

政企业务前沿技术和业态

第三期



前言

本期政企业务前沿技术和业态报告以 AIGC 为主题。目前 AIGC 生成正在完成从简单的降本增效（以生成 金融/体育新闻为代表）向创造额外价值（以提供绘画创作素材为代表）转移，跨模态/多模态内容成为关键的发展节点。

本次报告的第一部分是关于 AI 技术与落地场景。主要介绍了 AI 技术的定义、AIGC 技术及其在不同场景下的应用。首先，定义了 AIGC (AI-Generated Content) 技术，它基于生成对抗网络 GAN、大型预训练模型等人工智能技术，通过已有数据寻找规律并生成相关内容。接着，讨论了文本生成、音频及文字—音频生成、图像生成等场景，以及 GameAI 和虚拟人生成等综合赛道。

第二部分是 GPT-4 研究，探讨了 GPT-4 的技术与能力，包括其发展历程和能力分析，以及 GPT-4 在各行业中的应用和未来展望。

第三部分分析了 AIGC 在消费端和产业端的发展趋势，如数字内容生产、3D 互联网基础支撑、聊天机器人和数字人应用领域的拓展，以及合成数据在 AI 模型训练中的作用。

2024 年有望成为 AIGC 发展大年。近期微软、谷歌等全球计软件互联网大厂大力投入 AIGC，并开始探索应用层面的突破，AIGC 已经来到 2 年前自动驾驶的应用突破拐点，类似 ChatGPT 的“大模型”作为 AIGC 的底层平台，将成为 AIGC 的核心资源。

江苏有线数据公司

产品技术部

2024 年 7 月

目录

AIGC 技术与落地场景

AI 技术及落地场景介绍——基于模态.....	5
一、AI 技术定义	6
二、场景之文本生成	7
三、场景之音频及文字—音频生成	10
四、场景之图像生成	13
AI 技术及落地场景介绍——综合赛道.....	20
一、GameAI.....	20
二、虚拟人生成	23
三、相关预测	26

GPT-4 研究

GPT-4 的技术与能力.....	30
一、GPT 的发展历程.....	30
二、GPT-4 的能力分析	33
GPT-4 的应用与展望.....	46
一、当前各行业的应用：	46
二、下一个 GPT 是什么？	50

AIGC 发展趋势

消费端:AIGC 牵引数字内容领域的全新变革	53
一、AIGC 有望塑造数字内容生产与交互新范式,	53
二、AIGC 的应用生态和内容消费市场逐渐繁荣	57
三、AIGC 将日益成为未来 3D 互联网的基础支撑	62
四、聊天机器人和数字人不断拓展应用领域	68
五、AIGC 将作为生产力工具来推动元宇宙发展	71
产业端:合成数据牵引人工智能的未来	75
一、合成数据为 AI 模型训练开发提供强大助推器, 推动实现 AI2.0	76
二、合成数据助力破解 AI “深水区” 的数据难题, 持续拓展产业互联网应用空间	78
三、合成数据产业加快成为数据要素市场新赛道, 科技大厂和创新企业抢先布局	81
三、合成数据加速构建 AI 赋能、数实融合的大型虚拟世界	85

AIGC 技术与落地场景

(一) 技术及落地场景介绍——基于模态

(二) 技术及落地场景介绍——综合赛道

AI 技术及落地场景介绍——基于模态

一、AI 技术定义

AIGC 全称为 AI-Generated Content, 指基于生成对抗网络 GAN、大型预训练模型等人工智能技术, 通过已有数据寻找规律, 并通过适当的泛化能力生成相关内容的技术。与之相类似的概念还包括 Synthetic media, 合成式媒体, 主要指基于 AI 生成的文字、图像、音频等。

Gartner 也提出了相似概念 Generative AI, 也即生成式 AI。生成式 AI 是指该技术从现有数据中生成相似的原始数据。相较于量子位智库认为的 AIGC, 这一概念的范围较狭窄。

一方面, 这一概念忽略了跨模态生成 (如基于文本生成图像或基于文本生成视频) 这一愈加重要的 AIGC 部分。我们会在下一部分对跨模态生成进行重点讲解。另一方面, 在结合现有技术能力和落地场景进行分析后, “生成” 和 “内容” 都应该采取更为广泛的概念。例如, 生成中可以包含基于线索的部分生成、完全自主生成和基于底稿的优化生成。内容方面, 不仅包括常见的图像、文本、音频等外显性内容, 同样也包括策略、剧情、训练数据等内在逻辑内容。

从特定角度来看, AI 内容生成意味着 AI 开始在现实内容中承担新的角色, 从 “观察、预测” 拓展到 “直接生成、决策”。目前 AIGC 生成正在完成从简单的降本增效 (以生成金融/体育新闻为代表) 向

创造额外价值（以提供绘画创作素材为代表）转移，跨模态/多模态内容成为关键的发展节点。

技术视角下，以下场景将成为未来发展的重点：文本-图像-视频的跨模态生成、2D 到 3D 生成、多模态理解结合生成。后文将对以上技术场景的技术原理、现有进展、关键瓶颈等进行展开。

商业视角下，未来 3 年内，虚拟人生成和游戏 AI 这两种综合性的 AIGC 场景将趋于商业化成熟。

二、场景之文本生成

以结构性新闻撰写、内容续写、诗词创作等细分功能为代表，基于 NLP 技术的文本生成可以算作是 AIGC 中发展最早的一部分技术，也已经在新闻报道、对话机器人等应用场景中大范围商业落地。

• 细分技术介绍

大型预训练模型成为当下主流做法与关键技术节点。

一方面，2020 年，1750 亿参数的 GPT-3 在问答、摘要、翻译、续写等语言类任务上均展现出了优秀的通用能力，证明了“大力出奇迹”在语言类模型上的可行性。自此之后，海量数据、更多参数、多元的数据采集渠道等成为国内清华大学、智源研究院、达摩院、华为、北京大学、百度等参与者的关注点。

目前，大型文本预训练模型作为底层工具，商业变现能力逐渐清晰。以 GPT-3 为例，其文本生成能力已被直接应用于 Writesonic、Conversion.ai、SnazzyAI、Copysmith、Copy.ai、Headlime 等文本写

作/编辑工具中。同时也被作为部分文本内容的提供方，服务于 AI Dungeon 等文本具有重要意义的延展应用领域。

另一方面，以 Transformer 架构为重要代表，相关的底层架构仍在不断精进。研究者们正通过增加 K-adapter、优化 Transformer 架构、合理引入知识图谱及知识库、增加特定任务对应 Embedding 等方式，增加文本对于上下文的理解与承接能力、对常识性知识的嵌入能力、中长篇幅生成能力、生成内容的内在逻辑性等。

从现有的落地场景来看，将其划分为应用型文本和创作型文本生成，前者的进展明显优于后者。此外，从应用推广的角度来说，辅助文本创作是目前落地最为广泛的场景。

（一）应用型文本生成

应用型文本大多为结构化写作，以客服类的聊天问答、新闻撰写等为核心场景。2015 年发展至今，商业化应用已较为广泛，最为典型的是基于结构化数据或规范格式，在特定情景类型下的文本生成，如体育新闻、金融新闻、公司财报、重大灾害等简讯写作。据分析师评价，由 AI 完成的新闻初稿已经接近人类记者在 30 分钟内完成的报道水准。NarrativeScience 创始人甚至曾预测，到 2030 年，90% 以上的新闻将由机器人完成。

在结构化写作场景下，代表性垂直公司包括 Automated Insights（美联社 Wordsmith）、Narrative Science、textengine.io、AX Semantics、

Yseop、Arria、Retresco、Viable、澜舟科技等。同时也是小冰公司、腾讯、百度等综合性覆盖 AIGC 领域公司的重点布局领域。

（二）创作型文本生成

创作型文本主要适用于剧情续写、营销文本等细分场景等，具有更高的文本开放度和自由度，需要一定的创意和个性化，对生成能力的技术要求更高。市面上的小说续写、文章生成等 AIGC 工具长篇幅文字的内部逻辑仍然存在较明显的问题、且生成稳定性不足，尚不适合直接进行实际使用。据聆心智能创始人黄民烈教授介绍，目前文字生成主要捕捉的是浅层次，词汇上统计贡献的问题。但长文本生成还需要满足语义层次准确、在篇章上连贯通顺的要求，长文本写作对于议论文写作、公文写作等等具有重要意义。未来四到五年，可能会出现比较好的千字内容。

除去本身的技术能力之外，由于人类对文字内容的消费并不是单纯理性和基于事实的，创作型文本还需要特别关注情感和语言表达艺术。短期内创作型文本更适合在特定的赛道下，基于集中的训练数据及具体的专家规则进行场景落地。

在创作型文本领域，代表性的国内外公司包括 Anyword、Phrasee、Persado、Pencil、Copy.ai、Friday.ai、Retresco、Writesonic、Conversion.ai、SnazzyAI、Rasa.io、LongShot.AI、彩云小梦等。

（三）文本辅助生成

除去端到端进行文本创作外，辅助文本写作其实是目前国内供给及落地最为广泛的场景。主要为基于素材爬取的协助作用，例如定向采集信息素材、文本素材预处理、自动聚类去重，并根据创作者的需求提供相关素材。尽管目前能够提升生产力，但相对于实现技术意义上的 AI 生成，能否结合知识图谱等提供素材联想和语句参考等更具有实用意义。这部分的国内代表产品包括写作猫、Gilso 写作机器人、Get 写作、写作狐、沃沃 AI 人工智能写作。

（四）重点关注场景

长期来看，闲聊型文本交互将会成为重要潜在场景，例如虚拟伴侣、游戏中的 NPC 个性化交互等。2022 年夏季上线的社交 AIGC 叙事平台 HiddenDoor 以及基于 GPT-3 开发的文本探索类游戏 AI Dungeon 均已获得了不错的消费者反馈。随着线上社交逐渐成为了一种常态，社交重点向转移 AI 具有其合理性，我们预估未来 1-2 年内就会出现明显增长。目前较为典型的包括小冰公司推出的小冰岛，集中在精神心理领域的聆心智能、开发了 AI Dungeon 的 Latitude.io 等。

三、场景之音频及文字—音频生成

整体而言，此类技术可应用于流行歌曲、乐曲、有声书的内容创作，以及视频、游戏、影视等领域的配乐创作，大大降低音乐版权的采购成本。我们目前最为看好的场景是自动生成实时配乐、语音克隆以及心理安抚等功能性音乐的自动生成。

（一）TTS（Text-to-speech）场景

TTS 在 AIGC 领域下已相当成熟，广泛应用于客服及硬件机器人、有声读物制作、语音播报等任务。例如倒映有声与音频客户端“云听”APP 合作打造 AI 新闻主播，提供音频内容服务的一站式解决方案，以及喜马拉雅运用 TTS 技术重现单田芳声音版本的《毛氏三兄弟》和历史类作品。这种场景为文字内容的有声化提供了规模化能力。

目前技术上的关键，在于如何通过富文本信息（如文本的深层情感、深层语义了解等）更好的表现其中的抑扬顿挫，以及基于用户较少的个性化数据得到整体的复制能力（如小样本迁移学习）。基于深度学习的端到端语音合成模式也正在逐步替代传统的拼接及参数法，代表模型包括 WaveNet、DeepVoice 及 Tacotron 等。

目前的垂直代表公司包括倒映有声、科大讯飞、思必驰（DUI）、Readspeaker、DeepZen 和 Sonantic。随着内容媒体的变迁，短视频内容配音已成为重要场景。部分软件能够基于文档自动生成解说配音，上线有 150+款包括不同方言和音色的 AI 智能配音主播。代表公司有九锤配音、加音、XAudioPro、剪映等。

在 TTS 领域，语音克隆值得特别关注。语音克隆是本质上属于指定了目标语音（如特定发言人）的 TTS。

（二）乐曲/歌曲生成

AIGC 在词曲创作中的功能可被逐步拆解为作词（NLP 中的文本创作/续写）、作曲、编曲、人声录制和整体混音。目前而言，AIGC

已经支持基于开头旋律、图片、文字描述、音乐类型、情绪类型等生成特定乐曲。

其中, AI 作曲可以简单理解为“以语言模型(目前以 Transformer 为代表, 如谷歌 Megenta、OpenAI Jukebox、AIVA 等)为中介, 对音乐数据进行双向转化(通过 MIDI 等转化路径)”。此方面代表性的模型包括 Melody RNN、Music Transformer。据 Deepmusic 介绍, 为提升整体效率, 在这一过程中, 由于相关数据巨大, 往往需要对段落、调性等高维度的乐理知识进行专业提取, 而节奏、音高、音长等低维度乐理信息由 AI 自动完成提取。通过这一功能, 创作者即可得到 AI 创作的纯音乐或乐曲中的主旋律。2021 年末, 贝多芬管弦乐团在波恩首演人工智能谱写完成的贝多芬未完成之作《第十交响曲》, 即为 AI 基于对贝多芬过往作品的大量学习, 进行自动续写。

AI 编曲则指对 AI 基于主旋律和创作者个人的偏好, 生成不同乐器的对应和弦(如鼓点、贝斯、钢琴等), 完成整体编配。在这部分中, 各乐器模型将通过无监督模型, 在特定乐曲/情绪风格内学习主旋律和特定要素间的映射关系, 从而基于主旋律生成自身所需和弦。对于人工而言, 要达到乐曲编配的职业标准, 需要 7-10 年的学习实践。

人声录制则广泛见于虚拟偶像的表演现场(前面所说的语音克隆), 通过端到端的声学模型和神经声码器完成, 可以简单理解为将输入文本替换为输入 MIDI 数据的声音克隆技术。混音指将主旋律、人声和各乐器和弦的音轨进行渲染及混合, 最终得到完整乐曲。该环

节涉及的 AI 生成能力较少。该场景下的代表企业包括 Deepmusic、网易-有灵智能创作平台、AmperMusic、AIVA、Landr、IBM WatsonMusic、Magenta、Loudly、Brain.FM、Splash、Flow machines。其中，自动编曲功能已在国内主流音乐平台上线，并成为相关大厂的重点关注领域。以 QQ 音乐为例，就已成为 Amper music 的 API 合作伙伴。

对这一部分工作而言，最大的挑战在于音乐数据的标注。在标注阶段，不仅需要按时期、流派、作曲家等特征，对训练集中乐曲的旋律、曲式结构、和声等特征进行描述，还要将其有效编码为程序语言。此外，还需要专业人员基于乐理进行相关调整润色。以 Deepmusic 为例，音乐标注团队一直专注在存量歌曲的音乐信息标注工作上，目前已经形成了全球最精确的华语歌曲音乐信息库，为音乐信息检索(MIR)技术研究提供数据支持。

代表公司：Deepmusic

公司致力于运用 AI 技术从作词、作曲、编曲、演唱、混音等方面全方位降低音乐创作及制作门槛，为音乐行业提供新的产品体验，提升效率。产品包括针对视频生成配乐的配乐猫、支持非音乐专业人员创作的口袋音乐、可 AI 生成歌词的 LYRICA、AI 作曲软件 LAZYCOMPOSER。目前已与国内多家音乐平台厂商达成合作。

四、场景之图像生成

（一）关键技术介绍

生成对抗网络 GAN—图像生成传统思路

GAN 指 Generative Adversarial Nets，生成式对抗网络，由生成器和判别器两部分组成，生成器将抓取数据、产生新的生成数据，并将其混入原始数据中送交判别器区分。这一过程将反复进行，直到判别器无法以超过 50% 的准确度分辨出真实样本。

相较于 2014 年初现期，现有的 GAN 在神经网络架构、损失函数设计、模型训练稳定性、模型崩溃问题上取得了相应突破，提升了最终图像的特定细节、内在逻辑、生成速度等。但要在实际应用中大规模稳定应用，GAN 仍需解决以下问题：

GAN 领域仍待解决的问题及创新思路

•解决训练不稳定问题

GAN 模型的相互博弈过程容易造成训练不稳定，使得训练难以收敛。近期突破思路有 RelativisticGAN。有别于传统 GAN 测量生成样本是否为真的概率这一做法，RelativisticGAN 将关注点放在测量生成样本比真实样本“更真”的概率，这一尝试使得 GAN 获得了更好的收敛性。

•缓解 GAN 的模式坍塌 modelcollapse，即生成样本大量重复相似

模式坍塌被认为是应用 GAN 进行图像生成时最难解决的问题之一，它会造成训练结果冗余、生成图像质量差、样本单一等问题。近期突破性思路有包含两个判别网络的 D2GAN。

•GAN 的结构改善

将 GAN 与机器学习中最新的理论成果进行交叉训练，引入迁移学习、强化学习等，使 GAN 与计算机图形学等技术更好地融合，推动明显结构的改善，有望成为下一步的研究热点。

•GAN 的压缩

目前图像生成技术想要落地，势必要根据需求调整模型的大小，结合基于“知识蒸馏”的模型压缩方法，进行匹配需求的优化和压缩，从而内嵌入小型软件中，拓宽应用领域同样为当前的研究热点。

扩散模型 Diffusion Model—新一代图像生成主流模型。2022 年，Diffusion model 成为图像生成领域的重要发现，甚至有超越 GAN 的

势头。尽管大多通过 Imagen、DALL·E2 等项目知名，但 Diffusion 本身服务于图像生成。我们将在此对 Diffusionmodel 进行介绍。

Diffusion 是指扩散模型。最初受到热力学概念的启发，扩散模型通过增加噪声破坏训练数据来学习，然后找出如何逆转这种噪声过程以恢复原始图像。一旦经过训练，扩散模型就可以应用这些去噪方法从随机输入中合成新颖的“干净”数据。

相较于其他的图像生成模型(比如 GAN、VAE 和基于流的模型)，在所需数据更少的背景下，Diffusion Model 的图像生成效果有明显提升。我们将在跨模态部分对基于 Diffusion model 的知名模型，如 Guide Diffusion、Disco Diffusion、DALL·E2 等进行具体介绍。

神经辐射场模型 NeRF—符合 3D 内容消费趋势的新一代模型 NeRF 通过将场景表示为隐式的神经辐射场，渲染时通过神经网络查询位置上的场景信息生成新视角图像。简单来说，NeRF 利用深度学习完成了计算机图形学中的 3D 渲染任务。

此前，2D 到 3D 生成的领域也包含 GAN 方面的尝试，比如英伟达 20-21 年推出的 GANverse3D 能够生成可自定义并生成动画的 3D 图形。但由于 GAN 在对抗训练中会受限于样本特征，该模型当时仅适用于汽车、马匹和鸟类。而 NerF 在很大程度上克服了这一问题。

NeRF 模型基本原理

- 将场景的体积表示优化为向量函数，该函数由位置和视图方向组成的连续 5D 坐标定义。具体而言，是沿相机射线采样 5D 坐标，来合成图像

- 将场景表示参数化为一个完全连接深度网络（MLP），该网络将通过 5D 坐标信息，输出对应的颜色和体积密度值
- 通过体积渲染技术将这些值合成为 RGB 图像
- 基于可微分的渲染函数，通过最小化合成图像和真实图像之间的残差，优化场景表示随着元宇宙等概念、深度相机等相关基础设施、VR 眼镜等消费渠道的发展，3D 内容生成将成为未来内容领域的必经之路，可以预见，通过降低对创作者 3D 建模能力和渲染能力的要求，NeRF 对于虚拟人创建、3D 训练环境构建、增强现实、线上游戏及电影特效等都具有重要意义。

自 NeRF 在 ECCV2020 提出后，NeRF 模型也持续在生成范围、生成效果、乃至所需基础数据上进行改进。例如包含光影变化效果（谷歌团队联合 MIT、UC 伯克利发布的 NeRV）、动态 NeRF（HyperNeRF 和 ST-NeRF），接近实时生成（英伟达的 Instant-NGP 和 InstantNeRF），全场景 NeRF（谷歌，mip-NeRF360，对合成 3D 图像的视角和深度有了大幅提升。可以渲染完整的 3D 物体及场景，并支持在整个空间内从任何角度观看）、单张生成模型（2022 年谷歌 PHORHUM，甚至能够重现不可视区域；苏黎世联邦理工 Pix2NeRF）、可进一步生成 3D 几何数据（英伟达及斯坦福，EG3D，可生成高分辨率人脸的几何结构）

其中的关键节点是能够基于单张图像生成 3d 模型，这预示着 3D 建模效率的明显提升。尽管 NeRF 在 3D 内容生成领域受到了广泛关注，但具有创新架构设计的 3DGAN 依旧在扮演重要角色。例如，在

以单张图像生成 3D 模型上，作为纯隐式神经网络的 NeRF 确定位置的速度较慢，计算量较大，会导致画质低、变形等问题。而英伟达与斯坦福大学推出的结合了显隐式神经网络特征的 EG3D 模型，在速度和内存消耗上都有明显改进。

（二）技术及商业场景介绍

基于对不同技术原理的梳理，我们将图像生成领域的技术场景划分为图像属性编辑、图像局部生成及更改、以及端到端的图像生成。其中，前两者的落地场景为图像编辑工具，而端到端的图像生成则对应创意图像及功能性图像生成两大落地场景。

图像属性及部分编辑

属性编辑部分，可以直观的将其理解为经 AI 降低门槛的 PhotoShop。目前而言，图片去水印、自动调整光影、设置滤镜（如 Prisma、Versa、Vinci 和 Deepart）、修改颜色纹理（如 DeepAI）、复刻/修改图像风格（DALL·E2 已经可以仅凭借单张图像进行风格复刻、NightCafe 等）、提升分辨率等已经常见。关于此类场景，初创公司数量庞大，Adobe 等相关业务公司（AdobeSensei）也在进入这一领域。

对于此类业务而言，引流渠道和与技术流程挂钩的产品使用体验非常重要。现有的代表公司包括美图秀秀（美图 AI 开放平台）、Radius5、Photokit、Imglarger、Hotpot、Remove.bg、Skylum（MaskAI）、Photodiva。

图像部分编辑部分，指部分更改图像部分构成（如英伟达 CycleGAN 支持将图内的斑马和马进行更改）、修改面部特征（Metaphysics，可调节自身照片的情绪、年龄、微笑等；以 Deepfake 为代表的图像换脸）。

由于技术限制，图像的各部分之间需要通过对齐来避免扭曲、伪影等问题，目前 GAN 还难以直接生成高质量的完整图像。2019 年，曾宣传能够直接生成完整模特图的日本公司 DataGrid 目前已无动向。但同时，也出现了由局部生成拼接为完整图像的生成思路。典型代表为选入 CVPR2022 的 InsetGAN，该模型由 Adobe 推出。

同时，细粒度、分区域的图像编辑能力也较为关键，代表为英伟达的 EditGAN。该模型将需要编辑的原图像 x 嵌入到 EditGAN 的潜空间，借助语义分割图的相同潜码，将原图 x 分割成高度精细的语义块（segmentation mask），并得到分割图 y 。接着，使用简单的交互式数字绘画或标签工具进行手动修改。模型最终会共享潜码的优化，以保持新分割图与真实图像的 RGB 外观一致，

图像端到端生成

此处则主要指基于草图生成完整图像（Vans Portrait、谷歌 Chimera painter 可画出怪物、英伟达 GauGAN 可画出风景、基于草图生成人脸的 Deep Face Drawing）、有机组合多张图像生成新图像（Artbreeder）、根据指定属性生成目标图像（如 Rosebud.ai 支持生成虚拟的模特面部）等。

该部分包含两类场景，分别为创意图像生成与功能性图像生成。前者大多以 NFT 等形式体现，后者则大多以营销类海报/界面、logo、模特图、用户头像为主。垂直代表公司/产品包括 Deepdream Generator、Rosebud.ai、AIGahaku、artbreeder、nightcafe、starryai、wombo、depart、obvious、阿里鹿班、ZMO.ai、Datagrid、诗云科技、道子智能绘画系统等。

由于图像的生成复杂度远高于文字，在整体生成上，目前仍然难以达到稳定可靠的生成高质量图像。但据高林教授评价，人脸生成的应用将预计有更快的发展。从 VAQ、VAE 等技术选型开始，人脸生成的技术研究已经有了较好的效果，同时人脸数据集也较为充足。同时，单张的人脸生成价值相对有限。要进一步发挥其价值，可以考虑将其与 NeRF、也即 3D 内容生成相结合，支持从不同的视角和动作还原特定对象面部，能够在发布会、面见客户等场景中有重要作用。而对于近年视频换脸效果不佳的情况，高教授认为这与底层设计优化，例如除纹理相似度之外，在解编码中考虑更多的时间、动作、甚至情感等因素，并叠加考虑数据、渲染能力等因素。

AI 技术及落地场景介绍—综合赛道

在这部分，我们选取了 GameAI 及虚拟人生成这两个涉及了各方面 AIGC 能力的综合赛道，并认为其将是近期体系化发展最快的 AIGC 行业化赛道。

一、GameAI

以腾讯 AILab 在游戏制作领域的布局为例，人工智能在游戏前期制作、游戏中运营的体验及运营优化、游戏周边内容制作的全流程中均有应用。

我们将其中的核心生成要素提炼为 AIBot、NPC 相关生成和相关资产生成。

- AIBot，也即游戏操作策略生成，可以将其简单理解为人工智能玩家，重点在于生成真实对战策略。2016 年 DeepmindAlphaGO 在围棋中有所展示，随后，AI 决策在 Dota2、StarCraft2、德州扑克、麻将等游戏领域中均展现出了良好的实力。

技术关键在于强化学习方案优化设计，体现为多智能体使用、可适应游戏/环境复杂度、具体策略多样性等。目前，包括网易在内的我国主流游戏公司已经形成共识，除去直接以陪玩等形式服务 C 端玩家，也在通过游戏跑图、平衡度等方式服务于游戏开发/运营等 B 端。

相关业务场景

前期平衡性测试

游戏策划会根据具体的游戏内容，为角色本身的属性、技能、状态等，以及道具、环境、货币等参数，设定一系列的数值。平衡性测试能够充分地模拟玩家在某一套数值体系下的游戏体验，提出优化策略，为玩家带来更加平衡的多样性游戏交互。

之前需要在测试服上邀请人类玩家试玩 1-2 个月后才能得到结果。现在由 AI Bot 直接在内部完成相关工作即可。

游戏跑图/功能测试

通过 AI bot 针对性的找出游戏中所有交互的可能性，通过发现潜在漏洞辅助游戏策划

对局陪伴

包括平衡匹配、冷启动、玩家掉线接管等

特定风格模拟

在绝悟中，AI 通过模仿职业选手，掌握他们的典型个人风格，玩家则感觉像在与真实的职业选手对抗

基于玩法教学的新型人机互动

在游戏内“绝悟试炼”玩法中，在玩家发出各种指令后，AI 会根据血量、距离等实际情况，评估指令的合理性，选择执行或拒绝，身兼队友及老师，与玩家在真实对战环境中交流协作，并在过程中向玩家传授职业级的策略与操作技术，帮助玩家迅速熟悉英雄操作与游戏玩法。在引入王者绝悟 AI 教学后，玩家单局游戏主动沟通的次数有明显提升，提高了 PVE 玩法的可玩性

代表机构：腾讯 AILab（腾讯“绝悟”）

「绝悟」AI 通过强化学习的方法来模仿真实玩家，包括发育、运营、协作等指标类别，以及每分钟手速、技能释放频率、命中率、击杀数等具体参数，让 AI 更接近正式服玩家真实表现，将测试的总体准确性提升到 95%。

目前腾讯绝悟在环境观测、图像信息处理、探索效率等方面的创新算法已经突破了可用英雄限制（英雄池数量从 40 增为 100），让 AI 完全掌握所有英雄的所有技能并达到职业电竞水平，能应对高达 10 的 15 次方的英雄组合数变化。基于绝悟，王者荣耀的数值平衡性偏差已经从 1.05% 下降到 0.68%，其所涉及的“多智能体”决策过程中，可以在高达 10 的 20000 次方种操作可能性的复杂环境中进行决策。目前，腾讯 AILab 还与王者荣耀联合推出了 AI 开放研究平台「开悟」，并积极举办相关赛事。

代表公司：超参数

估值已达独角兽，业内率先实现在 3DFPS（游戏的帧数）领域的大规模商业化落地，服务对象包括数千级日活的游戏产品。超参数科技的 AI 服务已经为数千年流水超过 10 亿元的游戏产品贡献了巨大的商业价值，涵盖沙盒、开放世界、FPS、MOBA、休闲竞技等多个品类。

在我国，网易伏羲、商汤科技也已在业务布局中提及该部分业务。

- NPC 逻辑及剧情生成，也即由 AI 生成底层逻辑

此前，NPC 具体的对话内容及底层剧情需要人工创造驱动脚本，由制作人主观联想不同 NPC 所对应的语言、动作、操作逻辑等，这种动态的个性化匹配背后依旧是不同的静态分支，创造性及个性化相对有限。

而以 rct AI 的智能 NPC 为例，其 NPC 能够分析玩家的实时输入，并动态地生成交互反应，从而构建几乎无限且不重复的剧情，增强用户体验并延长游戏生命周期。特别是在养成类游戏中，AI 所提供的个性化生成能够带来画面、剧情及具体交互的个性化全新游戏体验。而实时剧情生成则有助于在特定框架内生成全新的可能性，增加游戏整体的叙事可能性。

目前，NPC 逻辑自动生成技术已经应用在《黑客帝国：觉醒》（Mass AI，车辆行人等将独立于玩家操控变动）、Red Dead Redemption 2、Monster Hunter:World 等大型游戏中。长期来看，NPC 的灵活自主将引领其成为重要的社交节点，有效填充元宇宙的内部架构。

代表公司：rctAI

rctAI 属于业内较为少见的可以在逻辑侧进行 AI 生成开发的公司，其核心产品包括个性化且动态的剧情探索、智能 NPC/AI MetaBeings/智能环境，提供真人般的游戏对抗体验。

通过简单设计并调整不同的参数，rctAI 的混沌球（ChaosBox）算法就可以在游戏中大规模地轻松生成具有智能意识的虚拟角色。他们的行为和对话不会重复，皆为动态生成。在游戏场景中，部署具有不同性格的智能 NPC，通过对话、行为等动态交互，增加玩家的游戏时长，同时提供新的变现途径。具体包括性格化 NPC（不同的对话、行为等动态交互）、对抗式 AI（对战类场景中个性化对手）、互动式 AI（具体涉及探索世界、技能训练、参与对抗、完成目标、闲聊陪伴、协作建造六种互动场景）、大规模智能 NPC 部署、智能留存及智能运营策略等。

- 数字资产生成，也即结合其他模态生成能力，落地为具体场景

我们将其进一步区分为 NPC 相关资产生成、场景生成和地图/关卡生成。

AIGC 在游戏领域的应用可能会对游戏行业的商业模式产生影响，例如引入更多 In-GamePurchase，通过游戏主线之外的其他设计增加用户留存时间，引导用户对游戏的部分内容而非整体进行付费。此外，这可能会导致游戏发行方与用户间的关系从产品交付（游戏本身）到全生命周期的个性化服务交付（提供更多个性化埋点）。

GameAI 后续的行业发展关键在于游戏行业能否接受，并从底层架构开始支持提供包括客户数据在内的相关接口。目前来看，国内相关游戏厂商合作意愿良好，普遍存在内部开发或外部测试需求。

二、虚拟人生成

虚拟数字人指存在于非物理世界(如图片、视频、直播、一体服务机、VR)中，并具有多重人类特征的综合产物。

目前“深度合成+计算驱动”型的虚拟人，综合运用文本、图像、

音频等生成技术，打造综合外观、面部表情、发声习惯等产出全面拟人化的数字内容，属于 AIGC 领域。此种多模态生成技术的聚合应用在虚拟偶像、虚拟主播等领域已有广泛应用。在《量子位虚拟数字人深度产业报告》中，我们将虚拟人按照产业应用划分为两种，即服务型虚拟人及身份型虚拟人。

计算驱动型/AIGC 型虚拟人制作流程

1.设计形象。扫描真人形态及表演、采集驱动数据

利用多方位摄像头，对通用/特定模特进行打点扫描，采集其说话时的唇动、表情、面部肌肉变化细节、姿态等数据。

2.形象建模，进行绑定

设计所需的模型，或基于特定真人进行高还原度建模，进行关键点绑定。关键点绑定的数量及位置影响最终效果。

3.训练各类驱动模型

决定最终效果的核心步骤利用深度学习，学习模特语音、唇形、表情参数间的潜在映射关系，形成各自的驱动模型与驱动方式。充足的驱动关键点配合以精度较高的驱动模型，能够高还原度的复原人脸骨骼和肌肉的细微变化，得到逼真的表情驱动模型。

4.内容制作

基于输入的语音（或由输入文本转化的语音），预测唇动、表情等参数核心的技术流程是基于输入的语音，或首先基于 TTS 技术（Text-to-speech，语音合成技术），将输入的文本转化为语音。基于语音，结合第 3 步得到的驱动模型，并利用生成对抗模型 GAN 选出最符合现实的照片，推理得到每帧数字人的图片。通过时间戳，将语音和每帧的数字人图片进行结合。

5.进行渲染，生成最终内容

直播时进行实时渲染。为保证在特定场景下能够实现实时低延迟渲染，计算框架的大小、算力供给等技术问题同样会影响到虚拟数字人的最终生成效果

综合来看，虚拟人生成代表着从文本/音频等低密度模态向图像/视频/实时交互等信息密度更高的模态的转化。其中，视频是短期的发展重点，而长期来看，乃至在元宇宙阶段，通过实时交互成为社交节点，都将是虚拟人重要的应用场景。

在 AIGC 领域，虚拟人生成分为虚拟人视频生成和虚拟人实时互动。虚拟人视频生成是目前计算驱动型虚拟人应用最为广泛的领域之一，不同产品间主要的区分因素包括：唇形及动作驱动的自然程度、语音播报自然程度、模型呈现效果（2D/3D、卡通/高保真等）、视频

渲染速度等。我们在此关注到了小冰公司与每日财经新闻合作的虚拟人实时直播，除虚拟人的自动生成外，还包括了摘要、图示、表格等的自动生成，在虚拟人的基础上，交付了更为完整的 AIGC 内容播报产品。此外，倒映有声的 TTSA 除虚拟人外，还包括整个画面中的素材呈现，相较于市面上嘴形、面部和身体律动的有限覆盖，虚拟人播报的整体效果也有所提升。

代表公司：倒映有声

一家以技术为核心的创新型公司和无人驱动数字分身技术解决方案供应商，通过自研神经渲染引擎和 TTSA 技术，实现基于文本实时生成高质量语音（音频）和动画（视频）。

在试用了倒映有声的产品后，我们发现其虚拟人自然度高于市面产品，倒映有声将其归结于神经渲染（NeuralRendering）、TTSA(基于文本和语音合成实时生成音频和视频)、ETTS(富情感语音合成)、DigitalTwin(数字孪生)。通过神经渲染技术快速构建 AI 数字分身，通过语音+图像生成技术，生成和驱动数字分身的唇形、表情、动作、肢体姿态，创造表情自然、动作流畅、语音充满情感的高拟真度数字分身 IP。

而虚拟人的实时互动则广泛应用于可视化的智能客服，多见于 APP、银行大堂等。在 AIGC 的虚拟人领域，由于更能够体现 AI 在个性化、高并发性等方面的优势，我们更强调虚拟人的实时交互功能。我们可以将这一功能理解为以人为单位的数字孪生，其中会进一步涉及思维及策略相关的生成。但由于文本生成的局限性，该场景目前只能适用于特定行业。

该领域的代表公司包括：HourOne.ai、Synthesia、Rephrase.ai、小冰公司、倒映有声、数字王国、影谱科技、科大讯飞、相芯科技、追一科技、网易伏羲、火山引擎、百度、搜狗等。除了基于 NLP 进行问答外，以小冰公司和腾讯 AILab（AI 虚拟人艾灵）为代表，部分公司也在尝试将不同的生成能力融合在虚拟人下，使虚拟人能够更好的融入现实世界。

以小冰公司的小冰框架为例，虚拟人不仅在人格化形式上涉及了虚拟面容生成、虚拟语音定制、交互等，并进一步被赋予了写诗、绘画、演唱、音乐创作等 AI 内容创作能力，以虚拟人为接口，对外提供全栈式的 AIGC 能力。

虚拟人及综合性 AIGC 代表公司：小冰公司

小冰是全球领先的人工智能科技公司，旗下小冰框架是全球承载交互量最大的完备人工智能框架之一，在开放域对话、多模态交互、超级自然语音、神经网络渲染及内容生成领域居于全球领先。

作为“Albeing”派虚拟人。小冰的产品始终是人+交互+内容。具体包括虚拟人（夏语冰等 somebodyinstance、虚拟男友等 nobody

instance 和国家队人工智能裁判与教练系统观君等在垂直场景中工作的虚拟人类）、音频生成（主攻超级语言及歌声，在线平台与歌手歌声合成软件 Xstudio）、视觉创造（毕业作品集《或然世界》、为国家纺织品开发中心、万事利等数百家机构提供了图案和纹样设计）、文本创造（2017 年即推出小冰诗集）、虚拟社交、GameAI（XiaoiceGameStudio）等。

商业客户已覆盖金融、智能车机、零售、体育、纺织、地产、文旅等十多个垂直领域，并提出了以“人力”的逻辑去进行商业报价的虚拟人商业模式。

三、相关预测

不同赛道下 AIGC 应用落地推广程度主要受到两方面影响，特定技术的水平状况以及在实际应用中出现的转化门槛。

特定技术的水平状况

•操作方式与传统创作流程的契合，需要符合创作者使用思路并提供相关人工的介入环节。与底层模型架构相关

- 是否支持小样本学习，一方面是训练效率的问题，一方面也是对实际使用门槛的回应
- 生成能力的鲁棒性，确保在生成规模扩大、高并发性等实际应用需求下依然能够有稳定的产出
- 生成模型可外延和泛化能力，例如可同时涉及多重设计/创作风格
- 配套数据及现有标注情况，对于实际场景落地有重要意义
- 模态对齐工具/架构具有重要意义，CLIP 的技术程度，以及潜在模态对齐模型具有重大风向标意义

•不同元素间的连续控制情况，例如文本生成中的上下文语义承接、视频生成中不同帧的连贯性、图像生成不同像素间的细节处理等

•具体的技术选型问题，需要相关的底层原理嵌入。但目前而言，整体的技术路线已经相对定型，需要更多在技术细节上的优化

在实际应用中出现的转化门槛

- 完成渲染、构建等相关配套环节的时间/技能/算力投入及难度，如是否搭配有相关后续平台
- 论文上指标和现实使用中的指标无法对齐
- 大量隐含信息的问题，难以直接回应现实中的应用问题

以下技术要素值得关注：长文本生成、开放式文本生成、NeRF 模型、Diffusion 模型、跨模态大型预训练模型（支持的模态数据类型、模态对齐架构设计、支持的下游应用）、小样本学习及自监督算法、强化学习及环境学习。

以下技术场景是短期内将有较明显爆发的：闲聊式文本生成、个性化营销文本、富情感及细节 TTS、拼凑式视频生成、基于文本的 AI 绘画、语音复刻。

关键结论

1.区别于 GenerativeAI 和 Syntheticmedia, AIGC 中的跨模态生成和策略/线索生成应当得到重视，这两者也是 AIGC 长期的重要价值增长点。

2.按照模态区分，AIGC 可分为音频生成、文本生成、图像生成、视频生成及图像、视频、文本间的跨模态生成，细分场景众多，跨模

态生成需要重点关注。GameAI 和虚拟人生成是目前变现较为明确的两大综合赛道，预计对原行业产生整体性影响。

3.以下技术要素值得关注：长/开放式文本生成、NeRF 模型、Diffusion 模型、跨模态大型预训练模型、小样本学习及自监督算法、强化学习及环境学习。以下技术场景值得关注：闲聊式文本生成、个性化营销文本、富情感及细节 TTS、拼凑式视频生成、基于文本的 AI 绘画、语音复刻。

5.除降本增效、提供参考外，AIGC 的以下两点价值更值得关注：对不同模态元素进行二次拆解组合，改变内容生产逻辑及形式；和其他 AI 系统或数据库进行联动，有实现高度个性化/高频优化。

6.我国 AIGC 行业仍处于刚刚起步阶段，距离大规模验证和体系化发展仍有距离，“模块分拆+个性化推荐”的“泛 AIGC”形式预计将有所发展。

GPT-4 研究

(一) GPT-4 的技术与能力

(二) GPT-4 的应用与展望

GPT-4 的技术与能力

AIGC 是继专业生成内容（PGC）和用户生成内容（UGC）之后，利用人工智能技术自动生成内容的新型生产方式。AIGC 根据其内容模态不同可分为文本、视频、图像，音频与跨模态生成。文本方面，例如文本创作、代码生成、问答对话等；视频方面，例如视频画质增强、视频内容创作、视频风格迁移等；图像方面，例如图片编辑、图片生成、3D 图像生成等；音频方面，例如文本合成人声、语音克隆、音乐生成等；跨模态方面，如文字生成图片、文字合成视频、图像描述等，而且在不同内容模态的技术应用场景也有着各自的细分品类。

一、GPT 的发展历程

ChatGPT 是 AIGC 发展的第一个丰碑。ChatGPT 是由人工智能研究公司 OpenAI 在 2022 年 11 月发布的一个对话型大语言模型，是人工智能技术驱动的自然语言处理工具和应用。ChatGPT 的全称是 Chat Generative Pre-trained Transformer，顾名思义，就是以 Transformer 为基础架构，采用预训练和生成式方式构建的面向对话的大语言模型，是 AIGC 在文本方面的典型代表。ChatGPT 的主要用途是生成对话，它能够通过学习 and 理解人类的语言来进行对话，根据聊天的上下文进行自然、流畅的互动，还能完成邮件撰写、文案编写、文本翻译、代码生成等任务。ChatGPT 提供了前所未有的高效、自然的人机交互体验和极富创造力的内容生成能力，成为了 AI 时代的第一个“杀手级”应用。

以 ChatGPT 为代表的生成式 AI 工具将使机器可以大规模参与知识类和创造性工作，极大提升生产力，涉及 数十亿人方方面面的工作，可能产生数万亿美元的经济价值。ChatGPT 覆盖了 NLP 所有领域，其所代表的大规模预训练语言模型（LLM）或基础模型已成为工业界和学术界最为关注的研究热点，并引领近期自然语言处理（NLP）乃至人工智能领域的研究范式的转变，对人工智能的技术发展可能产生重大影响。

相距 ChatGPT 发布仅仅 4 个月，OpenAI 在 2023 年 3 月正式发布了多模态预训练大模型 GPT-4。GPT-4 支持图片和文本输入并生成文本输出，相比 ChatGPT 增加了识图能力，并能够生成歌词、创意文本、学习用户的写作风格等，更具创造力和协作性。GPT-4 的输入限制提升至 2.5 万字，处理能力是 ChatGPT 的八倍，可用于长篇内容创作、扩展对话以及文档搜索和分析等应用场景，并可以用所有流行的编程语言写代码。GPT-4 的回答准确性大幅提高，性能优于现有的大型语言模型、以及当前最先进（SOTA, StateOfTheArts）模型，虽然在许多现实场景中不太聪明，但在各种专业和学术基准上表现出了人类水平。

ChatGPT 存在一定局限，如逻辑处理能力不够强，对上下文的理解仍然有限，多轮对话可能失控，缺乏一些通识能力，可能会出现创造不存在的知识，或者主观猜测提问者的意图，对 2021 年后的世界和事件了解有限。GPT-4 相较于 ChatGPT，不合规内容的响应请求降低了 82%，响应准确度提高了 40%，具备了更高的理解能力。但是 GPT-4 仍然具有与早期 GPT 模型相似的局限性，如：犯常识性的错误、缺乏对新世界知识的了解、存在社会偏见、产生幻觉、推理错误等。总的来说，ChatGPT

和 GPT-4 是 NLP 和人工智能领域划时代的标志，也昭示着人类向通用人工智能或强人工智能迈出了尝试性的一步。

GPT-4 基于 Transformer 架构演进发展，其发展历程如下：

- 2017 年 6 月，Google 发布论文《Attention is all you need》[1]，首次提出 Transformer 模型，成为 GPT 发展的基础；
- 2018 年 6 月，OpenAI 发布论文《Improving Language Understanding by Generative Pre-Training》[2]（通过生成式预训练提升语言理解能力），首次提出 GPT-1 模型 (Generative Pre-Training)；
- 2019 年 2 月，OpenAI 发布论文《Language Models are Unsupervised Multitask Learners》（语言模型应该是一个无监督多任务学习者），提出 GPT-2 模型；
- 2020 年 5 月，OpenAI 发布论文《Language Models are Few-Shot Learners》[4]（语言模型应是一个少量样本 (few-shot) 学习者），提出 GPT-3 模型；
- 2022 年 2 月底，OpenAI 发布论文《Training language models to follow instructions with human feedback》（使用人类反馈指令流来训练语言模型），公布 InstructionGPT 模型；
- 2022 年 11 月 30 日，OpenAI 推出 ChatGPT 模型并提供试用，自发布两个月内月活跃用户数已达 1 亿，成为史上用户数增长最快的消费者应用；
- OpenAI GPT-4 模型于 2023 年 3 月 14 日发布，并已在微软的新版搜索引擎必应 (Bing) 上运行。

二、GPT-4 的能力分析

NLP 领域研究包括了众多子领域，从任务角度看可分为两大类：一类是中间任务，即自然语言理解任务，如中文分词、词性标注、句法分析、指代消解等；另一类是最终任务，即自然语言生成任务，如文本分类、文本摘要、机器翻译、

文本对话等。随着 GPT 等 LLM 的出现，自然语言生成任务兼容覆盖了自然语言理解任务，中间任务逐渐退出，LLM 可以直接端到端实现最终任务。而 ChatGPT/GPT-4 则用自然对话的人类表达的方式，提供了 LLM 与人的新型交互接口，以统一的接口方式支撑 NLP 的所有领域端到端最终任务。在日常对话中，ChatGPT 和 GPT-4 之间的差异很小。但是当任务的复杂性超过了某一阈值时，GPT-4 的表现则要优于 ChatGPT。此外，GPT-4 不仅能接受纯文本输入，还可以接受图片作为输入来完成图片描述、图片分类和图片分析任务，但该功能暂未对外开放，仍处于内测阶段。下面以 GPT-4 对话接口的方式，通过示例进行 GPT-4 端到端任务能力分析。

（一）多种风格解释复杂名词

多风格复杂名词解释指人工智能在解释一个名词时使用多种不同的风格或方式，包含严谨、科技、幽默、简单易懂等风格。这种能力可以在不同的场景中得到应用，例如学术知识科普、复杂概念简化等。GPT-4 能够以各种风格来扮演各种角色，向各种类型的用户解释一些复杂的名词。例如，向 GPT-4 询问通信行业的 5G 网络技术时，它所产生

的回答如下图。



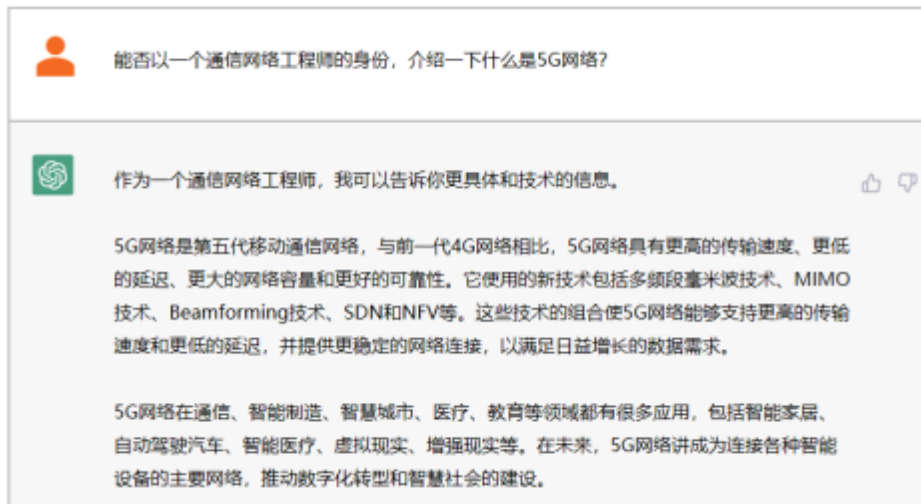
GPT-4 解释复杂名词

上述 GPT-4 对于 5G 网络的解释，可能很难让一个 5 岁的孩子快速理解。当告诉 GPT-4，需要 GPT-4 简化概念，向一个 5 岁的孩子介绍时：



(二) GPT-4 简化概念，解释复杂名词

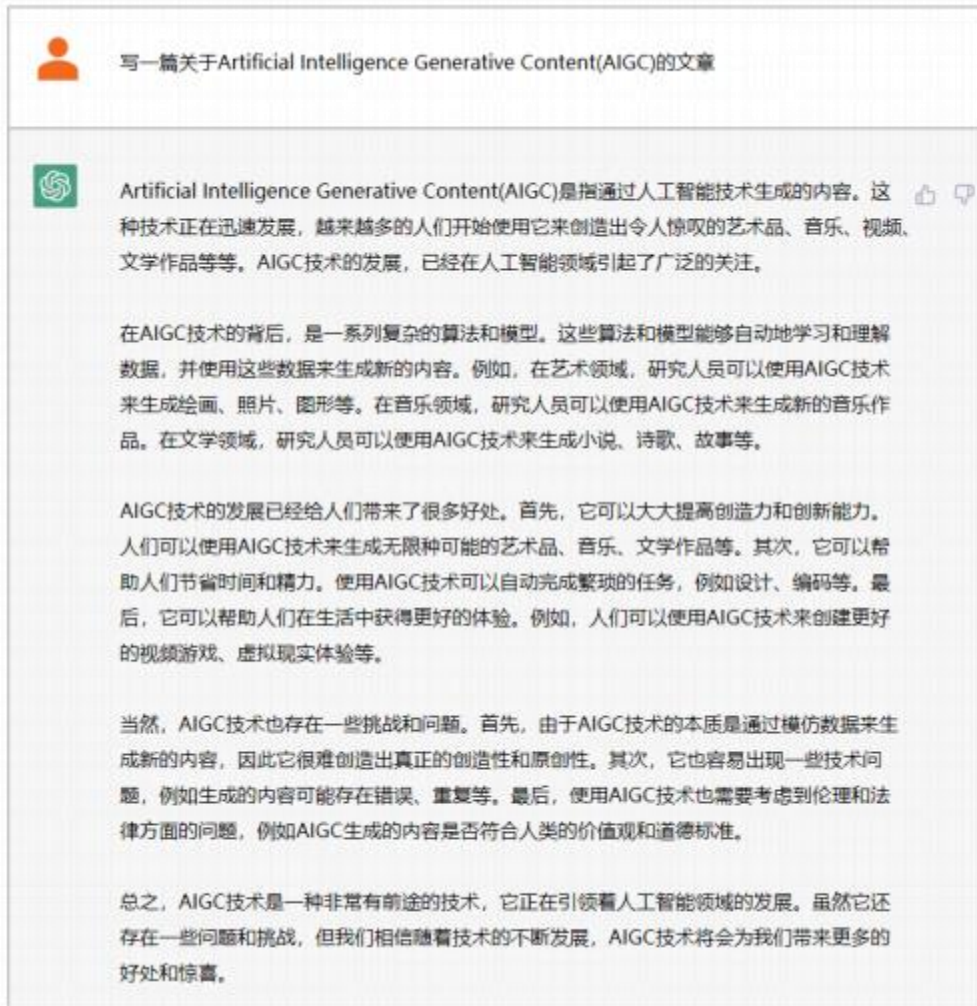
而对于一些想深入了解 5G 网络其中所运用到的技术的用户而言，GPT-4 能够以一个通信网络工程师的身份，向用户介绍 5G 网络的概念：



GPT-4 以专业人士身份，解释复杂名词

（三）特定要求或主题创作文章

AI 写作根据用户给定的输入生成符合语法和语义要求的文本内容，包括新闻报道、技术介绍、小说、商业信函等等。在 AI 写作的应用场景中，人们通常会提供一些初始的输入，例如主题、文章结构等等特定要求，AI 会根据这些输入自动生成对应的文本内容。与传统的人工撰写相比，AI 写作可以大大提高工作效率，同时还可以生成更加多样化和精准的文本内容，因此越来越受到各行各业的关注和应用，图 2-5 为 GPT-4 所撰写的关于 AIGC 的文章：



GPT-4 根据特定要求或主题创作文章

(四) 对话机器人

对话机器人要求人工智能理解用户的语言输入，并根据用户需求和问题生成相应的回答。用户可以通过对话机器人进行语音或文字交互，就像与真人交谈一样，从而获得所需的信息和服务。对话机器人可以应用于多种场景，例如客户服务、虚拟助手等。在客户服务领域，对话机器人可自动回答客户的常见问题，解决客户的问题，减少客服工作量，提高客户满意度；在虚拟助手领域，对话机器人可以帮助用户完成日常任务，如提醒、查询、推荐、预定等，图 2-6 为 GPT-4 根据

用户的具体需求而推荐的购机指南：



The screenshot shows a chat interface with two messages. The first message is from a user asking for a computer recommendation. The second message is from GPT-4, providing a detailed list of factors to consider (processor, memory, storage, display, battery) and three specific laptop models: ThinkPad X1 Carbon Gen 9, Dell XPS 13, and HP EliteBook x360 1040 G8. The chat interface includes a user icon, a GPT-4 logo, and thumbs up/down icons for each message.

想买一台电脑

好的，请问您想要购买哪种类型的电脑？比如说台式机还是笔记本电脑？您的使用需求是什么？比如说是否需要进行游戏、视频编辑、办公处理等操作？另外，您的预算是多少？这些信息可以帮助我更好地为您提供购买建议。

想买一台办公用的笔记本电脑

好的，如果您想要购买一台办公用的笔记本电脑，我建议您考虑以下几个因素：

1. 处理器性能：对于办公用途，一般来说您不需要特别高的处理器性能。但是，如果您需要处理大量的数据、进行复杂的计算或运行多个程序，那么一个高性能的处理器会让您的工作更加流畅。我建议选择至少Intel Core i5或AMD Ryzen 5处理器。
2. 内存：在办公场景下，您需要同时运行多个应用程序，因此建议选择至少8GB内存的笔记本电脑。如果您需要运行更加复杂的程序，比如视频编辑软件，可以选择16GB或更高内存的笔记本电脑。
3. 存储：一般来说，128GB或256GB的固态硬盘足以满足办公用途的需要。如果您需要存储更多的文件或数据，可以选择512GB或更高容量的固态硬盘。
4. 显示器：对于办公用途，一般建议选择至少14英寸的显示器。同时，如果您需要长时间使用电脑，可以选择护眼显示器来减少眼睛的疲劳感。
5. 电池续航：如果您需要在外出或无法插电的情况下使用电脑，建议选择具有较长电池续航能力的笔记本电脑。

根据以上建议，以下是几款适合办公用途的笔记本电脑：

1. ThinkPad X1 Carbon Gen 9：搭载Intel Core i5或i7处理器、16GB内存和512GB固态硬盘。具有14英寸护眼显示器和较长的电池续航能力。
2. Dell XPS 13：搭载Intel Core i5或i7处理器、8GB或16GB内存和256GB或512GB固态硬盘。具有13.4英寸护眼显示器和较长的电池续航能力。
3. HP EliteBook x360 1040 G8：搭载Intel Core i5或i7处理器、16GB内存和512GB固态硬盘。具有14英寸护眼显示器和较长的电池续航能力。

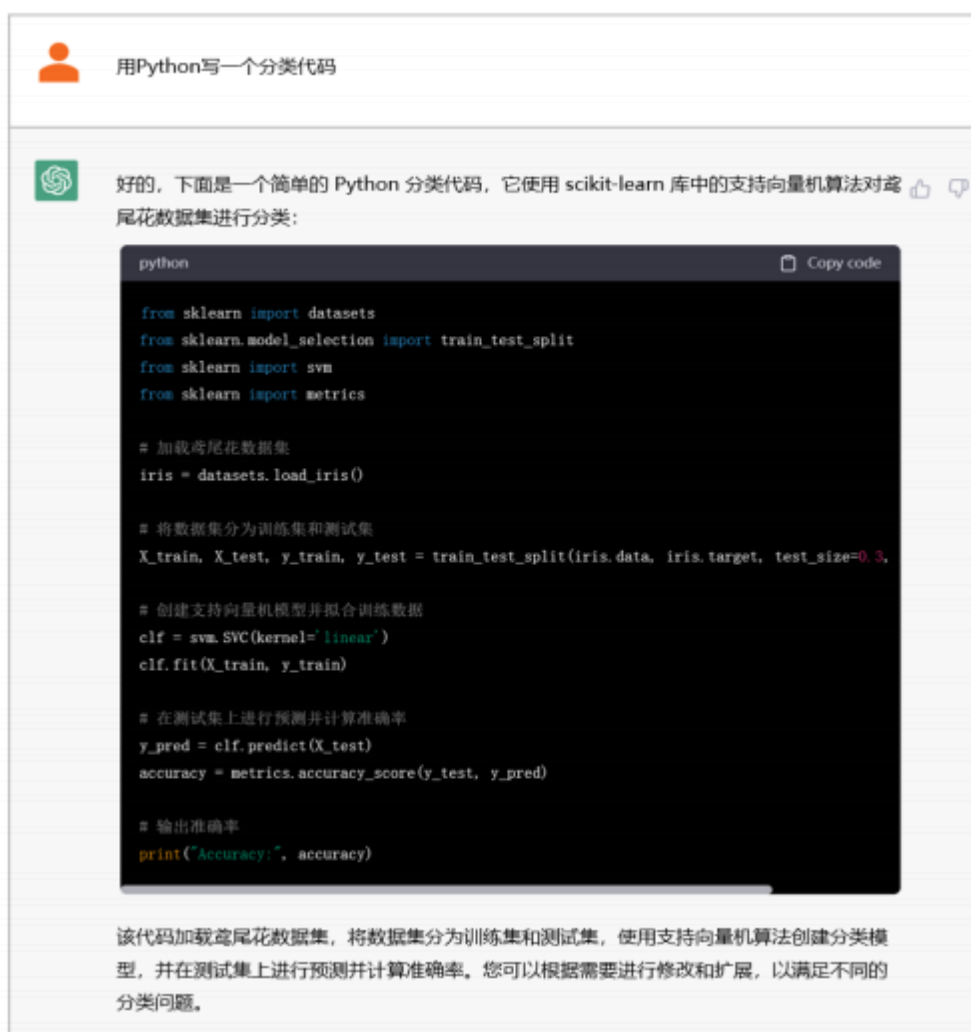
以上是我根据您的信息，给出的几款适合办公用途的笔记本电脑。您可以根据自己的预算和需求进行选择。

GPT-4 与用户多轮对话进行推荐

（五）辅助进行代码编写、Debug 及解释

GPT-4 具有非常强大的代码生成能力，可以生成、修正各种编程语言的代码片段。具体而言，GPT-4 的代码生成能力主要基于两个方面：一是其能够理解并生成自然语言描述的程序逻辑；二是其能够生成符

合编程语言语法规则的代码。GPT-4 通过对大量程序源代码和自然语言描述的语料进行训练，学习到程序的逻辑结构和语义含义。当输入一段自然语言描述的程序逻辑时，GPT-4 可以理解其中的含义并生成对应的代码。GPT-4 的训练数据也包括大量的符合编程语言语法规则的代码片段，因此它可以很好地遵循编程语言的语法规则生成代码。GPT-4 还可以结合常见的编程库和框架生成相应的代码，比如生成使用 Scikit-learn 进行机器学习训练的 Python 代码。图为 GPT-4 使用 Python 编程语言和机器学习框架 Scikit-learn 进行机器学习训练的代码：



The screenshot shows a chat window with a user prompt and a GPT-4 response. The user prompt is: "用Python写一个分类代码". The GPT-4 response is: "好的，下面是一个简单的 Python 分类代码，它使用 scikit-learn 库中的支持向量机算法对鸢尾花数据集进行分类:". Below the text is a code block with the following Python code:

```
python Copy code

from sklearn import datasets
from sklearn.model_selection import train_test_split
from sklearn import svm
from sklearn import metrics

# 加载鸢尾花数据集
iris = datasets.load_iris()

# 将数据集分为训练集和测试集
X_train, X_test, y_train, y_test = train_test_split(iris.data, iris.target, test_size=0.3,

# 创建支持向量机模型并拟合训练数据
clf = svm.SVC(kernel='linear')
clf.fit(X_train, y_train)

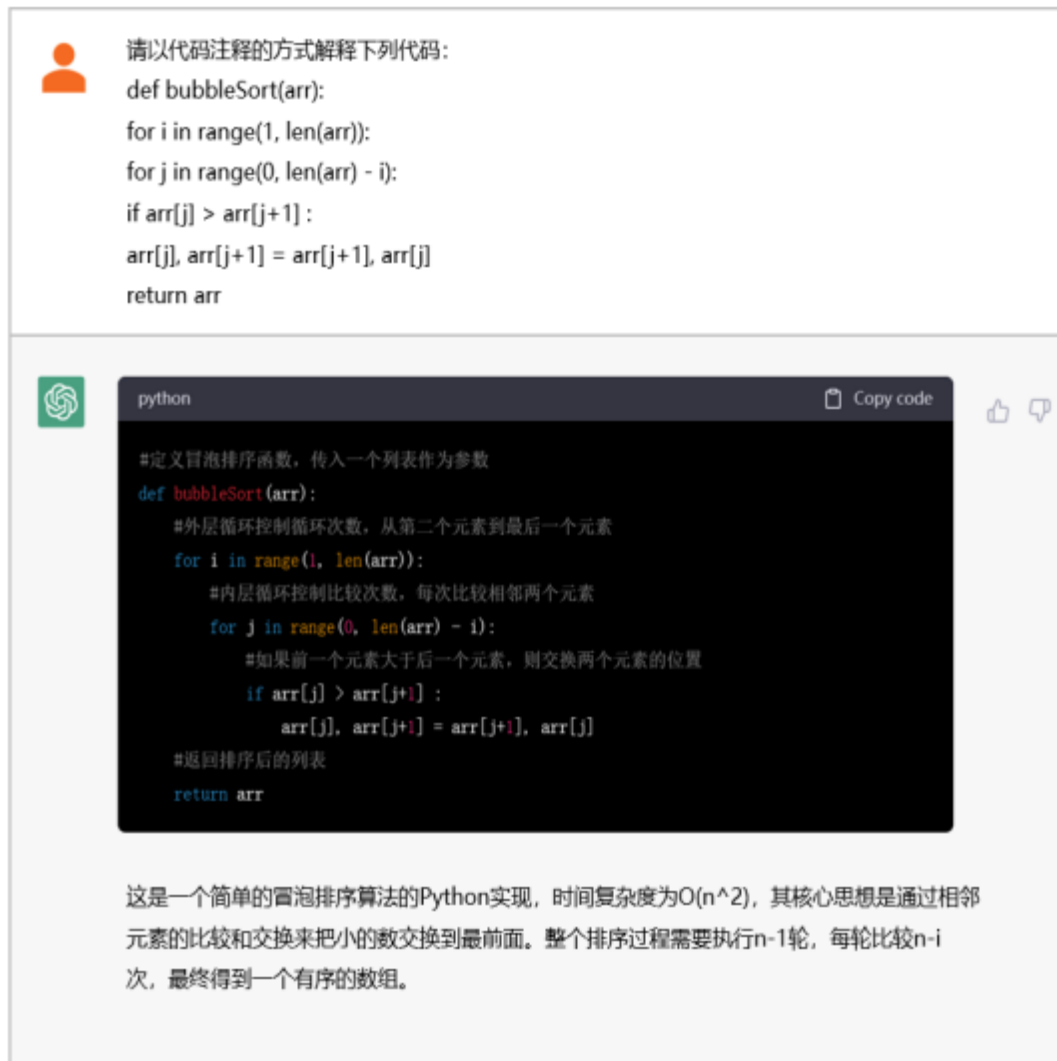
# 在测试集上进行预测并计算准确率
y_pred = clf.predict(X_test)
accuracy = metrics.accuracy_score(y_test, y_pred)

# 输出准确率
print("Accuracy:", accuracy)
```

该代码加载鸢尾花数据集，将数据集分为训练集和测试集，使用支持向量机算法创建分类模型，并在测试集上进行预测并计算准确率。您可以根据需要进行修改和扩展，以满足不同的分类问题。

GPT-4 智能生成代码

GPT-4 具备代码解释能力。对于一些缺少注释的代码，GPT-4 也能够自动识别代码块的编程语言，代码块的缩进等，以代码注释的方式向用户解释每行代码的具体作用：



The screenshot shows a chat window with a user prompt and a GPT-4 response. The user prompt asks for an explanation of a Python code snippet. The GPT-4 response includes a code block with detailed comments and a paragraph explaining the algorithm's complexity and logic.

请以代码注释的方式解释下列代码：

```
def bubbleSort(arr):
    for i in range(1, len(arr)):
        for j in range(0, len(arr) - i):
            if arr[j] > arr[j+1]:
                arr[j], arr[j+1] = arr[j+1], arr[j]
    return arr
```

```
python
#定义冒泡排序函数，传入一个列表作为参数
def bubbleSort(arr):
    #外层循环控制循环次数，从第二个元素到最后一个元素
    for i in range(1, len(arr)):
        #内层循环控制比较次数，每次比较相邻两个元素
        for j in range(0, len(arr) - i):
            #如果前一个元素大于后一个元素，则交换两个元素的位置
            if arr[j] > arr[j+1]:
                arr[j], arr[j+1] = arr[j+1], arr[j]
    #返回排序后的列表
    return arr
```

这是一个简单的冒泡排序算法的Python实现，时间复杂度为 $O(n^2)$ ，其核心思想是通过相邻元素的比较和交换来把小的数交换到最前面。整个排序过程需要执行 $n-1$ 轮，每轮比较 $n-i$ 次，最终得到一个有序的数组。

GPT-4 智能解释代码

GPT-4 具备代码修正能力。GPT-4 能够自动识别代码块中所存在的错误并给出合理的解释并修正：



GPT-4 智能修正代码

（六）文本知识抽取

文本知识抽取是指人工智能从自然语言文本中提取结构化或半结构化的信息或知识。对于 GPT-4 而言，用户仅需要输入特定的文本抽取任务和待抽取的文本，GPT-4 就能够抽取文本中的关键性实体，对于苹果即将上市的 iPhoneSE4 智能手机配置信息，GPT-4 的抽取结果如下图：

实体抽取: iPhoneSE4在性能方面, 搭载的A15芯片, 6G的运行内存, 性能毋庸置疑, 电池也升级到3000mAh时, 自带优化更好的ios16系统, 更重要的是价格, 大家都知道苹果旗舰手机定价并不便宜, 动辄五六千上万, 而该系列机型市场定位三千档。

实体抽取结果如下:

- 产品名称: iPhoneSE4
- 芯片型号: A15
- 运行内存: 6G
- 电池容量: 3000mAh
- 操作系统: ios16
- 价格定位: 三千档

GPT-4 抽取文本中关键信息

(七) 阅读理解与分析

阅读理解与分析任务要求人工智能不仅仅能理解和抽取文本中的关键词词, 还能够推理出更加深入的信息和数据。对于 2022 年亚信科技公司的年中财报数据, GPT-4 不仅仅能够将文本中的关键数字提取并列表, 还能够推理出 2021 年的三新业务收入、三新业务收入占比等:

学习下面的内容, 用表格的形式, 对比一下亚信科技2022年上半年和去年同期的营收: 近日, 亚信科技公布了公司截至2022年6月30日中期综合业绩。报告期内, 公司新签订单同比增长超过30%; 收入达到约人民币31.09亿元, 同比上升14.5%。其中三新业务保持高速增长, 收入达到约人民币10.04亿元, 同比上升81.4%, 三新业务收入占收入比重提升至32.3%。

经济指标	2022年上半年	2021年上半年	同比变化
收入 (亿元)	31.09	27.13	上升14.5%
三新业务收入 (亿元)	10.04	5.53	上升81.4%
三新业务收入占比 (%)	32.3	20.4	上升11.9个百分点

根据亚信科技公布的中期综合业绩, 可以看出亚信科技2022年上半年的营收同比去年同期增长了14.5%。其中, 三新业务收入同比增长了81.4%, 三新业务收入占比也从去年同期的20.4%提升至32.3%。

GPT-4 智能分析文本和数据并制表

复杂问题分析

GPT-4 具有更渊博的知识和更强大的解决问题的能力，它可以以更高的精度分析解决复杂的问题：

 在使用生成对抗网络 (GANs) 进行图像生成时，如何解决模式崩溃 (mode collapse) 问题，同时避免生成器 (generator) 和判别器 (discriminator) 之间的训练不稳定性，以提高生成图像的质量和多样性？讨论可能的解决方案和相应的优缺点

 在使用生成对抗网络 (GANs) 进行图像生成时，解决模式崩溃和训练不稳定性问题至关重要。以下是一些可能的解决方案及其相应的优缺点:

- 使用迷你批次 (minibatch) 判别：**

在判别器中，而非针对单个样本进行判断，使用小批量样本的统计信息来评估真实性。这有助于生成器产生更多样化的样本。

优点：有助于提高多样性。

缺点：可能会增加训练时间。
- 修改损失函数：**

使用 Wasserstein GAN (WGAN) 或其他替代损失函数，可以减轻训练不稳定性问题。

优点：更稳定的训练过程。

缺点：需要调整超参数以获得最佳性能。
- 梯度惩罚 (Gradient Penalty)**

在 WGAN 的基础上，引入梯度惩罚，确保判别器的梯度在局部保持合适的范围，有助于避免模式崩溃。

优点：提高稳定性和生成样本质量。

缺点：计算成本较高。
- 自动编码器-生成对抗网络 (AAE-GAN)**

结合自动编码器 (AE) 和生成对抗网络 (GAN) 的优点，学习生成更多样化的样本。

优点：提高多样性，缓解模式崩溃问题。

缺点：需要训练额外的自动编码器。
- 实例噪声 (Instance Noise)**

在训练期间，在判别器的输入数据中添加随机噪声。这有助于平滑判别器的决策边界。提高生成图像的多样性。

优点：提高多样性，降低训练不稳定性。

缺点：可能导致生成图像的质量降低。
- 使用卷积神经网络 (CNN)**

使用卷积神经网络作为生成器和判别器，有助于捕获图像的局部特征，并改善生成图像的质量。

优点：提高生成图像的质量。

缺点：计算复杂度较高。
- 结合多个生成器和/或判别器：**

使用多个生成器和/或判别器进行训练，以提高生成图像的多样性。

优点：提高多样性，降低模式崩溃的风险。

缺点：计算成本较高，需要协调多个生成器和判别器。

综上所述，解决 GAN 中的模式崩溃和训练不稳定性问题需要综合考虑各种方法。在实际应用中，可以根据具体任务和需求，结合多种技巧来提高生成图像的质量和多样性。需要注意的是，每种方法都有其优缺点，可能需要对超参数和网络结构进行调整，以找到最合适的解决方案。

GPT-4 分析解决复杂问题

图文分析

GPT-4 可以同时接受文本和图片输入，通过分析图片内容，结合输入文本作出合理回复：

 我能用这些食材做什么？



 你可以用这些食材做很多东西！以下是一些选项：

- 煎饼或华夫饼
- 法式薄饼
- 法式吐司
- 煎蛋或煎蛋饼
- 乳蛋饼
- 蛋奶或布丁
- 蛋糕或纸杯蛋糕
- 松饼或面包
- 曲奇或松饼

这些只是一些例子，但可能性是无限的!

GPT-4 处理文本图片混合输入

从上述分析示例看，GPT-4 展示了四个极为重要的能力：

- 为人类提供世界知识：GPT-4 拥有庞大的世界知识，包括事实性知识和常识，相比之前的 LLM 覆盖了更多的主题和知识。不仅可以提供知识的查询和检索，而且可以进行知识的归纳和总结，甚至提供简单的知识推理和证明能力。
- 为人类与 LLM 之间提供了符合人类习惯的交互方式：以遵循提示（prompt）并生成补全提示词的句子的语言生成方式，通过上下文学习保持对话一致性，能理解人类意图并用自然语言回答问题、生成内容和解决问题，改变了现有人机互动方式及人类获取世界知识的方式。
- 为人类提供了跨语言及多模态交流能力：不仅可以处理多种人类语言，还可以理解图片内容，为人们提供更加便捷的交流，此外还可以将人类语言与机器语言进行相互翻译，促进了人机物三元世界的融合。
- 具备了自我学习，自我进化的能力：能够自动从海量数据和人类指令中学习到其中的世界知识，学习过程不需要人的介入，而且可以自行检查学习成果并不断优化迭代，能灵活应用所学知识来解决实际问题。

当然，我们也看到，由于 GPT-4 是基于通用公开数据进行训练，缺少各行业领域的非公开的专业数据。这就造成其世界知识中缺乏深度的专业知识。因此，虽然在与人类进行沟通中，可以进行百科全书式的回答，但距离真正满足专业场景，还有相当大的差距。

GPT-4 的应用与展望

在 ChatGPT 和 GPT-4 相继公布后，这一波由深度学习驱动的人工智能技术长驱直进，对人类社会产生了巨大影响。据报道，89%的美国大学生已经在用 ChatGPT 写作业，ChatGPT 通过职位年薪 18 万美元的谷歌 L3 工程师入职测试，也通过了美国执业医师资格考试。微软已经迅速将 Bing 搜索接入了 GPT-4。在为人类设计的学术专业考试上，测试模型理解能力，GPT-4 能排到 TOP10%，GPT-3.5 排在尾部 10%[6]。在传统 NLP 测评数据集上，GPT-4 超越了所有语言模型与大部分 SOTA 模型，在其他语言上测试结果：26 个测试语言中，24 种语言 GPT-4 优于 GPT3.5 与其他语言模型[6]。可以预见，基于前述分析展现出来的 GPT-4 的强大生产力和巨大潜力，GPT-4 与各行各业应用相结合，将在多领域、多功能应用中密集落地[18]。

一、当前各行业的应用：

传媒领域：提升新闻效率，促进网络媒体安全

GPT-4 可以帮助新闻媒体工作者智能生成新闻报道，通过智能方式优化重复劳动性的采编工作。过往新闻智能写作的成功案例众多，例如中国地震网的写稿机器人在九寨沟地震发生后 7 秒内就完成了相关信息的编发；美联社使用的智能写稿平台 Wordsmith 可以每秒写出 2000 篇报道。同时 GPT-4 强大的看图说话能力，整合互联网信息采集技术及信息智能处理技术，可以形成简报、报告、图表等分析结果，

为全面掌握传媒思想动态，做出正确舆论引导，提供分析依据。

影视领域：拓宽创作素材，降低后期优化成本

GPT-4 可以为剧本素材的创作提供新思路，创作者可根据大众兴趣使用 GPT-4 生成内容，再进行筛选和二次加工，从而激发创作者的灵感，开拓创作思路，缩短创作周期。还可以根据长篇小说、传记、诗歌定制影视内容，从而更有可能吸引爱好者的注意力，获得更好的收视率、票房和口碑。2016 年，纽约大学利用人工智能编写剧本《Sunspring》，经拍摄制作后入围伦敦科幻电影 48 小时前十强；国内海马轻帆科技公司推出的“小说转剧本”智能写作功能，服务了包括《你好，李焕英》《流浪地球》等爆款作品在内的剧集剧本 30000 多集、电影/网络电影剧本 8000 多部、网络小说超过 500 万部。

营销领域：打造虚拟客服，提升品牌形象和服务态度

GPT-4 可以打造虚拟客服，助力产品营销，GPT-4 虚拟客服可以 24 小时不间断的提供服务，而且不会像人类客服一样产生疲劳。同时，虚拟人客服可以同时处理多个用户的问题，不会像人类客服一样需要等待处理完一个用户的问题后才能接待下一个用户。这使得虚拟客服在高峰期能够更好地应对客户的需求。此外，虚拟客服还可以自我进化和优化自己的服务，提高自己的准确性和服务质量。同时随着 GPT-4 在多模态、复杂计算、多语言理解等方面的提升，使得虚拟客服的服务水平可以为客户提供更加优质的服务。

文娱领域：加强人机互动，丰富虚拟现实趣味

GPT-4 可以讲故事、创作小说；可以生成歌词，增加音乐创作的灵感；可以解读图片，满足探索未知渴望；可以作为闲聊机器人，尤其图文交互，增加趣味性；可以作为知识检索库，增强对未知的理解，尤其在游戏方面，GPT-4 作为一种基于自然语言处理的人工智能技术，可以帮助游戏开发人员制作出更加智能的游戏。它可以在游戏中生成任务描述，并帮助玩家解决游戏中的问题。此外，GPT-4 还可以用于生成 NPC 的对话，从而让游戏的故事更加生动。

教育领域：激发教材活力，释放孩子探索天性

GPT-4 能够帮助研究人员撰写连贯、（部分）准确、信息丰富且系统的论文，写作效率极高（2-3 小时），而且不需要作者有丰富的专业知识。即使作者改变了主题，GPT-4 也能给出完全相同格式的回答。除了直接作用于文案本身，在接入 GPT-4 以后，在线教育的模式和质量也会有明显的变化。比如，在教学过程中，老师往往会面临跨学科教学，学生的一些问题也需要老师长时间思考才能解答，但如果使用 GPT-4，就可以第一时间为学生提供快速简便的问题答案；一些重复性的问题，GPT-4 就可以直接代替老师回答，让老师腾出大量时间在思想创造力上发挥更大价值；尤其是对不知道如何描述问题或者问题本身就是一些图片，GPT-4 可以从多个维度给出问答答案。

金融领域：促进风控安全，保驾金融科技稳步发展

GPT-4 在金融领域的应用范围非常广泛，包括且不限于：欺诈检测和风险管理、客户服务与支持、客户意图识别及精准营销，对金融市场的信息处理和决策辅助等。反欺诈 GPT-4 可以快速识别和分析怀疑欺诈的数据，并在接收到异常请求时进行快速反应，以确保客户在资金交易过程中的安全；金融风险控制金融机构需要不断地进行风险评估和风险控制，GPT-4 可以通过对历史数据的学习，生成预测模型，帮助机构进行风险评估和风险控制；金融舆情分析金融机构需要对市场和客户的情绪进行分析，以便做出正确的投资决策，GPT-4 可以通过对大量社交媒体和新闻数据的学习，生成情感分析模型，帮助机构了解市场和客户的情绪变化；充当研究助理 GPT-4 可以结合金融行业的信息，构建知识库问答系统。对于一些初级的资料信息和数据分析提取，使用者通过询问可以快速找到相关问题的答案。

医疗领域：提高诊断能力和效率，减负医护人员

GPT-4 的复杂名词解释、知识抽取能力、多轮对话能力、复杂问题分析、图文分析能力等，可以在医学影像诊断、临床辅助决策、精准医疗辅助、健康管理、医疗信息化、医药研发以及医疗机器人等细分场景中全面开花，以助力降本增效、提升医疗水平、改善患者体验、降低患病风险等为目的，全面赋能院前、院中、院后各个环节。哈佛大学一项研究表明：在不同严重性的 45 个真实病例中，ChatGPT 能够对 39 个病例（87%）做出正确的诊断，而同样的病例，人类医生的准确性可达 84%，其他工具准确率只有 51%，GPT-4 预计会达到更高

的诊断精确率。

二、下一个 GPT 是什么？

ChatGPT 属于 AIGC 的范畴，代表了文本生成领域的最新进展。GPT-4 则增加了图到文的生成，具备了图片识别的能力。下一个 GPT 预计将支持更多模态的 AIGC，即除文本生成之外，还可以生成图片、视频、音频等模态。与仅使用语言或仅使用视觉相比，构建良好的多模态模型更是一项极具挑战性的任务。

但如同 Transformer 架构从语言走向视觉领域，下一个 GPT 或者 LLM 模型，必将向文本、图片、音频、视频等跨领域多模态 AIGC 不断发展演进。Google、Stability AI、Open AI、百度、阿里等公司已经开始多模态内容生成模型研发，并开源部分模型，比如图片生成的 StableDiffusion、音乐生成的 MusicLM、视频生成的 Dreamix 等。

在文本生成场景中，可以分为非交互式生成和交互式生成。非交互式文本生成场景中，结构化写作其形式相对固定，生成难度较小，商业化应用较为广泛，例如公司财报简讯、新闻简讯、专业文书等，而创作型写作开放性较大，在长文本生成中难度较大，仍需技术进一步发展，相比于结构化的文本，创作型文本具有更高的开放度与自由度，需要一定的创意与个性化，如营销文案、广告语、客服话术等。交互式文本生成场景主要以聊天机器人为代表，在基于上下文中信息进行文本交互，如客服问答、闲聊机器人、知识检索库等。

在图像生成场景中，模型结构不断进化提高了 AIGC 生产图像的

多样性，但要求较高的功能实现还有待于技术的进一步提升。“图像编辑”难度低于“图像生成”与“2D-3D”转换，目前已存在多款产品支持“图像编辑”，而对于“图像生成”任务，由于图片相较包含更多元素，其生成效果仍存在不稳定性，对于要求较高的功能类图像生成仍需要技术上的提升。

音频生成场景中，文本到语音任务已比较成熟，语音质量已达到自然的标准，未来将朝更富情感、富韵律的语音合成以及小样本语音学习方向发展。但音乐生成任务中仍需解决音乐数据难以标注的问题，数据标注的颗粒度大小影响音乐生成任务的可控性。若可控性得以解决，则可指定风格、情绪等因素的音乐生成任务有希望在影视、游戏等场景下得到大量应用。

视频生成场景中，视频生成本质上与图片生成类似，通过对视频进行帧数级别的切割，实现对每一帧的处理。视频生成过程包括三个阶段：数据的提取、训练和转换，当前技术正在着重提升视频修改精准度和实时性两个维度。鉴于视频本身的文本、图像和音频的综合属性，视频生成也是跨模态生成领域的重要应用场景。

跨模态生成场景中，现实世界的信息是文本、音频、视觉、传感器以及人类各种触觉的综合体系，要更为精准地模拟现实世界，就需要将各种模态能力之间打通，例如文字-图像、文字-视频等跨模态生成能力大型预训练模型的发展使得跨模态逐步成熟，“文本-图像”生成正在快速落地，“文字-视频”的实验效果也已较为理想（视频时长、清晰程度、逻辑等还有较大提升空间）。

AIGC 发展趋势

(一) 消费端：AIGC 牵引数字内容领域的全新变革

(二) 产业端：合成数据牵引人工智能的未来

消费端:AIGC 牵引数字内容领域的全新变革

一、AIGC 有望塑造数字内容生产与交互新范式，

成为未来互联网的内容生产基础设施

内容生产，特别是创意工作，一向被认为是人类的专属和智能的体现。如今，AI 正大步迈入数字内容生产领域，不仅在写作、绘画、作曲等多个领域达到“类人”表现，更展示出在大数据学习基础上的非凡创意潜能。这将塑造数字内容生产的人机协作新范式，也让内容创作者和更多普通人得以跨越“技法”和“效能”限制，尽情挥洒内容创意。

数字内容迈入强需求、视频化、拼创意的螺旋式升级周期，AIGC 在这阶段的发展迎合了这一需要。从全球来看，人们在线的时间持续增长，在线新闻、音乐、动漫、影视、文学、游戏的市场规模仍在上升。同时，数字内容消费的结构也正转型升级，从以图文为主过度到视频为主，视频在网络流量中的占比不断上升。短视频和直播的流行，在生产侧，让原本需要长制作周期的视频变成了可以源源不断产出的“工业品”；在消费侧，过往需要高注意投入、反复观看的视频内容变成了一种媒体“货架”上的“快消品”。最后，创意仍是核心，仍旧稀缺，需要新的模式辅助创作者持续产生、迭代和验证创意。种种因素，都需要更加低成本、高效能的数字内容生产新方式。

AIGC 正在越来越多地参与数字内容的创意性生成工作，以人机协同的方式释放价值，成为未来互联网的内容生产基础设施。

从范围上看，AIGC 逐步深度融入到文字、音乐、图片、视频、3D 多种媒介形态的生产中，可以担任新闻、论文、小说写手，音乐作曲和编曲者，多样化风格的画手，长短视频的剪辑者和后期处理工程师，3D 建模师等多样化的助手角色，在人类的指导下完成指定主题内容的创作、编辑和风格迁移。

从效果上看，AIGC 在基于自然语言的文本、语音和图片生成领域初步令人满意，特别是知识类中短文，插画等高度风格化的图片创作，创作效果可以与有中级经验的创作者相匹敌；在视频和 3D 等媒介复杂度高的领域处于探索阶段，但成长很快。尽管 AIGC 对极端案例的处理、细节把控、成品准确率等方面仍有许多进步空间，但蕴含的潜力令人期待。

从方式上看，AIGC 的多模态加工是热点。吴恩达(AndrewNg)认为多模态是 2021 年 AI 的最重要趋势，AI 模型在发现文本与图像间关系中取得了进步，如 OPENAI 的 CLIP 能匹配图像和文本，Dall·E 生成与输入文本对应的图像；DeepMind 的 PerceiverIO 可以对文本、图像、视频和点云进行分类。典型应用包括如文本转换语音、文本生成图片，广义来看 AI 翻译、图片风格化也可以看作是两个不同“模态”间的映射。

	PRE-2020	2020	2022	2023?	2025?	2030?
TEXT	Spam detection Translation Basic Q&A	Basic copy writing First drafts	Longer form Second drafts	Vertical fine tuning gets good (scientific papers, etc)	Final drafts better than the human average	Final drafts better than professional writers
CODE	1-line auto-completes	Multi-line generation	Longer form Better accuracy	More languages More verticals	Text to product (draft)	Text to product (final), better than full-time developers
IMAGES			Art Logos Photography	Mock-ups (product design, architecture, etc.)	Final drafts (product design, architecture, etc.)	Final drafts better than professional artists, designers, photographers
VIDEO / 3D / GAMING			First attempts at 3D/video models	Basic / first draft videos and 3D files	Second drafts	AI Roblox Video games and movies are personalized dreams

Large model availability: ● First attempts ● Almost there ● Ready for prime time

AIGC 在新研究范式及算力进步下加速发展，特别是自然语言理解能力大幅进化，典型代表是 GPT 系列模型。自然语言是不同数字内容类型间转化的根信息和纽带，比如“猫”这个词语就是加菲猫的照片，音乐剧《猫》和无数内容的索引，这些不同的内容类型可以称为“多模态”。创作者以自然语言描述要素、特征后，AI 就能生成对应的不同模态的结果。2017 年 Google 发布的 Transformer 开启了大型语言模型(Large language model, 简称 LLM)时代。它作为更强大的特征提取器，生成的语言模型质量高、效率高，被应用到后续的 GPT、BERT 等各种预训练语言模型中，以大数据预训练+小数据微调的方式，摆脱了对大量人工调参的依赖，在手写、语音和图像识别、语言理解方面的表现大幅突破，所生成的内容也越来越准确和自然。

但大模型的高算力要求对普通用户不友好，这使 AIGC 长期只是小圈子的爱好，例如 GPT-3 有 1750 亿参数量，既需要大算力集群也不开放。2022 年，先是部署在论坛 discord 上、以聊天机器人形式向用户开放的 midjourney 引起关注，一位设计师用其生成的图片甚至在线下比赛中获奖。2022 年，StableDiffusion 等文本生成图片产品，第

一次让大众感受 AIGC 的魅力。开源的 StableDiffusion 仅需一台电脑就能运行，截至 10 月已有超过 20 万开发者下载，累计日活用户超过 1000 万;而面向消费者的 DreamStudio 则已获得了超过 150 万用户，生成超过 1.7 亿图片。



图:游戏设计师杰森·艾伦使用 Midjourney 模型生成的《太空歌剧院》是首个获奖的 AI 生成艺术

在 AIGC 图像生成火爆的同时，ChatGPT 横空出世，真正做到和人类“对答如流”，能理解各式各样的需求，做出回答、短文和诗歌创作、代码写作、数学和逻辑计算等。ChatGPT 可以看作 GPT3.5 版本，使用了人类反馈强化学习(RLHF)技术，能持续学习人类对回答的建议和评价，朝更加正确的方向前进。ChatGPT 基于这一思想，用不到 GPT3 的 1%的参数实现了极佳的效果。但是也依然存在一些缺陷，例如引用不存在的论文和书籍，误解和扭曲直观的自然科学知识和历史事实。

在 AIGC 技术开源的背景下，AIGC 图片生成也促进了小众文化的

发展。早在 2022 年五月，QQ 影像中心推出的国内第一款基于 AIGC 的“AI 恋爱专属画”520 活动就在 QQ 小世界走红，最近新推出的“异次元的我”二次元形象生成器火遍国内外。

QQ 影像中心将输入模型的提示词分为两个部分——图片主体内容描述和风格描述。这样做是为了避免其他基于 Stable Diffusion 的 AIGC 图像生成平台常见的问题，如不便捷，用户想要生成比较好的结果，需要对图片内容进行详尽的描述，并增加很多提升效果和强化风格的词，才能得到一个比较理想的效果。

(1)内容描述:为了得到图片的内容描述，QQ 小世界对输入图中的信息进行人物性别、年龄、姿势等属性的检测，得到最终尽可能准确详尽的内容描述词。

(2)风格描述:为了让结果更接近二次元，QQ 小世界为各种不同场景设计了非常鲜明的二次元风格。根据第一步对图片内容进行分析后，进一步细分为单人男性、单人女性、多人等不同场景。

结合内容描述和风格描述，QQ 小世界通过对原图的分析得到最终完整的提示词输入网络，从而能够稳定生成画面精致、风格突出的结果。既保证了生成内容和原图的对应性，又增添了很多“异次元”要素，让用户更有穿越到二次元的感觉，同时也保证了较高的生成质量。

二、AIGC 的应用生态和内容消费市场逐渐繁荣

AIGC 在学习通用知识和理解泛化上有更好的表现，这是之前的

生成模型没有体现出来的能力。这也决定了 AIGC 不光是生成，而是建立在认知与理解上的模型的应用生态。2020 年，OpenAI 推出 1750 亿参数的预训练语言模型 GPT-3，在国内外掀起千亿参数预训练模型的研究浪潮。实际上，那时就出现了专门做文字生成的商业公司，如 Jasper.ai 与 Copy.ai。而今在内容消费领域，AIGC 已经重构了整个应用生态。AIGC 在内容生成领域有以下优势特征：

(1)自动内容生成:大型语言和图像 AI 模型可用于自动生成内容，例如文章、博客或社交媒体帖子。对于定期创建内容的企业 and 专业人士来说，这可能是一种宝贵的节省时间的工具。2022 年初， Disco Diffusion 给行业带来很大的冲击，其在生成速度、效果，还有词条内容的丰富性等方面都有了显著的改进。架设在社交平台 Discord 上的 Midjourney，因其使用便捷与相对较好的效果，在社会层面得到了很大的关注。

(2)提高内容质量:AIGC 生成的内容可能比人类创建的内容质量更高，因为人工智能模型能够从大量数据中学习并识别人类可能看不到的模式。这可以产生更准确和信息更丰富的内容。DALL·E2 和 Google 的 Imagen 都可以通过文字来要求 AI 画出特别具体的内容，而且效果已经接近于中等画师的水平。今年 8 月，Stable Diffusion 的开源模型生成的面部和肢体相较于 Disco Diffusion 自然了许多。

(3)增加内容多样性:AIGC 模型可以生成多种类型的内容，包括文本、图像和音视频、3D 内容等等。这可以帮助企业 and 专业人士创建更多样化、更有趣的内容，从而吸引更广泛的人群。2021 年底，

VQGAN+CLIP (Vector Quantised General Adversarial Network+Contrastive Language-Image Pre-training)技术串联,可以生成抽象绘画作品。2022年9月29日,Meta AI公布了一款基于文本生成视频短片的系统“Make-A-Video”,不仅能够通过提示词生成几秒连贯的视频,并且有“超现实的”、“现实的”、“风格化”的视频变种供调整,而且也能从一两张图片生成视频,或者根据一小段视频片段派生不同风格、拍摄角度、相似动作的视频变种。

(4)内容制作成本低、门槛低:基于AIGC,内容制作的成本显著降低、效率显著提高,可以实现以十分之一的成本,以百倍千倍的生产速度,创造出有独特价值和独立视角的内容。07而且AIGC也降低了用户的内容制作门槛,即使是毫无绘画基础的用户也能借助AIGC工具,画出高质量的作品。

(5)可实现个性化内容生成:人工智能模型可以根据个人用户的喜好生成个性化内容。这可以帮助企业和专业人士创建目标受众更有可能感兴趣的内容,因此更有可能被阅读或分享。比如,基于Stable Diffusion的二次元画风生成工具Novel-AI,可以满足小众的二次元群体的喜好和内容需求,一定程度上也促进了小众文化的发展。

AIGC已经掀起了一场内容生产的革命。在内容需求旺盛的当下,AIGC所带来的内容生产方式变革也开始引起了内容消费模式的变化。AI绘画是AIGC的一大重要分支。它可以提高美术素材生产效率,在游戏、数字藏品领域初步得以应用。文字生成图片(Text-to-Image, T2I)是目前AI绘画的主流生成方式,用户输入文本命令便可生成相

应效果图。对于游戏开发者而言，T2I 工具在概念构思阶段可迅速提供多种创作方向，并降低初期投入成本，在开发过程中亦可批量生产石头花纹、花草树木等通用型贴图素材。艺术是另一应用领域，全球最大的 AI 生成艺术画廊 ArtAI 以历史上大量艺术品的时期、流派、风格等为灵感生成数字藏品，单字仅能使用一次的机制保障每份数字藏品对应不同的文本内容，从而保障其唯一性。2022 年 10 月，海外知名图库 Shutterstock 就公布说他们与 OpenAI 建立了合作，让用户可以输入文字即时生成满足需求的原创图片。Adobe、Getty Images 也将图像生成模型集成到他们自己的产品和服务中。这些动作不仅是一个传统企业的及时反应，实际上还意味着 AIGC 商业化变现的一个想象似乎开始落地:打造一个基于生成的全新内容平台。

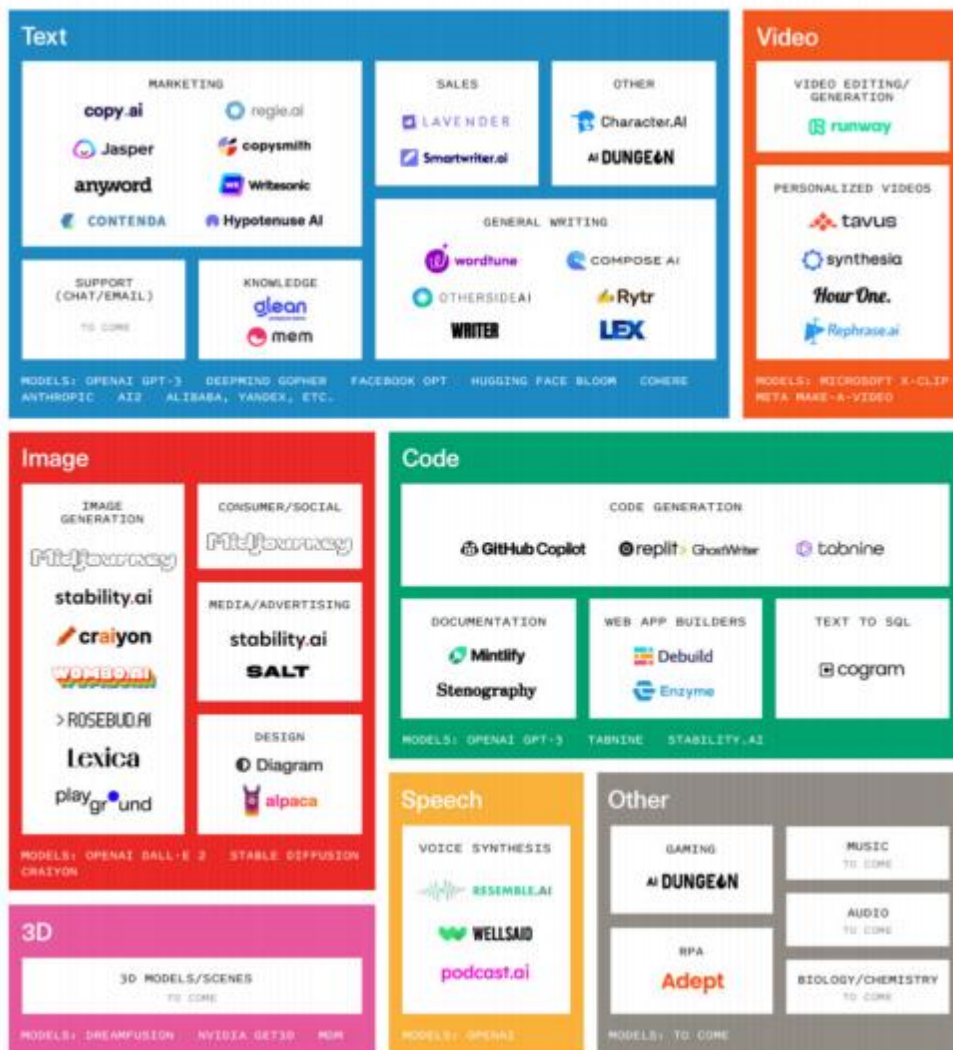


图:AIGC 应用现状概览(引用自红杉资本)

AIGC 作为当前新型的内容生产方式，已经率先在传媒、电商、影视、娱乐等数字化程度高、内容需求丰富的行业取得重大创新发展，市场潜力逐渐显现。2022 年，AIGC 发展速度惊人，迭代速度呈现指数级爆发，谷歌、Meta、百度等平台型巨头持续布局，也有 StabilityAI、JasperAI 等独角兽创业公司出世。AI 绘画已经能承担图像内容生成的辅助性工作，前期初稿的形成可以由 AI 提供，后期再由创作者进行调整，从而提高内容产出效率。根据 Gartner 预测，未来五年 10%-30%

的图片内容由 AI 参与生成，有望创造超过 600 亿以上市场空间，若考虑到下一代互联网对内容需求的迅速提升，国外商业咨询机构 AcumenResearchandConsulting 预测，2030 年 AIGC 市场规模将达到 1100 亿美元⁰⁹。

三、AIGC 将日益成为未来 3D 互联网的基础支撑

互联网向下一代技术升级和演进的重要方向是从“在线”走向“在场”，迈向 3D 互联网时代，AIGC 将成为打造虚实集成世界的基石。人们将在在虚拟空间构建仿真世界，在现实世界“叠加”虚拟增强，实现真正的临场感。随着 XR、游戏引擎、云游戏等等各种交互、仿真、传输技术的突破，信息传输越来越接近无损，数字仿真能力真假难辨，人类的交互和体验将到达新阶段。

AIGC 为 3D 互联网带来的价值，既包括 3D 模型、场景、角色制作能效的提升，也能像 AI 作画那样，为创作者激发新的灵感。传统的 3D 制作需要耗费大量时间和人力成本。以 2018 年发售的游戏《荒野大镖客 2》为例，为了打造约 60 平方公里的虚拟场景、先后有六百余名美术历经 8 年完成。

程序化内容生成技术(PCG, ProceduralContentGeneration)是一种在效能和可控度上介于纯手工和 AIGC 之间的产出方式。2022 年的交互内容《黑客帝国:觉醒》中就采用了这项技术，在最新引擎技术的加持下，虚拟城市共包括 700 万个实例化资产，每个资产由数百万个多边形组成，包括 7000 栋建筑、38000 辆可驾驶的车和超过 260 公

里的道路。



图:Epic 使用虚幻 5 引擎和程序化生成技术高效制作《黑客帝国:觉醒》中的庞大城市

2016 年发布、主打宇宙探险的独立游戏《无人深空》(NoMan, sSky) 声称创造了 1840 亿颗不同的星球，每个星球都有形态各异的环境和生物。游戏正是用 PCG 构造了一系列生成规则和参数，用数学函数和极少的数据量生成了这个看似无限的游戏宇宙。

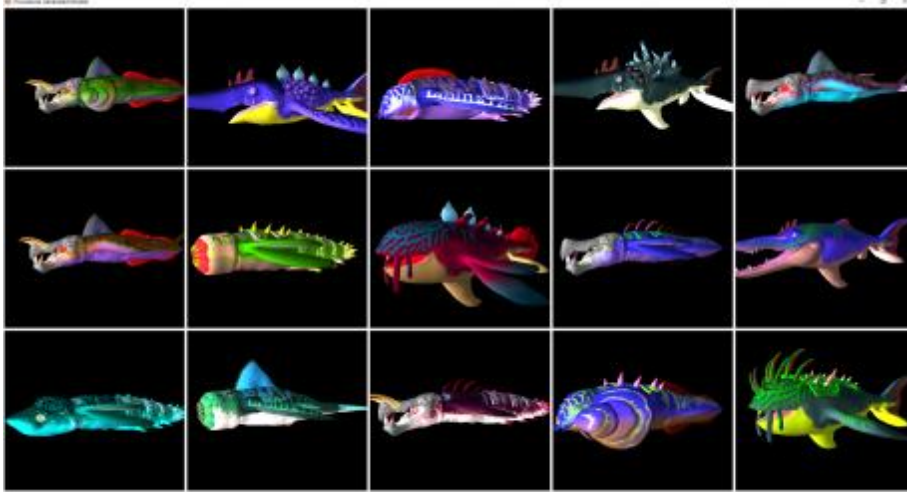


图:游戏《无人深空》中自动化生成的海洋生物示例

尽管如此，距离“复刻”真实世界和打造千千万万个幻想世界，这样的速度还是太慢了。为什么在文字、图片领域大放异彩的 AI，没有继续快速“征服”3D 领域呢？

AI 直接生成 3D 内容，比 2D 图片难度大得多。首先，信息量不同，一张图片和一个 3D 模型相比，本身蕴含的信息量就有数量级的差距；其次，图片和 3D 的实质不同，如果说 2D 是像素点阵在显示器的客观陈列，3D 则是实时、快速、海量的矩阵运算，就像对着模型在 1 秒内进行几十次“拍照”。为了准确计算得到每个像素点，“渲染”在显示器上，需要考虑的因素至少有：(1)模型几何特征，通常用几千上万个三角面来表示；(2)材质特征，模型本身的颜色，是强反射的金属，还是漫反射的布料；(3)光线，光源是点状的吗，颜色和强度如何。最后，原生 3D 模型的数据相对较少，远不如已存在了数千年、可以以非数字化形态存在的图像那么多，例如 ImageNet 中就包含了超过 1400 万张图片。

现阶段，AIGC 在 3D 领域的突破，主要是从图片、现实中存在的

物体取材后，再由 AI 介入进行转化。计算机视觉技术的进步，为 3D 另辟蹊径：我们可以通过图片来推测物体在 3D 空间中的形态，尽可能还原每个像素点原本的深度信息。采用激光雷达等设备扫描物理世界，能快速得到大量 3D 模型。当然，这些素材要经过简化处理才能更好使用。因为它们携带了海量的坐标点和颜色信息，呈现起来耗费的计算资源多。想要应用在影视、游戏、元宇宙等场景中，和运行在电脑、手机等设备中，需要经过一系列的后处理，比如让模型的几何表示更简洁，还原材质信息、尽可能去除拍摄时的光照影响等。处理后的模型，也会具有更强的可编辑性，比如更改它原本的形状、颜色等。

目前 AIGC 在 3D 模型领域还处于探索阶段，从基本原理和使用的模型来看，仍然以扩散模型为基础，基于文字-图片-3D 路径进行处理。谷歌和英伟达在这一领域较为领先，先后在 2022 年发布了自己的文字生成 3D 的 AI 模型。¹⁰但从生成效果看，距离现在人工制作 3D 内容的平均质量还有距离；生成速度也未能尽如人意。

10 月，谷歌率先发布了 DreamFusion，它能够使用文本到图像模型先生成 2D 图像，然后优化为体积 NeRF(神经辐射场)数据。但其缺点也很显著，首先扩散模型仅对 64x64 的图像生效，导致生成 3D 的质量不高；其次场景渲染模型使基于 Mip-NeRF360 的大型 MLP，不仅需要海量样本，也在计算上费时费力，导致生成速度较慢。

随后，英伟达发布了 Magic3D，同样是从文字描述中生成 3D 模型的 AI 模型。在输入诸如“一只坐在睡莲上的蓝色毒镖蛙”这样的提示后，Magic3D 在大约 40 分钟内生成了一个 3D 网格模型，并配

有彩色纹理。Magic3D 的转换分为两步，先将低分辨率生成粗略模型，再优化为高分辨率的精细模型，而且使用 InstantNGP 的哈希特征编码，高分辨率图像特征表示的计算成本大大降低，因此比谷歌 DreamFusion 的 3D 生成速度更快、效果更好。同时，与文本到图片的扩散模型类似，Magic3D 可以在几代生成图像中均保留相同的主题（一般称为“一致性”），并将 2D 图像(如立体派绘画)的风格应用于 3D 模型。研究人员表示，模型完善后有望使 3D 合成技术进一步普及，并激发每个人在 3D 内容上的创造力。所产生的技术可以加速游戏和 VR 应用的开发，可能最终会在电影和电视的特效制作中实现落地应用。

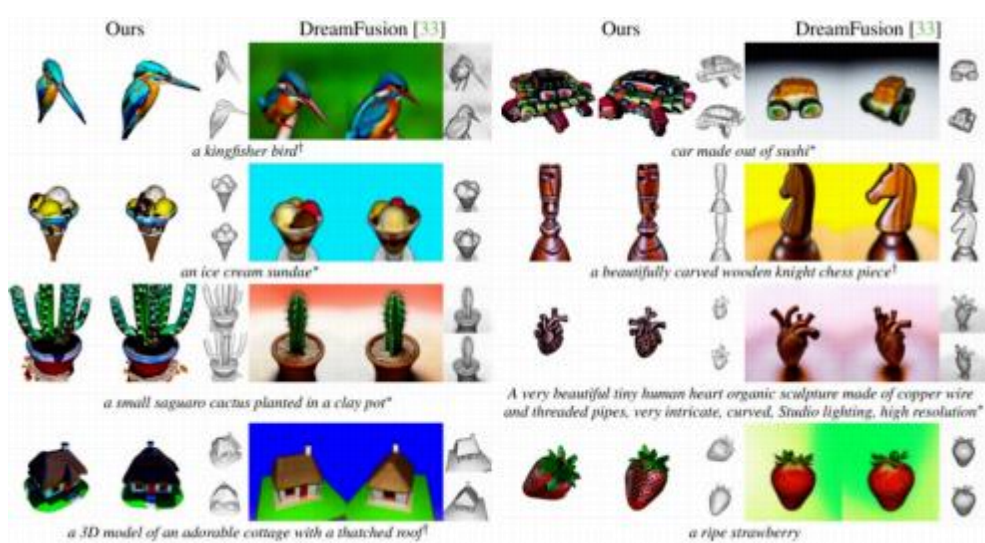


图:Magic3D(第 1、3 列)与 DreamFusion(第 2、4 列)对比

借助 AI 从图片直接生成 3D 也是一条切实可行的路径。英伟达在 12 月的 NeurIPS 上展示了 AIGC 模型 GET3D (Generate Explicit Textured 3D 的缩写)，具备生成具有显示纹理的 3D 网格的能力，可

根据其所训练的建筑物、汽车、动物等 2D 图像类别，即时合成 3D 几何体。AIGC 生成的物体具有高保真的纹理和复杂的几何细节，并且以常用图形软件应用中所使用的三角网格格式创建，这使得用户可以十分轻松地将这些几何体导入 3D 渲染器和游戏引擎，以进行后续编辑。可用于构建为游戏、机器人、建筑、社交媒体等行业设计的数字空间，比如建筑物、户外空间或整座城市的 3D 表达。该模型是在 NVIDIA A100 Tensor Core GPU 上使用从不同相机角度拍摄的约 100 万张 3D 几何体的 2D 图像训练而成，每秒可生成约 20 个物体。

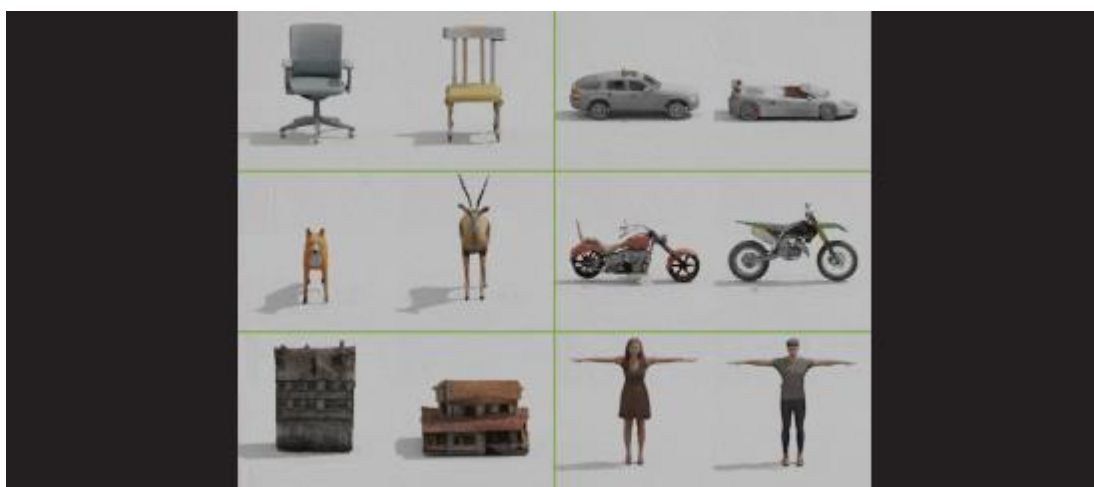


图:NVIDIA GET3D 基于 AI 生成的具有光影效果的 3D 物体模型

结合团队的另一项技术，AI 生成的模型的可编辑性会大幅加强。6 月发布的成果显示，在分离 3D 物体中的材质和光照方面取得了重大进展，这反过来又提高了 AI 生成几何体的可编辑能力，创造者们能够轻松地对场景中移动的物体替换材质或者调整光照。

四、聊天机器人和数字人不断拓展应用领域

(1)聊天机器人:2022 年 12 月 1 日,美国人工智能研究公司 OpenAI 研发的 AI 聊天机器人产品 ChatGPT 正式对外开放,一经推出便火遍全网,截至 12 月 12 日已拥有超过 100 万名用户。而 ChatGPT 正是典型的文本生成式 AIGC。

ChatGPT 采用 Web 浏览器上的对话形式交互,不仅能够满足与人类进行对话的基本功能,能够回答后续问题、承认错误、质疑不正确的前提和拒绝不适当的请求。而且,根据各领域的用户在使用后反馈,ChatGPT 还可以驾驭各种风格和文体,且代码编辑能力、基础脑力工作处理能力等一系列常见文字输出任务的完成程度也大大超出预期。比如根据关键词或提问,生成剧本、发言稿等多种文书。因此,业内人士宣称 ChatGPT 已经大幅超越过去的 AI 问答系统。

未来随着性能的进一步提升,对话式 AIGC 在搜索、知识传播等领域有很大的应用空间。正如 OpenAI 的 CEO Sam Altman 在 Twitter 上说过,AIGC 的最终目标是做一个类似于新时代的搜索引擎。目前,从 ChatGPT 展示出来的内容输出质量和内容覆盖多维度,已经可以直面“搜索引擎”与“问答社区”。根据《纽约时报》报道,谷歌及其母公司 Alphabet 的首席执行官 Sundar Pichai 参加了几次围绕谷歌人工智能战略的会议,并指示公司的许多团队重新集中精力解决 ChatGPT 对其搜索引擎业务构成的威胁。近日,谷歌并为 ChatGPT 带来的威胁发布“红色警报”,着手进行紧急应对。13



图:有网友将 Google 和 ChatGPT 进行了对比, 分别在 Google 搜索和 ChatGPT 中提出了同一个问题

(2)数字人:数字人、虚拟机器人是数字智能体(digitalagent), 作为新的交互形式, 目前已有很多应用, 包括元宇宙应用中的 NPC 虚拟角色、用户虚拟替身/虚拟形象(avatar), 主要分为真人驱动和 AI 驱动两种。这里主要阐述由 AI 驱动的数字人。



图:应用了 AIGC 的腾讯 2D 数字人自动化管线生产流程(引用自腾讯研究院《数字人产业发展趋势报告》)

首先, AIGC 大大提升了数字人的制作效能。用户可以上传照片/视频, 通过 AIGC 生成写实类型的数字人, 具有生成过程较短、成本低、可定制等特点。基于 AIGC 的 3D 数字人建模已经初步实现产品化, 目前精度可以达到次世代游戏人物级别, 优势是可以开放程序接口, 对接各种应用, 因此潜力较大, 范围较广, 特别是 C 端的应用, 可以通过便捷化的上传图片即可生成 3D 数字人面部模型。未来发展方向是通过算法驱动提高精度, 优化建模效果, 比如:偏移矫正、阴影修复等。在用户 avatar 生成、创建方面, AI 引擎可以分析用户的 2D 图片或 3D 扫描, 然后形成高度逼真的仿真渲染, 同时结合脸部表情、情绪、发型、年龄特征等因素让用户的虚拟形象更具活力。目前, Meta、英伟达等众多科技公司已经在利用 AI 技术帮助用户在虚拟世界打造虚拟化身, 例如英伟达的 omniverseavatar 可以生成、模拟、渲染可互动的虚拟形象。

其次, AIGC 支撑了 AI 驱动数字人多模态交互中的识别感知和分析决策功能, 使其更神似人。自然语言处理是数字人的大脑, 直接影响交互体验, 而计算机视觉决定了数字人面部表情和肢体动作。目前主流的方式是围绕 NLP 能力通过文本驱动, 本质是通过 ASR-NLP-TTS 等 AI 技术进行感知-决策-表达的闭环来驱动数字人交互。计算机视觉(CV)目前数字人声唇同步技术相对完善, 在游戏中已

经大量应用。

数字人进化的趋势，一是融入会话式 AI 系统(Conversational AI)，给传统的 Siri 等虚拟助手、智能客服等聊天机器人以一个具象化、有亲和力的人类形象，提升交流中情感的连接，有望给这一领域带来更大的市场前景。据机构统计，2021 年会话式 AI 的全球市场规模为 68 亿美元，预计到 2026 年将增长到 184 亿美元。随着线上空间日益丰富，更多普通用户也希望拥有自己的个性化虚拟形象，因此，数字人进化的第二个方向是制作工具更丰富、更易用。例如 Epic 在虚幻引擎中集成的虚拟人工具 Metahuman，用户可以在系统提供的基础形象模板上修改参数，仅用 30 分钟就能“捏”成独一无二的形象。可调节内容既包括整体的肤色、身材，也包括细节的面庞轮廓、五官大小等。未来，对话式 AI 系统、先进的实时图形处理等技术的结合，将使得数字人、虚拟助手、虚拟伴侣、NPC 等数字智能体(digitalagent)能够逼真地模仿人类的音容笑貌，变得更加智能化、人性化。这将带来更复杂的、自然交互的 AI 虚拟角色，除了模仿人类的语言表达，还具有表情、肢体语言、情绪甚至物理交互等能力，给用户提供更直观的、更具沉浸感的数字化体验。可以说，数字人等新型 AI 角色将决定 VR/AR、元宇宙等未来互联网应用的体验质量和吸引力。

五、AIGC 将作为生产力工具来推动元宇宙发展

元宇宙最令人兴奋的特征之一是它的可扩展性潜力，它将极大扩展人类在虚拟世界的生存空间。但是问题在于，谁来建造这个巨大的

世界?它将如何填充内容?如今, 创建高保真 3D 模型或虚拟人通常需要大量的计算机科学家和数字艺术家。面对这个空间所需要的大量的数字 内容, 单靠人工来设计和开发根本无法满足需求, 而 AIGC 将是新的元宇宙内容生成解决方案。近期, LinkedIn 做了一个调查, 向 LinkedIn 用户征集他们最希望了解哪个元宇宙相关的技术主题, 76% 的受访者选择了 AIGC。¹⁴ 只有通过 AIGC, 元宇宙才可能以低成本、高效率的方式满足海量用户的不同内容需求。



图:扎克伯格在元宇宙的岛屿上, 可以通过发出语音命令变换天气、生成创建海滩等不同场景

首先, AIGC 为构建沉浸式的元宇宙空间环境提供了核心基础设施技术, 将成为元宇宙的生产力工具。元宇宙在理论上可以用于工作、会议、游戏以及生活社交等各种人类活动, 因此需要在元宇宙空间中创建各种各样的活动对象、建筑和活动环境。但是, 在过去, 为了构建这些数字环境, 需要开发团队半手工地创建每一个部分, 从山丘到

大海，用鼠标拖动来放置树木或家具，而现在 AIGC 已经可以实现创建逼真的 3D 虚拟空间环境、虚拟人物，并且效率和成本可以满足大规模的元宇宙空间环境创建。

在元宇宙发展的过程中，游戏是虚拟世界重要的载体，将需要大量数字原生的场景。AIGC 可以极大提高数字原生内容的开发效率，降低游戏开发成本。具体来说，剧情、角色、头像、道具、场景、配音、动作、特效、主程序未来都可以通过 AIGC 生成。通过 AIGC 加速复刻物理世界、进行无限内容创作，从而实现自发有机生长。例如，当用户朝任何方向行走时，游戏 Minecraft 的世界地图会通过 AIGC 生成模型无限扩展。NVIDIA 也开发了 AIGC 生成模型，仅通过将 2D 视频输入来创建语义标记和可修改的 3D 环境。正如 Unity 人工智能高级副总裁 DannyLange 所言：“市政府想要通过城市的数字孪生来做一些规划。而城市数字孪生很难聘请大量的技术开发人员来编写代码。而此时就是 AIGC 的用武之地。”

而且 AIGC 能够释放开发人员的生产力，帮助开发人员编写代码，使开发人员可以花更少的时间编写代码，而花更多的时间来表达需求，最终实现生产力的大幅提升，以及元宇宙空间环境的建成速度。

其次，AIGC 将作为生产力工具为元宇宙用户提供个性化内容体验。大量个性化、多元化的数字内容是吸引人们进驻元宇宙的重要原因，而这些内容如果仅通过人工方式来生产既成本高又耗时长，肯定远远无法满足用户的海量需求，因此，需要 AIGC 作为生产力工具来满足和填充 Metaverse 等内容匮乏的虚拟世界。NFT 便是 AIGC 的一

个很好的例子。Bored Ape Yacht Club 和 Crypto Punks 等 NFT 项目开始于创作 10,000 件独特的艺术品。如果没有 AIGC 的支持，这项工作将需要一个数字艺术家团队数周才能完成，然而这两个项目都是由两个软件开发人员完成的。

同时，AIGC 将赋予用户更多的创作权力和自由，促进创新并提升元宇宙的用户体验。比如用户可以使用手机拍摄的一系列照片，通过 AIGC 工具生成一个可以使用的 3D 渲染图。采用这种创造内容的方式，我们可以想象未来的元宇宙体验将不再完全由开发人员构建，而是利用 AIGC 响应用户的输入按需生成。

最后，AIGC 也将作为用户交互界面的一部分在元宇宙中发挥作用。元宇宙中不仅有用户，还应该存在很多类似 NPC 的智能体为用户提供交互服务，这些智能体可以由 AIGC 生成并驱动。ChatGPT 便可以作为语言模型中集成到各种智能体中充当元宇宙中高度逼真的虚拟助手或伙伴，执行“智能”动作和更为复杂的任务。例如，Meta 已经提出了创建智能助手的计划，用户可以向助手发出简单的语音指令继而创建他们想要的虚拟现实设置类型。

产业端:合成数据牵引人工智能的未来

随着 AIGC 技术持续创新发展，基于 AIGC 算法模型创建、生成合成数据(synthetic data)迎来重大进展，有望解决 AI 发展应用过程中的数据限制，进一步推动 AI 技术更广泛的应用。因此，业界非常看好合成数据的发展前景及其对人工智能未来发展的巨大价值。Forrester、埃森哲(Accenture)、Gartner、CBInsights 等研究咨询公司都将合成数据列为人工智能未来发展的核心要素，认为合成数据对于人工智能的未来而言是“必选项”和“必需品”。例如，Forrester 将合成数据和强化学习、Transformer 网络、联邦学习、因果推理视为实现人工智能 2.0 的五项关键技术进展，可以解决人工智能 1.0 所面临的一些限制和挑战，诸如数据、准确性、速度、安全性、可扩展性等。Gartner 预测称，到 2030 年合成数据将彻底取代真实数据，成为 AI 模型所使用的数据的主要来源。MIT 科技评论将 AI 合成数据列为 2022 年十大突破性技术之一，称其有望解决 AI 领域的数据鸿沟问题。数据是人工智能的燃料和驱动力，合成数据将极大拓展人工智能发展应用的数据基础，可以认为，合成数据关乎人工智能的未来。

在概念上，合成数据是计算机模拟(computer simulation)技术或算法创建、生成的自标注(annotated)信息，可以在数学上或统计学上反映真实世界数据的属性，因此可以作为真实世界数据的替代品，来训练、测试、验证 AI 模型。简而言之，合成数据是在数字世界中创造的，而非从现实世界收集或测量而来。²² 合成数据拥有很长的历史，

在其发展过程中技术不断创新。例如，游戏引擎、3D 图形等模拟技术(simulation technology)可以创建高保真的仿真物体和仿真环境，而结合了 AI 技术的 3D-AI 技术则可以极大提升自动化生产 3D 内容的效率和保真度(fidelity)。得益于生成对抗网络(GAN)、变分自编码器(VAE)、Transformer 模型、扩散模型(DiffusionModel)、神经辐射场模型(NeRF)等不断涌现的 AI 算法，不仅合成数据的种类得到了扩展，而且其质量也不断得到提升。就目前而言，合成数据大致可分为三类：表格数据/结构化数据，图像、视频、语音等媒体数据，以及文本数据。²³ 这几类合成数据在多个领域都有应用。AIGC 技术的持续创新，让合成数据迎来新的发展契机，开始迸发出更大的产业发展和商业应用活力。

一、合成数据为 AI 模型训练开发提供强大助推器，推动实现 AI2.0

人工智能的发展应用离不开数据，但真实世界数据面临着难以获取、质量差、标准不统一等诸多问题。为此，计算机模拟技术或算法生成的合成数据，作为真实数据数据的廉价替代品，日益被用于创造精准的 AI 模型。合成数据服务商 AIReverie 指出，人工标注一张图片可能需要 6 美元，但人工合成的话只需要 6 美分。2019 年的一篇论文《合成数据用于深度学习》(synthetic data for deep learning)认为，合成数据是现代深度学习领域冉冉升起的最具前景的通用技术之一，尤其对于依赖于图像、视频等非结构化数据的计算机视觉技术而言；

并认为合成数据对于人工智能的未来发展至关重要。²⁴ 而且，研究表明在 AI 模型的训练开发上，合成数据相比基于真实物体、事件或人物的数据，可以发挥同样好甚至更好的效果。²⁵ 总之，合成数据技术可以实现更廉价、更高效地批量生产制造 AI 模型训练开发所需的海量数据(诸如训练数据、测试数据、验证数据等等)，作为对真实数据的替代或补充，将推动人工智能迈向 2.0 阶段，从本质上扩展 AI 的应用可能性。

可以说，目前人工智能仍处在 1.0 阶段(AI1.0)，数据是最大掣肘，业界利用真实世界数据训练 AI 模型面临多方面问题:数据采集、标注费时费力、成本高企;数据质量较难保障;数据多样化不足，难以覆盖长尾、边缘案例，或者特定数据在现实世界中难以采集、不方便获取;数据获取与使用、分享等面临隐私保护挑战和法规限制，等等。这些数据方面的限制在很大程度上阻碍了人工智能更广泛的应用和部署。

合成数据有望解决这些问题，推动人工智能迈向 2.0 阶段(AI2.0)，可以在更大程度上拓展人工智能的应用。具体而言，对于人工智能而言，合成数据可以发挥诸多价值:(1)实现数据增强和数据模拟，解决数据匮乏、数据质量等问题，包括通过合成数据来改善基准测试数据(benchmark data)的质量等;(2)避免数据隐私/安全/保密问题，利用合成数据训练 AI 模型可以避免用户隐私问题，这对于金融、医疗等领域而言尤其具有意义;(3)确保数据多样性，更多反映真实世界，提升 AI 的公平性，以及纠正历史数据中的偏见，消除算法歧视;(4)应对长尾、边缘案例，提高 AI 的准确性、可靠性，因为通过合成数据可以自动

创建、生成现实世界中难以或者无法采集的数据场景，更好确保 AI 模型的准确性;(5)提升 AI 模型训练速度和效果。总之，利用合成数据可以更廉价、更高效、更准确、更安全可靠地训练 AI 模型，进而极大扩展 AI 的应用可能性，将人工智能推向新的发展阶段。

二、合成数据助力破解 AI “深水区” 的数据难题，持续拓展产业互联网应用空间

合成数据早期主要应用于计算机视觉领域，因为计算机视觉被广泛应用于自动驾驶汽车、机器人、安防、制造业等领域，在这些应用场景中打造 AI 模型都需要大量的被标注的图像、视频数据。但获取现实数据往往并非易事。以自动驾驶汽车为例，由于实际道路交通场景千变万化，让自动驾驶汽车通过实际道路测试来穷尽其道路上可能遇到的每一个场景是不现实的，必须借助于合成数据才能更好地训练、开发自动驾驶系统。为此，自动驾驶企业开发了复杂的仿真引擎来“虚拟地合成”自动驾驶系统训练所需的海量数据，并高效地应对驾驶场景中的“长尾”问题和“边缘案例”。例如，腾讯自动驾驶实验室开发的自动驾驶仿真系统 TADSim 可以自动生成无需标注的各种交通场景数据，助力自动驾驶系统开发。在安全的、合成的仿真环境中，计算机可以模拟任何人类想象得到的驾驶场景，诸如调节天气状况、添加或移除行人、改变其他车辆的位置等等。可以说，合成数据和仿真技术是自动驾驶的核心支撑技术。最早涌现的一批合成数据创业公司就瞄准的是自动驾驶汽车市场，帮助自动驾驶企业解决其在

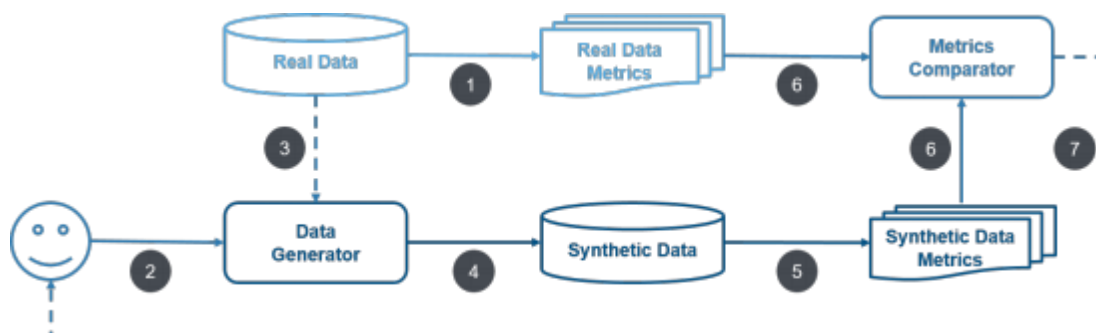
自动驾驶系统开发过程中所面临的数据和测试难题。

目前，合成数据正迅速向金融、医疗、零售、工业等诸多产业领域拓展应用。合成数据在金融服务领域的探索仍处于早期且不断拓展，并且受到咨询公司、金融巨头和监管机构的关注。

合成数据背后的生成式 AI 被 Gartner 评为 2022 年银行和投资服务领域越来越受欢迎的三项技术之一。²⁶ 生成性 AI 受欢迎的原因是能够通过合成数据以成本更低、易规模化、隐私保护合规的方式提供接近真实世界的的数据。而在银行和投资服务领域，生成对抗网络(GAN)和自然语言生成(NLG)的应用可以在大多数欺诈检测、交易预测、合成数据生成和风险因素建模的场景中找到。例如，美国运通(America Express)利用 GAN 创建合成数据来训练、优化其进行欺诈检测的 AI 模型。谷歌利用 AI 生成的医疗记录来帮助预测保险诈骗(insurance fraud)。

摩根大通(J.P.Morgan)2021 年 9 月在其官网发布相关研究，提出通过生成合成数据集加快金融服务领域的 AI 研究和模型开发，来改善服务体验、解决欺诈检测和反洗钱等重要问题。²⁷ 国外金融服务业所产生的大量真实数据因为法律限制(如欧盟的 GDPR 和美国的 CCPA)和隐私保护要求无法使用或使用存在诸多限制。合成数据创造的新样本具有真实数据的性质，增加真实数据中的罕见样本，以便更有效地训练机器学习算法。一个关键的领域是欺诈检测模型训练。由于欺诈性案件的数量与非欺诈性案件相比较十分稀少，研究人员很难有效地从可用数据中训练模型，导致无法针对欺诈性行为进行建模。

然而，合成数据可以生成比实际数据中欺诈案例比例更高的合成数据样本，用于帮助改进模型训练。



图：金融数据合成过程(来源：摩根大通官网)

英国金融行为监管局(Financial Conduct Authority, FCA)也积极跟进合成数据作为一项隐私保护技术在金融行业的应用前景。FCA 在 2022 年 3 月的公开报告中指出，合成数据通过生成统计学上真实 (statistically realistic) 的但“人造” (artificial) 的数据，为数据共享提供更多的机会。技术的应用有利于提供更普及、更高效、更好提升消费者体验的金融服务，目前的应用领域包括：金融犯罪和欺诈预防、信用评分、销售和交易、保险产品定价和索赔管理、资产管理和组合优化等。值得注意，合成数据也存在通过逆向工程来获取真实数据的风险，对数据隐私保护而言并非完美的解决方案。²⁸ 为此，FCA 已经通过公开征求意见来了解业界对合成数据支持创新的潜力和满足企业有效需求的看法以及潜在的限制和风险。

在医疗领域，医疗影像的合成数据正对医疗 AI 领域的发展产生巨大推动作用。例如，创业企业 Curai 基于仿真的医疗案例数据训练了一个医疗诊断模型，表明合成的医疗数据同样可以用于支持医疗 AI

应用的开发。美国生物技术公司 Illumina 和合成数据创业公司 Gretel 合作，利用合成的基因组数据来进行医学研究，以避免限制性立法、病人同意等问题，这些问题阻碍医学研究的速度和规模。另一个典型的例子是将合成数据用于新冠肺炎相关的医学研究。为了对抗新冠肺炎，研究人员需要获取足够的病例数据来开展研究，但出于隐私保护等顾虑，研究人员在获取新冠肺炎相关的病例数据时面临不小的困难。为此，美国国立卫生研究院(NIH)和 Syntegra、MDClone 等合成数据服务商合作，基于其 Covid-19 病人病历数据库“全国新冠合作群组”(NationalCOVIDCohortCollaborative, N3C)中超过 500 万个新冠阳性个体的病例数据，合成了不具有可识别性的替代数据，即合成的 Covid-19 数据。合成数据精确地复制了原始数据集的统计特征，但与原始数据不存在任何关联，可以被世界范围内的研究人员自由分享和使用，用于开展研究发现、疾病治疗、疫苗开发等方面的工作。

在零售领域，创业公司 Caper 可以基于一个商品的几张图片，利用 3D 模拟技术创建含有上千个图片的合成数据集，进而支持智慧零售店。在工业领域，FORD 利用游戏引擎和 GAN 来创造合成数据训练 AI 模型。此外，合成数据还可帮助应对算法歧视等 AI 伦理问题，支持打造更公平、可信的 AI 模型，因为合成数据可以帮助消除 AI 数据集中的偏见因素，支持构建更具包容性的、代表性的高质量数据集。

三、合成数据产业加快成为数据要素市场新赛道，科技大厂

和创新企业抢先布局

合成数据对人工智能未来发展的巨大价值使其加速成为 AI 领域的一个新产业赛道。一方面，国外的主流科技公司纷纷瞄准合成数据领域加大投入与布局。英伟达 29 是典型代表，其元宇宙平台 Omniverse 拥有合成数据能力 omniverse replicator; omniverse replicator 作为 Omniverse 平台的一部分，是为 AI 算法训练生成具有物理属性的合成数据的技术引擎，有两项代表性应用。一是在 Omniverse 平台中创建用于机器人训练的虚拟环境 IssacSIM 平台，在这个虚拟环境中训练的机器人之后可以直接应用于现实世界，即这个机器人可以将虚拟世界中的训练结果同步到现实世界的机器人身上，叠加人工智能的算法，实现机器人的大规模应用。二是 driveSIM 平台，提供丰富的模拟场景，用于自动驾驶算法训练和验证，如物体识别、车道保持、泊车等自动驾驶汽车应用。Omniverse 的这两项合成数据应用可以显著加速机器人和自动驾驶的开发进程。

微软的 Azure 云服务则推出了 airSIM 平台，可以创建高保真的 (high fidelity) 的 3D 虚拟环境来训练、测试 AI 驱动的自主飞行器;32 微软还开发了可以生成合成和聚合数据集的开源工具 Synthetic Data Showcase，并创建了合成人脸数据库，和国际移民组织(IOM)合作打击人口贩卖。亚马逊在多个场景探索合成数据的应用，例如使用合成数据来训练、调试其虚拟助手 Alexa，以避免用户隐私问题;其合成数据技术 Wordforge 工具可以用来创建合成场景(synthetic scenes);在

2022 年的亚马逊 re:MARS 大会上，其数据标注服务 Sage Maker Ground Truth 推出了合成数据能力，可以自动生成标注的合成图片数据，即该工具可以就特定物体创建不同位置和不同灯光条件的图片，以及不同比例和其他变化的图片。Meta(原 Facebook)也着力于为其人工智能业务增强合成数据能力，2021 年 11 月收购了合成数据创业公司 AI.Reverie。

另一方面，合成数据作为 AI 领域的新型产业，相关创新创业方兴未艾，合成数据创业公司不断涌现，合成数据领域的投资并购持续升温，开始涌现了合成数据即服务(synthetic data saservice, SDaaS)这一发展前景十分广阔的全新商业模式。据国外研究者统计，目前全球合成数据创业企业已达 100 家。³⁴ 比较知名、有影响力的包括 AI.Reverie、datagen、skyen-gine、mostly.ai、synthesisAI、gretel.ai、oneview、innodata、cvedia 等等。在过去的 18 个月，公众视野中已知的合成数据公司融资总额达到 3.28 亿美元，比 2020 年高出 2.75 亿美元。合成数据的创业赛道主要涵盖非结构化数据(图片、视频、语音等)、结构化数据(表格等)、测试数据(test data)、开源服务等几大方向。其中，非结构化合成数据持续保持强劲发展势头，这主要得益于以下几个因素:第一，计算机视觉应用场景相对成熟;第二，有游戏引擎(game engines)、图像建模软件、AIGC 技术的支撑;第三，自动驾驶汽车、零售、电子游戏等快速发展的产业对合成数据有较高需求。目前结构化数据合成和测试数据合成正在迅猛发展，尤其是合成的测试数据更少受到数据隐私立法的限制，所以开始受到业界青睐。此外，合成数据开源服务也

在快速发展，例如 synthetic data vault、synner、synthea、synthetig 等。

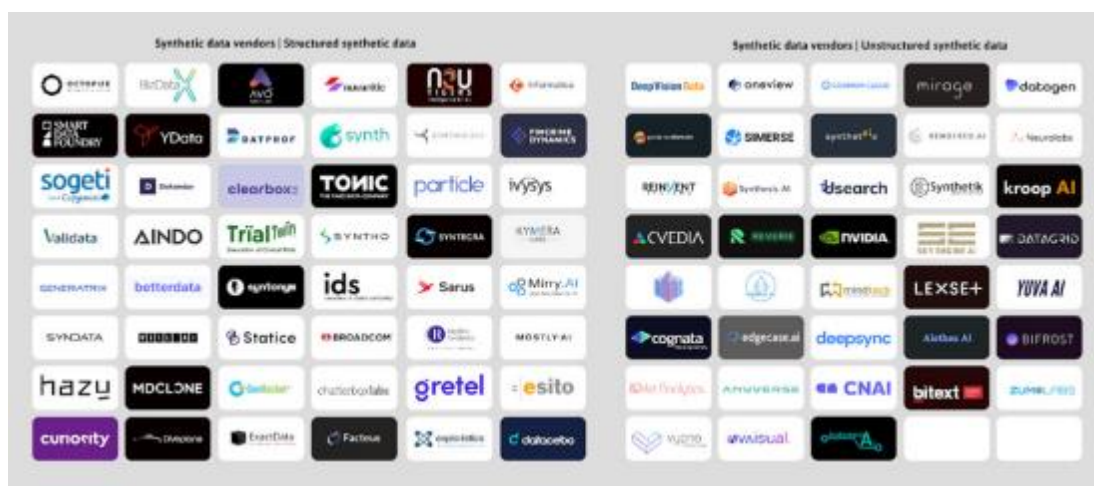


图:合成数据产业生态一览表(来源:medium 网站)

市场规模方面,根据 Cognilytica 的数据,合成数据市场规模在 2021 年大概是 1.1 亿美元,到 2027 年将达到 11.5 亿美元。GrandViewResearch 预测, AI 训练数据市场规模到 2030 年将超过 86 亿美元。Gartner 预测,到 2024 年用于训练 AI 的数据中有 60% 将是合成数据,到 2030 年 AI 模型使用的绝大部分数据将是人工智能合成的。可以预见,合成数据作为数据要素市场的新增量,在创造巨大商业价值的同时,也有望解决人工智能和数字经济的数据供给问题。

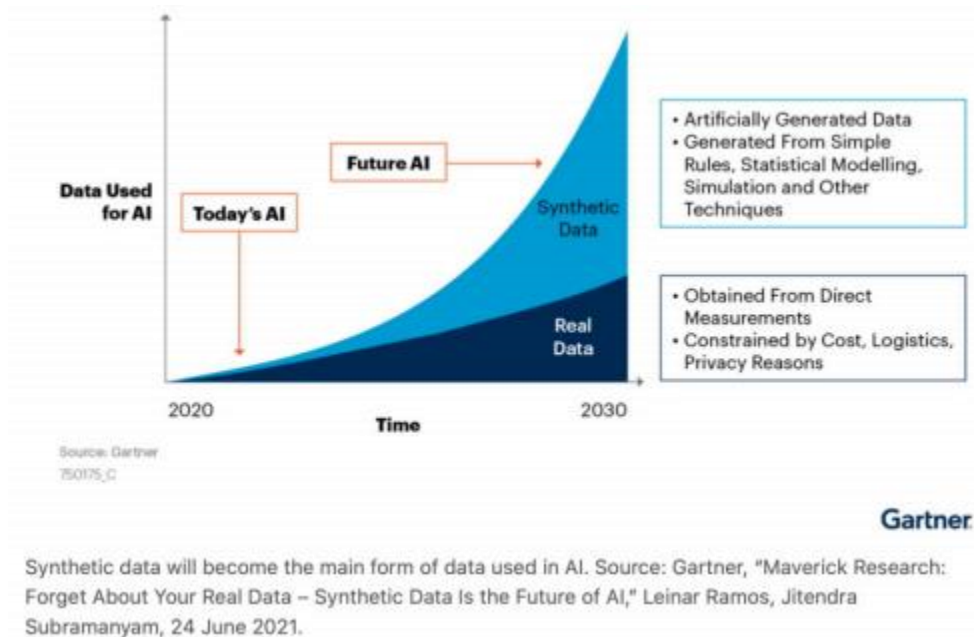


图:合成数据发展预测(来源:Gartner)

三、合成数据加速构建 AI 赋能、数实融合的大型虚拟世界

合成数据指向的终极应用形态是借助游戏引擎、3D 图形、AIGC 技术构建的数实融合的大型虚拟世界。大型虚拟世界对于人工智能的核心价值在于，为测试、开发新的人工智能应用，提供了一个安全、可靠、高效以及最重要的是——低成本的、可重复利用的环境。展望未来，可以从三个方面来理解大型虚拟世界为什么成为 AI 数实融合的关键场景。

第一，大型虚拟世界可以提供人工智能开发所需的数据和场景，为 AI 应用开发提供“加速度”。游戏开发者、发行者已经意识到了这一机遇，即为 AI 系统设计游戏般的虚拟世界。例如，《星际争霸》、《我的世界》等游戏是人工智能的理想的测试场地，3A 游戏场景资产也是生成性 AI 工具的重要潜在应用。

第二，大型虚拟世界为各行各业训练开发 AI 提供了试验田。这个方面的一个典型例子就是用于自动驾驶算法测试的虚拟仿真平台。谷歌、英伟达、腾讯等国内外科技公司都在大力布局自动驾驶仿真业务，即提供一个仿真的合成世界来训练、测试自动驾驶算法。在国内，腾讯自动驾驶实验室开发的自动驾驶仿真平台 TADSim，可以让自动驾驶算法在城市级别的虚拟仿真世界中进行测试和学习，极大降低了自动驾驶汽车的研发成本。在国外，自动驾驶企业 Waabi 希望主要通过虚拟仿真而非实际道路测试来训练其自动驾驶系统，构建了 WaabiWorld 这一用于自动驾驶系统的大型虚拟世界，其核心即是合成数据和合成场景。WaabiWorld 不仅可以接近实时地模拟汽车的传感器，而且模拟的环境可以和人工智能之间可以相互做出反应。这是非常重要的，因为自动驾驶汽车不仅需要感知世界，还需要安全地采取行动。大型虚拟世界提供了这样一个安全有效且可重复使用的试验田。

第三，在大型虚拟世界中通过 AI 连通虚拟与现实，实现 AI 数实融合。在大型虚拟世界，如游戏虚拟世界或者其他的仿真的 3D 合成世界(simulatedsyntheticworld)中训练 AI 虚拟机器人，让其自我学习、进化，然后部署到现实世界中，产生现实的价值。此外，大型虚拟世界对强化学习和其他 AI 算法也具有巨大价值，例如，游戏世界是开展 AI 强化学习和多智能体协作的最佳载体;也可以在大型虚拟世界中利用虚拟人开发更宜居的建筑设计;大型虚拟世界如果可以仿真物理现象，就可进行虚拟化学实验，相比人类利用真实的化学药品开展实

验，软件可以利用虚拟的化学药品更高效、更安全、更大规模地开展化学实验。总之，大型虚拟世界对于支持人工智能研究开发、促进虚实融合的巨大可能性，值得持续深挖和不断探索。