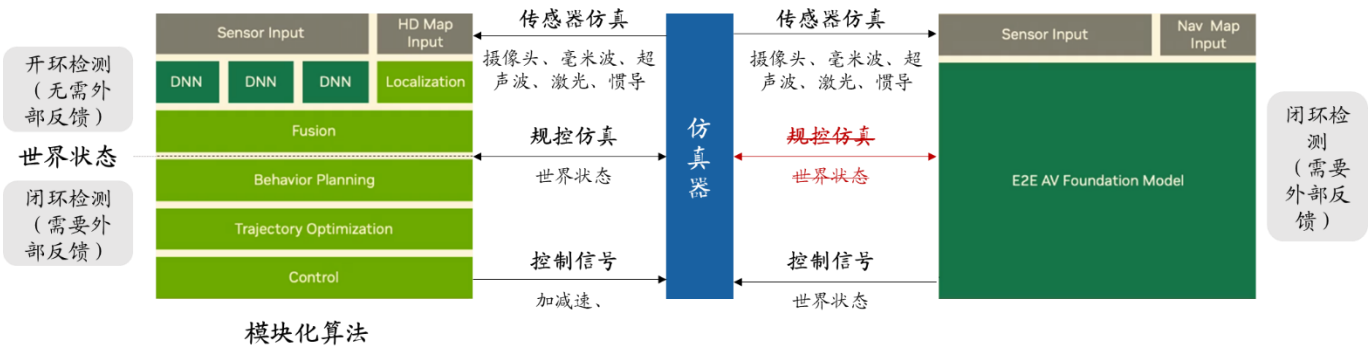
**图31：世界模型可以用来实现多种自动驾驶任务**




资料来源：《World Models for Autonomous Driving: An Initial Survey》（Yanchen Guan 等）

**自动驾驶进入深水区，端到端逐步成为未来方向，世界模型重要性凸显。一方面，**随着自动驾驶走入深水区，玩家对数据的要求日益提升，厂家希望数据能够模拟复杂交通流、具有丰富的场景、广泛收集各类长尾场景、并且具备 3D 标注信息。而现实状态下，数据的采集成本居高不下，部分危险的场景如车祸等难以采集，长尾场景稀缺，同时 3D 标注的成本高昂，因此采用合成数据来助力自动驾驶模型训练测试成为颇具前景的发展方向，而世界模型恰为良好的场景生成和预测器。**另一方面，**随着端到端自动驾驶成为未来的发展方向，开发者需要依靠数据将驾驶知识赋予模型，数据需求会伴随模型体量的增加而扩大。此外更重要的影响在于，在仿真和验证环节，传统的模块化算法时代可以对感知和规控模块分别进行验证，感知端可以进行开环的检测（即将感知的结果和带有标注的真实世界状况直接对比即可，不需要反馈和迭代），规控环节可以依靠仿真工具，将世界的状况（各类场景）输入，通过环境的变化来给与模型反馈，进而闭环的（外部环境可以根据智能体的输出变化而改变，形成反馈）验证规控算法的性能。这其中，感知环节更注重仿真环境的逼真性，而规控环节更注重逻辑的丰富度。在端到端时代，感知和规控合二为一，这要求仿真工具既可以逼真的还原外部环境，同时能够给与模型反馈实现闭环测试，尽管 Nerf、3D 高斯等等算法层出不穷，但能够很好的做到自动驾驶全过程完整的闭环测试亦难度较高，而世界模型则能够很好的应对类似的场景。

**图32：端到端时代，仿真算法提出新的要求，世界模型有望成为解决方案**




资料来源：英伟达 GTC 大会《Building the Next-Generation of Autonomous Vehicles in Simulation》、开源证券研究所

## 2.4、 自动驾驶各方势力发力研究世界模型

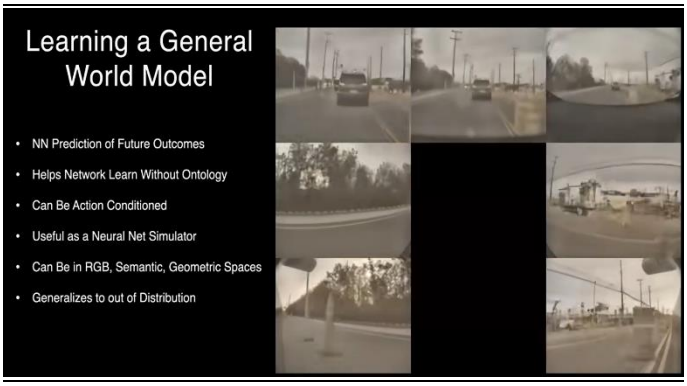
- **特斯拉的 World Model:**

**特斯拉世界模型可预测驾驶场景下的未来发展。**特斯拉在 2023CVPR 对其端到端模型进行了简单的介绍，希望能够构建一个完整的 4D 神经网络，能够理解世界运行的规律。具体而言，世界模型可以根据过去的视频预测未来场景的演化，形成几大功能：（1）预测未来；（2）在没有本体实体的情况下帮助网络学习；（3）行动本身可以作为生成的条件；（4）自车的行为会影响生成的效果，比如左转右转会分别生成不同的视角。（5）可以用于仿真；（6）可以生成图像、几何空间的信息、语义信息等；（7）泛化性比较好。

**经过训练，模型具备了一定程度对物理世界的理解。**特斯拉发现网络可以联合预测汽车周围 8 个摄像头的信息；同时各个摄像头的颜色保持一致，表明可以更好地预测传感器的特性；此外尽管开发者没有显式的要求它以三维或者非三维的方式进行计算，网络即自行理解了三维空间的概念，视频中运动的物体也具有一致性，

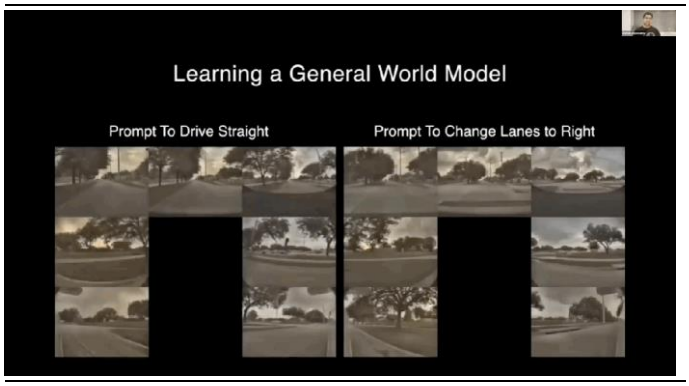
通过自然语言的提示，模型可改变视角；其可根据要求以相同的起点生成不同的结局；对视频语料的适应性好，可以通过行驶记录、油管或者自己手机中的数据来训练这个模型。相比游戏引擎所生成的仿真场景，这样的世界模拟器可以表示一些很难用显式系统描述的事物，如物体的意图和行为等。

**图33：特斯拉的世界模型可实现对未来的预测**




资料来源：智能车参考公众号

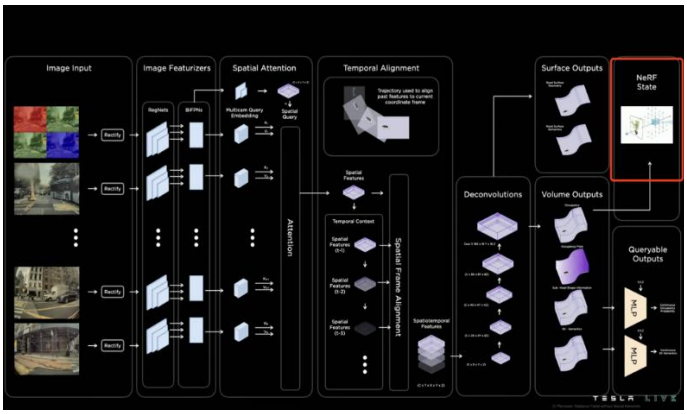
**图34：可以用自然语言对模型生成内容进行控制**




资料来源：智能车参考公众号

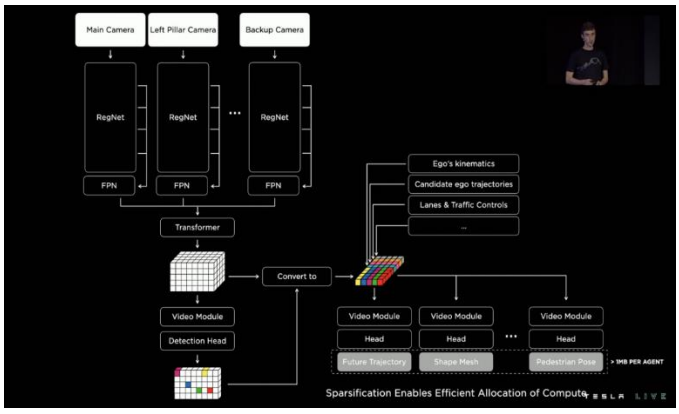
**从感知基础模型窥探特斯拉世界模型端倪。**特斯拉在 2023 年 CVPR 的演讲中介绍到了其感知基础模型的构建方式，算法中先将外部的信息经过特征提取网络进行压缩和特征提取，送入基于 Transformer 的模型，构建对于 4D 的时空环境的理解。之后根据不同的任务需求，加入不同的解码器或者其他算法模块来实现不同任务，如自动驾驶加入表面输出、体积网格输出等，供后续的模块使用，机器人也是同样的道理。并且我们看到特斯拉的模型中不光包含多摄像头视角的视频信息，还包含里程计、自车轨迹等信息，当然特斯拉也提到，在如何将里程信息和视频信息很好的融合，以及如何提升模型的时空理解能力方面，亦做了大量工作。而我们看到这样的基础模型和玩家们构建的世界模型出现了诸多相似之处。

**图35：特斯拉构筑了基础模型架构实现嵌入式 AI 功能**




资料来源：42 号车库公众号

**图36：特斯拉模型不光输入视频信息，还包含里程计等**



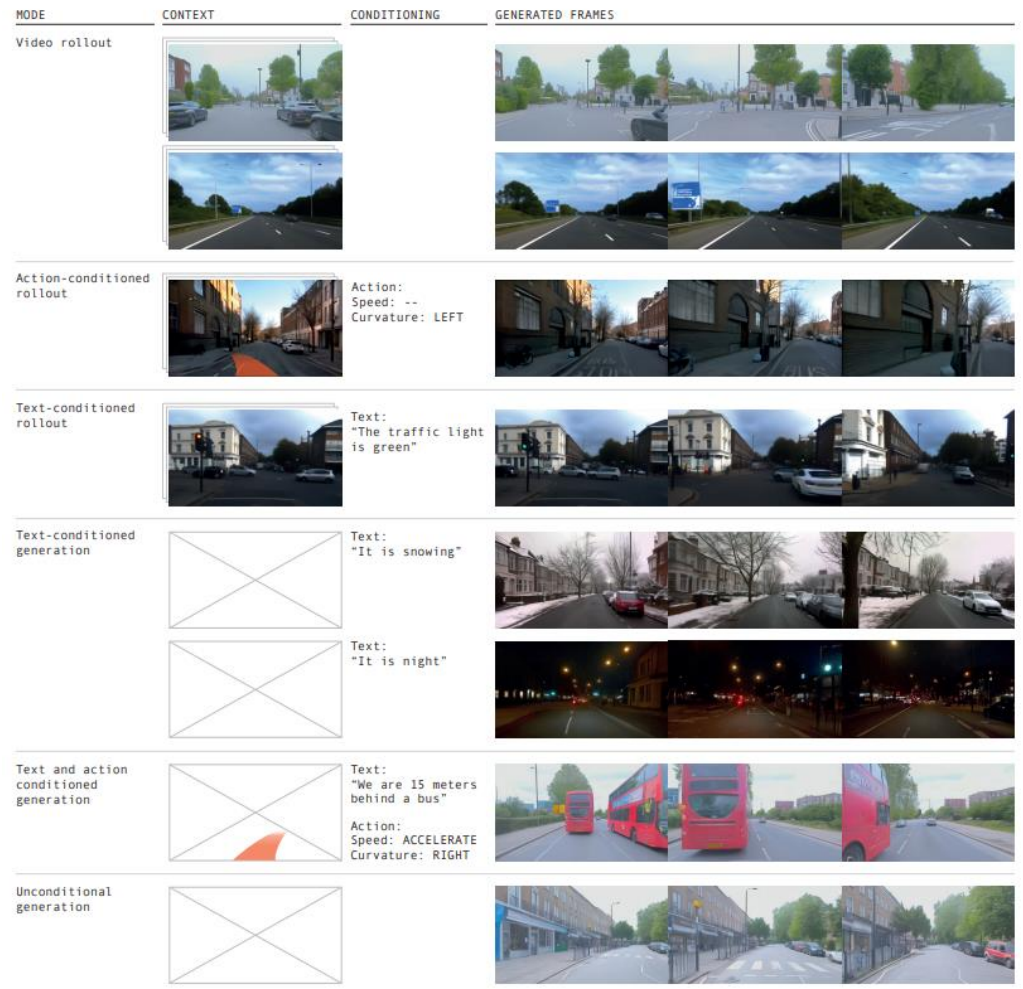

资料来源：42 号车库公众号

- **Wayve：GAIA-1**

**GAIA-1 亦可实现对场景的理解。**英国的端到端自动驾驶公司 Wayve.ai 在 2023 年发布了 GAIA-1 模型，它可以依靠视频、文本和动作的输入生成逼真的视频。模型可以生成分钟级的视频，同时可以生成多种合理的未来，模拟多种场景：如与道路使用者的交互、自车行为的改变等，同时可依靠文字来对视频中元素进行精细粒度的控制，帮助自动驾驶模型的训练和仿真。

**多模态数据训练后的模型亦呈现出对驾驶场景出人意料的认知。**GAIA-1 模型呈现出一些有趣的特点：（1）学习到了高级结构和场景动态：可以生成连贯的场景，其中的对象位于合理的位置并且展示出合理的交互状态，如路灯、道路规则、让路等，表明模型不仅记住统计模式，还理解世界上物体的排列和基本规则。（2）拥有强泛化性和创造性：可以产生训练集中尚未明确出现的对象和场景。（3）拥有情景意识：可以根据上下文的信息生成连贯的动作和响应，并展示出对 3D 几何的理解以及道路使用者决策过程中的因果关系的理解，如可反应道路不平整引起的视角俯仰等作用。

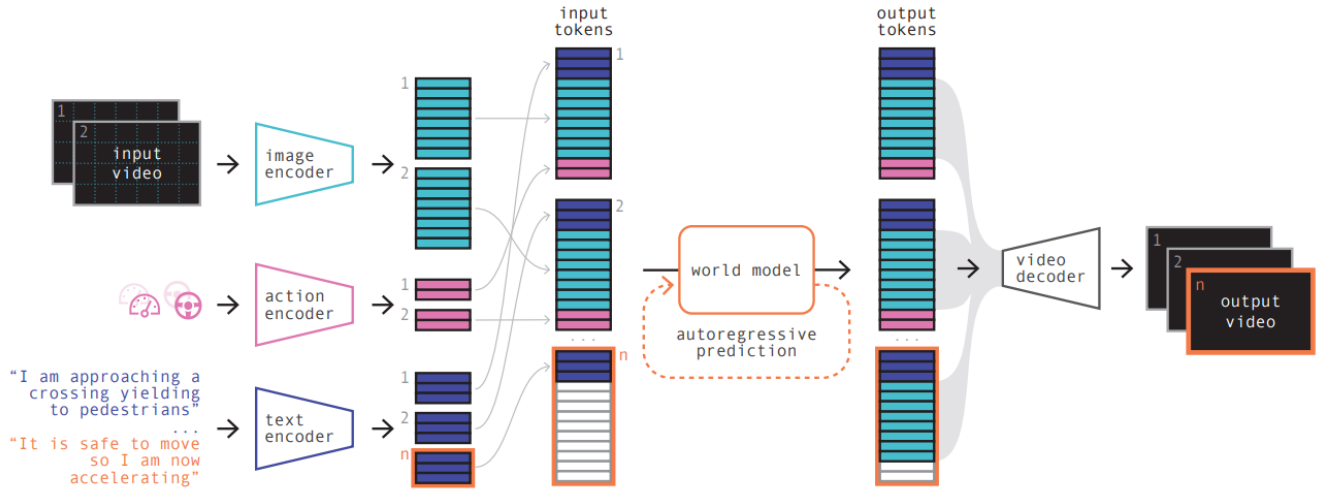
**图37：Wayve 的 GAIA-1 可生成各类交通场景以及相互之间的交互作用**




资料来源：《GAIA-1: A Generative World Model for Autonomous Driving》（Anthony Hu 等）

**自回归+扩散模型形成模型的骨架。**GAIA-1 的构建采用了自回归的世界模型和视频扩散模型结合的方式，原理上和大语言模型以及各类视频生成模型相似，采用向量化的表示，将预测未来转变成预测下一个 Token 的任务。首先采用不同的编码器将图像、文本、动作编码和压缩变成各类 Token，之后将其送入以 Transformer 为基础构成的世界模型中来对未来进行预测，最后将预测好的向量送入以扩散模型为主体的视频解码器将其输出成为视频。

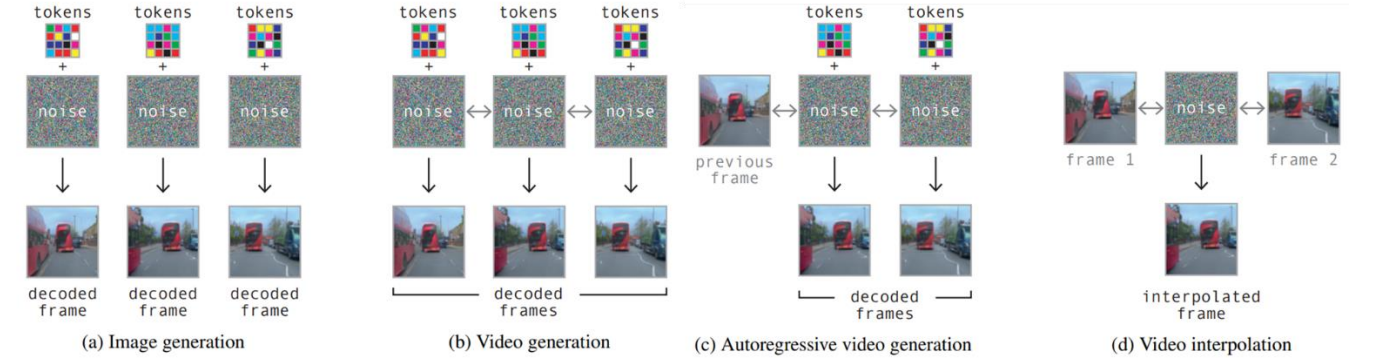
**图38：GAIA-1 采用多模态数据进行训练，以自回归模型和扩散模型作为整体算法的骨架**




资料来源：《GAIA-1: A Generative World Model for Autonomous Driving》（Anthony Hu 等）

**模型架构特殊设计来增强生成视频的帧间一致性。**在模型架构方面，GAIA-1 采用了专门的设计的图像分词器（Image Tokenizer），以保证在对视频信息进行切分的时候，能够有效压缩原始图像信息减少冗余和噪声，同时能够让压缩后的内容包含较多的语义信息而非高频像素信号。世界模型方面，采用因果掩码的方式让模型尝试预测未来。此外在视频解码器方面，GAIA-1 采用视频去噪扩散模型，并让模型在扩散过程中对帧序列去噪（多帧同步生成），从而提高了视频输出的时间一致性。此外对视频解码器在多个任务上进行训练如图像生成、视频生成、自回归解码和视频插值等，因为同时训练会提升单一任务的性能。

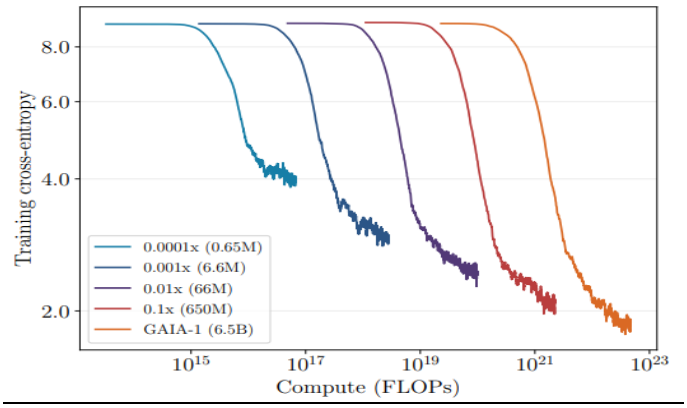
**图39：模型可以实现各类不同功能如图像生成、视频生成、自回归视频生成、视频插帧等**




资料来源：《GAIA-1: A Generative World Model for Autonomous Driving》（Anthony Hu 等）

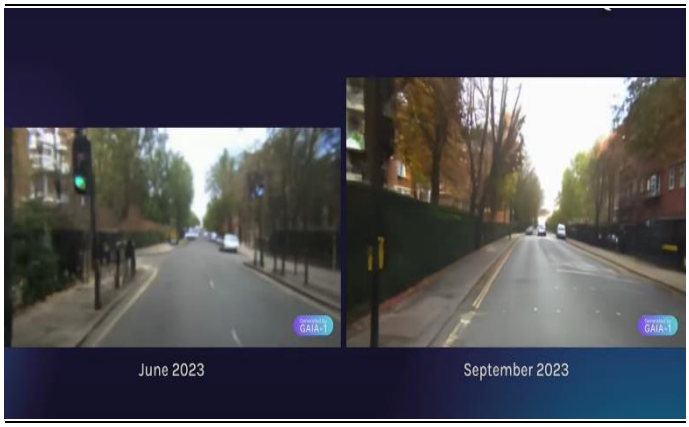
**Wayve 训练的 GAIA-1 呈现出良好的规模效应和潜力。**GAIA-1 在数据方面采用了 2019-2023 年在英国伦敦收集的 4700 小时，25Hz 的专有驾驶数据，对应大约 4.2 亿张图像。模型总共约 90 亿参数，其中视觉编码器部分（Image tokenizer）拥有 3 亿参数，在 32 张英伟达 A100GPU 上训练了 4 天；世界模型拥有 65 亿参数，在 64 张英伟达 A100 GPU 上训练了 15 天；视频解码器拥有 26 亿参数，在 32 张英伟达 A100 上训练了 15 天。模型显示出显著的规模效应，当将模型的参数量增加，或者将模型的计算量增加，都可以显著提升模型的性能，表明模型仍然具有较好的潜力。

图40：当模型的算力和参数量提升时效果显著提升




资料来源：《GAIA-1: A Generative World Model for Autonomous Driving》（Anthony Hu 等）

图41：2023.9 采用的更大参数模型效果显著优于 2023.6

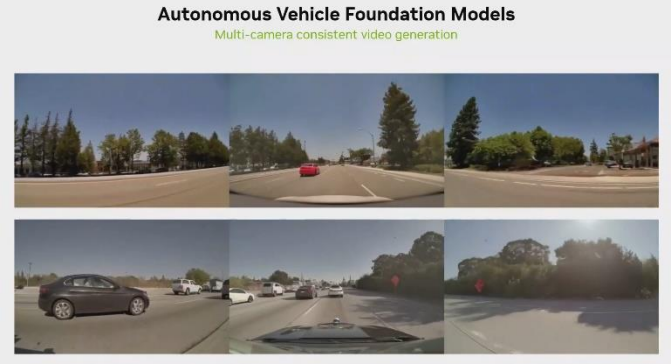



资料来源：wayve 官网

➢ **英伟达：**

**英伟达的基础模型基于多模态数据训练，可生成逼真且灵活变化的驾驶场景视频。**英伟达在近期 2024 年 GTC 大会上也展示了其在世界模型领域的新进展，通过将多模态数据（传感器参数、文字如天气等、自车行为、2D/3D 检测框、道路布局等；Token 化的传感器感知数据）输入模型训练并让模型预测未来驾驶场景，自动驾驶基础模型可以稳定生成多个摄像头拍摄到的驾驶场景演变，效果逼真。此外通过语言提示词也可以使得模型呈现的场景灵活变化，如告诉模型视角为前视摄像头，汽车正行驶在雪天的道路上，两侧道路的树木被雪覆盖，道路上也有雪散落，模型可以生成逼真的相应驾驶场景。

图42：英伟达的自动驾驶基础模型可生成多摄视角




资料来源：2024 春季 GTC 大会《Accelerating the Shift to AI-Defined Vehicles 》

图43：模型亦可根据提示词生成多种驾驶场景

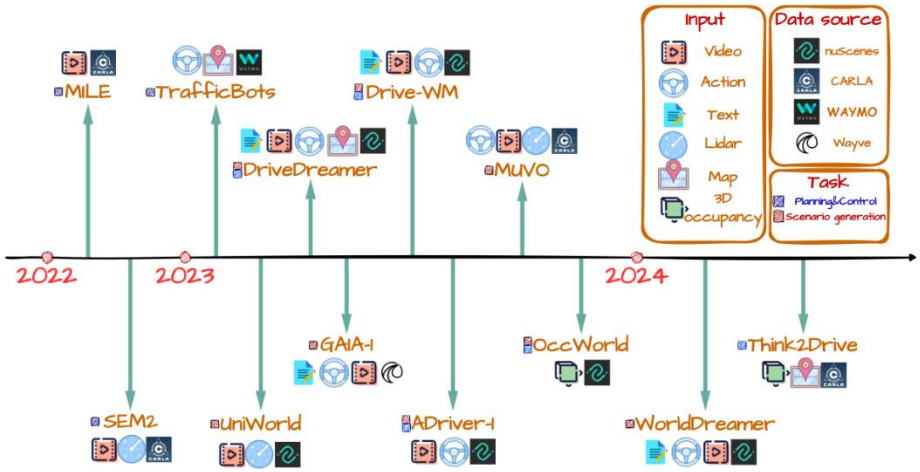

资料来源：2024 春季 GTC 大会《Accelerating the Shift to AI-Defined Vehicles 》

➢ **学术界：**

**学术界的自动驾驶世界模型方案层出不穷。**学术界尤其是 GPT 出现后在 2023 年和 2024 年涌现出一大批关于构建自动驾驶世界模型的方案。AI 创业公司极佳科技和清华大学联合推出的 DriveDreamer 在架构上和 GAIA-1 类似，但在输入端包含更多模态的数据如高精地图等，可实现更加深入的理解，更加精确的控制驾驶场景的生成，同时还增加了 ActionFormer 模块来使得模型可输出控制信号，之后的 DriveDreamer v2 结合了大语言模型和新的架构来增强模型的可塑性和一致性；此外双方还联合推出 WorldDreamer，采用了新的架构结合 Transformer 和 Diffusion 优势进行视频生成。旷世科技、早稻田大学、中科大等机构的论文中提出了 ADriver-1 模

型，使用多模态大语言模型作为主干网络，以自回归的方式输出控制信号，使用视频潜在扩散模型来生成未来的视频输出。德国信息技术研究中心和 KTI 推出的 MUVO 模型则包含三个部分，首先将视频、点云信息进行解码压缩以及特征融合，并和编码过的行为数据一同输入基于 Transformer 构建的世界模型，对未来进行预测，之后通过不同的解码器分别输出 3D 占用网络、激光雷达点云、图像。中科院自动化所提出的 DriveWM 采用多图联合建模提升了所生成驾驶场景的质量。

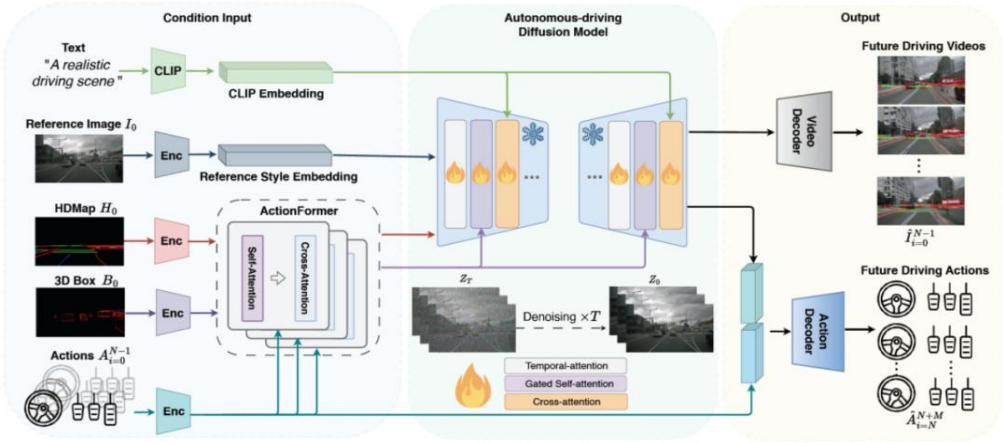
**图44：学术界的自动驾驶世界模型如雨后春笋**




资料来源：《World Models for Autonomous Driving: An Initial Survey》（Yanchen Guan 等）

**DriveDreamer 构建世界模型可实现未来场景的生成以及驾驶员可能相应产生的动作。**以极佳科技和清华大学联合推出的 DriveDreamer 为例，模型主要采用注意力机制和 Diffusion 模型构建。可对驾驶场景实现全面的理解，集成了多模态的输入数据如文本、视频、高精度地图、3D 检测框、驾驶行为等，可以实现可控的驾驶视频生成和预测未来的驾驶行为。同时 DriveDreamer 还可以与驾驶场景互动，根据输入的驾驶动作预测不同的未来驾驶视频。

**图45：DriverDreamer 采用扩散模型和注意力机制构建**




资料来源：《DriveDreamer: Towards Real-world-driven World Models for Autonomous Driving》（Xiaofeng Wang 等）

图46：DriveDreamer 呈现出较强的场景和动作输出能力

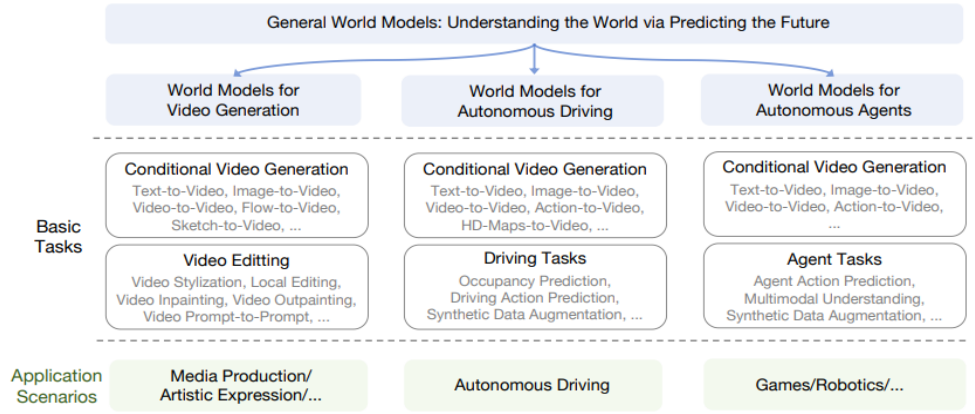

资料来源：《DriveDreamer: Towards Real-world-driven World Models for Autonomous Driving》（Xiaofeng Wang 等）

# 3、 世界模型、视频生成殊途同归，自动驾驶有望加速

## 3.1、 模拟世界，自动驾驶世界模型、视频生成模型以及具身智能拥有相似的目标

**面向相似的目标，多种任务殊途同归。**OpenAI 给自己的 Sora 模型起名叫做 World simulator（世界模拟器），无独有偶，视频生成公司 Runway 在接连发布了 Gen-1 和 Gen-2 视频生成软件后，表示将进军通用世界模型的构建，以更好的理解和预测视觉世界及其动态，文生图新星 Midjourney 亦表示将进军世界模型。在自动驾驶领域包括特斯拉、Wayve、英伟达等公司均通过视频或多模态的数据训练来构建自己的世界模型，学术界也涌现出诸多相似想法。在此之外，具身智能领域也出现诸多以世界模型为基础的算法。无论自动驾驶、具身智能还是视频生成均出现相同的目标：让模型理解世界的基础规律，长时间稳定的对未来进行预测，最终面向不同任务采用不同的形式将这个对未来的“预测”进行表达，如视频生成领域即通过解码器生成各类视频，自动驾驶领域即通过解码器来预测各类自动驾驶的任务，具身智能领域则通过解码器生成自身需要完成的各类动作。深层次而言，让模型理解世界运行的物理规律，进行因果层面的推理，提升泛化能力，提高计算效率并开发出有效的评估体系，是所有世界模型一致的目标。

图47：视频生成、自动驾驶、具身智能面向相似的目标，有望获得世界模型赋能




资料来源：《Is Sora a World Simulator? A Comprehensive Survey on General World Models and Beyond》（Zheng Zhu 等）

## 3.2、 模型架构相似，产业发展有望加速

**产业迈向相似的路径，行业发展有望加速。**我们看到视频生成领域的模型结构和训练方式和自动驾驶领域呈现出诸多相似之处：（1）通常会采用编码器将复杂外部世界获取的数据进行编码、压缩、抽象成为低维度的向量，这个过程如何进行数据的压缩、编码和分割通常会被精心设计。（2）通常会采用 Transformer 或者其他模型来分析和学习这些序列在不同的时间和空间维度的关系，进而实现对下一个时间段的情况进行预测。（3）通过不同类型的解码器将之前生成的潜在空间的向量解码成为我们所需要的信息形式，如视频、点云、甚至执行器的控制信息。而实际上诸多玩家需要解决的问题也高度相似，即如何将复杂的世界压缩、精简、分割，进而让模型能够学习到“世界”的知识；如何长时间、稳定的生成前后一致的视频；如何促进规模化，我们看到前述 Sora 采用的方案也有异曲同工之处，而后续更加开放的 OpenSora 亦有类似的架构，产业发展有望加速。

## 3.3、 集结最优秀人才和资源，Sora 和世界模型有望相互促进，加速发展

**集结优秀人才和资源，产业发展有望加速。**我们看到在 Sora 的启发下，OpenSora、Vidu 等视频生成工具迭出，效果亦不俗，OpenSora 可以生成 16 秒时长的视频，分辨率最高可达 720P，支持任何宽高比、不同分辨率和时长的文本到图像、文本到视频、图像到视频、视频到视频和无限长视频的生成需求，团队采用改进化的 STDiT 架构，参考 SD3 增强模型稳定性，最终形成效果优异的视频生成工具，可见龙头公司的引领作用不可小觑。大模型开发和自动驾驶汇集 AI 领域诸多优秀人才和资源，相似的开发目标和过程有望让产业互相借鉴，加速产业发展，自动驾驶亦将受益于世界模型的发展，迈上新台阶。同时我们也看到，自动驾驶和大模型的能力需求日益趋同，未来拥有大模型开发能力的玩家，在自动驾驶领域有望持续获得先机。

图48：OpenSora 可生成优质视频素材

资料来源：机器之心公众号

图49：Vidu 亦出现，效果惊艳


资料来源：观察者网

## 4、 推荐及受益标的

当前技术的进步无疑将一步步帮助自动驾驶实现真正的落地，有关的零部件、整车玩家有望持续受益。推荐标的：长安汽车、比亚迪、长城汽车、德赛西威、经纬恒润-W、均胜电子、华阳集团、美格智能、华测导航等。受益标的：小鹏汽车-W、理想汽车-W、蔚来-SW、中科创达等。

表2：推荐及受益公司盈利预测与估值

| 股票代码 | 公司简称 | 最新收盘价（元） | 总市值（亿元） | EPS | | | P/E | | | 评级 |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | | 2024E | 2025E | 2026E | 2024E | 2025E | 2026E | |
| 000625.SZ | 长安汽车 | 14.37 | 1249.44 | 0.81 | 1.03 | 1.23 | 15.6 | 12.2 | 10.2 | 买入 |
| 002594.SZ | 比亚迪 | 222.87 | 6305.75 | 19.80 | 16.50 | 14.00 | 19.4 | 16.1 | 13.7 | 买入 |
| 601633.SH | 长城汽车 | 27.40 | 2015.12 | 1.21 | 1.55 | 1.79 | 19.5 | 15.2 | 13.2 | 买入 |
| 9868.HK | 小鹏汽车-W | 34.25 | 647.18 | -4.90 | -1.90 | 0.30 | | | 116.9 | 增持 |
| 2015.HK | 理想汽车-W | 99.90 | 2119.90 | 8.40 | 11.90 | 14.60 | 12.7 | 8.9 | 7.2 | 增持 |
| 9866.HK | 蔚来-SW | 42.20 | 875.03 | -10.20 | -6.80 | -6.60 | | | | 增持 |
| 002920.SZ | 德赛西威 | 108.04 | 599.63 | 3.98 | 5.42 | 6.73 | 27.2 | 19.9 | 16.1 | 买入 |
| 688326.SH | 经纬恒润-W | 60.43 | 72.51 | 0.08 | 1.54 | 3.4 | 805.7 | 39.2 | 17.8 | 买入 |
| 600699.SH | 均胜电子 | 16.71 | 235.39 | 1.03 | 1.38 | 1.68 | 16.2 | 12.1 | 9.9 | 买入 |
| 002906.SZ | 华阳集团 | 28.81 | 151.09 | 1.13 | 1.54 | 1.92 | 25.6 | 18.7 | 15.0 | 买入 |
| 002881.SZ | 美格智能 | 21.29 | 55.71 | 0.57 | 0.72 | 0.91 | 37.4 | 29.5 | 23.5 | 买入 |
| 300627.SZ | 华测导航 | 30.74 | 167.53 | 1.03 | 1.29 | 1.56 | 29.8 | 23.8 | 19.7 | 买入 |
| 300496.SZ | 中科创达 | 46.83 | 215.42 | 1.32 | 1.72 | 2.24 | 35.5 | 27.2 | 20.9 | 买入 |

数据来源：Wind、开源证券研究所（注：收盘日期为 2024 年 5 月 20 日，盈利预测均来自开源证券研究所，港股上市公司收盘价单位为港币，2024 年 5 月 20 日汇率 港币：人民币=0.9273）

## 5、 风险提示

技术进步不及预期：自动驾驶与人工智能深度融合，技术难度高，对人才、资源等要求高，可能会有技术进步不及预期的风险。

市场需求不及预期：如果自动驾驶性能提升迟缓，则无法有效拉动客户需求，同时不同价格带的用户对于自动驾驶的需求亦不同，市场需求有不及预期可能。

重大事故导致行业发展受挫：自动驾驶目前仍然不成熟，如遇到消费者预期与自动驾驶能力不符，或将带来意外事故，影响整个行业发展进度。

## 特别声明

《证券期货投资者适当性管理办法》、《证券经营机构投资者适当性管理实施指引（试行）》已于2017年7月1日起正式实施。根据上述规定，开源证券评定此研报的风险等级为R4（中高风险），因此通过公共平台推送的研报其适用的投资者类别仅限定为专业投资者及风险承受能力为C4、C5的普通投资者。若您并非专业投资者及风险承受能力为C4、C5的普通投资者，请取消阅读，请勿收藏、接收或使用本研报中的任何信息。
因此受限于访问权限的设置，若给您造成不便，烦请见谅！感谢您给予的理解与配合。

## 分析师承诺

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。负责准备本报告的分析师获取报酬的评判因素包括研究的质量和准确性、客户的反馈、竞争性因素以及开源证券股份有限公司的整体收益。所有研究分析师或工作人员保证他们报酬的任何一部分不曾与，不与，也将不会与本报告中具体的推荐意见或观点有直接或间接的联系。

## 股票投资评级说明

| | 评级 | 说明 |
|---|---|---|
| 证券评级 | 买入（Buy） | 预计相对强于市场表现 20%以上； |
| | 增持（outperform） | 预计相对强于市场表现 5%～20%； |
| | 中性（Neutral） | 预计相对市场表现在－5%～＋5%之间波动； |
| | 减持（underperform） | 预计相对弱于市场表现 5%以下。 |
| 行业评级 | 看好（overweight） | 预计行业超越整体市场表现； |
| | 中性（Neutral） | 预计行业与整体市场表现基本持平； |
| | 看淡（underperform） | 预计行业弱于整体市场表现。 |
| 备注：评级标准为以报告日后的 6~12 个月内，证券相对于市场基准指数的涨跌幅表现，其中 A 股基准指数为沪深 300 指数、港股基准指数为恒生指数、新三板基准指数为三板成指（针对协议转让标的）或三板做市指数（针对做市转让标的）、美股基准指数为标普 500 或纳斯达克综合指数。我们在此提醒您，不同证券研究机构采用不同的评级术语及评级标准。我们采用的是相对评级体系，表示投资的相对比重建议；投资者买入或者卖出证券的决定取决于个人的实际情况，比如当前的持仓结构以及其他需要考虑的因素。投资者应阅读整篇报告，以获取比较完整的观点与信息，不应仅仅依靠投资评级来推断结论。 | | |

## 分析、估值方法的局限性说明

本报告所包含的分析基于各种假设，不同假设可能导致分析结果出现重大不同。本报告采用的各种估值方法及模型均有其局限性，估值结果不保证所涉及证券能够在该价格交易。

## 法律声明

开源证券股份有限公司是经中国证监会批准设立的证券经营机构，已具备证券投资咨询业务资格。

本报告仅供开源证券股份有限公司（以下简称“本公司”）的机构或个人客户（以下简称“客户”）使用。本公司不会因接收人收到本报告而视其为客户。本报告是发送给开源证券客户的，属于商业秘密材料，只有开源证券客户才能参考或使用，如接收人并非开源证券客户，请及时退回并删除。

本报告是基于本公司认为可靠的已公开信息，但本公司不保证该等信息的准确性或完整性。本报告所载的资料、工具、意见及推测只提供给客户作参考之用，并非作为或被视为出售或购买证券或其他金融工具的邀请或向人做出邀请。本报告所载的资料、意见及推测仅反映本公司于发布本报告当日的判断，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波动。在不同时期，本公司可发出与本报告所载资料、意见及推测不一致的报告。客户应当考虑到本公司可能存在可能影响本报告客观性的利益冲突，不应视本报告为做出投资决策的唯一因素。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。本公司未确保本报告充分考虑到个别客户特殊的投资目标、财务状况或需要。本公司建议客户应考虑本报告的任何意见或建议是否符合其特定状况，以及（若有必要）咨询独立投资顾问。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。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。若本报告的接收人非本公司的客户，应在基于本报告做出任何投资决定或就本报告要求任何解释前咨询独立投资顾问。

本报告可能附带其它网站的地址或超级链接，对于可能涉及的开源证券网站以外的地址或超级链接，开源证券不对其内容负责。本报告提供这些地址或超级链接的目的纯粹是为了客户使用方便，链接网站的内容不构成本报告的任何部分，客户需自行承担浏览这些网站的费用或风险。

开源证券在法律允许的情况下可参与、投资或持有本报告涉及的证券或进行证券交易，或向本报告涉及的公司提供或争取提供包括投资银行业务在内的服务或业务支持。开源证券可能与本报告涉及的公司之间存在业务关系，并无需事先或在获得业务关系后通知客户。



## 开源证券研究所

**上海**
地址：上海市浦东新区世纪大道1788号陆家嘴金控广场1号楼10层
邮编：200120
邮箱：research@kysec.cn

**深圳**
地址：深圳市福田区金田路2030号卓越世纪中心1号楼45层
邮编：518000
邮箱：research@kysec.cn

**北京**
地址：北京市西城区西直门外大街18号金贸大厦C2座9层
邮编：100044
邮箱：research@kysec.cn

**西安**
地址：西安市高新区锦业路1号都市之门B座5层
邮编：710065
邮箱：research@kysec.cn