

生成式人工智能

作者：丁磊

前言

人从出生开始，就在不断通过视觉、听觉、嗅觉、味觉、触觉等各种方式认识这个世界。我们通过不停地与外界接触、学习，逐渐长大成人，再通过专业课程的学习在某些方面获得一技之长从而立足于社会，并试图改造世界。

“硅基”AI（人工智能）也按照类似的模式成长，但是相比于“碳基”人，它在速度方面极具优势。AI经历了从最初的机器学习到神经网络，再到Transformer模型的发展，2022年底ChatGPT以及2023年初GPT-4横空出世，引燃了公众对生成式AI的关注，其中最让人激动的就是AI大模型已经初步具备了人类的通识和逻辑能力——这恰恰是之前的AI所缺失的。此前，无论是AlphaGo还是AlphaFold，最多只能称作其各自领域的“专家”，而ChatGPT是通用的。

正如OpenAI首席科学家、ChatGPT背后的技术大佬伊利亚·苏茨克维(Ilya Sutskever)所说，GPT（生成式预训练模型）学习的是“世界模型”。他将互联网文本称作世界的映射，因此，将海量互联网文本作为学习语料的GPT学习到的就是整个世界。在我们认识世界的同时，GPT模型也以惊人的算力，快速地获取我们数年甚至数十年才能拥有的认知，即将成为一个接近成年人思维水平的“世界模型”。

不仅如此，已具备了“世界模型”能力的GPT还能够生成“万物”。当然，如苏茨克维所说，这里的万物指的是世界万物在数字空间的映射，包括文本、图片、音频、视频、剧本、代码、方案、设计图等一切和我们生产、生活息息相关的事物。因为GPT模型在一定程度上可能已经具备了成年人的通识和逻辑，所以我们只需要拿特定专业领域的数据对其再做训练（称为“微调”），它就可以成为独当一面的专业人才，可能成为艺术家、设计师、程序员、工程师或广告优化师、供应链专家、客服人员等。这也许就是生成式AI或者说AIGC（AI generated content，人工智能生成内容），带给我们的核心价值。

在AI技术大爆炸的今天，生成式AI处在高速发展阶段，技术和应用领域日新月异，因此我们非常有必要系统地了解生成式AI。在这样的背景下，本书将系统介绍生成式AI的原理与模型，同时也将对其在行业场景中的应用展开论述，将理论和实际相结合，让大家从本源上了解ChatGPT里程碑式存在的意义。结合作者二十余年AI领域研究与工作的经验，本书会为读者指明方向。尤其值得一提的是，本书既在理论上解释了数字媒体即虚拟世界的生成式AI，又探讨了生成式AI如何服务和赋能实体经济。在当前的存量经济时代，通过生成式AI重新定义生产力，助力行业更新发展，在存量里促增长，具有尤为重要的意义。

如图0-1所示，我们用图表明本书所覆盖的知识领域：X轴是生成式大模型的维度，对应的是第二章“AIGC的底层逻辑”，我们将了解“用什么去生成”；Y轴是数字媒体形态的维度，对应的是第三章“功能分析：AIGC能生成什么内容？”，我们将了解生成式AI能“生成什么”；Z轴是行业职能的维度，对应的是第四章“商业落地：AIGC的产业应用与前景”，我们将了解用生成式AI可以“做什么事”。三个轴所形成的空间里的每个点都有其特定含义，例如：通过GPT-4模型生成代码用在生产上，通过Stable Diffusion模型生成图片用在营销上。除了这三章，第一章会带领读者初识生成式AI，第五章则探讨生成式AI是否会取代大量的工作岗位，以及我们应该如何主动应对。

图0-1 本书所覆盖的知识领域

希望任何一个不想在生成式AI时代落伍的人，在阅读本书之后，都能理解生成式AI的底层逻辑和实际应用，也希望本书对他们的工作和生活有所助益。“万物皆可生成”的时代已经来临，理解AI、训练AI、使用AI，甚至和AI一起工作，对每个人来说或将无法避免。未来已来，让我们一起出发！

第一章 初识生成式人工智能

现象级产品ChatGPT的横空出世带火了AIGC和它背后的生成式AI，让不少人对使用AI工具跃跃欲试。在本章，我们会一起进入AI的产业国度，从决策式AI跃迁至生成式AI，对比这两种人工智能模型的异同，深度挖掘它们的应用场景和商业潜力，同时还会探讨“内容皆可生成”的生成式AI及其核心价值。在概览生成式AI后，我们会把视角转向具体工具，深入解析大众已经熟知的ChatGPT，了解这项“奇妙工具”背后的原理和发展历程。若你对生成式AI一知半解，可以在本章的内容中初步认识它，也能了解到近期最值得关注的生成式AI工具。

纵观AI产业版图

如果要选出2023年最热的几个话题，ChatGPT一定榜上有名。2023年初，ChatGPT席卷全球并成为流量热点，人们都在前赴后继地挖掘ChatGPT的各种潜能，探讨其未来发展趋势，甚至是与人类的关系。作为“人工智能家族”的大热应用，以ChatGPT为首的各大人工智能应用开始被越来越多的人关注，也引发了人们的更多思考。

人工智能技术被称为当代三大尖端技术之一，近年来在人们生活中的“存在感”也越来越强，这都是产业飞速发展的结果。想要清晰地了解以ChatGPT为代表的新兴智能技术，完整地认识人工智能，我们可以先从其产业版图的发展和现状入手。

其实，人工智能的发展、传播和被接受是经过了一段漫长的寒冬的。十多年前，它还只是一个不被人看好的小众领域，但是现在，它却已经成了街头巷尾的热点谈资，几乎任何事情都可以和人工智能联系在一起。短短十多年间，世界发生了天翻地覆的变化，新数据不断涌现，各种问题层出不穷，直到现在，人工智能的春天才算是真的到来了，各个领域都急需人工智能的帮助。这也是为什么人工智能的行业应用范围如此广阔，人工智能市场更是如一块一望无际的辽阔土地，有待进一步开发。如图1-1，这是一份人工智能的行业应用版图，不同的行业领域（零售、金融、医疗和教育等）与不同的职能方向（营销、风控和安全等）共同构成了一个人工智能应用矩阵，对于每个行业中的相关职能，人工智能都可以找到应用场景，例如在零售行业的供应链、营销、客服等方面以及金融行业的研发、营销、客服、风控等方面都已经有人工智能落地实践（图中蓝色表示）。但是，现在的人工智能只填充了广阔的行业领域中的一部分，还有更多没尝试和拓展的行业以及职能中的应用场景。

□

图1-1 人工智能的行业应用版图

从产业的视角来看，人工智能包括基础层、技术层和应用层。其中，基础层是人工智能产业的基础，为人工智能提供数据及算力支撑；技术层是人工智能产业的核心，主要包括各类模型和算法的研发和升级；应用层则是人工智能面向特定场景需求而形成的软硬件产品或解决方案。那么，人工智能的产业规模发展至何种程度了呢？英国德勤(Deloitte)的报告中预测，世界的人工智能产业规模会从2017年的6900亿美元增长至2025年的64000亿美元，2017—2025年的复合增长率将达到32.10%，整体呈现出飞速攀升的趋势。另外，人工智能近几年成了各个行业在进行投资的热门选择。人工智能完全称得上是风头正劲，受万人追捧，为经济带来了十分显著的增量。

在产业应用上，人工智能发展到今天，我们能看到其在各个行业都有用武之地：制造业、零售业、金融业、医疗卫生行业……它在一定程度上改变了组织的运转方式，使其可以更快更好地解决遇到的问题，并压低各类成本。站在消费者的角度，人工智能的出现也为广大的用户群体带来了更多的选择。总的来说，人工智能可以看作一块已开始被打磨的原石，露出了它璀璨的一角，它在推动世界经济发展的同时，也将深层次地改变人类的生活。为了进一步了解AI产业版图，下面我们从两种不同的AI——决策式AI和生成式AI谈起。

决策式AI和生成式AI

人工智能可从不同的维度进行划分。如果按其模型来划分（人工智能是由模型支撑的）可以分为决策式AI和生成式AI。

决策式AI（也被称作判别式AI）学习数据中的条件概率分布，即一个样本归属于特定类别的概率，再对新的场景进行判断、分析和预测。决策式AI有几个主要的应用领域：人脸识别、推荐系统、风控系统、其他智能决策系统、机器人、自动驾驶。例如在人脸识别领域，决策式AI对实时获取的人脸图像进行特征信息提取，再与人脸库中的特征数据匹配，从而实现人脸识别。再例如，决策式AI可以通过学习电商平台上海量用户的消费行为数据，制订最合适的推荐方案，尽可能提升平台交易量。

生成式AI则学习数据中的联合概率分布，即数据中多个变量组成的向量的概率分布，对已有的数据进行

总结归纳，并在此基础上使用深度学习技术等，创作模仿式、缝合式的内容，相当于自动生成全新的内容。生成式AI可生成的内容形式十分多样，包括文本、图片、音频和视频等。例如，我们输入一段小说情节的简单描述，生成式AI便可以帮我们生成一篇完整的小说内容；再例如，生成式AI可以生成人物照片，而照片中的人物在现实世界中是完全不存在的。如图1-2，它展示的是国外一个网站生成的“不存在的人”的照片。

总的来说，不管是哪种类型的模型，它的基础逻辑是一致的：AI模型从本质上来说是一个函数，要想找到函数准确的表达式，只靠逻辑是难以推导的，这个函数其实是被训练出来的。我们通过喂给机器已有的数据，让机器从数据中寻找最符合数据规律的函数。所以当有新的数据需要进行预测或生成时，机器就能够通过这个函数，预测或生成新数据所对应的结果。

图1-2 “不存在”的人

图片来源：<https://generated.photos/faces>

决策式AI和生成式AI作为AI模型的两个主要分支，顾名思义，在诸多方面都有相异之处。

从宏观角度来看，决策式AI是一种用于决策的技术，它利用机器学习、深度学习和计算机视觉等技术来处理专业领域的问题，并帮助企业 and 组织优化决策。而生成式AI则是一种用于自动生成新内容的AI技术，它可以使用语言模型、图像模型和深度学习等技术，自动生成新的文本、图片、音频和视频内容。因此，决策式AI可以说是在对人类的决策过程进行模仿，但生成式AI就聚焦在创作新内容上。

而从微观上看，这两类技术的区别就更加明晰了，我们就从技术路径、成熟程度、应用方向这三个角度来挖掘其深层次的不同（表1-1）。

表1-1 决策式AI和生成式AI的对比

从技术路径来看，决策式AI的主要工作是对已有数据“打标签”，对不同类别的数据做区别，最简单的例子如区分猫和狗、草莓和苹果等，干的主要是“判断是不是”和“区分是这个还是那个”的活儿。生成式AI就不一样了，它会在归纳分析已有的数据后，再“创作”出新的内容，如在看了很多狗的图片后，生成式AI再创作出一只新的狗的图片，实现“举一反三”。

从成熟程度看，决策式AI的应用更为成熟，已然在互联网、零售、金融、制造等行业展开应用，极大地提升了企业的工作效率。而生成式AI的“年岁更小”，2014年至今发展迅猛，堪称指数级的爆发，已在文本和图片生成等应用内落地。

从应用方向来看，决策式AI在人脸识别、推荐系统、风控系统、机器人、自动驾驶中都已经有了成熟的应用，非常贴合日常生活。生成式AI则在内容创作、人机交互、产品设计等领域展现出巨大潜力。

我们来举一些生活中的例子，以更深入地了解两者在日常生活中的应用。喜欢购物的读者都知道，你在购买某一类产品后，购物平台会自动给你呈现诸多同类或相关商品。这件事的背后就是，电商平台会根据用户常看的商品，剖析用户和商品的关联，从而有针对性地为用户推荐内容，而这项功能就应用了决策式AI技术。从2003年开始，亚马逊就将此技术应用到了电商领域，推荐的商品精准地匹配用户需求，可以极大地降低用户的搜索次数，并因此增加产品的销售额。由此你可能会发现，平台似乎比你更清楚你需要什么，自然而然，自己的消费额也跟着上去了。平台也凭借这个功能，让更多用户心甘情愿地掏了腰包，来获取更广阔的商业价值。

根据行业数据统计，在亚马逊的收入中约有40%来自个性化推荐系统，而推荐系统每年能给网飞(Netflix)带来10亿美元以上的产值。除了电商平台，新闻、音乐、视频等平台，也会利用个性化推荐系统为用户推荐内容，在剖析用户的长期兴趣和短期兴趣后，将精细化内容推送给用户，并可以通过对用户的停留、观看时间、点赞、收藏等行为特征的实时分析，精准刻画用户画像，减少人工运营的介入，显著

提升用户黏性，这已将人工智能的价值凸显无遗。在自动驾驶领域，AI可进行智能分析、识别路况，渗透率稳步提升。自动驾驶汽车可以借助决策式AI技术，分析判别各种路况，对多种物体进行识别与跟踪，提升行车安全。无须人工干预的自动驾驶汽车虽然现在并不成熟，但随着技术的迭代升级，有望获得更大的市场潜力。

对于生成式AI，ChatGPT的出现让我们对其有了冲击式的关注和理解。因生成式AI功能强大、应用范围广泛，文字、图片、音视频内容相关的从业者在面对“强大助手”上线时，也会感觉到焦虑，恐被其取代。从可能性来讲，它可以进行文字生成语音、图像智能编辑、视频智能剪辑、文字续写或纠错等十分多样的工作，让大家摆脱机械劳动，把时间花在创造性工作上，给文字作者、翻译人员、插画师、视频剪辑师等带来极大的支持。不仅如此，生成式AI还能胜任部分由设计师、程序员甚至专业工程师从事的设计与编程类工作，在提升工作效率的同时让这些专业人士更能发挥所长，减少在初级工作上的时间投入。与此同时，生成式AI对于从业人员的素质和技能，也提出了新的要求。总的来说，决策式AI和生成式AI均可以帮助用户推进部分工作，如事件决策、创作内容等。可以说，人工智能的合理利用有助于提升客户体验，帮助企业降本增效，并抓住新的商业机会。

如前文所述，数据和模型分属人工智能产业的基础层和技术层，无论是决策式AI还是生成式AI的应用都离不开数据和模型，下面我们进一步了解“大数据”和“大模型”是如何重塑人工智能版图的。

从大数据到大模型

无论是决策式AI还是生成式AI，以其现在的功能和潜力，都能为人类做很多工作，未来甚至有点万能，那么这么万能的技术，是怎么被“训练”出来的呢？这就要说到大数据了，决策式AI和生成式AI，其实都离不开用大量数据来训练模型。

对于大数据，大众已经比较熟悉。顾名思义，大数据指的是海量的数据，但大数据并没有看上去这么简单，它还有多样性和高速增长的特性。图1-3展示了从2017年到2025年全球数据总量的增长趋势及预测情况。收集、存储、处理和分析各种形式和来源的大数据，可以帮助企业和组织迅速获得有价值的信息，并做出正确的决策，它还可以用于商业活动的改善，如此能提升工作效率，降低工作成本，并推动企业实现更大的增长。就如人类通过经历各类事件来积累经验一般，在人工智能领域，我们通过大量的数据来训练模型。

图1-3 2017—2025年全球数据总量增长趋势及预测情况

数据来源：国际数据公司发布的白皮书《数据时代2025》

而随着深度学习的落地和发展，模型本身所需的存储空间在近年有了显著增长，最初的GPT就有1.17亿个参数，ChatGPT有1750亿个参数，最新的GPT-4参数数量更多，有报道称可能达到1T（即10000亿），但OpenAI公司其实并没有公布具体的参数数量，这些拥有海量参数的模型都被称为“大模型”。如图1-4，它展示了大模型参数数量变化趋势。这里我们提到了深度学习，这是一种受人脑的生物神经网络机制启发，并模仿人脑来解释、处理数据的机器学习技术，它能自动对数据进行特征提取、识别、决策和生成。你可能觉得这个词有点耳熟，其实它大规模地应用于自然语言处理(NLP)、计算机视觉、机器翻译等领域。深度学习的出现，为很多领域的工作带来了前所未有的精度和效率。人工智能行业也因深度学习收获了前所未有的发展速度，整个人工智能领域的发展都曾被它带动。

图1-4 大模型参数数量变化趋势

大模型能分析处理海量的数据，在解决问题上取得更好的效果。本书的“主角”生成式AI就是大模型的产物。近年来，大模型在越来越多的行业和消费类应用中崭露头角，原因主要是它能够迅速有效地处理海量的数据，帮助个人和企业提升效率。大模型与人工智能技术相辅相成，随着人工智能技术的发展，大模型也会持续发展进步。另外，生活中日益普及的5G网络和弹性计算等基础设施，也会给大模型的发

展创造更多可能性，使其成为不可或缺的内容生成工具。

生成式AI市场格局

2021年，高德纳咨询公司(Gartner)就曾预测，至2023年将有20%的内容被生成式AI创建，至2025年生成式AI产生的数据将占有所有数据的10%（2021年不到1%）。2022年9月，红杉资本官网发布的文章《生成式AI：充满创造力的新世界》预测，生成式AI将产生数万亿美元的经济价值。据预测，2025年，国内生成式AI应用规模有望突破2000亿元，国内传媒领域应用空间超1000亿元。而且，生成式AI“八面玲珑”，它的应用场景十分广泛，目前不仅应用于文本、图片、音视频、游戏等数字媒体，还可以应用于制造业、建筑业等实体行业。

在文本生成方面，生成式AI可以通过语言模型、神经网络和深度学习技术，快速创建大量有助于改善客户体验的内容，如新闻资讯、剧本、营销文本、智能客服等。其中作为经典应用的AI生成营销文本、智能客服等都已许多行业广泛地应用；AI生成新闻资讯和剧本等功能大家也可以期待一下，或许以后结合了ChatGPT等突破性的模型，文字性工作真的能依靠它变得轻松不少。

在图片生成方面，生成式AI可以通过计算机视觉来分析图片，生成营销素材、设计方案和艺术作品等，帮助节省人力成本和时间。另外，生成式AI还能在音频生成、视频生成和跨模态生成领域大展拳脚。

在音频生成方面，生成式AI可以帮助使用者更好地分析、编辑和生成音频文件，从而帮助创作出优秀的音频作品。例如，克隆真人的语音、文本生成特定语音、作曲编曲等，生成式AI都能代替人类去做，并均已广泛应用于市场。

视频生成也是生成式AI的重要应用，它可以帮助使用者生成高质量的视频，如检测和删除特定片段、跟踪剪辑、生成特效、合成视频等。另外，大火的AI数字人也是它的“拿手绝活”。在李安执导的《双子杀手》中，工作人员就用AI创造了一个数字人物小克。威尔·史密斯在数字技术的帮助下同时出演了50岁特工亨利和23岁特工小克，该片实现了真实明星“年轻版”的数字化制作。

在跨模态生成中，生成式AI能够根据文字生成创意图片、根据图片生成视频、根据文字生成视频，或根据图片或视频生成文字。对想象力丰富的朋友，或者影视行业从业者来说，这称得上是“工作神器”。图1-5就是一个根据文字“panda in a space suit”（穿着宇航服的熊猫）生成图片的例子。在游戏方面，生成式AI可以用于游戏开发，实现自动化的游戏设计，同时能够实现更好的游戏体验，如人工智能NPC（非玩家控制角色）等，说不定以后你玩的游戏就有人工智能的深度参与。

□

图1-5 根据“panda in a space suit”生成的图片

生成式AI不光在这些数字经济领域广泛应用，在实体领域的潜力也非常大，如在建筑业等巨型垂直实体领域中，生成式AI所生成内容就不再局限于图片和文字，而是进入了信息形式更为丰富的3D（三维）设计领域。例如构建数字建筑模型时，生成式AI能帮助建筑师们产出3D建筑模型，让他们更好地理解项目。建筑师们能够使用AI图像生成应用来丰富建筑设计方案的细节，假如建筑师们向应用中输入较为初级的建筑设计方案，AI就能够在初级设计的基础上，继续产出较为细化的设计方案，以此来深化设计。建筑师们还可以随手绘制一个潦草的建筑场景线图，让人工智能来生成对应的建筑实景图。我们可以想象，随着手绘信息的增加，生成式AI输出的实景图也越来越稳定。图1-6所示的就是利用AI图像生成工具生成的建筑设计图。

□

图1-6 由AI图像生成工具生成的建筑设计图

图片来源：<https://stability.ai/blog/stablediffusion2-1-release7-dec-2022>

技术的浪潮层叠翻涌，人工智能已成为人类社会冲向未来世界的战舰，产业前景十分广阔。生成式AI更是一个突破性的产业发展方向，它不仅能给数字媒体和虚拟空间带来价值，还能促进实体行业的发展，

在提升行业效率的同时优化原有的流程，创造出新的价值增长点，可以说是实体行业升级不可多得的机遇。

聚焦AIGC：内容皆可生成

当下，世人的目光被ChatGPT、GPT-4这些AIGC深深吸引。而在清楚地认识这些新事物之前，我们需要梳理一下它们的历史脉络，其实在数年硝烟弥漫的“内容大战”中，我们已经悄然经历了多种内容形式的迭代：PGC(professional generated content)、UGC(user generated content)和AIUGC(artificially intelligent UGC)。PGC即“专业生产内容”，主要指具备专业背景的内容生产者所创造的内容；UGC则为“用户生产内容”，其内容的源头更偏大众化，人人都可作为用户进行内容生产；AIUGC则为人工智能与UGC的结合，人工智能参与到了用户创作内容的过程中。如今，在三度更迭之后，AIGC正式来袭。与PGC、UGC和AIUGC不同的是，在AIGC的概念中，“无生命的”人工智能成了完全的内容源头，“无生命主体”成了为人类创作内容的生产者。人工智能在人类社会的应用又取得了颠覆性的突破，透出了不同于以往的炫目光彩，吸引着人们不断探索。如图1-7，从PGC、UGC、AIUGC到AIGC，所对应的内容数量呈逐渐增加的趋势。

图1-7 内容创作的四个发展阶段

从字面上看，AIGC就是利用人工智能自动生成内容的生产方式，它可以在生成式AI模型、训练数据等的基础上，生成文本、图片、音频、视频、代码等多样化内容，这种快速的内容生产方式给市场注入了令人兴奋的新鲜血液。AIGC的出现，使得各行各业都受益，使得人们的生活更加便捷。但在发展得如火如荼的同时，AIGC又引发了我们对其更深层次的思考。

AIGC开启了新一轮的内容生产革命，它在多样性、质量、效率三个方面推动了内容生产大踏步前进。AIGC的出现，既可以满足消费型内容亟待扩充的需求，也可以快速产出多样化的内容形态，迎合多种细分场景，以AI作者的身份助力商业化浪潮的翻涌。或许我们现在正在看的某张图片、某段视频就是AI作者的“作品”，而我们却不自知。

下面我们就围绕AIGC，对文本、图片、视频等不同的内容形式展开论述，看看AIGC究竟是如何“长袖善舞”，在各个内容形式中发挥作用的。

文本生成

AIGC生成文本目前主要被应用于新闻的撰写、给定格式的撰写、风格改写以及聊天对话，GPT是主流的文本生成模型之一。

GPT的“学名”是生成式预训练模型(generative pre-training transformer)，这是一种用来分析和预测语言的人工智能模型，它可以帮助我们进行自然语言处理，例如机器翻译、自动文摘和快速问答。GPT的厉害之处是，它可以在文本中自动学习概念性内容，并自动预测下一段内容。也就是说，它可以根据上下文记住概念，并能够在短时间内直接输出相关内容。

GPT背后的基础模型是一种新型的机器学习技术，它可以帮助我们分析大量的自然语言数据。它背靠一个大型神经网络，通过在已有文本库中找到有关自然语言的规律来学习。GPT无须人工设计特定的自然语言处理系统，可以根据已有文本，自动生成语法正确、内容相关的文本。有这样一个“神器”，很多内容就可以借助它的力量来完成了！

GPT的发展目前经历了GPT-1、GPT-2、GPT-3、GPT-3.5和GPT-4几个阶段。对于GPT-1模型，我们可以这么理解：先使用海量没有进行标注的语料，预训练出一个语言模型，而后对语言模型进行微调，使之应用于特定的语言任务中。GPT-2则在GPT-1的基础上进行了多任务的训练，使用了更大的数据集，提升了语言处理能力。GPT-3则在训练的参数量、训练数据和训练费用上都高于前两者，能完成更加复杂的任务。

OpenAI推出的ChatGPT是GPT-3.5的延伸，这是一款聊天机器人程序，能通过学习和理解人类的语言与人类对话，还能实现视频脚本撰写、营销文案写作、文本翻译、代码编写等功能。例如它在代码理解和

编写方面的能力，就在程序员圈引起了广泛的关注：它可以看懂你输入的代码片段，帮你解读其中的含义，甚至可以根据你的要求帮你编写一段完整的代码。如此强大的能力，几乎颠覆了人们的认知，并引发了诸多关于“AI替代人类”的相关讨论。

而当人们还沉浸在ChatGPT带来的无限遐想中时，就在2023年3月，OpenAI推出了史上最强大的模型——GPT-4。它在文学、医学、法律、数学、物理和程序设计等不同领域表现出很高的熟练程度，各方面能力已全面超越ChatGPT。不仅如此，它还能够将多个领域的概念和技能统一起来，并能够理解一些复杂概念。OpenAI在官网上演示了这样一个示例：向GPT-4展示一张图片（图1-8），并询问图中有什么有趣的地方。而GPT-4的回答相当精妙：这幅图的有趣之处在于，把一个大而过时的VGA（视频图形阵列）接口插入一个小而现代化的智能手机充电端口，这是十分荒谬的。GPT-4俨然拥有一个普通人的正常思维。

要想深刻了解AI技术的发展，我们就需要到推动主体——企业中去。主打AI文本生成的Jasper公司位于美国加利福尼亚州，通过其产品的文本生成功能，用户可以轻松完成生成Instagram（照片墙）标题，编写TikTok（抖音国际版）视频脚本、广告营销文本、电子邮件内容等略显烧脑的重复性工作。AI文本生成功能一经推出，便给社交媒体、跨境电商、视频制作等多个新兴行业带来了巨大的颠覆力量。

□

图1-8 一张“有趣”的图片

图片来源：<https://openai.com/research/gpt-4>

除了Jasper以外，OpenAI更是近期谈论AI时不可绕过的热门企业。OpenAI是一家AI研究公司，成立于2015年，它旨在促进人工智能的安全可控发展。我们前文中提到的GPT这类卓越的自然语言处理模型，就是OpenAI首创推出的，这也使得OpenAI一跃成为AI行业的佼佼者。在自己进行技术创新之外，OpenAI也通过与微软等行业巨头的合作，将AI的应用推向更高的层次，这也将为人类的日常生活带来丰富的可能性。

由于GPT有基于英文语料库且不开源的局限，国内的技术人员也在探索我们自有的自然语言处理模型。2020年11月中旬，北京智源人工智能研究院和清华大学研究团队就合作推出了中文预训练模型——清源CPM(Chinese Pretrained Models)，我们也有了自主研发的类似于GPT的模型。

图片生成

你是否尝试过用AI生成图片呢？谈到AI生成图片，你第一时间又会想到哪个程序呢？你所使用的程序，很可能背后是由Diffusion（扩散）模型来进行技术支撑的。Diffusion模型是一种新兴的AI技术，它的灵感来源于物理学中的扩散现象：通过对图片不断加入噪声来生成一张模糊的图片，这个过程类似于墨水滴入水池的扩散过程；再通过深度学习神经网络学习模糊的图片并还原成原始图片的逆扩散过程，实现生成图片的功能。目前，Diffusion模型在视觉艺术和设计相关领域非常受欢迎。

Stability AI是一家全球领先的AI研究型企业，致力于开发前沿的人工智能模型。2022年，由该公司与另外两家初创公司共同研发的Stable Diffusion模型发布，可以真正实现“一秒出图”，这个“一秒”不是夸张的代指，而是真正的事实。这就意味着你可以借助AI，实现自己瑰丽的梦境，复原宏大的想象，也可以为自己的小说配上极富幻想感的插图，不论它们有多超现实，你都可以通过AI把它们呈现在大家的眼前，让想象不再孤独。

2022年，AI绘图突然大热，随着DALL·E2、Stable Diffusion、Midjourney等图像生成领域现象级应用的纷纷兴起，AI绘画就像一阵旋风，首先在国外引起了不小的风浪，社交平台上出现了大量的AI绘画相关尝试和讨论。很快这场旋风就从国外刮到国内，引起了国内用户的广泛关注。这些应用到底有着怎样惊奇的功能，而它们背后又有哪些企业在推动这场AI绘画“旋风”呢？

首先我们把目光放到Midjourney身上（图1-9），这是由同名研究实验室开发的AI绘画工具。在AI绘画领

域，Midjourney降低了艺术绘画创作的门槛，用户只需要输入文字描述，计算机就会自动生成一张作品。Midjourney采用了深度学习模型，能够自动为用户生成高质量的绘画作品，包括素描、油画等，让用户的使用更加方便。

毫不夸张地说，Stable Diffusion模型是掀起AI绘画热潮的源头之一，Stable Diffusion本身及基于它开发的绘画工具，让AI绘画引爆了舆论热潮。而其背后的公司Stability AI在AI绘画模型爆火前的估值为1亿美元，爆火后的估值则为10亿美元，狂涨10倍，足见AI技术产出的大众化程序有多么强大的市场潜力。

图1-9 Midjourney官网

与此同时，也有其他公司在AI绘画赛道“另辟蹊径”。如一家成立时间不到两年的公司PromptBase，主营业务为销售AI绘画工具的提示词，将提示词复制到Midjourney、Stable Diffusion等AI绘画平台，可以实现精准快速的图像生成，让用户在探索提示词上少走弯路。

若把目光转向国内，百度集团旗下的人工智能产品文心一格也在2022年8月宣布，用户只需要输入一段文字，并选择作画风格，文心一格就可以快速生成一幅画作。它以百度飞桨深度学习平台、文心大模型等技术为支撑，通过对海量优质图文的学习，经过多次迭代升级，如今已具备了更强的中文内容语义理解能力以及高质量图像生成能力，进一步满足国内用户对AI绘画的需求。

视频生成

AIGC视频生成，是一种基于人工智能的视频制作技术，它能够根据用户提供的文字提示，自动生成视频内容，而且还能够根据不同的需求调整视频的参数，以达到最佳效果。这在某种程度上是AIGC图片生成的延伸，视频生成的目标是生成连续图片（每张图片即一帧）的序列，它可以使用深度学习技术来生成高质量视频和动态内容，从而极大地提高视频的制作速度，也能够让视频内容更加逼真生动。

AIGC视频生成已经在很多行业得到了应用，并取得了不错的效果。学校可以使用AI视频生成技术来制作动画片或教学视频，医院也可以使用AI视频生成技术来模拟手术过程，帮助外科医生更好地理解手术流程。我们体验过的视频游戏、虚拟现实(VR)、视频会议等，都可能与AIGC视频生成的技术有关。

在AIGC视频生成技术逐渐成熟后，不少新兴科技公司也开始使用人工智能技术来进行影视制作，传统的影视制作方法与人工智能技术强强联合，能实现大规模的动态图像处理、自动剪辑、自动字幕添加、智能特效设计等，在影视制作中也能极大地解放人力和物力，压低制作成本。

AI影视制作的案例颇多，如电脑艺术家格伦·马歇尔(GlennMarshall)的人工智能电影《乌鸦》(The Crow)就获得了2022年戛纳短片电影节评审团奖。《乌鸦》的基础是视频网站上的短片Painted，马歇尔将其输入OpenAI创建的神经网络中，然后指导另一个模型生成图像，这样就生成了一段关于“荒凉风景中的乌鸦”的视频。在电影《速度与激情7》中，剧组将虚拟演员“放置”到视频中，实现虚拟与现实的完美融合，减轻人物和场景的限制，实现更多可能。这种效果是怎样实现的呢？这涉及多重技术支持：首先从之前的镜头中选择拍摄所需的动作和表情，建立数字成像模型，再渲染出虚拟的人物；在替身演员拍摄完肢体动作后，还会对脸部进行替代。通过这种方式，逝去的保罗·沃克在电影《速度与激情7》中“重生”，为影迷带来了慰藉。

在AIGC视频制作赛道同样有很多“明星企业”。2023年2月6日，人工智能初创公司Runway官网宣布推出AI视频生成模型Gen-1，给竞争已十分激烈的AIGC赛道又添了一把熊熊烈火。Gen-1究竟有什么令人惊叹之处呢？它采用了最新的深度学习编码技术，可以将数据转化为精美的3D图像和视频，还能根据文字脚本、图片、视频剪辑等进行自动内容生成，创造出真实感十足的3D场景，帮助用户体验真实世界中所不能触及的情景，比如现在无法实现的太空旅行、历史重现等，小说中的“穿越”情节可以在现实中上演，给生活带来了无尽想象和无限可能。此外，Runway还提到会不断改进Gen-1，让其以更低的成本和更快的速度，生成更精彩的内容，为人类提供无尽的创意。

欢迎访问：电子书学习和下载网站 (<https://www.shgis.com>)

文档名称：《生成式人工智能》丁磊.epub

请登录 <https://shgis.com/post/5122.html> 下载完整文档。

手机端请扫码查看：

