甲子光年
JAZZYEAR

2024

AIGC视频生成：走向AI创生时代

视频生成的技术演进、范式重塑与商业化路径探索

出品机构：甲子光年智库
研究团队：张一甲、宋涛
发布时间：2024.03

*刘瑶、小麦对本报告亦有贡献。

甲子光年
JAZZYEAR

“一类人有一类人原力觉醒的方式。

物理学家想学习上帝；

数学家想反抗上帝；

哲学家认为自己就是上帝；

生物学家想造上帝的反……

工程师说都不用，我们再造一个。”

——《甲小姐：站在两个世界之间》甲子光年 2017.10

目录

CONTENTS

甲子光年
JAZZYEAR

1.1 Sora让文生视频迎来“GPT-3”时刻

OpenAI发布文生视频模型Sora，堪称视频生成领域的“GPT-3”时刻

春节假期甚至还未结束，Sora已引发全民关注

“Sora”一词在微信指数及百度指数的关注度快速上升

“炸裂”视频效果成为讨论热点

- 效果逼真：普通人一时难以分辨
- 时长感人：60秒高清视频生成
- “百万” 剪辑：堪比专业的镜头语言
- 多模态：文字、图片、视频皆可生成视频

数据来源：百度指数、微信指数，甲子光年智库整理，2024年；

1.2 Sora的展现效果

Sora模型展现自身超强视频生成及剪辑能力，超出其他竞品一个段位

能力项		Sora	其他模型
基本视频生成	视频时长	60秒	20秒以内
	视频长宽比	1920*1080之间的任意尺寸	固定尺寸比例，例如16:9，9:16，1:1等
	视频清晰度	1080p	部分upscale后达到4k
多模态生成	语言理解能力	强	弱
	文本生成视频	支持	支持
	图片生成视频	强	支持
	视频生成视频	支持	支持
视频编辑	文本编辑视频	支持	支持
	扩展视频	向前/向后扩展	仅支持向后
	视频的无缝连接	支持	不支持
独特模拟能力	3D一致性	强	弱或不支持
	远程相干性和物体持久性	强	弱
	世界交互	强	弱
	数字世界模拟	支持	不支持

其他模型情况

模型	Gen-2	pika1.0	Stable Video Diffusion	Emu Video	W.A.L.T
开发团队	Runway	Pika Labs	Stablity AI	Meta	李飞飞及其学生团队、谷歌
时间	2023年11月	2023年11月	2023年11月	2023年11月	2023年12月
长度	4-18秒	3-7秒	2-4秒	4秒	3秒
分辨率	768*448，1536*896，4096*2160	1280*720 2560*1440	576*1024	512*512	512*896
是否开源	非开源	非开源	开源	非开源	非开源

Sora的语言理解能力更强，可将简短的用户提示转换为更长的详细描述

Sora还可以生成图片，最高可达到2048*2048分辨率

Sora通过插帧技术，实现完全不同主题和场景构图的视频之间的流畅自然的过渡效果

Sora可生成具有动态摄像机运动效果的视频，随着摄像机的移动和旋转，人和场景元素在三维空间中保持一致移动

Sora可以对短期和长期依赖关系进行建模，保持各个主体的时空连贯性和一致性

Sora以简单的方式模拟影响世界状态的行为，比如一个人吃完汉堡可以在上面留下咬痕

Sora还能够模拟人工过程，比如视频游戏，同时通过基本策略控制玩家，同时以高保真度渲染世界及其动态

数据来源：甲子光年智库梳理，2024年；

1.2 Sora的展现效果

大模型训练的“暴力美学”在视频生成领域再次涌现卓越特性

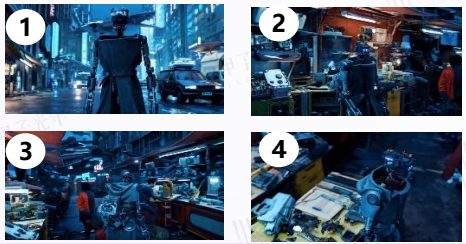
- OpenAI发现视频模型在大规模训练时表现出许多有趣的“涌现”能力，使Sora能够从物理世界中模拟人、动物和环境。值得一提的是OpenAI官网所说的“they are purely phenomena of scale”——它们纯粹是“规模现象”，这再一次验证了“暴力美学”。

文/图像/视频生视频的功能

3D一致性：确保景别切换时运镜的连贯

以上四个镜头由远及近，保证了视频镜头中人和场景的一致性，是其他AI生成视频中少见的。

远程相关性和物体持久性



以上四个镜头在同一视频中生成，包括机器人的多个角度。

与世界互动：Sora有时可以用简单的方式模拟影响世界状况的动作

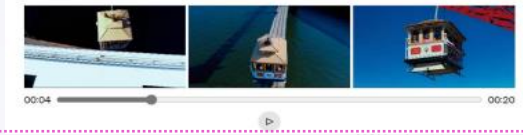
画家可以在画布上留下新的笔触，并随着时间的推移而持续存在。

模拟数字世界

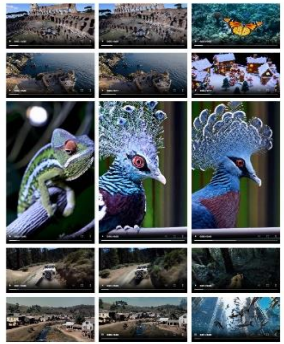
例如，Sora可以同时通过基本策略控制《我的世界》中的玩家，同时以高保真度渲染世界及其动态。

视频剪辑功能

基于时空双维度的视频扩展



不同主题场景视频的无缝连接



一键进行风格渲染

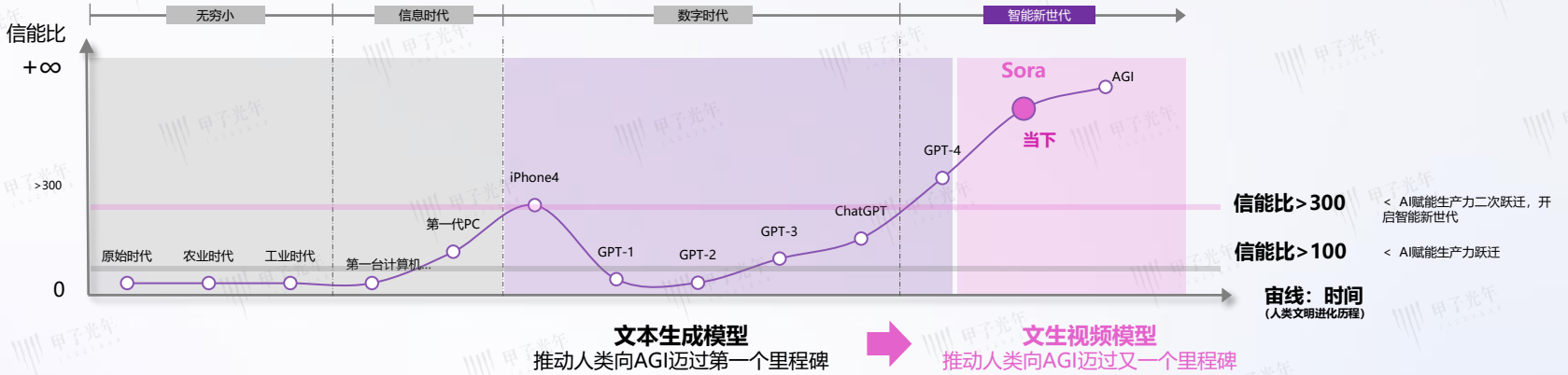
数据来源：公开资料，Open AI技术文件，甲子光年整理，2024年；

1.3 Sora的出现意味着AGI的又一个里程碑时刻

Sora意味着scaling law（规模法则）再次验证，推动文生视频进入“GPT-3”时刻

- **Scaling law（规模法则）的再次验证：**虽然Sora并不十全十美，但它通过scaling law和原有模型拉开了差距，为视频生成领域提供了另一条可以走通的路线，推动行业进入全新的阶段。
- **文生视频的“GPT-3”时刻：**从发展阶段类比，Sora更像文本模型的GPT-3时刻。ChatGPT让人类看到实现AGI的雏形，Sora让实现AGI的目标又进一步。

智能新世代：Sora向AGI再进一步



备注说明：
信能比，是甲子光年智库发明的概念，反映单位能源所能驾驭的信息量。信能比通过单位时间内产生/传输/使用/存储的信息量除以单位时间内所消耗的能源量计算得出，反映单位能源所能调用的信息量水平的高低。
信能比可以体现数据智能技术的先进性和能源效率的高效性：它能够反映整个社会数字化、智能化水平的高低；它能体现能源体系的可持续发展能力；它能反映生产力的高低和生产效率的提升；它能体现社会经济发展的先进性、创新性、可持续性。

1.4 Sora开启“明牌游戏”，推动AIGC应用时间轴进一步被压缩

历史反复表明，一旦先行者模式验证，后来者整体的应用进程时间表将加快

- 先行者往往要花费大量时间精力试错，一旦模式跑通，“明牌游戏”就开启了。后来者会有更好的参考系和聚焦方向。ChatGPT后续的文本生成模型进展就说明了这一点。
- 过去一年，AI文本生成和图像生成相继走向成熟，Sora的发布意味着视频生成应用走向成熟的时间比原先预计的更早出现，AIGC已经加速迈入视频生成阶段。
- 对此，甲子光年智库更新了生成式AI技术的成熟应用进程时间表。2024年可实现根据文本提示生成初版短视频，2025年有望实现根据文本生成初版长视频，并在视频制作环节真实使用落地。

图1：AIGC用户偏好使用的大模型产品类型

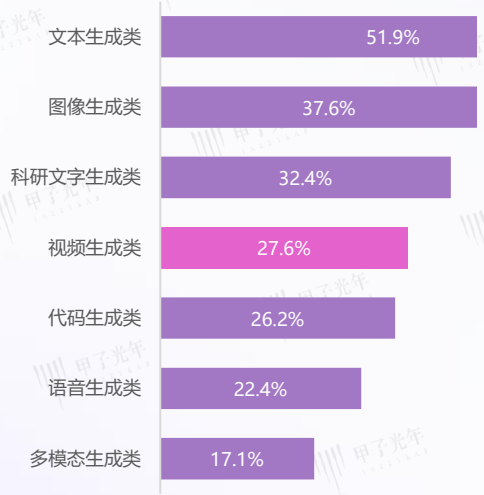


图2：生成式AI技术的成熟应用进程时间表

大模型成熟难度：初级尝试　接近成熟　成熟应用

领域类型	2020年之前	2020年	2022年	2023年	2024年E	2025年E	2030年E
文本领域	诈骗垃圾信息识别 翻译 基础问答回应	基础文案撰写 初稿	更长的文本 二稿	垂直领域的文案 撰写实现可精调（论文等）	终稿，水平接近人类平均值	终稿，水平高于人类平均值	终稿，水平高于专业写手
代码领域	单行代码补足	多行代码生成	更长的代码 更精确的表达	支持更多语种 领域更垂直	根据文本生成初版应用程序	根据文本生成初版应用程序	根据文本生成终版应用程序，比全职开发者水平更高
图像领域			艺术 图标 摄影	模仿（产品设计、建筑等）	终稿（海报设计、产品设计等）	终稿（产品设计、建筑等）	终稿，水平高于专职艺术家、设计师等
视频/3D/游戏领域				视频和3D文件的基础版/初稿	根据文本生成初版的短视频	根据文本生成初版的长视频，并实际应用于制作环节	AI版Roblox可依个人梦想定制的游戏与电影

数据来源：红杉资本，甲子光年智库，2024年；

1.5 Sora验证视频生成的新技术范式

Sora的出现意味着视频生成的DiT技术路线得到有力验证

- 视频生成技术路线在过去主要有两条，一条是基于Transformer的路线，以Phenaki为代表，第二条是Diffusion Model（扩散模型）路线，该路线在2023年是主流路线，诞生了Meta的Make-A-Video、英伟达的Video LDM，Runway的Gen1、Gen2，字节的MagicVideo等代表性产品。
- **Sora的发布，对Transformer + Diffusion Model（DiT）路线进行了成果瞩目的验证。**

图1：AIGC视频生成的技术演进路径

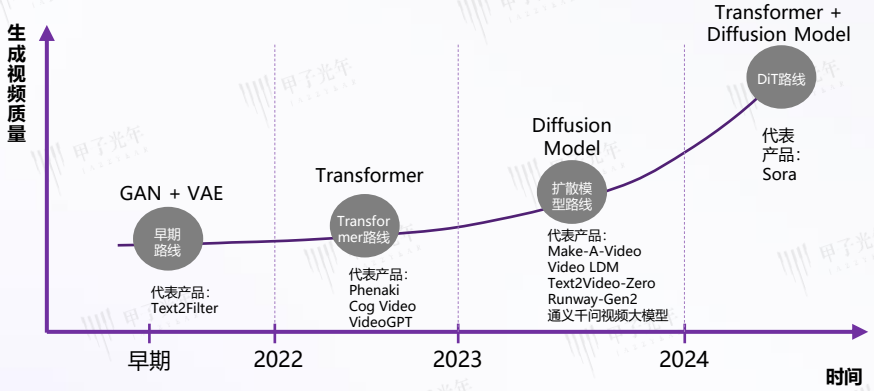


图2：Sora技术优势与竞品的对比情况

数据来源：国盛证券、硅基研究室，甲子光年智库梳理，2024年；

1.6 Sora的技术原理

Patch（时空编码思路）+DiT（Diffusion和Transformer模型的结合）+Scaling Law（规模效应）

- Sora模型将视频压缩到低维空间（latent space），并使用时空补丁（Spacetime latent patches）来表示视频。这个过程类似于将文本转换为Token表示，而视频则转换为patches表示。Sora模型主要在压缩的低维空间进行训练，并使用解码器将低维空间映射回像素空间，以生成视频。
- Sora使用了diffusion模型，给定输入的噪声块+文本prompt，它被训练来预测原始的“干净”分块。
- Sora是diffusion transformer，而transformer在各个领域都表现出显著的规模效应。

图：业内推测出的Sora技术架构图