

# AIGC

## 智能创作时代

杜 雨 张孜铭◎著



从机器学习到智能创造  
生成式人工智能，一场科技与艺术的碰撞

朱嘉明

著名经济学家

童 晨

IDG资本合伙人

刘天池

中央戏剧学院教授

乔尼·森德

芬兰图尔库大学客座教授

梁 柱

腾讯音乐娱乐集团CEO

袁佛玉

百度集团副总裁

联袂推荐

中国出版集团  
中 译 出 版 社



# AIGC

## 智能创作时代

杜雨 张孜铭◎著



从机器学习到智能创造  
生成式人工智能，一场科技与艺术的碰撞

朱嘉明  
著名经济学家  
童晨  
IDG资本合伙人

刘天池  
中央戏剧学院教授  
乔尼·森德  
芬兰图尔库大学客座教授

梁柱  
腾讯音乐娱乐集团CEO  
袁佛玉  
百度集团副总裁

联袂推荐

中国出版集团  
中译出版社

**AIGC：智能创作时代**

杜雨 张孜铭◎著

中国出版集团

中译出版社

## 版权信息

图书在版编目（CIP）数据

AIGC：智能创作时代/杜雨，张孜铭著.--北京：中译出版社，2023.3

ISBN 978-7-5001-7345-8

I. ①A... II. ①杜...②张... III. ①人工智能-经济学-通俗读物 IV. ①TP18-05

中国版本图书馆CIP数据核字（2023）第018660号

**AIGC：智能创作时代**

**AIGC：ZHINENG CHUANGZUO SHIDAI**

著 者：杜雨 张孜铭

策划编辑：于宇 田玉肖 薛宇

责任编辑：于宇 龙彬彬 方荟文 李晟月 朱小兰

文字编辑：田玉肖 华楠楠 李梦琳

营销编辑：马萱 纪菁菁

出版发行：中译出版社

地 址：北京市西城区新街口外大街28号102号楼4层

电 话：（010）68002494（编辑部）

邮 编：100088

电子邮箱：book@ctph.com.cn

网 址：<http://www.ctph.com.cn>

印 刷：北京顶佳世纪印刷有限公司

经 销：新华书店

规 格：710mm×1000mm 1/16

印 张：16.25

字 数：171千字

版 次：2023年3月第1版

印 次：2023年3月第1次印刷

ISBN 978-7-5001-7345-8 定价：69.00元

版权所有 侵权必究

中译出版社

# 代序 AIGC和智能数字化新时代

——媲美新石器时代的文明范式转型

数字时代：“代码即法律”（Code is law）

——劳伦斯·莱斯格（Lawrence Lessig）

智能时代：向量和模型构成一切（Vector and models rule it all）

——朱嘉明

2022年，在集群式和聚变式的科技革命中，人工智能生成内容（AIGC，AI Generated Content）后来居上，以超出人们预期的速度成为科技革命历史上的重大事件，迅速催生了全新的科技革命系统、格局和生态，进而深刻改变了思想、经济、政治和社会的演进模式。

第一，AIGC的意义是实现人工智能“内容”生成。人们主观的感觉、认知、思想、创造和表达，以及人文科学、艺术和自然科学都要以具有实质性的内容作为基础和前提。所以，没有内容就没有人类文明。进入互联网时代后，产生了所谓专业生成内容（PGC），也出现了以此作为职业获得报酬的职业生成内容（OGC）。与此同时，“用户生成内容”（UGC）的概念和技术也逐渐发展，由此形成了用户内容生态。

内容生产赋予了Web2.0的成熟和Web3.0时代的来临。相较于PGC和OGC、UGC，AIGC通过人工智能技术实现内容生成，并在生成中注入了“创作”，意味着自然智能所“独有”和“垄断”的写作、绘画、音乐、教育等创造性工作的历史走向终结。内容生成的四个阶段如图0-1所示。

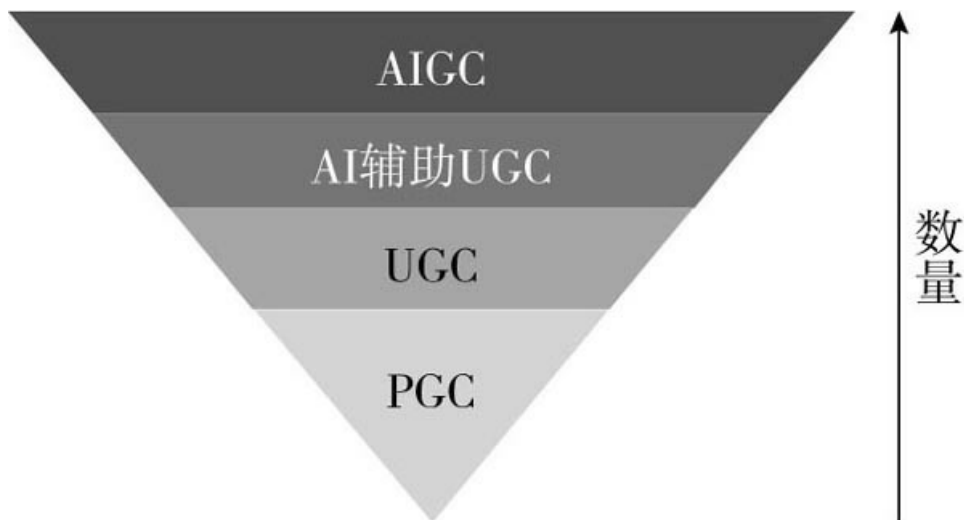


图0-1 内容生成的四个阶段

第二，AIGC的核心技术价值是实现了“自然语言”与人工智能的融合。自然语言是一个包括词法、词性、句法、语义的体系，也是不断演进的动态体系。代表AIGC最新进展的是由OpenAI公司开发的ChatGPT（Chat Generative Pre-trained Transformer）。它完成了机器学习算法发展中，自然语言处理领域的历史性跨越，即通过大规模预训练模型，形成人工智能技术理解自然语言和文本生成能力，可以生成文字、语音、代码、图像、视频，且能完成脚本编写、文案撰写、翻译等任务。这是人类文明史上翻天覆地的革命，开启了任何阶层、任何职业都可以以任何自然语言和人工智能交流，并且生产出从美术作品到学术论文的多样化内容产品。在这样的过程中，AIGC“异化”为一种理解、超越和生成各种自然语言文本的超级“系统”。

第三，AIGC的绝对优势是其逻辑能力。是否存在可以逐渐发展的逻辑推理能力是人工智能与生俱来的挑战。AIGC之所以迅速发展，是因为AIGC基于代码、云计算、技术操控数据、模式识别，以及通过机器对文本内容进行描述、分辨、分类和解释，实现了基于语言模型提示学习的推理，甚至是知识增强的推理，构建了坚实的“底层逻辑”。不仅如此，AIGC具备基于准确和规模化数据，形成包括学习、抉择、尝试、修正、推理，甚至根据环境反馈调整并修正自己行为的能力；它可以突破线性思维框架并实现非线性推理，也可以通过归纳、演绎、分析，实现对复杂逻辑关系的描述。可以毫不夸张地说，AIGC已经并继续改变着21世纪逻辑学的面貌。

第四，AIGC实现了机器学习的集大成。21世纪的机器学习演化到了深度学习（Deep learning）阶段。深度学习可以更有效地利用数据特征，形成深度学习算法，解决更为复杂的场景挑战。2014年生成对抗网络（GAN）的出现，加速了深度学习在AIGC领域的应用。AIGC实现了机器学习的集大成（图0-2）。



逻辑回归	决策树	随机森林	支持向量机	神经网络	KNN	朴素贝叶斯	梯度boosting	K-Means 聚类	层次聚类	DBSCAN 聚类	AP 聚类	谱聚类	PCA	t-SNE	SVD
								聚类				降维			
监督学习								非监督学习							
机器学习常用算法															
深度学习							强化学习								
CNN	RNN	AE	GAN	Transformer	Q-learning										
					SARSA										
					PPO										
					DQN										
					DDPG										
					A3C										

图0-2 机器学习常用算法

资料来源：程序员zhenguo（2023），“梳理机器学习常用算法（含深度学习）”

第五，AIGC开创了“模型”主导内容生成的时代。人类将跑步进入传统人类内容创作和人工智能内容生成并行的时代，进而进入后者逐渐走向主导位置的时代。这意味着传统人类内容创作互动模式转换为AIGC模型互动模式。2022年是重要的历史拐点（图0-3）。

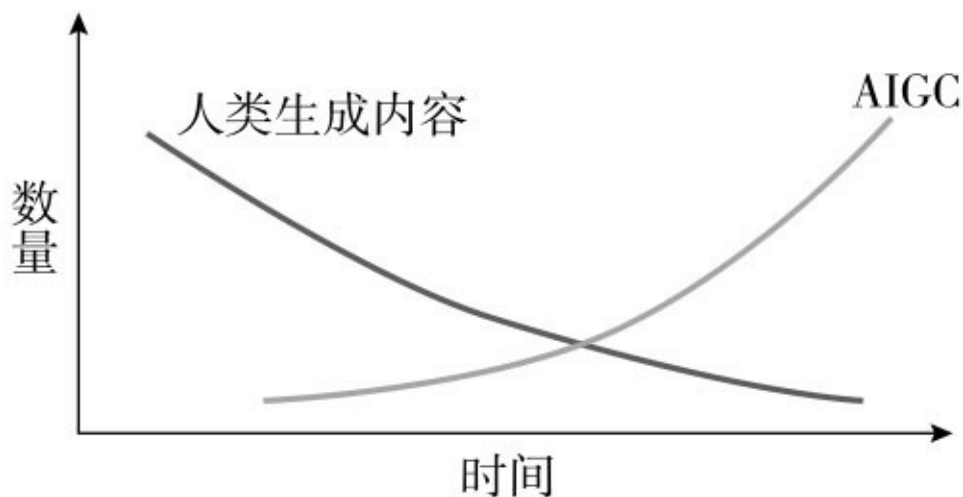


图0-3 人类生成内容向AIGC转换趋势

而在自然语言处理（NLP）系统中，“Transformer”是一种融入注意力机制和神经网络模型领域的主流模型和关键技术。Transformer具有将所处理的任何文字和句子“向量”或者“矢量”化，最大限度反映精准意义的能力。

总之，没有Transformer，就没有NLP的突破；没有大模型化的AIGC，ChatGPT升级就没有可能。多种重要、高效的Transformer的集合模型如图0-4所示。

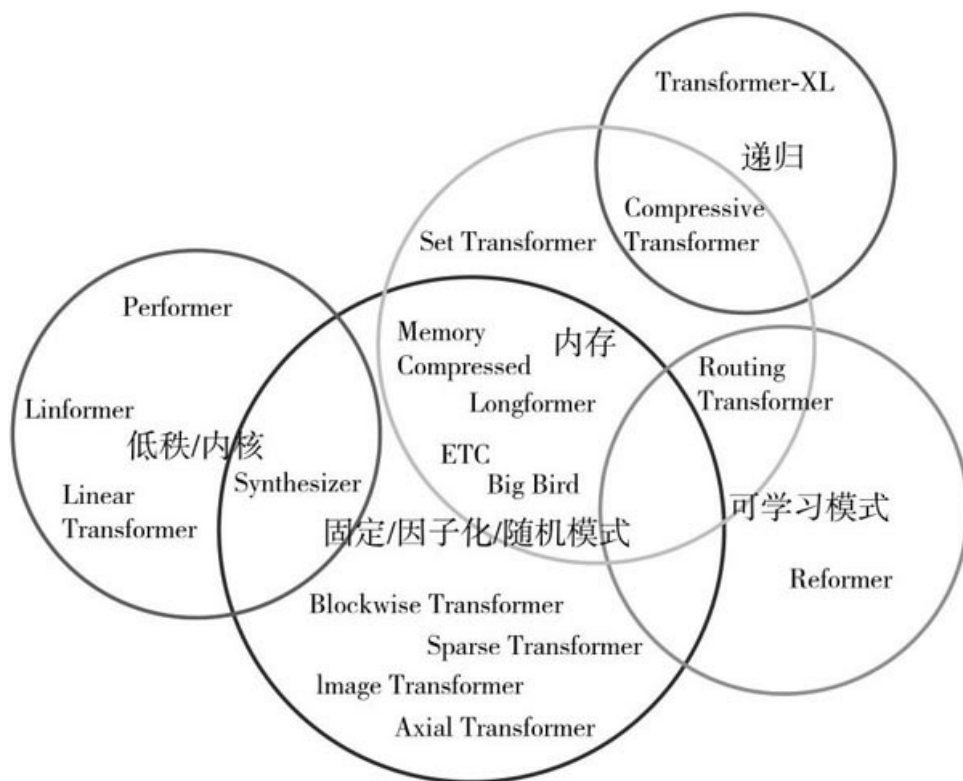


图0-4 多种重要、高效的Transformers的集合模型

资料来源: Tay et al (2022), “Efficient Transformers: A Survey”, doi: 10.48550/arXiv.2009.06732

第六, AIGC开放性创造力的重要来源是扩散 (Diffusion) 模型。扩散模型的概念最早在2015年的论文《利用非均衡热力学的深度非监督学习》 (*Deep Unsupervised Learning Using Nonequilibrium Thermodynamics*) 中被提出。<sup>[1]</sup>2020年, 论文《去噪扩散概率模型》 (*Denoising Diffusion Probabilistic Models*) 中提出DDPM模型用于图像生成。<sup>[2]</sup>从技术的角度来看, 扩散模型是一个潜在变量 (Latent Variable) 模型, 通过马尔可夫链 (Markov chain) 映射到潜在空间。<sup>[3]</sup>一般来说, AIGC因为吸纳和依赖扩散模型, 而拥有开放性创造力。

2021年8月, 斯坦福大学联合众多学者撰写论文, 将基于Transformer架构等的模型称为“基础模型” (Foundation model), 也常译作大模型。Transformer推动了AI整个范式的转变 (图0-5)。

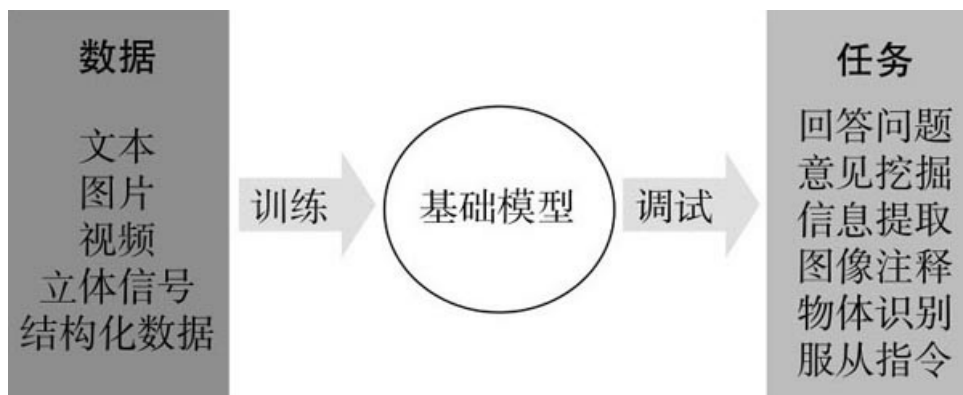


图0-5 基础模型“Transformer”

资料来源：Bommasani et al（2022），“On the Opportunities and Risks of Foundation Models”，doi: 10.48550/arXiv.2108.07258

第七，AIGC的进化是参数以几何级数扩展为基础。AIGC的训练过程，就是调整变量和优化参数的过程。所以，参数的规模是重要前提。聊天机器人ChatGPT的问世，标志着AIGC形成以Transformer为架构的大型语言模型（Large Language Model，简称LLM）机器学习系统，通过自主地从数据中学习，在对大量的文本数据集进行训练后，可以输出复杂的、类人的作品。

AIGC形成的学习能力取决于参数的规模。GPT-2大约有15亿个参数，而GPT-3最大的模型有1 750亿个参数，上升了两个数量级。而且，它不仅参数规模更大，训练所需的数据也更多。根据媒体报道但还未被证实的消息，GPT-4的参数可能将达到100万亿规模（图0-6）。

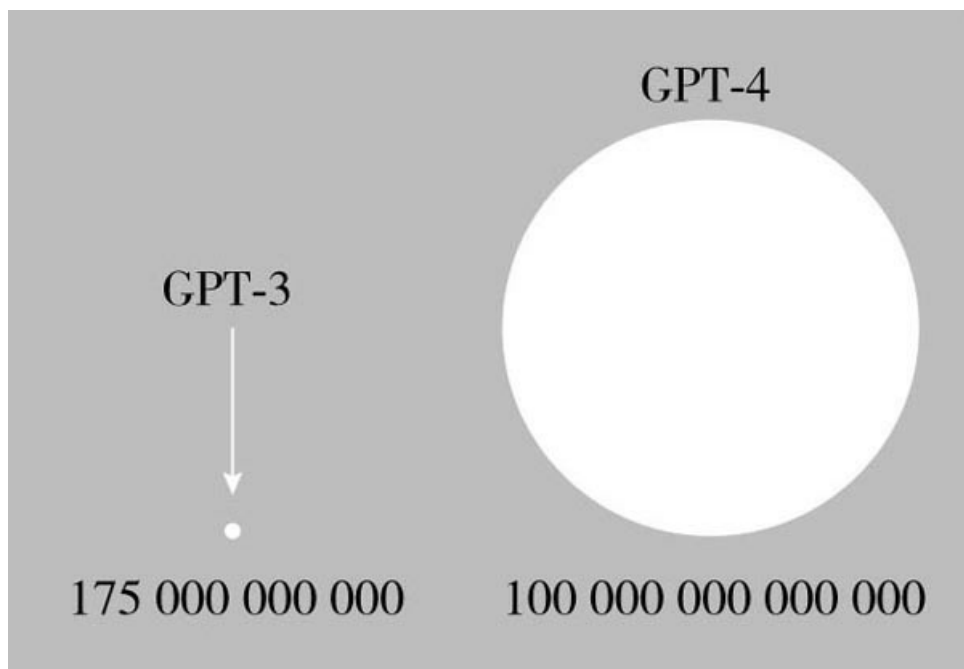


图0-6 GPT-4的参数规模

根据学界经验，神经网络的学习能力和模型的参数规模呈正相关。人类的大脑皮层有140多亿个神经细胞，每个神经细胞又有3万多个突触。所以，大脑皮层的突触总数超过100万个。所谓的神经细胞就是通过这些突触相互建立联系。假设GPT-4实现100万亿参数规模，堪比人的大脑，意味着它达到与人类大脑神经触点规模的同等水平。

第八，AIGC的算力需求呈现显著增长。数据、算法、算力是人工智能的稳定三要素。根据OpenAI分析，自2012年以来，6年间AI算力需求增长约30万倍（图0-7）。

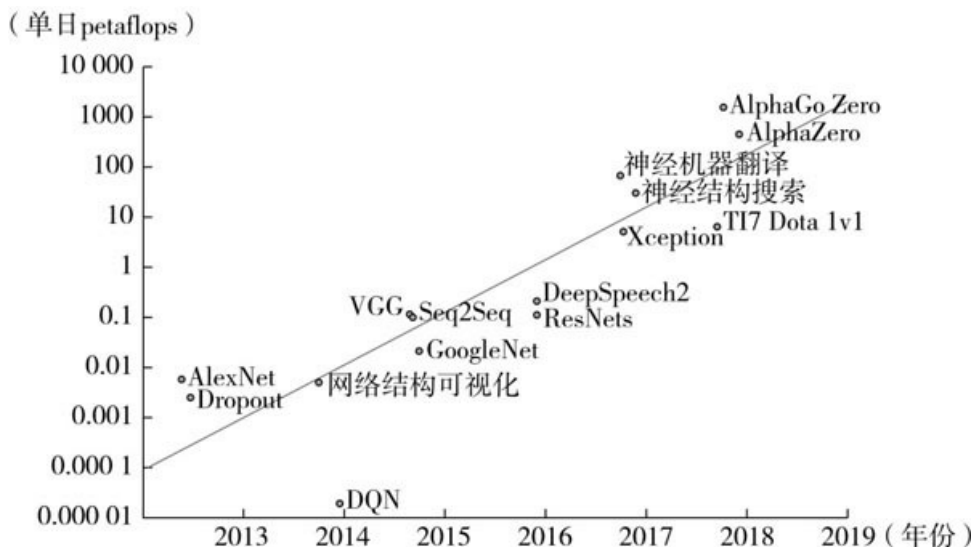


图0-7 从AlexNet到AlphaGo Zero：30万倍的运算量增长

资料来源：OpenAI（2018），“AI and Compute”，<https://openai.com/blog/ai-and-compute/>

在可以预见的未来，在摩尔定律（Moore’s Law）已走向失效的情况下，AI模型所需算力被预测每100天翻一倍，也就是“5年后AI所需算力超100万倍”。<sup>[4]</sup>造成这样需求的根本原因是AI的算力不再是传统算力，而是“智能算力”，是以多维度的“向量”集合作为算力基本单位。

第九，AIGC和硬技术相辅相成。从广义上讲，AIGC的硬技术是AI芯片，而且是经过特殊设计和定制的AI芯片。AI芯片需要实现CPU、GPU、FPGA和DSP共存。随着AIGC的发展，计算技术的发展不再仅仅依靠通用芯片在制程工艺上的创新，而是结合多种创新方式，形成智能计算和计算智能技术。例如，根据应用需求重新审视芯片、硬件和软件的协同创新，即思考和探索新的计算架构，满足日益巨大、复杂、多元的各种计算场景。其间，量子计算会得到突破性发展。

第十，AIGC将为区块链、NFT、Web3.0和元宇宙带来深层改变。AIGC不可枯竭的创造资源和能力，将从根本上改变目前的NFT概念生态。Web3.0结合区块链、智能合约、加密货币等技术，实现去中心化理念，而AIGC是满足这个目标的最佳工具和模式。

没有悬念，在Web3.0的环境下，AIGC内容将出现指数级增长。元宇宙的本质是社会系统、信息系统、物理环境形态通过数字构成了一个动态耦合的大系统，需要大量的数字内容来支撑，人工来设计和开发根本无法满足需求，AIGC可以最终完善元宇宙生态的底层基础设施。随着AIGC技术的逐渐成熟，传统人类形态不可能进入元宇宙这样的虚拟世界。未来的元宇宙主体将是虚拟人，即经过AIGC技术，特别融合ChatGPT技术，以代码形式呈现的模型化的虚拟人。

简言之，区块链、NFT、Web3.0，将赋予AIGC进化的契机。AIGC的进化，将加速广义数字孪生形态与物理形态的平行世界形成。

第十一，AIGC催生出全新的产业体系和商业化特征。AIGC利用人工智能学习各类数据自动生成内容，不仅能帮助提高内容生成的效率，还能提高内容的多样性。文字生成、图片绘制、视频剪辑、游戏内容生成皆可由AI替代，并正在加速实现，使得AIGC进而渗透和改造传统产业结构。在产业生态方面，AIGC领域正在加速形成三层产业生态并持续创新发展，正走向模型即服务（MaaS）的未来（图0-8）。

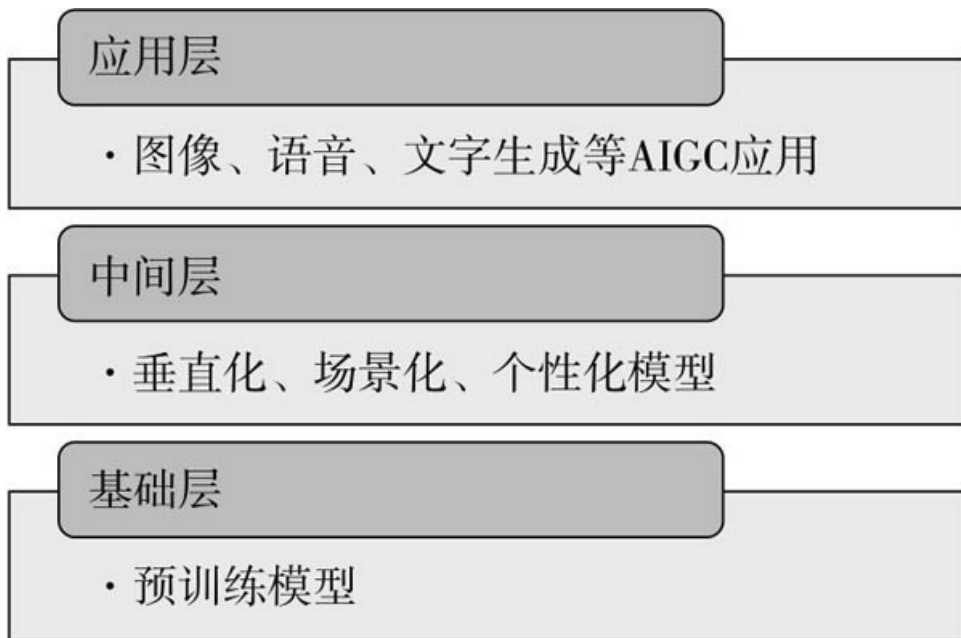


图0-8 AIGC产业生态

资料来源：腾讯《AIGC发展趋势报告》，2023年1月31日发布

伴随AIGC生成算法的优化与改进，AIGC对于普通人来说也不再是一种遥不可及的尖端技术。AIGC在文字、图像、音频、游戏和代码生成中的商业模型渐显。2B（to Business的简称）将是AIGC的主要商业模式，因为它有助于B端提高效率和降低成本，以填补数字鸿沟。但可以预见，由于AIGC“原住民”的成长，2C（to Consumer的简称）的商业模式将接踵而来。根据有关机构预测，2030年的AIGC市场规模将超过万亿人民币，其产业规模生态如图0-9所示。

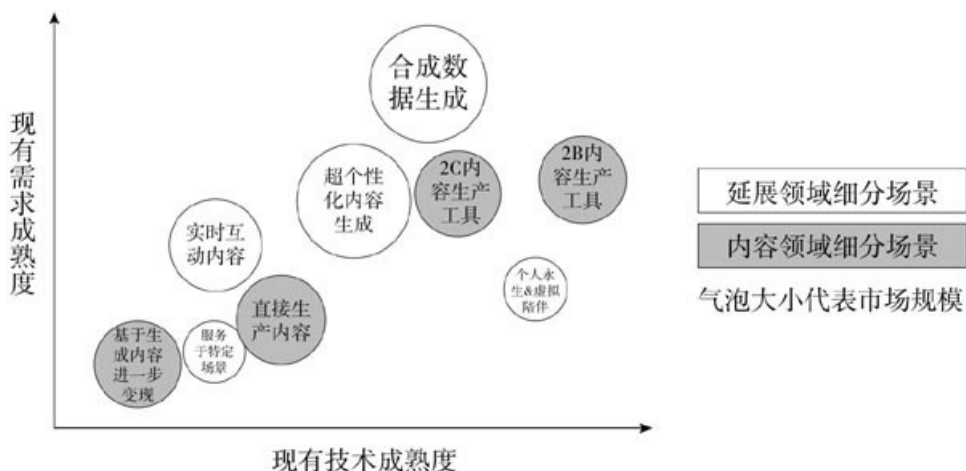


图0-9 AIGC产业规模生态分布

资料来源：陈李，张良卫（2023），“ChatGPT：又一个“人形机器人”，东吴证券，  
[https://www.nxny.com/report/view\\_5185573.html](https://www.nxny.com/report/view_5185573.html)

现在，AIGC，特别是在语言模型领域的全方位竞争已经开始。所以，发生了微软对OpenAI的大规模投资，因为有这样一种说法：“微软下个十年的想象力，藏在ChatGPT里。”近日，谷歌宣布推出基于“对话应用语言模型”（LaMDA）的Bard，实现其搜索引擎将包括人工智能驱动功能。ChatGPT刺激谷歌开始“创新者困境”突围。未来很可能出现Bard和ChatGPT的对决或共存，也就是LaMDA和GPT-3.5的对决和共存，构成AIGC竞争和自然垄断的新生态。

在这样的新兴产业构造和商业模式下，就业市场会发生根本性改变：其一，专业职场重组，相当多的职业可能衰落和消亡；其二，原本支持IT和AI产业的码农队伍面临严重萎缩。因为AIGC将极大地刺激全球外包模式并取代码农。

第十二，AIGC的法律影响和监管。虽然AIGC这样的新技术提供了很多希望，但也会给法律、社会和监管带来挑战。在中国，继2022年1月国家互联网信息办公室、工业和信息化部、公安部、国家市场监督管理总局联合发布《互联网信息服务算法推荐管理规定》后，2022年11月，国家互联网信息办公室再次会同工业和信息化部、公安部联合发布《互联网信息服务深度合成管理规定》。该规定的第五章第二十三条，对“深度合成技术”内涵做了规定：“利用深度学习、虚拟现实等生成合成类算法制作文本、图像、音频、视频、虚拟场景等网络信息的技术。”但可以预见，由于AIGC技术日趋复杂，并将得到高速发展，国家很难避免监管缺乏专业性和滞后性。

第十三，AIGC正在引领人类加速逼近“科技奇点”。现在，人工智能已经接管世界；世界正在经历一波人工智能驱动的全球思想、文化、经济、社会和政治的转型浪潮。AIGC呈现指数级的发展增速，开始重塑各个行业乃至全球的“数字化转型”。说到底，这就是以AIGC为代表，以ChatGPT为标志的转型。这一切，在2023年会有长足的发展，特别是在资本和财富效益领域。<sup>[5]</sup>



如果说，2022年8月的AI绘画作品《太空歌剧院》（*Théâtre D'opéra Spatial*）推动AIGC进入大众视野，那么，ChatGPT的底层模型GPT-3.5是一个划时代的产物。它之前常见的语言模型（BERT/BART/T5）的区别几乎是导弹与弓箭的区别。现在，呼之欲出的GPT-4，很可能通过图灵测试。<sup>[6]</sup>如果是这样，不仅意味着GPT-4系统可以改造人类的思想和创作能力，形成人工智能超越专业化族群和大众化趋势，而且意味着这个系统开始具备人类思维能力，并有可能在某些方面和越来越多的方面替代人类。<sup>[7]</sup>

特别值得关注的是被称为“人工智能激进变革先锋”的BLOOM（大型开放科学获取多语言模型）的诞生。从2021年3月11日到2022年7月6日，60个国家和250多个机构的1 000多名研究人员，在法国巴黎南部的超级计算机上整整训练了117天，创造了BLOOM。这无疑是一场意义深远的历史变革的前奏。

斯坦福大学心理学和计算机科学助理教授丹尼尔·亚明斯（Daniel Yamins）说过：“人工智能网络并没有直接模仿大脑，但最终看起来却像大脑一样，这在某种意义上表明，人工智能和自然之间似乎发生了某种趋同演化。”<sup>[8]</sup>

2005年，雷·库茨维尔（Ray Kurzweil，1948—）的巨著《奇点临近：当计算机智能超越人类》（*The Singularity Is Near: When Humans Transcend Biology*）出版。该书通过推算奇异点指数方程，得出了这样一个结论：“在2045年左右，世界会出现一个奇异点。这件事必然是人类在某项重要科技上，突然有了爆炸性的突破，而这项科技将完全颠覆现有的人类社会。它不是像手机这种小的奇异点，而是可以和人类诞生对等的超大奇异点，甚至大到可以改变整个地球所有生命的运作模式。”

现在处于狂飙发展状态的AIGC，一方面已经开始呈指数形式膨胀，另一方面其“溢出效应”正在改变人类本身。在这个过程中，所有原本看来离散和随机的科技创新和科技革命成果，都开始了向AIGC技术的收敛，人工智能正在形成自我发育和完善的内在机制，推动人类社会快速超越数字化时代，进入数字化和智能化时代，逼近可能发生在2045年的“科技奇点”。

朱嘉明

2023年2月8日

[1] Sohl-Dickstein et al（2015），“Deep Unsupervised Learning Using Nonequilibrium Thermodynamics”，doi: <https://doi.org/10.48550/arXiv.1503.03585>.

[2] Ho et al（2020），“Denoising Diffusion Probabilistic Models”，doi: 10.48550/arXiv.2006.11239.

[3] 马尔可夫链的命名来自俄国数学家安德雷·马尔可夫（Andrey Andreyevich Markov，1856—1922），定义为概率论和数理统计中具有马尔可夫性质，且存在于离散的指数集和状态空间内的随机过程。马尔可夫链可能具有不可约性、常返性、周期性和遍历性。

[4] 新智元《5年后AI所需算力超100万倍》，2023年1月31日，发表于北京。

[5] AI产业在2022年接近3 874.5亿美元，预计到2029年将超过13 943亿美元，可谓市场机会巨大。2023年，全球企业在人工智能方面的支出将突破5 000亿美元。

[6] 根据韩国IT媒体报道，自2022年11月中旬开始，业界已经传出了GPT-4全面通过了图灵测试的说法。

[7] 根据Metaverse Post 消息，ChatGPT通过了美国宾大沃顿商学院MBA的考试。如果消息属实，近乎完成图灵试验。<https://mpost.io/zh-CN/chatgpt-passes-the-whartonmba-exam/>.

[8] Anne Trafton (2021), “Artificial intelligence sheds light on how the brain processes language”, <https://news.mit.edu/2021/artificial-intelligence-brain-language-1025>.

## 前言 从机器学习到智能创造

不知道你有没有想过这样一个问题：是什么让我们得以思考？

从如同一张白纸的婴儿，成长为洞悉世事的成人，正是长辈的教诲和十年寒窗塑造了我们如今的思考力。学习，似乎就是智能形成的最大奥秘。

人类崇尚智能，向往智能，并不断利用智能改造世界。走过农业革命，迈过工业革命，迎来信息革命，一次又一次对生产力的改造让人们相信，人类的智能最终也能创造出人工的智能。

数十年前，图灵抛出的时代之问“机器能思考吗？”将人工智能从科幻拉至现实，奠定了后续人工智能发展的基础。之后，无数计算机科学的先驱开始解构人类智能的形成，希望找到赋予机器智能的蛛丝马迹。正如塞巴斯蒂安·特伦所言：“人工智能更像是一门人文学科。其本质在于尝试理解人类的智能与认知。”如同人类通过学习获得智能一样，自20世纪80年代起，机器学习成为人工智能发展的重要力量。

机器学习让计算机从数据中汲取知识，并按照人类所期望的，按部就班执行各种任务。机器学习在造福人类的同时，似乎也暴露出了一些问题，这样的人工智能并非人类最终期望的模样，它缺少了人类“智能”二字所涵盖的基本特质——创造力。这个问题就好像电影《我，机器人》中所演绎的一样，主角曾与机器人展开了激烈的辩论，面对“机器人能写出交响乐吗？”“机器人能把画布变成美丽的艺术品吗？”等一连串提问，机器人只能讥讽一句：“难道你会？”这也让创造力成为区分人类与机器最本质的标准之一。

面对庐山雄壮的瀑布时，李白写出“飞流直下三千尺，疑是银河落九天”的千古绝句，感慨眼前的壮丽美景；偶遇北宋繁荣热闹的街景时，张择端绘制出《清明上河图》这样的传世名画，记录下当时的市井风光与淳朴民风；邂逅汉阳江口的知音时，伯牙谱写出《高山流水》，拉近了秋夜里两位知己彼此的心灵。我们写诗，我们作画，我们谱曲，我们尽情发挥着创造力去描绘我们的所见所闻，我们因此成为人类的一分子，这既是智能的意义，也是我们生活的意义。

但是，人类的创造力真的不能赋予机器创造力吗？答案显然是否定的。

在埃米尔·博雷尔1913年发表的《静态力学与不可逆性》论文中，曾提出这样一个思想实验：假设猴子学会了随意按下打字机的按钮，当无限只猴子在无限台打字机上随机乱敲，并持续无限久的时间，在某个时刻，将会有猴子能打出莎士比亚的全部著作。虽然最初这只是一个说明概率理论的例子，但它也诠释了机器具备创造力的可能性。只不过具备的条件过于苛刻，需要在随机性上叠加无穷的时间量度。

在科学家们的不懈努力下，这个时间量度被从无限缩减至了有限。随着深度学习的发展和大模型的广泛应用，生成型人工智能已经走向成熟，人们沿着机器学习的路，探索出了如今的智能创造。在智能创作时代，机器能够写诗，能够作画，能够谱曲，甚至能够与人类自然流畅地对话。人工智能生成内容（AIGC）将带来一场深刻的生产力变革，而这场变革也会影响人们工作与生活的方方面面。本书希望通过生动的比喻和有趣的案例，用浅显易懂的语言，让每个人都能真切地参与到这一次轰轰烈烈的科技革命中，一起迎接全新的智能创作时代。

本书由杜雨、张铭铭负责统筹和编写，其他对本书内容做出贡献的编写者包括：胡宇桐、张之耀参与编写第一章、第六章第二节；李悦莹协助制作第一章、第二章部分图表；庞舜心参与编写第三章、第六章第一节；袁誉铭参与编写第四章第一、第二、第四节；刘子源参与编写第四章第

三节；段靖宇参与编写第五章第三节；郭雨萍、王芸参与编写第六章第二、第三节。

在本书的编写过程中，感谢未知和QQ（Quadratic Acceleration Quantum）大家庭所有成员一直以来对我们的鼓舞，感谢中译出版社和所有好友对本书的支持，感谢陈定媛、相子恒、徐臻哲为本书编校提供的建议和帮助。

编者

**2023**年**1**月

- 第一章 AIGC：内容生产力的大变革
  - 第一节 从PGC、UGC到AIGC
  - 第二节 人工智能赋能内容创作的四大模态
  - 第三节 AIGC助力元宇宙和Web3.0
- 第二章 AIGC的技术思想
  - 第一节 前AIGC时代的技术奠基
  - 第二节 早期AIGC的尝试：GAN
  - 第三节 AI绘画的推动者：Diffusion模型
  - 第四节 大模型的重要基建：Transformer
- 第三章 AIGC的职能应用
  - 第一节 AIGC与产品研发
  - 第二节 AIGC与市场营销
  - 第三节 AIGC与管理协作
- 第四章 AIGC的行业应用
  - 第一节 AIGC资讯行业应用
  - 第二节 AIGC影视行业应用
  - 第三节 AIGC电商行业应用
  - 第四节 AIGC教育行业应用
  - 第五节 AIGC金融行业应用
  - 第六节 AIGC医疗行业应用
- 第五章 AIGC的产业地图
  - 第一节 产业上游：数据服务
  - 第二节 产业中游：算法模型
  - 第三节 产业下游：应用拓展
- 第六章 AIGC的未来
  - 第一节 AIGC的技术趋势
  - 第二节 AIGC时代的参与主体
  - 第三节 AIGC的风险与监管
- 附录一 AIGC产业地图标的公司列表（部分）
- 附录二 AIGC术语及解释
- 附录三 AIGC大事记
- 后记
- 专家推荐

## 第一章 AIGC：内容生产力的大变革

AIGC如何从生产力角度促进当今数字经济的发展？

万物的智能成本无限降低，人类的生产力与创造力得到解放。

——山姆·阿尔特曼（Sam Altman）

人工智能经历了从科幻小说走向现实应用的漫长历程，如今已走进人们的日常生活。几十年前，科学家的普遍观念也许如阿达·洛夫莱斯（Ada Lovelace）所言：“机器不会自命不凡地创造任何事物，它只能根据我们能够给出的任何指令完成任务。”计算机科学的先驱也许预料到了人工智能的迅猛发展，但我们相信他们依然会对今天人工智能取得的成就感到震惊。

自工业革命以来，“是否具备创造力”就被视为人类和机器最本质的区别之一。然而，今天的人工智能却打破了持续数百年的铁律。人工智能可以表现出与人类一样的智慧与创意，例如撰写诗歌、创作绘画、谱写乐曲，而人类创造出的智能又将反哺人类自身的智能。AIGC（Artificial Intelligence Generated Content，人工智能生成内容）的兴起极大地解放了人类的内容生产力，将数字文明送入智能创作时代。我们有幸处于时代浪潮之巅，见证由技术进步带来的全新变革。下面就让我们一起走进AIGC的世界，探索智能创作时代的无限可能。

第一节 从PGC、UGC到AIGC

生产力是推进社会变革的根本动力，而生产工具则是衡量生产力发展水平的客观尺度，也是划分经济时代的物质标志。从钻木取火到机器大生产，生产力的发展推动了从农业社会到工业社会的社会跃迁。自第三次科技革命之后，互联网成为连接人类社会的主要媒介，内容则是人们生产和消费的主要产品。互联网经历了Web1.0、Web2.0、Web3.0与元宇宙时代，不同互联网形态下也孕育了相辅相成的内容生产方式，并一直沿用至今。表1-1呈现了内容生产方式从PGC（Professional-Generated Content，专业生成内容）到UGC（User-Generated Content，用户生成内容），再到AIGC的发展历程。下面就让我们一起来了解一下每个内容生产时代的特点与故事吧！

表1-1 从PGC到UGC，再到AIGC的发展历程

互联网形态	Web1.0	Web2.0	Web3.0 与元宇宙
内容生产方式	PGC（专业生产）	UGC（用户生产）	AIGC（AI生产）
生产主体	专业人	非专业人	非人
核心特点	内容质量高	内容丰富度高	内容生产效率高

一、PGC：专家创作时代

20世纪90年代，伴随着万维网的诞生与推广，互联网领域迎来了投资创业的热潮，正式进入了Web1.0阶段。在这个阶段，一种基于“信息经济”的全新商业模式孕育而生，互联网技术提供商不仅提供技术服务，还能从生产与组织内容的流量曝光中获得收益。此时的互联网是静态互联网，大多数用户只能在网上浏览和读取信息，内容的创建与发布只掌握在极少数专家手中。不过，这里的专家未必是内容领域的专家，他们只是通过专业的方式将信息聚合在一起，便利地提供给用户浏览，门户网站、浏览器、搜索引擎是当时最主要的产品。通过专业方式聚合、筛选并呈现出来的内容大多具有专业性，是由专业人士生产的高质量内容，这种内容生产方式被称为PGC。雅虎的综合指南网站以及亚马逊的互联网电影资料库

（IMDb）就是典型的产品代表，前者提供包含互联网图文内容的查询工具，后者则聚合了优质的电影、电视节目等视频内容的相关信息。

在Web1.0阶段，虽然互联网上的主要内容大多是由专家生产的，可以说是专家创作的时代，但后来诸多内容平台、互联网媒体机构、知识付费公司的创立与发展，才真正促使现在普遍意义上PGC概念的形成。现在的PGC主要是指由专家与专业机构负责生产内容，因为他们具备专业的内容生产能力，能够保证内容的专业性。对于内容本身是否专业或许有不同的

评价标准，但人们更多是从创作主体的性质来界定内容生产方式是否属于PGC。根据创作主体过往的作品质量，人们可以更统一地界定内容的“专业性”。创作者一般会根据明确的用户需求对内容进行加工，借助高质量内容本身的原创性和价值赚取收益，例如版权作品、在线课程的销售等。而高价值的内容也会收获更多用户的关注，在获得一定流量的基础上，通过广告等方式进行变现也是常见路径之一。直至今日，这种最早出现的互联网内容生产方式依然陪伴在我们左右，无论是爱奇艺、腾讯视频采购的影视剧综，还是36氪、虎嗅等专业媒体平台的新闻报道，抑或是得到、网易云课堂等平台的音视频课程，都属于PGC的范畴。

PGC虽然具有高质量、易变现、针对性强等优势，但也存在着明显的不足。因为专业的质量要求往往导致这类内容创作门槛高、制作周期长，由此带来了产量不足、多样性有限的问题。此外，由于生产成本低，采购平台或用户通常需要支付相对较高的成本来获取内容，从而导致普通用户的日常高频次、多样化的内容消费需求无法得到满足。基于上述原因，互联网需要新的内容生产形式来解决这些问题。

## 二、UGC：用户创作时代

伴随着互联网的发展和网民数量的增多，用户对多样化和个性化内容的需求也日渐增加。同时，许多用户也不再满足于单向的内容接收，而是希望自己也能够参与到内容的创作之中。21世纪初，众多社交媒体的出现迎合了这一需求，也宣告了互联网演化到了Web2.0形态——平台互联网。在Web2.0阶段，用户不仅是内容的消费者，也是内容的创作者，每一位用户的创造力都得到了前所未有的彰显。虽然PGC内容生产方式依然存在，但井喷式增长的UGC内容生产方式已成为时代的趋势所向。所谓UGC，指的是由所有普通用户生产内容，这些内容具有多样化的特征，并借由推荐系统等平台工具触达与内容匹配、具有相应个性化需求的用户。专业与否早已不是互联网内容创作的门槛，非专业人士也可以创作出大众喜欢的内容，这也让互联网迎来了用户创作时代。

在用户创作时代，整个互联网的内容丰富度都大大提升。贴吧、豆瓣等论坛平台上，志同道合的用户可以自由交流，一起探讨感兴趣的电影与书籍；微信、微博等社交平台上，每个人都可以用图文记录自己的生活，同时也能了解到他人的生活；抖音、快手等自媒体平台上，用户可以拍摄并上传自己创作的短视频，在获取大众关注的同时，还能获得各种流量变现的奖励。各类内容平台的角逐，也逐渐从高质量PGC内容的生产，转向有利于UGC创作者生态的构建。

与PGC类似，UGC突出的内容优势也必然伴随着不可避免的痛点，极其丰富的内容背后存在着内容质量参差不齐的问题，平台方需要投入大量精力和成本去进行创作者教育、内容审核、版权把控等方面的工作。此外，虽然从平台层面，内容生产供给的问题得到了解



决，但对于每个创作者个体而言，依然面临着内容质量、原创程度和更新频率的不可能三角，即上述三个方面不可能同时做到。相较于PGC的团队工作，UGC的创作者很多都是单打独斗，难以在保证内容质量、原创程度的情况下还能兼顾更新频率。而与此同时，创作者数量的增多使竞争变得更加激烈，许多创作者不得不选择降低质量、洗稿抄袭等捷径，用高频率的更新留住关注者。长此以往，健康的创作生态将遭到破坏，这种创作者的窘境呼吁着内容生产方式的全新变革，生产效率的提升已迫在眉睫。

### 三、AIGC：智能创作时代

面对互联网内容生产效率提升的迫切需求，人们突发奇想，是否能够利用人工智能去辅助内容生产呢？这种继PGC、UGC之后形成的、完全由人工智能生成内容的创作形式被称为AIGC。正如人们最初眺望Web3.0时构想的“语义网”一样，未来的互联网应该是更加智能的互联网，它不仅能够读懂各种语义信息，还能从信息角度解放人类的生产力。即便后来区块链技术的蓬勃发展改变了Web3.0的指代，元宇宙也展现出互联网浩瀚的未来，但内容的价值确权和虚拟空间的发展仍然需要更高效的内容生产方式，AIGC也就凝聚了人们对于未来的期待。

让人工智能这样的非人机器学会创作绝非易事，科学家在过往做了诸多尝试，并将这一研究领域称为生成式人工智能（Generative AI），主要研究人工智能如何被用于创建文本、音频、图像、视频等各种模态的信息。为了便于理解，本书并不打算对生成式人工智能和AIGC的概念加以区分，在后续的内容中将全部以AIGC作为指代。

最初的AIGC通常基于小模型展开，这类模型一般需要特殊的标注数据训练，以解决特定的场景任务，通用性较差，很难被迁移，而且高度依赖人工调参。后来，这种形式的AIGC逐渐被基于大数据量、大参数量、强算法的大模型（Foundation Model）取代，这种形式的AIGC无须经过调整或只经过少量微调（Fine-tuning）就可以迁移到多种生成任务。

2014年诞生的GAN（Generative Adversarial Networks，生成对抗网络）是AIGC早期转向大模型的重要尝试，它利用生成器和判别器的相互对抗并结合其他技术模块，可以实现各种模态内容的生成。而到了2017年，Transformer（变换器）架构的提出，使得深度学习模型参数在后续的发展中得以突破1亿大关，这种基于超大参数规模的大模型，为AIGC领域带来了前所未有的机遇。此后，各种类型的AIGC应用开始涌现，但并未获得全社会的广泛关注。

2022年下半年，两个重要事件激发了人们对AIGC的关注。2022年8月，美国科罗拉多州博览会上，数字艺术类冠军颁发给了由AI自动生成并经由Photoshop润色的画作《太空歌剧

院》，该消息一经发布就引起了轩然大波。该画作兼具古典神韵和太空的深邃奥妙，如此恢宏细腻的画风很难让人相信它是由AI自动生成的作品，而它夺得冠军的结果也大大冲击了人们过往对于“人工智能的创造力远逊于人”的固有认知，自此彻底引爆了人们对于AIGC的兴趣与讨论，AIGC也从看似遥远的概念逐步以生动有趣的方式走入人们的生活，带来了过去令人难以想象的丰富体验。

2022年11月30日，OpenAI发布了名为ChatGPT的超级AI对话模型，再次引爆了人们对于AIGC的讨论热潮。ChatGPT不仅可以清晰地理解用户的问题，还能如同人类一般流畅地回答用户的问题，并完成一些复杂任务，包括按照特定文风撰写诗歌、假扮特定角色对话、修改错误代码等。此外，ChatGPT还表现出一些人类特质，例如承认自己的错误，按照设定的道德准则拒绝不怀好意的请求等。ChatGPT一上线，就引发网民争相体验，到处都是体验与探讨ChatGPT的文章和视频。但也有不少人对此表示担忧，担心作家、画家、程序员等职业在未来都将被人工智能所取代。

虽然存在这些担忧，但人类的创造物终究会帮助人类自身的发展，AIGC无疑是一种生产力的变革，将世界送入智能创作时代。在智能创作时代，创作者生产力的提升主要表现为三个方面：

- 代替创作中的重复环节，提升创作效率。
- 将创意与创作相分离，内容创作者可以从人工智能的生成作品中找寻灵感与思路。
- 综合海量预训练的数据和模型中引入的随机性，有利于拓展创新的边界，创作者可以生产出过去无法想出的杰出创意。

即便如此，AIGC也并非完美无缺的，“人工智能生成的内容如何确定版权归属”“AIGC是否会被不法分子利用，生成具有风险性的内容或用于违法犯罪活动”等一系列问题都是现在人们争论的焦点。目前，学界与业界在尝试从各个方面解决这些问题。但不管怎样，AIGC的迅猛发展已成不可逆转之势，智能创作时代的序幕正在缓缓拉开。

## 第二节 人工智能赋能内容创作的四大模态

本节将从文本、音频、图像、视频四大模态角度介绍人工智能赋能内容创作的相关案例。不过，为了更全面地介绍不同模态内容的生成应用，本节提供的案例将不仅仅包括引起本次AIGC热潮的大模型应用，还包括利用传统小模型的相关生成应用。

### 一、AI文本生成

2014年，在洛杉矶地震发生三分钟后，《洛杉矶时报》就立刻发表了一篇相关报道。《洛杉矶时报》之所以能够在这么短的时间内完成这一创作壮举，是因为公司早在2011年就开始研发名为Quakebot的自动化新闻生成机器人，它可以根据美国地质调查局产生的数据自动撰写文章。这些新闻媒体机构最初撰稿借助的AI工具大多是外部采购的，而在智能创作时代的背景下，许多媒体机构已经开发了内部AI，比如英国广播公司（BBC）的“Juicer”、《华盛顿邮报》的“Heliograf”，而彭博社发布的内容有近三分之一是由一个叫“Cyborg”的系统生成的。<sup>[1]</sup>

中国媒体在AI撰稿领域也有相关尝试。例如，2016年5月，四川绵阳发生4.3级地震时，中国地震台网开发的地震信息播报机器人在6秒内写出了560字的快速报道；2017年8月，当四川省阿坝州九寨沟县发生7.0级地震时，该机器人不仅翔实地撰写了有关地震发生地及周边的人口聚集情况、地形地貌特征、当地地震发生历史及发生时的天气情况等基本信息，还配有5张图片，全过程不超过25秒；在后续的余震报道中，该机器的最快发布速度仅为5秒。<sup>[2]</sup>

以上便是AI进行结构化写作的典型范例，虽然上述案例都与新闻撰写相关，但AI在文本生成领域的应用绝不仅限于此。AI文本生成的方式大体分为两类：非交互式文本生成与交互式文本生成。非交互式文本生成的主要应用方向包括结构化写作（如标题生成与新闻播报）、非结构化写作（如剧情续写与营销文本）、辅助性写作。其中，辅助性写作主要包括相关内容推荐及润色帮助，通常不被认为是严格意义上的AIGC。交互式文本生成则多用于虚拟男/女友、心理咨询、文本交互游戏等涉及互动的场景。

前文提到的新闻播报就属于结构化写作，通常具有比较强的规律性，能够在有高度结构化的数据作为输入的情况下生成文章。同时，AI不具备个人色彩，行文相对严谨、客观，因此在地震信息播报、体育快讯报道、公司年报数据、股市讯息等领域具有较大优势。国内许多知名媒体旗下都有这种类型的AI小编，包括新华社的“快笔小新”、第一财经的“DT稿王”、《南方都市报》的“小南”、封面新闻的“小封”、腾讯财经的“Dreamwriter”，以及今日

头条的“Xiaomingbot”等。

AI结构化写作还可以被用于生成自动标题与摘要，它可以通过自然语言处理（Natural Language Processing，简称NLP）对一篇纯文本内容进行读取与加工，从而生成标题与摘要。以Github上标题生成的GPT2-NewsTitle项目为例，输入文本内容：“今日，中国三条重要高铁干线——兰新高铁、贵广铁路和南广铁路将开通运营。其中兰新高铁是中国首条高原高铁，全长1 776公里，最高票价658元。贵广铁路最高票价320元，南广铁路最高票价206.5元，这两条线路大大缩短西南与各地的时空距离。”可以得到AI返回的标题：“中国‘高铁版图’再扩容，三条重要高铁今日开通”。<sup>[3]</sup>提炼的标题简约而精准，具有很高的实用价值。

而相较于这种结构化写作，非结构化写作会有难度。非结构化写作任务，比如诗歌、小说/剧情续写、营销文本等，都需要一定的创意与个性化，然而即便如此，AI也展现出了令人惊叹的写作潜力。

以诗歌为例，2017年微软推出的人工智能虚拟机器人“小冰”出版了人类史上第一部AI编写的诗集《阳光失了玻璃窗》，其中包含139首现代诗。诸如“而人生是萍水相逢/在不提防的时候降临/你和我一同住在我的梦中/偶然的梦/这样的肆意并不常见/用一天经历一世的欢喜”，虽然在逻辑性上有所欠缺，但整体上富有韵律与情感，同时带有意象的朦胧感。

你如果对此感到好奇，不妨前往小冰写诗的网站亲自尝试。在首页，就会看到一则有趣的声明：“小冰宣布放弃她创作的诗歌版权，所以你可以任意发表最终的作品，甚至不必提及她参与了你的创作。”这段声明让人不禁好奇，这两年看到的很多现代诗会不会都是AI创作的？

按照官网提示，点击“马上开始”便会来到输入“灵感”的页面，页面上设置了上传图片或提示性文字的位置。你可以上传一张提供创作灵感的照片，就好像诗人会触景生情、吟诗作对一样，人工智能同样需要观景而抒怀。例如，我们上传了一张在海边拍摄的夕阳照片（图1-1），等待了大约10秒钟处理时间，便可以看到小冰写诗处理过程的展现界面。

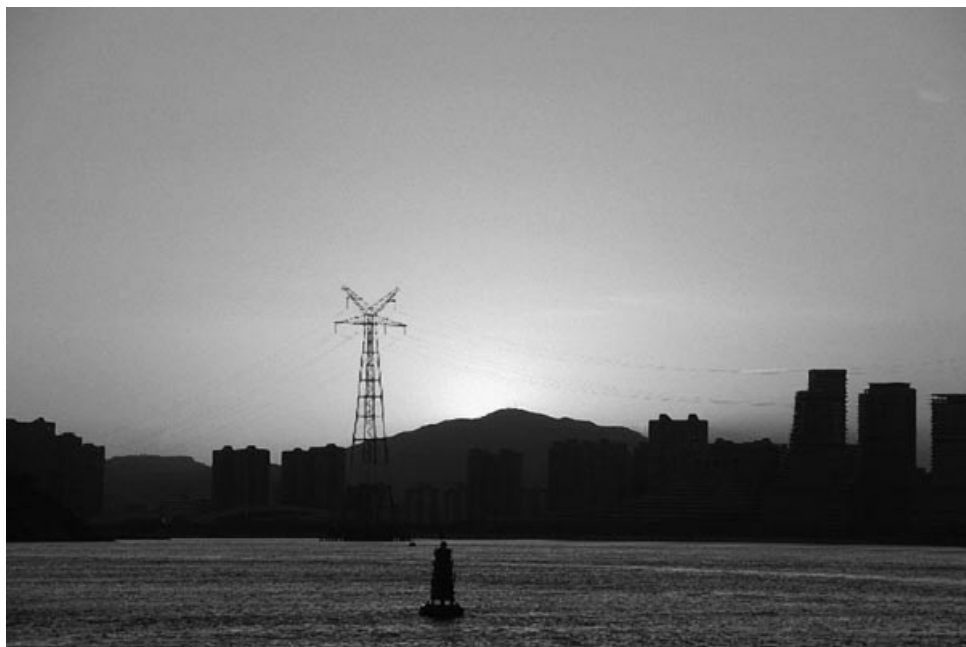


图1-1 海边日落图（摄于2022年7月9日）

在经历完意象抽取、灵感激发、文学风格模型构思、首句试写、诗句迭代和完成全篇的流程后，小冰生成了一首十四行诗，我们从中截取两段分享给各位读者。

每一条温水下的微风  
你会瞧见她的时候却皱起眉  
青春就是人生的美酒  
虽然是梦中的幻境

喝的是人们认识的人  
纵使千万人的美酒会化成灰烬  
乘你的眼睛里藏着深情  
又如天空徘徊

读罢，我们仿佛看见一位伫立在夕阳下的诗人，举起手中的酒杯，对天吟咏，感慨着青春易逝、物是人非。而当我们点击“复制初稿”进行粘贴时，连带诗歌一起粘贴过来的，除了再次出现的“放弃版权声明”，还有这样一段话：“未来世界，每个人类创作者的身边，都将有一个人工智能少女小冰，而你今天已经拥有。”看到这里，很难不让人幻想未来世界人类与机器携手创作的画面。

除了诗歌，AI也能进行故事、剧本和小说的写作。在2016年的伦敦科幻电影节上诞生了人类史上第一部由AI撰写剧本的电影《阳春》（*Sunspring*）。这部影片的机器人编剧“本杰明”由纽约大学研究人员开发，虽然影片只有9分钟，但本杰明在写作前经过了上千部科幻电影的训练学习，包括经典影片《2001太空漫游》《超时空圣战》《第五元素》等。<sup>[4]</sup>

2021年10月初，美国热门流媒体平台网飞（Netflix）与知名喜剧人基顿·帕蒂（Keaton Patti）在YouTube上合作发布了一部AI剧本创作的电影《谜题先生希望你少活一点》（*Mr. Puzzles Wants You to Be Less Alive*）。AI被基顿·帕蒂强迫着“观看”了超过40万个小时的恐怖电影剧本之后，创作出了这部电影作品，并收获了用户的广泛关注。截至2022年12月11日，该电影在YouTube上的播放量已超过420万，远高于网飞频道其他视频的播放量。

在这部电影中，我们能够看到向《电锯惊魂》《德州电锯杀人狂》《猛鬼街》等知名恐怖电影致敬的画面。不过，真正赋予这部影片讨论度的，并非其中的恐怖元素，而是作为一部恐怖影片，它的笑点非常密集，“我爸会花钱赎我，但我妈不会”“请不要杀我，我有好几个家庭”“他醉了，但被清醒所困扰”等金句频出，很难不让人印象深刻。评论区有网友的感叹道出了很多人的心声：“怎么恐怖元素没抓住，喜剧精髓倒是拿捏死了。”整部影片充满了毫无逻辑的荒诞设定和出其不意的笑点，也不怪乎有人感叹：“真正可怕的是，这些机器人已经掌握了人类的幽默感。”

同样令人一边惊叹AI智慧、一边忍俊不禁的还有Botnik Studios公司研发的AI机器人的作品。AI机器人在拜读了《哈利·波特》整套小说后写出了续集《哈利·波特与看起来像一堆灰烬的肖像》（*Harry Potter and the Portrait of What Looked Like a Large Pile of Ash*），故事情节异想天开，比如“怀孕的伏地魔”“他看到了哈利，然后立刻开始吃赫敏的全家”“罗恩打算变成一只蜘蛛”。<sup>[5]</sup>

当然，这种脑洞大开的故事类写作并不局限于非交互式文本生成类型，交互式文本生成也可以撰写故事。2017年万圣节期间，MIT媒体实验室推出了一个讲恐怖故事的人工智能系统“Shelly”，它可以生成恐怖故事的开头，然后与人类读者合作把一个个令人毛骨悚然的故事续写下去。Shelly每隔一个小时就会在推特上发布一个新故事开篇，当有人回应、故事足

够受欢迎时，Shelly 就会回复新的句子，让故事继续下去。<sup>[6]</sup>

这种交互式的故事写作模式也可以用来制作文字类冒险游戏。2021年，开发者尼克·沃尔顿（Nick Walton）就曾推出过这样一款游戏《AI地下城 2》（*AI Dungeon 2*）。这是一款能够无限生成的文字冒险游戏，玩家可以从生成的几个设定和角色中进行选择，然后用动词开头的祈使句来输入行动，游戏AI会自动根据输入的行动生成相应的故事。比如，你可以选择作为一个出生在奇幻世界的骑士踏上旅程，在旅途中，你可能遇到王国的巫师、吐火的巨龙、被绑架的公主，等等。

除了上面介绍的这些应用，最令人印象深刻的交互式文本内容生成应用要属ChatGPT。与前文所展示的例子不同，ChatGPT可以同时作为问答、聊天及创作AI存在，它的使用场景日常且多样，融合了文案生成、小说续写、代码生成、代码漏洞修复、在线问诊等场景，甚至展现出了超越搜索引擎的潜力。下面我们来直观感受一下ChatGPT的强大功能（后文ChatGPT针对问题的回答均用AI表示）。

（1）首先，让我们询问一下它知不知道自己是什么，看看它会如何作答。

问：ChatGPT是什么？

**AI：** ChatGPT 是一个由OpenAI开发的聊天机器人模型，它是基于大型语言模型GPT-3训练的，可以自然地对话，并回答用户的问题。它能够以人类般流畅的方式进行交流，并能够模仿各种风格和口吻。ChatGPT是一个非常强大的工具，可以用来辅助客服、创建聊天机器人等。

（2）关于本书的写作貌似没什么思路了，让我们尝试着从ChatGPT这里获得启发。

问：我想写一本关于AIGC的书，该怎么拟定大纲框架？

**AI：** 第一部分：AIGC的概述

- 介绍AIGC是什么，它的基本概念和特点。
- 讲述AIGC的发展历程，以及它的技术背景。

第二部分：AIGC的应用

- 讨论AIGC在各个领域的应用情况。
- 举例说明AIGC在客服、语音识别和翻译等方面的具体应用。

第三部分：AIGC的未来

- 展望AIGC的发展趋势。
- 探讨AIGC在未来可能的应用场景。

#### 第四部分：总结

(3) 除了写作辅导，我们也来试一下它能不能教人表白。

问：我想要在2022年结束前，和我喜欢的人告白，请帮我想一段诚挚的告白的话。

**AI：**亲爱的，在这个即将结束的2022年，我想对你说：我喜欢你，喜欢你的一切。我喜欢和你在一起的感觉，感觉你陪伴的温暖。你是我一生中遇见的最特别的人，我希望我们能一直在一起，共度余生，请让我成为你的伴侣，与你共同走过这段旅程。

(4) 不会发朋友圈？不会写贺卡？也可以试试求助人工智能！

问：请帮我写一条朋友圈，表达我失眠的心情，同时体现我的文学才华。

**AI：**朋友们，今晚我又失眠了。我坐在床上，面对着漆黑的天花板，想象着各种奇怪的事情，挣扎着想要入睡，却始终如故。失眠之苦，唯有身在其中的人才能体会。

我想起了莎士比亚的名句：“夜深人静，空气中弥漫着一种诡异的气息。”是的，失眠之夜，就像漆黑的夜空，让人感到不安和孤独。

朋友们，你们有没有好的方法，能帮助我入睡呢？敬请指教。

这还不是最有趣的。最有趣的是，ChatGPT具有很强的上下文连接与多轮对话的能力。ChatGPT还具有记忆能力，当向它追问“能不能再夸张一点”时，它可以在之前生成的文案基础上进行改写。这种连续对话大幅提高了人机交互体验。

问：能不能再夸张一点？

**AI：**朋友们，今晚我又失眠了。我翻来覆去，把被子打成了螺旋形，绷得紧紧的，就像拧紧了弹簧一样。失眠之苦，实在是令人难以忍受。

我想起了莎士比亚先生的名言：“夜深人静，空气中弥漫着一种诡异的气息。”是的，失眠之夜，就像漆黑的夜空，让人感到不安和孤独，我想，失眠之人，应该像受伤的勇士一样，坚强地挺过难关。我们要不屈不挠，不懈地努力，才能摆脱失眠的折磨，才能重新健康地入睡。

朋友们，你们有没有好的方法，能帮助我入睡呢？敬请指教。

(5) 当不会撰写代码的时候，你也可以尝试求助于ChatGPT。

问：如何在LaTeX上写一个积分方程？

**AI：**在LaTeX中，可以使用`\int`命令来表示积分。例如，要在LaTeX中写出下面的积分方程：
$$\int_0^1 x^2 dx$$

可以使用如下的LaTeX代码：

```
\int_0^1 x^2 dx
```



完整的LaTeX代码如下：

```
\documentclass{article}
\usepackage{amsmath}

\begin{document}

\[
\int_0^1 x^2 dx
\]

\end{document}
```

不只是辅助编程，ChatGPT甚至在一些测试中可以直接修改代码，并给出可以直接复制的代码示例，这在未来将极大减轻程序员的工作强度。

人类在想方设法地释放AI的力量。例如，产品经理试图引导ChatGPT生成产品需求文档，训练ChatGPT根据自己之前写的周报自主修改周报。根据用户Jiayuan@Tisoga在推特上分享的案例，ChatGPT目前生成的产品需求文档内容完备且带有技术方案，未来产品经理需要做的可能只是和ChatGPT聊一下自己的想法以及项目相关背景，让AI生成一份草稿文档，再进行修改润色即可。

同样，用户也可以通过不断的提问引导ChatGPT，层层递进完成从研究框架到研究报告的输出。先让ChatGPT生成报告大纲，再通过类似专家访谈一样一步步提问让ChatGPT补充报告内容，并在这个过程中根据其答案进行衍生提问，不断完善它的逻辑。例如，按照上面第（2）个回答中的写作大纲，我们可以让ChatGPT继续写下去：什么是AIGC？它的基本概念与特点是什么？

面对这样强大的功能，很难不让人幻想AI生成文本的未来：程序员、研究员、产品经理等涉及重复性工作的脑力劳动者可能都将被AI取代，这些职业可能都演变成了新的职业——提示词（Prompt）工程师，目的就是帮助人类更好地与AI互动。

## 二、AI音频生成

目前，AIGC在音频生成领域已经相当成熟，并广泛应用于有声读物制作、语音播报、

短视频配音、音乐合成等领域。AI音频生成主要分为两种类型：语音合成与歌曲生成，这两种类型都有许多经典案例。

在语音合成领域，喜马拉雅曾采集著名评书表演艺术大师单田芳生前的演出声音，运用文本转语音（Text to Speech，简称TTS）技术，推出单田芳声音重现版的《毛氏三兄弟》和历史类作品。在QQ浏览器首页的“免费小说”频道中的听书功能模块，用户也可以选择自己喜欢的AI语音包进行播放，语音包有六种AI音色可供选择：清朗男声、标准男声、软萌音、御姐音、东北女声、温柔淑女音，并且合成的语音节奏分明、情绪自然，能够很好地解放双眼。

除了语音读书，短视频配音也是一个常见的音频生成应用领域。“注意看，这个男人叫小帅。”短视频平台的很多电影解说都伴随这句话开始，随后很可能还会听到女主角“小美”的名字。抑扬顿挫的男声搭配一些电影的高潮情节画面，再加上相似的解说套路和背景音乐，这其实也是AI生成语音的典型应用，用户只需3~5分钟就可以看完一部“电影”。当然，语音合成不仅可以应用于说话语音，也可以应用于唱歌语音，歌手歌声合成软件X studio就能够为用户提供具有不同音色和唱腔的虚拟歌声。

而对于AI歌曲生成，在OpenAI发布的最新项目MuseNet中，用户可以使用AI生成多达10种乐器演奏的歌曲，甚至还可以制作多达15种风格的音乐，模仿莫扎特和肖邦等古典作曲家、Lady Gaga等当代艺术家，也可以模仿电子游戏音乐等类型。

除了直接生成音乐，AI歌曲在实际应用中常用来自动作词。“醒来灿烂星光透过了窗台，海岸线连接了那片山川大海。涌动梦境边缘像是空旷舞台，在眼前忽然展开。”看到这段文字，你的脑海中是否浮现出星河璀璨、山川河海一望无际的绚丽景象呢？这段颇具画面感和动态美的歌词正是由网易新开发的人工智能所创作。

网易伏羲利用自主研发的“有灵智能创作平台”，让AI学会人类语言组织的基本逻辑。借助大规模的语料训练，用户可以仅凭借输入预设风格、标签、情绪和韵脚便可以得到一首极富韵律美和意境感的歌词。例如，在设定好古风的预设风格之后，加入“夜晚”“梧桐”“叶落”“深秋”“乡愁”等标签并选定江阳韵，便得到了由人工智能创作的歌词。

一阵秋风 夜晚微凉  
不远处竹影悠长  
梧桐路旁 谁留下幽幽暗香  
那一片片梧桐叶落我心上  
孤灯 暗夜风霜

梧桐 心悲凉  
低语 道不尽半世情伤  
深深秋雨 让人惆怅  
天上弯弯的月亮  
思念的人儿在他乡  
这样的夜晚 夜太漫长  
心爱的姑娘你在何方  
斜阳晚夜 雪落西窗  
留下一根琴弦 唱着你的忧伤  
梧桐雨巷 人影茫茫

“孤灯，暗夜风霜”渲染了萧瑟凄凉的异乡秋景，“天上弯弯的月亮，思念的人儿在他乡”刻画了身在异乡的有情人不得相见的哀婉与凄苦，用简单的意象却能营造出如此意境，实在是令人为之惊叹，人工智能生成音乐远比我们想象的更加熟练灵活。当然，除了根据伴奏配歌词，人工智能同样可以根据编写好的歌词编曲。网易天音就是这样一个一站式的音乐编曲平台，不过，编曲的生成相对于歌词生成会更有难度，一般需要拥有一定的乐理基础、能够根据和弦谱进行编曲微调的编辑。

此外，AI歌曲生成还有一些更有趣的玩法，比如腾讯在2020年携手明星王俊凯推出了AI歌姬“艾灵”：当用户选择关键词后，可以输入个人的名字或昵称，AI便能自动生成带有用户名字的歌词，并会生成歌声与王俊凯共同演唱。

### 三、AI图像生成

你是否在生活中使用过修图软件？如果使用过，那么很有可能在你未曾注意到的时候，就已经在接触AI生成图像了，比如去除水印、添加滤镜等都属于广义上AI图像生成的范畴。

目前，AIGC在图像生成方面有两种最成熟的落地使用场景：图像编辑工具与图像自主生成。图像编辑工具的功能包括去除水印、提高分辨率、特定滤镜等。图像自主生成其实就是近期兴起的AI绘画，包括创意图像生成（随机或按照特定属性生成画作）与功能性图像生成（生成logo、模特图、营销海报等）。

2022年下半年，AI绘图无疑成为热门应用，不少人都乐此不疲地在自己的朋友圈分享各种形式的AI绘画作品。从参与感与可玩度来看，AI绘画大致可以分为三类：借助文字描述

生成图像、借助已有图像生成新图像，以及两者的结合版。

当被问及周围最早一批使用AI绘画软件的用户为什么喜欢AI绘画时，有人这样回答道：“我小时候就喜欢画画，但天赋实在有限，家里觉得既然走不了艺考，还是好好学习更重要，就没有花太多精力在上面。但现在，AI绘画实现了我曾经的梦想。”曾经，那些因为各种各样原因放弃绘画或没有学绘画的人，在这个时代也能仅凭输入几个词语、一段文字，就能得到一张还不错的绘画作品。如图1-2所示，在AI绘画工具Stable Diffusion上输入“一座复古未来主义的空中浮岛”的英文，便可以得到一张生动的图片。

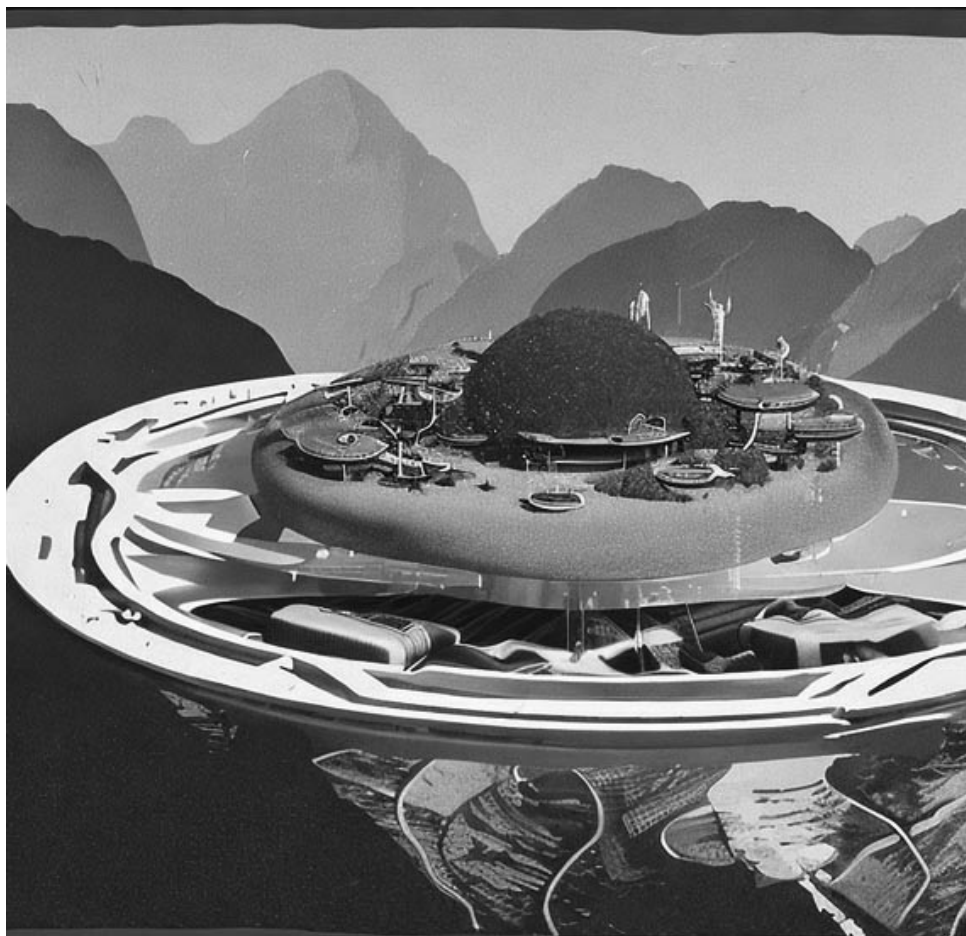


图1-2 “一座复古未来主义的空中浮岛”生成图像

生成来源：Stable Diffusion

你是否也觉得这很神奇，仿佛魔法一般？事实上，从文本到图像的生成真的有“咒语”存在，这个“咒语”就是被用来激发创作与思考的提示词。提示词可以是一个问题、一个主题、一个想法或一个概念，在AI绘画的语境下可以简单理解为“喂给”AI进行创作的一组灵感词组，通常是对自己设想作品的简要描述。

现在流行的国外AI绘画工具Stable Diffusion、DALL·E 2、Midjourney等，以及国内AI绘画工具文心一格、意间AI绘画、AI Creator等，都会在创作时引导你输入“咒语”。如果你暂时缺乏灵感，有些平台也会提供“自动生成”选项，让AI帮你自主搭配，然后在其基础上进行你想要的修改。

如此一来，AI降低了普通人参与艺术创作的门槛，让没有绘画基础的人也能通过文字描述表达自己的创作灵感，满足自己的创作欲望。比如，我想得到一幅中国风的山水画，我可以这样输入提示词：水、林木、云雾、山石、溪流、山峦、霞光、水墨画、中国风、低饱和。AI成功读取了我的“咒语”，然后返回了我下面这幅画（图1-3）。



图1-3 中国山水画生成图像

生成来源：Midjourney

如果你对画家及其作画风格有所了解，你还可以在编写“咒语”时加入这些画家的名字进行画风定制。AI绘画工具不只支持知名画家如达·芬奇、梵高、毕加索等的画风，还支持众多现代画家的画风。假如你想要复古神秘的画风，可以尝试加入英国插画师汤姆·巴肖（Tom Bagshaw）的名字；想要CG（计算机动画）人物画，则可以加入代表性画师

Artgerm、阮佳（Ruan Jia）的名字。为了方便读者直观地感受融入了特定风格生成画作的效果，我们利用Jasper.AI生成了具有张大千与梵高画风的画作（图1-4和图1-5）。



图1-4 “轻舟已过万重山”生成图像（张大千风格）

生成来源：Jasper.AI



图1-5 “手捧玫瑰花的少女”生成图像（梵高风格）

生成来源：Jasper.AI

AI的能力超乎你的想象，除了一键构图与风格调整，它甚至可以辨别2D与3D，满足用户的精细化定制需求。例如，当我们想在人物画上生成小狗时，DALL·E 2会把小狗画入画中，如图1-6所示。



图1-6 AI生成画中的二次元小狗

生成来源：DALL·E 2

而当我们想要把一只3D小狗画在座位上的时候，DALL·E 2便生成了一只真实的、三次元的小狗，如图1-7所示。





图1-7 AI生成座位上的三次元小狗

生成来源：DALL·E 2

伴随着AI绘画技术的逐渐成熟，AI插画也被用作一些具有功能性的场景中。例如，2022年6月11日，著名杂志《经济学人》首次采用了AI插画作为封面，作品名为《AI的新边界》（*AI's new frontier*）。在封面油画风格的分割色块背后，有着一张具备少量机械特征的人脸，预示着AI将以全新的面貌出现在我们面前，拓展人类技术的新边界。

除了可能提高封面插图类的设计效率外，AI绘画目前也被用于游戏开发环节，包括前期的场景与人物图辅助等，此外也有部分游戏工作者正在探索基于Stable Diffusion生成游戏资产，比如游戏图标及游戏内的道具。

虽然AI绘画对内容生产力的提升具有很大帮助，但与此同时也引发了许多人的忧虑，许多艺术家担心AI绘画可能会因为训练样本的选取而剽窃自己的作品元素，也担心这些AI生成的作品被用于一些欺骗性的用途，危害到人类自身。

针对AI与人类的辩题，Midjourney的创始人大卫·霍尔兹（David Holz）这样评价：“AI是水，而非老虎。水固然危险，但你可以学着游泳，可以造舟，可以造堤坝，还能借此发电；水固然危险，却是文明的驱动力，人类之所以进步，正是因为我们知道如何与水相处并利用好它。水给予更多的是机会。”<sup>[7]</sup>

艺术家是否买单尚且不论，投资人已经开始竞相押注。2022年10月17日，Stable

Diffusion的母公司Stability AI宣布完成1.01亿美元融资，成为估值10亿美金的超级独角兽。随后，不到一个月，另一家AI绘画平台Jasper.AI宣布完成了1.25亿美元A轮融资，估值达15亿美元，距离产品上线也不过18个月的时间。

从技术开发到应用落地固然有一定时间差，但值得惊喜的是，至少在图像生成领域，我们正看到日益成熟的应用场景以及商业化的可能性。

## 四、AI视频生成

目前，AI技术不仅可以生成图片，也能够生成序列帧，组成一个完整的视频。2022年10月，AI重置版《幻觉东京》发布。《幻觉东京》原本是一部记录日本亚文化人物的纪录片，作者将经过剪辑的短片交给AI美术大师，经过160小时生成3万张独立插画，再进行人工手动微调，连成了一部赛博朋克大幻想。虽然目前还只是在原脚本和视频的基础上，通过AI逐帧完成图片生成，但这让我们看到了AIGC参与到视频创作中的可能性。

当然，除了这种连接AI生成图片组成视频的生成方式，也有直接利用文字描述生成视频的方法。2022年9月，Meta推出的Make-A-Video工具就具有根据文本描述生成相应短视频的能力。Make-A-Video推出不久，谷歌就推出了主打高清生成的Imagen Video和主打更长视频内容生成的Phenaki。Imagen Video是由谷歌在2022年5月推出的AI绘图工具Imagen进化而来，它继承了Imagen对于文字的准确理解能力，能够生成1280×768分辨率、每秒24帧的高清视频片段。除了分辨率高以外，它还能理解并生成不同艺术风格的作品，比如水彩画风格、像素画风格、梵高风格。同时，它还能理解物体的3D结构，在旋转展示中不会变形。而谷歌推出的另一款AI视频生成工具Phenaki则可以根据200个词左右的提示语生成2分钟以上的长镜头，讲述一个完整的故事，并能根据提示语自由切换风格场景，让人人都能够成为导演。<sup>[8]</sup>

除了刚提到的这些新兴的视频AIGC技术，AIGC在视频生成方面的常见传统应用场景还包括视频属性编辑、视频自动剪辑及视频部分编辑。

视频属性编辑包括删除特定主体、生成特效、跟踪剪辑等，能够高效节省人力和时间。AI能够通过对画面人物的动态追踪，自动搜索人物，定位关键时间节点，极大提升剪辑效率。此外，AI还能够去除视频的拍摄抖动，修复视频画质。

视频自动剪辑是对特定片段进行检测及合成。2020年全国两会期间，《人民日报》创造性地推出“5G+AI”模式的新闻报道，打造的智慧平台iMedia、iMonitor、iNews等可以第一时间对素材进行智能处理，只需要短短几分钟，就能结合语音、人像、文字识别，从海量的视频资料中迅速生成剪辑视频片段，并自动匹配字幕。

此外，从广义上讲，AI主播也可以看作一种AIGC生成视频的应用，只不过是生成的音频内容去对应到虚拟人的口型与动作进行综合剪辑。2020年5月，由新华社与搜狗公司联合推出的身穿蓝白正装的“新小微”第一次亮相演播室。“新小微”是全球首位3D版AI合成主播，能够像真人一样走动和转身，并摆出各种复杂动作与姿态。同时，她还在不断迭代，比如从“职业微笑”到增加了许多微表情，从单一妆发到根据播报场景变更妆发等。<sup>[9]</sup>

事实上，“新小微”并非新华社推出的第一个AI主播，在她之前已经诞生了由“新小萌”“新小浩”等组成的AI合成主播家族。其中，中国首个AI合成女主播“新小萌”在2019年上岗时就惊艳了全球媒体圈，被外媒评价为“几乎可以以假乱真”。

除了新华社，中央广播电视总台、人民日报社等国家级媒体以及湖南卫视等省市媒体也在积极布局应用AI合成主播，先后推出了央视AI主播“AI王冠”、湖南卫视AI主持人“小漾”、安徽卫视AI主持人“安小豚”、央视网AI小编“小C”等。与“新小微”一样，这些AI主播在全国两会、冬季奥运会、冬季残奥会等重大活动期间被广泛应用，极大地提升了新闻产出与传播效率。

除此之外，像AI视频换脸这种视频部分编辑的形式从广义上来说也属于AIGC的范畴。2019年，一款名为ZAO的AI换脸软件刷屏各大社交网络，人们乐于把自己的脸替换进经典影视剧如《甄嬛传》《权力的游戏》，并进行分享。虽然ZAO后来因为侵犯个人隐私被下架，但类似的AI视频换脸软件却层出不穷，这在某种程度上也反映了人们的内容创作热情以及对于自由表达的欲望，但合法合规问题则是AI视频换脸长久发展需要重点解决的问题。

### 第三节 AIGC助力元宇宙和Web3.0

元宇宙与Web3.0的未来近在眼前，而AIGC作为全新的内容生产方式无疑能为这些美好的图景注入新的活力。本节将对AIGC如何助力元宇宙和Web3.0展开诸多畅想与探讨。

#### 一、AIGC在元宇宙方向的应用

起源于1992年科幻小说《雪崩》的“元宇宙”概念，在2021年伴随着“Roblox的上市”和“Facebook更名为Meta”两大事件掀起了前所未有的科技热潮。人们迫切希望打造一个与现实世界相平行、高度沉浸化的虚拟世界。这样一个“世界级”的工程项目，单靠人力创作可能难以做到尽善尽美，而AIGC的介入将可能大大提升元宇宙的构建效率。

##### 1.虚拟形象

就像《头号玩家》《失控玩家》等描绘元宇宙蓝图的科幻电影中表现的一样，每个人都需要通过自己的虚拟化身接入元宇宙之中。过去，人们想要生成这样的虚拟化身，大多是依靠系统内置好的几种模型，通过调整不同的五官、身材、服装等搭配方式，来生成属于自己的虚拟化身。从用户角度来看，这种生成方式不仅十分烦琐，而且也很难定向生成一些与自己真实形象相关联的特征，甚至一不小心还可能就和其他人撞形象，难以凸显个人的独特性。而从开发者的角度来看，想要生成足够多的虚拟化身不仅费时费力，而且也难以生成让大多数用户满意的化身形象。

AIGC有助于解决这一问题。例如，Roblox在2020年末就收购了初创公司Loom.ai，利用AI技术解决虚拟化身的生成问题。用户使用Loom.ai的组件，可以直接利用单张图片生成3D写实风格或卡通风格的虚拟化身形象。不仅如此，Loom.ai还可以提供精确的面部动画生成能力，可以让虚拟形象生成包括嘟嘴、皱眉等复杂的表情，让虚拟化身之间进行更加沉浸式的交流。

##### 2.虚拟物品

在元宇宙世界，除了虚拟化身之外，还有许多各式各样的虚拟物品。不少元宇宙选择同时利用PGC和UGC的方式来丰富这些模型，无论是专业机构还是感兴趣的个人，都可以借助官方提供的编辑器来创作各种类型的3D物品模型，或者在创作好3D物品模型后导入元宇宙世界。除了PGC和UGC外，虚拟物品的生成在未来同样可以借助AIGC来实现。例如，2022年9月底，谷歌就发布了文本生成3D模型的工具DreamFusion，而在此之后不久，英伟达也推出了类似工具Magic3D，并将DreamFusion视为直接对标，在生成速度和分辨率上都

实现了一定程度的提升。这些由文字生成的模型除了可以通过文本输入自动渲染3D模型，也可以提供额外的提示语对原有的模型进行修改、编辑。而在2022年12月，OpenAI也推出了自己的文本生成3D模型的工具Point-E。虽然该模型采用点云模型的生成方式，不能直接生成渲染完毕的3D模型，但它的生成速度可以达到谷歌DreamFusion的数百倍。

如果未来能将这些模型大规模落地推广并应用于元宇宙中，可以大大提高虚拟物品的生成效率。

### 3.虚拟场景

对于元宇宙来说，最重要的就是沉浸式的场景体验。而一个完善的元宇宙场景，可能不仅包括各式各样的虚拟物品，还包括场景内的背景音乐、与部分物品之间的交互效果，等等。对于背景音乐，可以利用AIGC进行音频生成；对于交互效果，也可以利用AIGC进行相关代码的生成。除了这种通过拼接不同类型元素的元宇宙场景生成，整个元宇宙场景中的每个元素，在未来可能都将使用AIGC去构建。例如，Meta在2022年初就公布了AIGC生成元宇宙场景的概念系统“Builder Bot”，用户在元宇宙中只要通过语音说出自己想要的环境，周围的虚拟空间就会自动生成相应的场景。在Meta AI公示的Demo中，用户说出“带我们去公园吧”，周围的环境就变成了公园；当用户说出“天上来一些云吧”，天上就会生成许多白云。如果未来能够广泛地实现这种形式的AIGC，将大大降低元宇宙场景的创作门槛，元宇宙的场景也能变得更丰富、更多样。

## 二、AIGC在Web3.0方向的应用

这里的Web3.0主要指基于区块链技术所构建的价值互联网。在Web3.0中，用户可以借助NFT（Non-Fungible Token，非同质化代币）将自己的创作物添加至代币上链，来确认自己对于创作物的创作权益。结合NFT的这一特质，AIGC在Web3.0方向也可以有诸多有趣的应用。

### 1.AIGC生成制作NFT

目前，许多NFT绑定的创作物都是以图像的形式存在。既然如此，自然也可以用AIGC的方式生成图像并制作成NFT，这样可以帮助NFT项目方更快捷地生成全套的NFT形象。除了为项目方提供创作工具外，普通人无须任何绘画基础也可以参与到图像类NFT的制作中，并通过销售NFT获得收益。

### 2.绘画风格确权

对于艺术创作者来说，绘画风格是艺术创作者的核心资产，但如何对这类资产进行确权 and 变现是困扰着很多创作者的问题。对于需要大量原画的游戏厂商来说，他们也希望采用特定创作者的绘画风格，并高效稳定地生产出大量满足游戏场景风格的插画。而通过“NFT+AIGC”的方式，就可以实现这种绘画风格的确权。艺术家可以将自己的绘画风格制作成NFT进行交易，而购买了绘画风格NFT的客户就可以利用AIGC批量生成该风格下的各种类型的插画，这就是AIGC在绘画风格确权方面的应用。

Hiimeta就是这样一个集艺术风格的确权、授权和使用为一体的AI艺术生态。在Hiimeta提供的工具内，用户可以先上传指定风格的原型素材，然后Hiimeta团队自主研发的算法会对整体风格、布局、纹理等基础元素，以及感情色彩、哲学思想等进行提炼，生成对应风格的“AI机器人”。购买了特定风格的AI机器人后，用户就可以生成具有相似风格的虚拟角色、自由插图、批量头像等。生成之后，用户还可以对细节进行微调处理。在保障创作者的版权基础上，这大大提升了艺术风格采购者的内容生产力。

### 3.结合 AIGC的GameFi

GameFi是一种结合区块链技术的游戏形式，通常会对游戏内的资产NFT化，并设计有一定的经济体系维持游戏的运转。在开发GameFi游戏过程中，人物、场景、动画甚至逻辑代码等都可以由AIGC创作，通过这种方式可以大大缩短游戏的创作周期，也能产生一些意想不到的创新应用。

Mirror World就是将AIGC相关技术应用在GameFi领域的代表性项目。Mirror World曾在2021年9月推出过首款可交互NFT产品：Mirror NFT，它创新性地让GameFi平台内的虚拟生命“Mirrors”活了过来。每一个Mirror NFT都具备独一无二的外形特征以及特定的语言模型，持有者第一次能够与自己的NFT自由地沟通与交流。在交流过程中，所有的对话数据均由AI生成，用户可以借此享受到有趣且无尽的对话体验。此外，依托于Y Combinator孵化的初创公司rct AI在游戏AI领域的不断探索，Mirror World在完成三款GameFi游戏开发后，根据自身在区块链技术领域的研究，研发出了一整套Mirror World Smart SDK，致力于帮助更多优质的游戏类项目在“低代码、零门槛”的前提下集成区块链与AI技术。

[1] 参考自<https://www.forbes.com/sites/calumchace/2020/08/24/the-impact-of-ai-on-journalism/?sh=2415c20f2c46>。

[2] 参考自<https://www.woshipm.com/ai/4797279.html>。

[3] 参考自<https://github.com/liucong/GPT2-NewsTitle>。

[4] 参考自<https://www.shuzix.com/4146.html>。

[5] 参考自[https://www.sohu.com/a/212719545\\_752829](https://www.sohu.com/a/212719545_752829)。

[6] 参考自<https://www.fastcompany.com/90148966/this-ai-writes-horror-stories-and-theyresurprisingly-scary>。

[7] 参考自<https://mp.weixin.qq.com/s/u4XYV8Tg6epHyBcKKgRRsg>。

[8] 参考自[https://mthepaper.cn/baijiahao\\_20196224](https://mthepaper.cn/baijiahao_20196224)。

[9] 参考自<https://www.jiemian.com/article/4419244.html>。

## 第二章 AIGC的技术思想

哪些技术思想对AIGC的演进做出了重要贡献？

人们总喜欢活在舒适区内，用粗暴的断言安慰自己，例如机器永远无法模仿人类的某些特性。但我给不了这样的安慰，因为我认为并不存在无法模仿的人类特性。

——艾伦·图灵（Alan Turing）

人工智能技术历经了漫长的演进过程，见证了基于规则、机器学习、深度学习、强化学习等领域的兴起。目前，人工智能技术在多模态和跨模态生成领域取得了傲人成绩。本章将回顾前AIGC时代各种奠基技术的相关思想，并在刨除复杂数学原理的基础上，用通俗易懂的语言对目前推动AIGC进行商业落地的重要技术和理念进行介绍。需要特别说明的是，为了帮助没有任何技术基础的读者理解本章内容，我们在模块拆解和技术诠释相关内容上可能会牺牲部分严谨性，想要完整细致地了解技术脉络的读者可以阅读相关科技文献。



## 第一节 前AIGC时代的技术奠基

### 一、图灵测试与人工智能的诞生

1950年，艾伦·图灵发表了一篇划时代的论文《计算机器与智能》（*Computing Machinery and Intelligence*），探讨了让机器具备人类一样智能的可能性。论文在开篇就抛出了一个有趣的问题：“机器能思考吗？”虽然在过去众多科幻作品中，对此已有诸多不同的解读，但在一篇严肃的科技论文中探讨这件事似乎是少见的。在论文里，图灵并没有一上来就解答这个问题，而是提出了一种模仿游戏，想要借助思想实验的方式，为确定“什么样的机器才是具备智能的”给出具有可操作性的定义方式。下面我们依照图灵的设计来模拟这样的游戏场景。

场景：小明、小红和小刚三个人决定一起来玩这个模仿游戏，小刚被关在密闭的屋子里，只能使用两台远程打印机分别与小红、小刚进行交流，但他并不知道每台打印机的背后是谁在回答他的问题。在游戏结束时，三个人的胜利目标是不同的。

·小明：在游戏结束后，需要根据提问和回答的记录，猜出每台远程打印机背后对应的是小红还是小刚。

·小红：尽可能地帮助小明猜对自己是小红。

·小刚：尽可能地干扰小明，让他以为自己才是小红。

对于小刚来说，一个很自然的游戏策略就是在回答时故意模仿小红，因此这个游戏被称为模仿游戏。现在，我们不妨微调一下这个游戏，把里面的人类“小刚”，更换成机器“小钢”。如果机器小钢能够借助预先设定好的程序模仿小红，并回答小明的问题，似乎也能让这个游戏进行下去。而图灵就在论文中提出，在用机器替换人类的情况下，根据小明这类角色回答错误概率有没有显著增加，可以评估这个替换的机器是否具备智能，这也就是著名的“图灵测试”（图2-1）。虽然“图灵测试”作为一种简易的思想实验存在着诸多缺陷，但它第一次让人们能够确切地想象出具备智能的机器是什么样子，而不仅仅停留在科幻的虚无中，为后世围绕人工智能展开科学实践指引了方向。

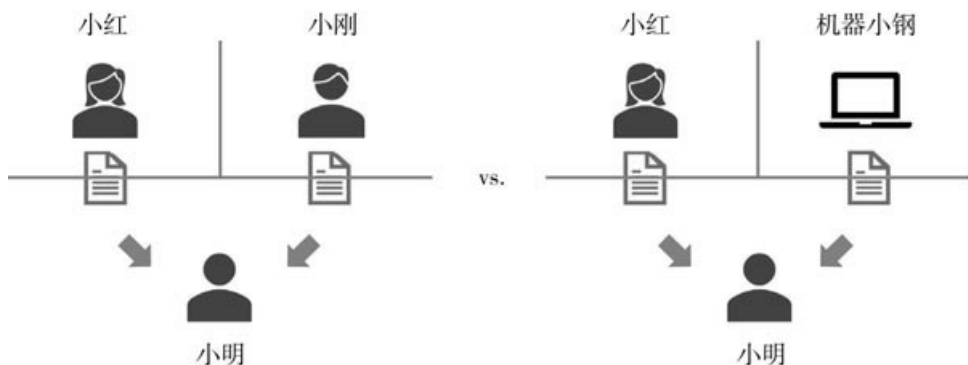


图2-1 图灵测试最初版本示意图

虽然此时图灵已经从理论角度给出了机器拥有智能的可能性，但是让“人工智能”这个科学领域正式形成的是1956年在美国达特茅斯学院举行的人工智能夏季研讨会。这次会议的组织方包括后来的图灵奖获得者马文·明斯基（Marvin Lee Minsky）和约翰·麦卡锡（John McCarthy）、信息论创始人香农（Claude Elwood Shannon）、IBM工程师罗彻斯特（Rochester），而其余参会者也均是后来对人工智能发展做出过重要贡献的科学家。通过达特茅斯会议，“人工智能”的名称和任务被真正界定下来，因而该会议也被广泛认为是人工智能诞生的标志，开启了人工智能领域曲折向上的技术发展之路。

## 二、符号主义、联结主义和行为主义

在人工智能诞生早期，就出现了“符号主义”和“联结主义”两种不同的发展流派，并都取得了一系列阶段性的成果。

符号主义认为人的智能来自逻辑推理，世界上所有信息都可以抽象为各种符号，而人类的认知过程可以看作运用逻辑规则操作这些符号的过程。在这样的前提假设下，如果计算机能够自动化地执行和人脑一样的各种规则，说不定就可以实现完全的智能化。由艾伦·纽厄尔（Allen Newell）在达特茅斯会议报告的逻辑理论家（Logic Theorist）项目就是符号主义早期的代表性成果，这个程序能够证明《数学原理》第二章52个定理中的38个，甚至找到了相对于原著更加精巧的证明方式。

而联结主义则认为，让机器模拟人类智能的关键不是去想办法实现跟思考有关的功能，而是应该模仿人脑的结构。联结主义把智能归结为人脑中神经元彼此联结成网络共同处理信息的结果，希望能够运用计算机模拟出神经网络的工作模式来打造人工智能，并在“人工智

能”领域正式形成之前就开始了各种尝试。1943年，神经科学家沃伦·麦卡洛克（Warren McCulloch）和数学家沃尔特·皮茨（Walter Pitts）按照神经元的结构和工作原理搭建了数学模型，奠定了人工神经网络的雏形。1958年，美国神经学家弗兰克·罗森布拉特（Frank Rosenblatt）发表了模拟人类学习过程的“感知器”算法，机器利用它就可以自主完成像分类这样的简单任务，后续算法杰出的实践效果掀起了第一次人工神经网络的热潮。

可以说，不论是符号主义还是联结主义，在人工智能诞生的前十余年，都取得了一个又一个令人震惊的成果，但好景不长，20世纪60年代末，人工智能的发展陷入瓶颈，人工智能的研究者遇到了很多难以克服的难题，其中包括两个最典型的难题：

- 受限的计算能力：当时计算机有限的内存和处理速度不足以支持AI算法的实际应用。

- 认知信息的匮乏：许多人工智能领域的应用需要大量认知信息，当时的数据库条件无法让程序获得如此丰富的信息源。

除了这些难题外，新兴研究成果针对符号主义和联结主义的批评也在一定程度上阻碍了人工智能的发展。对于符号主义，许多哲学家提出了各种各样的论断，试图证明人类的思考过程仅涉及少量的符号处理，大多是直觉性的，运用符号去模拟人类智能是徒劳的尝试。而对于联结主义，明斯基指出了感知器的致命缺陷：只能处理线性分类问题，连异或最简单的非线性分类问题无法得到支持，直接宣判了感知器的“死刑”。在整体研究进度受阻和成果难以落地的背景下，各大机构分别削减了对于人工智能研究的经费支持，人工智能的发展陷入了寒冬。

但这一切只是暂时的，新的机遇也在寒冬中酝酿。符号主义学者在20世纪70年代充分认识到了“知识”对于人工智能的重要性，不再过分追求当时难以实现的通用人工智能，而是将视野聚焦在较小的专业领域上，很大程度上缓解了计算能力受限和认知信息匮乏的问题，也让人工智能的程序变得实用起来。学者们试图利用“知识库+推理机”的结构，建设出可以解决专业领域问题的专家系统（图2-2）。在专家系统中，用户可以通过人机界面向系统提问，推理机会把用户输入的信息与知识库中各个规则的条件进行匹配，并把被匹配规则的结论存放到综合数据库中，然后呈现给用户。20世纪80年代，卡耐基梅隆大学为DEC公司研发的第一个专家系统XCON取得了巨大成功，在诞生初期平均每年能为公司节约4000万美元，这也使得全球各地的公司掀起了建设专家系统的知识革命。

而几乎在同一时期，联结主义也迎来了复兴。新型的神经网络结构及相关算法的普及为科研界注入了新的生机，适用于多层感知器的BP算法（误差反向传播算法），解决了非线性情况下的分类学习问题。至此，人工神经网络掀起了第二波发展热潮。

除了符号主义与联结主义，一种倡导“感知+行动”的行为主义流派也得到了较好的发

展。行为主义起源于控制论，强调模拟人在控制过程中的智能行为和动作，虽然它的起源也可以追溯到人工智能诞生时期，但一直未成为主流。在模拟人类智能方面，如果说符号主义是知其然且知其所以然，联结主义是知其然但不知其所以然，那么行为主义就是既不知其然也不知其所以然，因而行为主义在智能控制与智能机器人兴起的21世纪末才引起人们的广泛关注。<sup>[1]</sup>至此，符号主义、联结主义和行为主义便成为人工智能的三大经典流派，共同影响着后来人工智能的发展。

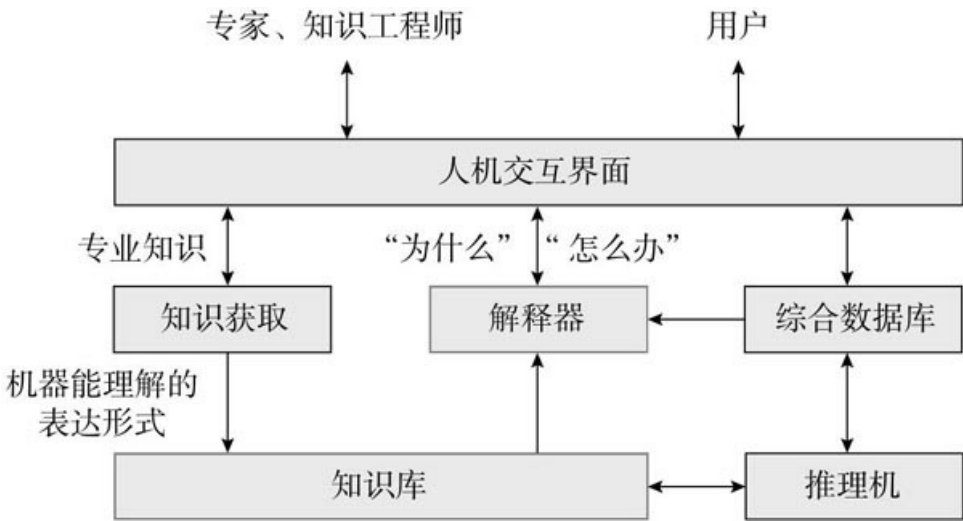


图2-2 专家系统结构图

### 三、机器学习

#### 1.机器学习的概念

1950年，图灵在他的论文《计算机器与智能》中提出了“学习机器”的概念，强调与其去编程模拟成人的大脑，还不如选择更简单的儿童大脑，通过辅之以惩罚和奖励的教学过程，让机器在学习后具备智能。此后，“机器学习”逐渐发展成为一个专门的细分研究领域，在人工智能领域占据了一席之地。

根据卡耐基梅隆大学计算机学院教授汤姆·米切尔（Tom Michell）的定义，机器学习是指“计算机程序能从经验E中学习，以解决某一任务T，并通过性能度量P，能够测定在解决T时机器在学习经验E后的表现提升”。这听起来似乎有些绕口，下面将通过人类学习的一个具

体例子来进行解释。

假如老师想让小明好好学习，在考试中取得好成绩，他可能会给小明布置很多习题，并观察每次的考试成绩以判断小明学习的效果。如果把这里的“小明学习”替换成“机器学习”，那么经验E就是反复的刷题过程，任务T就是参加考试，而性能度量P就是考试成绩。小明在反复刷题训练、参加考试的过程中，让成绩不断提升以达到预期的分数水平，当他走上高考考场时就能取得好成绩。而机器也是类似的，在经过反复的训练并达标后，执行任务时就可以取得比较好的性能，只不过这里的训练指的是输入数据。

综上所述，机器学习模型的训练过程可以分为以下四步。

- 数据获取：为机器提供用于学习的数据。
- 特征工程：提取出数据中的有效特征，并进行必要的转换。
- 模型训练：学习数据，并根据算法生成模型。
- 评估与应用：将训练好的模型应用在需要执行的任务上并评估其表现，如果取得了令人满意的效果就可以投入应用。

根据训练的方式，机器学习可以简单划分为监督学习和无监督学习。监督学习就好比小明每次做完题之后，老师都会对题目进行批改，让小明知道每道题是否答对。分类就是最经典的监督学习场景，机器先学习具备什么样特征的数据属于什么样的类别，然后当获取新的数据后，它就可以根据数据特征将数据划分到正确的类别。而无监督学习则好比老师把大量题目直接丢给小明，让小明在题海中自己发现题目规律，当题量足够大的时候，小明虽然不能完全理解每道题，但也会发现一些知识点的固定的选项表述。聚类是最经典的无监督学习场景，机器获得数据后并不知道每种特征的数据分别属于什么类别，而是根据数据特征之间的相似或相异等关系，自动把数据划分为几个类别。

## 2.感知器与神经网络

前文提到的感知器算法就是典型的监督学习的案例，它是人工神经网络的基础。为了方便读者理解后续与人工神经网络相关的内容，这里将刨除复杂的数学公式，对感知器算法的工作原理进行一个简单的介绍。首先，让我们想象一个具体的分类任务场景。

场景：小明在大学里选修了一门课程，这门课程并没有公布详细的合格评价标准，只知道平时的两次作业和一次考试会影响这门课程的通过与否，于是小明希望从往届的学长、学姐那里搜集他们的作业、考试及最终是否通过的相关数据，来帮助判断自己是否会在这门课程中挂科。在搜集完学长、学姐的数据后，小明决定先假设一个老师的评价标准：

·第一次作业 $\times 0.3$ +第二次作业 $\times 0.3$ +考试 $\times 0.4$ =课程评分。

·如果课程评分 $\geq 60$ ，则课程及格；否则课程挂科。

在写完假设的评价标准后，小明迫不及待地想把学长、学姐的成绩带入评价标准中，结果发现，按照现在的计算方式，所有的学长、学姐都不满60分，全都挂科了。这也就说明，作业和考试的评价系数假设得太小了，小明于是把它们调大了一些，但又发现，包括挂科的学长、学姐在内的所有人又全都及格了，这说明评价系数又调得太大了，需要调小一点。在反复调整的过程中，评价标准的公式会找到一组相对合适的系数，将学长、学姐是否挂科划分准确，此时，小明就可以输入自己的成绩来看看自己是合格还是挂科。

如果我们让程序来执行上面小明的工作流程，一个简易的感知器也就形成了（图2-3）。学长、学姐两次作业和考试的成绩就是三个输入节点，好比接收外界刺激信息的神经元。最终判断是否挂科的输出节点，也可以看作一个神经元，而根据分数情况算出合格与挂科的函数叫作激励函数（Activation Function）。输入节点和输出节点之间神经信号的通信就是由评价标准公式的计算来传递的，而传递信号的强弱就是作业和考试分数所对应系数的大小。通过将传递信号的强弱反复调整到一个合适的值，也就完成了模型的学习，可以用于分类等任务。

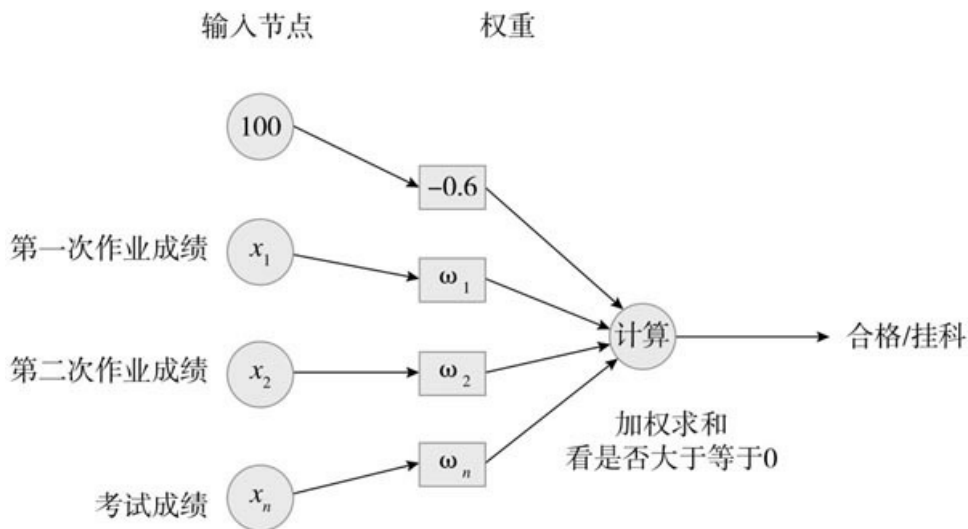


图2-3 简化版感知器结构示意图

而我们前面反复提及的人工神经网络，就可以看作一个多层的感知器。在人工神经网络中，除了和感知器一样拥有包括输入节点的输入层和包括计算出输出结果的输出层外，还加入了若干隐藏层。隐藏层中间的神经元节点可以与输入节点和输出节点一一相连，每条连接的链条上都有各自的权重系数，最终构成了一个网络的结构。

那么，为什么要加入隐藏层呢？我们不妨考虑一个更加复杂的课程成绩预测的例子。老师把课程评价的考查维度划分为态度得分、能力得分、创新得分，这三个分数会根据前面的两次作业、一次考试以及两次报告进行特定的处理得出，而这三个分数经过特定计算后会输出为课程的结果。那么，刚刚搭建的结构就不能满足这种情况了，因为每个特征可能要在模型中经过中间处理，不过，刚刚说的态度、能力、创新得分的例子只是为了方便理解进行的举例，实际上这个中间的过程是由隐藏层自行决定节点是什么样子的（图2-4）。

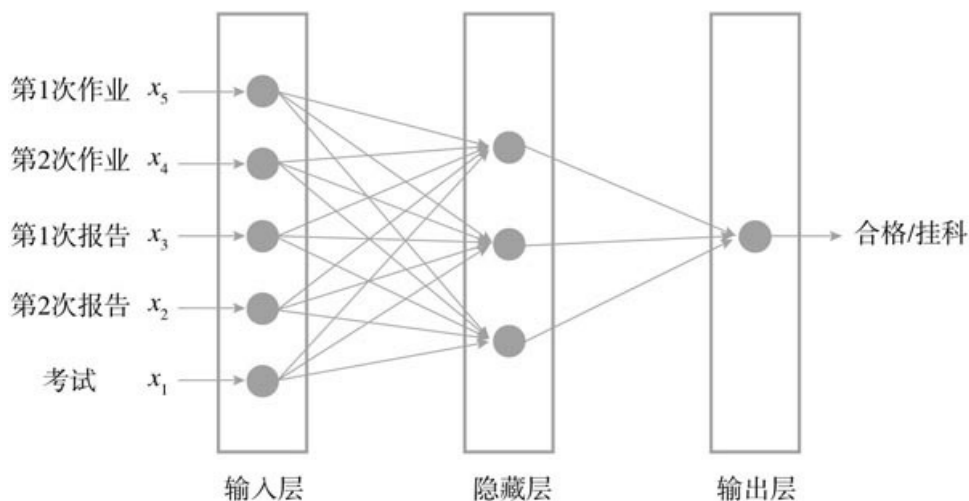


图2-4 简化的人工神经网络结构示意图

除此之外，在网络的结构下，激励函数也可以被替换成其他形式，以解决更加复杂的问题。与感知器一样，人工神经网络也需要在训练数据的过程中反复调整各神经元连接的权重，以完成模型的学习过程。而调整的依据是对比数据和模型的结果来查看神经网络有没有犯错误。如果在数据上存在误差，就相当于造成了损失，输出每个样本数据损失的函数叫作损失函数（Loss Function）。而所有的损失综合在一起的平均情况，会反应在代价函数

（Cost Function）里，描述训练这一个模型产生的错误代价。不过，需要注意的是，这里的代价并非越小越好。根据前面的例子，我们当然希望能好好利用先前学长、学姐给到的经验

数据，避免产生经验风险；但是，每一届的课程情况可能有变化，过多地利用过去经验产生结构复杂的模型，可能无法很好地适应新一届学生的情况，从而造成无法使用，产生结构风险。所以，调整后的合适标准应该是综合考虑经验风险和结构风险的结果。

## 四、强化学习

### 1.强化学习的概念

强化学习是机器学习除监督学习与无监督学习之外的又一领域，也可以与深度学习结合进行深度强化学习。区别于监督学习和无监督学习，强化学习并不是要对数据本身进行学习，而是在给定的数据环境下，让智能体学习如何选择一系列行动，来达成长期累计收益最大化的目标。强化学习本质上学习的是一套决策系统而非数据本身。它与监督学习、无监督学习的区别如表2-1所示。

表2-1 监督学习、无监督学习和强化学习对比

对比维度	监督学习	无监督学习	强化学习
学习对象	有标签数据	无标签数据	决策系统
学习反馈	直接反馈	无反馈	激励系统
应用场景	预测结果	寻找隐藏的结构	选择一系列行动

根据前文表述，强化学习听起来似乎是在玩一场游戏。环境就是游戏，智能体就是玩家，目标就是玩家在这个游戏中需要达成的核心任务，玩家需要不断地玩游戏来学习如何选择一系列行动达成游戏目标。在实际应用中，强化学习确实也广泛地应用在游戏领域，无论是棋牌这种简单游戏，还是《王者荣耀》、*Dota*、《星际争霸》等复杂游戏，都可以对强化学习加以应用。例如，谷歌旗下DeepMind公司研发出了围棋人工智能AlphaGo，它的训练过程就结合了强化学习的技术，它在2016年、2017年分别击败了李世石和柯洁两位围棋世界冠军，名噪一时。

### 2.强化学习的构成元素

强化学习系统的逻辑如图2-5所示，我们可以用一场《超级马里奥》游戏来分析图中的每个元素。



- 智能体（Agent）：人工智能操作的马里奥，它是这个游戏的主要玩家。
- 环境（Environment）：马里奥的游戏世界，马里奥在游戏里做出的任何选择都会得到游戏环境的反馈。
- 状态（State）：游戏环境内所有元素所处的状态，可能包括马里奥的位置、敌人的位置、障碍物的位置、金币数、马里奥的变身状态等，玩家的每次选择可能都会观测到状态的改变。
- 行动（Action）：马里奥可以做出的选择，可选的行动可能会随着状态的变化而变化，比如在平地的位置上可以选择左右移动或跳起，遇到右侧有障碍物时就无法选择向右的行动，获得火焰花道具变身后可以选发射火焰弹的行动等。
- 奖励（Reward）：马里奥在选择特定的行动后获得即时的反馈，通常与目标相关联。如果反馈是负向的，也可以被描述为惩罚。马里奥的游戏目标是到达终点通关，因而每次通过都可以获得奖励分数，而每次失败都会被扣除奖励分数。如果目标是获得尽量多的金币，奖励也可以与金币数量挂钩，这样训练出的马里奥AI不会去尝试通过终点，而是拼命在关卡里搜集金币。
- 目标（Goal）：在合理设置奖励后，目标应该可以被表示为最大化奖励之和，例如马里奥的通关次数最多。

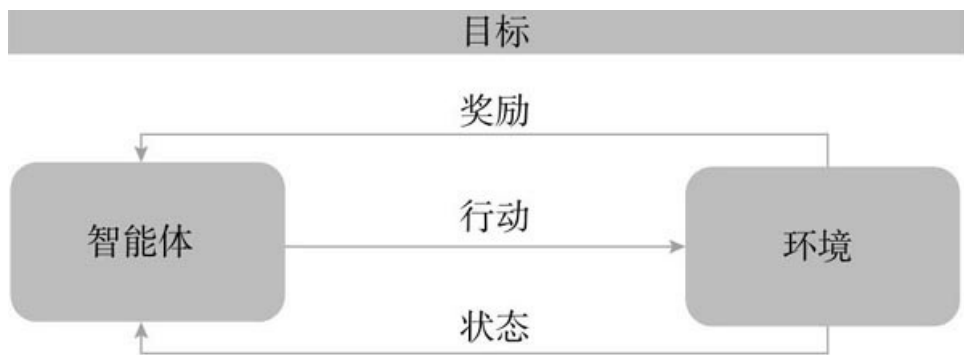


图2-5 强化学习构成元素及其关系

整个强化学习的过程，是为了学到好的策略（Policy），本质上就是学习在某个状态下应该选择什么样的行动，在刚刚的例子中就相当于马里奥的通关秘籍，输入马里奥每次的状态，秘籍会输出告诉你马里奥应该采取的行动，如此循环往复就能通关。因此，强化学习就是让人工智能通过不断的学习试错，找到合适的策略去选择一系列行动，来达成目标。在构建策略时，还有一个需要考虑的关键因素叫作价值（Value），它反映的是将来能够获得所有奖励的期望值。例如，马里奥为了达成目标，获得更多的奖励，所以应该选择多进入高价值的状态，并且在高价值状态下选择能够产生高价值的行动。

### 3.强化学习的训练过程

介绍完强化学习的基本概念，下面我们根据这些基本概念来描述一下强化学习算法的工作过程。

- 观测环境，获取环境的状态并确定可以做出的行动：马里奥目前在一个悬崖边上，系统读取了所有元素的状态，马里奥可以左右移动或者跳起。

- 根据策略准则，选择行动：策略里面显示，这种状态下左右移动和跳起的价值差不多，在差不多的情况下，马里奥应该向右走。

- 执行行动：马里奥在人工智能的指挥下向右走。

- 获得奖励或惩罚：马里奥掉下了悬崖，游戏失败，被扣除一定的奖励。

- 学习过去的经验，更新策略：在这个悬崖边向右走的价值较低，获得奖励的概率更低，人工智能知道后应该倾向于操作马里奥跳起或左走。

- 重复上述过程直到找到一个满意的最优策略。

综合上述过程，我们可以发现，强化学习其实可以看作一个从试错到反馈的过程，通过不断地试错，来找到一个合适的策略。不过，每一个行动的反馈其实都是有延迟的，大多数状态下，马里奥都不会因为跳起或左右移动而输掉游戏或赢得游戏，从而获得惩罚或奖励，但这并不代表这个行动就没有价值，因为未来的胜利或失败就是一系列行动所导致的，现在的行动会影响未来的奖励。不过，这也带来了一个问题：现在看起来价值最高、最优的行动真的就是最终最优的吗？是否可能只是因为没有充分地尝试采取其他行动呢？因而，对于很多强化学习的过程来说，我们通常会在没有充分尝试时，选择积极探索（Exploration），而充分尝试之后会选择倾向于直接利用（Exploitation）现有的价值信息，综合适应强化学习“试错”和“延迟反馈”两大特征。

当然，强化学习不仅可以用于游戏类人工智能的训练，许多AIGC的模型都结合了强化学习的技术，后文将对此展开详细介绍。

## 五、深度学习

### 1.深度学习的概念

经过前面对于机器学习的介绍，我们可以知道，特征的选取和处理对于模型训练是十分重要的，但在一些场景下，想要直接提取出合适的有效特征无疑是非常困难的，比如提取图片和句子的特征。在这种情况下，机器需要学习的并不是图片中的颜色数量、图形大小，或是句子中的词语数量等这种浅层次的特征，而是需要学习深藏在图片像素之间的复杂关系，或是句子中词语之间的上下文联系。人类无法自行处理这种深层特征的提取转换，而是需要由有深度的模型进行自动计算，采用的模型主要是复杂化了的神经网络，也被称为深度神经网络。

络。而所谓的深度学习，简单理解就是采用像深度神经网络这样有深度的层次结构进行机器学习的方法，是机器学习的一个子领域。深度学习与无监督学习、监督学习及强化学习的关系如图2-6所示。

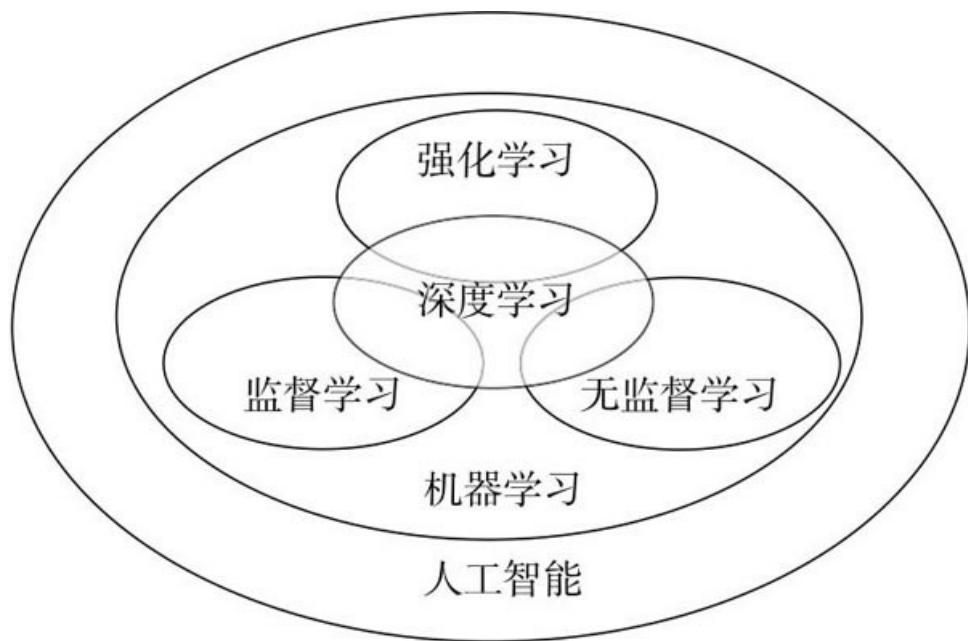


图2-6 深度学习与无监督学习、监督学习及强化学习的关系

资料来源：Yuxi Li（2018），“Deep Reinforcement Learning”，doi: 10.48550/arXiv.1810.06339

## 2.深度神经网络与一般神经网络的区别

根据前面的描述，可以得出深度神经网络和一般神经网络的四点区别：

- 深度神经网络具有更多的神经元。
- 深度神经网络层次更多、连接方式更复杂。
- 深度神经网络需要更庞大的计算能力加以支持。
- 深度神经网络能够自动提取特征。

结合这些特点，我们可以将深度学习运用在计算机视觉（Computer Vision，简称CV）、自然语言处理等涉及复杂特征的领域，后文中各类AIGC模型的主体基本上都是深度学习模型。

## 第二节 早期AIGC的尝试：GAN

GAN（生成对抗网络）诞生于2014年，是早期广泛应用于AIGC的算法之一，有诸多衍生形式，并至今仍被诸多AIGC应用所采用。GAN综合了深度学习和强化学习的思想，通过一个生成器和一个判别器的相互对抗，来实现图像或文字等元素的生成过程。原始的GAN并不要求生成器和判别器都是一个深度神经网络，但是在实践中通常都采用深度神经网络去构建GAN，下面将对它的构建原理进行介绍。

### 一、生成器

我们可以向生成器（Generator）输入包含一串随机数的向量，生成器会根据这一串随机数生成并输出图像或句子。向量里的每一个数字都会与生成的图像或句子的特征相关联。打一个并不严谨的比方，假设生成器收到的输入是[0.1, -0.5, 0.2 ... 0.9]，据此生成了一张小猫的图片，而第一个数是和小猫的颜色相关的，当你把0.1换成0.2时，小猫可能就从橘猫变成了白猫。因为随机数是可以随意构造的，因此我们就可以利用生成器生成各种各样的新图片。不过，和一般的神经网络一样，在生成之前会有提前训练的过程，我们需要准备一个全是各种各样小猫图片的数据集供生成器训练。

### 二、判别器

判别器（Discriminator）用于评价生成器生成的图像或句子到底看起来有多么真实。判别是否真实的方式也很简单，就是看这个图像或句子像不像来自生成器训练用的数据集，因为数据集是最真实的。我们可以向判别器输入一个生成的图像或句子，判别器会输出一个数值（也被称为得分）。一般来说，我们会使用0到1的区间来表示得分，如果这个图像或句子非常像数据集里的真实数据，得分就会靠近1；反之，得分就会靠近0。

### 三、生成对抗过程

以图像生成的过程为例，生成器就好像一个学习画画的学生，而判别器就是评价学生画作的老师。一开始，学生读一年级，他看了一堆小猫的图片，然后随便画了一只猫，老师看了看学生画的猫，说画得不够逼真，看不清小猫的两只眼睛，这就是最开始的生成器和判别器的交互过程。学生努力练习画画，终于画好了小猫的两只眼睛，老师一看说合格了，然后学生升到了二年级，老师也开始依照二年级的评价标准去评价学生的画作，相当于生成器和判别器的性能都提升了。升到二年级后，学生再拿出原来的小猫画作肯定就无法令老师满意了，老师会觉得画得不够真实，无法看清小猫的脸部轮廓，于是学生又反复练习修改，直到令老师满意，于是学生升到三年级。如此循环往复，学生画画的水平会越来越精湛，画作看

起来越来越真实。而老师判别画作的标准也会越来越严苛，督促学生完善画技，这就是生成器和判别器对抗过程的基本思想（图2-7）。而就具体的实现过程来说，可以把GAN的训练过程分为两个步骤。

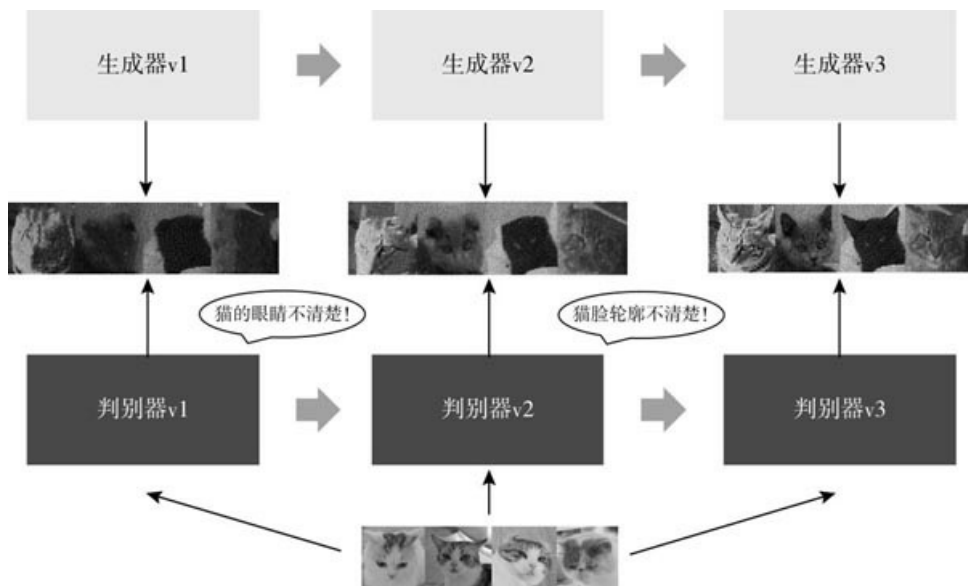


图2-7 生成对抗过程示意图

### 步骤一：固定生成器，更新判别器。

首先，生成器抽取一些包含一系列随机数的向量，输入生成器之中，生成器会生成一系列图片。这时，在生成器内部参数不变的情况下，判别器需要从生成器训练的数据集中抽取一部分图片，将它们和生成器生成的图片一起做学习训练。判别器需要调整内部参数，学习给真实的图片打高分，给生成器生成的假图片打低分。就好比如果想要让老师指导学生，首先需要对老师进行教学技能培训，让老师先学会评价标准，才能去教育和考查学生。

### 步骤二：固定判别器，更新生成器。

判别器训练好之后，保持内部参数不变，生成器需要调整内部参数进行训练，以学会如何在判别器那里取得高分。这个过程就像学生反复考试一样，在每次复盘自己的失误进行改进后，终有一天会达到老师的标准。

步骤一和步骤二交替反复进行，GAN最终就可能生成让人满意的作品。[\[2\]](#)

#### 四、GAN的AIGC应用

虽然GAN的一些变体也可以用于句子这种文本类信息的生成，但因为对于离散型数据的处理能力较差，AIGC应用最广泛的场景还是在图像之中，或是与图像相关的跨模态生成中。表2-2展示了GAN的一部分AIGC应用案例。[\[3\]](#)

表2-2 GAN的部分常见AIGC应用方式

类别	应用方式	描述
图像数据集的生成	手写数字图片数据集生成	提取手写数字笔迹特征，生成新的手写数字图像集合
	人脸图像集生成	提取人脸特征，生成新的人脸图像集合
	动物图像集生成	提取动物特征，生成新的动物图像集合
	动漫人物数据集生成	提取动漫人物特征，生成新的动漫人物图像集合

续表

类别	应用方式	描述
图像联想创作	面部正面视图生成	通过人脸局部或侧面的照片可以生成完整的面部正面视图
	人体新姿势生成	通过人体任意的一张照片生成具有全新姿势的照片
	照片转卡通头像	将真实照片转换成与其类似的卡通头像
	照片编辑	重建具有特定特征的面部照片，例如头发颜色和风格、面部表情甚至性别的变化
	不同年龄面部图片生成	生成同一个人不同年龄（从年轻到年老）的面部照片
	照片融合	将不同元素的照片进行混合，如田野、山脉等大型地理结构相融合
	服装生成	根据模特穿着服装的照片直接生成服装照片
	草图上色	对服装饰品线条草图进行上色
图像修复	分辨率增强	生成具有更高像素分辨率的输出图像
	照片填充	填补照片中因某种原因被删除的区域
多模态与跨模态生成	文本生成图片	输入对图片在颜色、对象、场景等方面的描述，生成完全符合要求并且十分逼真的图片
	文本生成语音	进行文本向语音的转换
	3D 物体生成	从多个角度将物体的二维图片生成三维模型，如通过椅子的多角度二维图像进行三维重建
	视频预测生成	预测视频中后续画面的内容，如海浪后续的波动、人行进的轨迹等
	游戏关卡生成	通过使用视频游戏关卡数据生成新的游戏关卡



### 第三节 AI绘画的推动者：Diffusion模型

#### 一、Diffusion模型的基本原理

Diffusion模型是一类应用于细粒度图像生成的模型，尤其是在跨模态图像的生成任务中，已逐渐替代GAN成为主流。在2022年美国科罗拉多州博览会艺术比赛中击败所有人类画家、斩获数字艺术类冠军的AI创作画作《太空歌剧院》的底层技术模型就涉及Diffusion模型。

传统的GAN虽然已经能较好地完成与图像相关的生成任务，但依然存在以下诸多问题。

- 需要同时训练生成器和判别器这两个深度神经网络，训练难度较大。
- 生成器的核心目标是骗过判别器，因而可能会选择走捷径，学到一些并不希望被学到的特征，模型并不稳定，有可能会生成奇怪的结果。
- 生成器生成的结果通常具备较差的多样性，因为具有多样性的结果不利于骗过判别器。

为了解决这些问题，Diffusion模型尝试使用一种更加简单的方法生成图像。大家是否记得老式电视机信号不好时屏幕上闪烁的雪花？这些雪花是随机、无序、混乱的，因而被称为噪声。当电视机信号不好的时候，屏幕上就会出现这些噪声点，信号越差就会出现越多的噪声点，直到最后屏幕完全被随机的噪声覆盖，图2-8就展示了这样一个在图像上增加噪声的演变过程。那么换一个角度去思考，既然任何一张图像都可以在不断添加噪声后，变成一张完全随机的噪声图像，那我们能不能将这个过程翻转，让神经网络学习这个噪声扩散的过程之后逆向扩散，把随机生成的噪声图像，逐渐转化为清晰的生成图像呢？Diffusion模型就是基于这个思想实现的。



图2-8 图片增加噪声的演变示意图

#### 二、CLIP模型与AI绘画

除了前文GAN部分提到的图像处理领域之外，Diffusion模型应用最广泛的领域就是AI绘画，并且迅速地表现出较大的商业潜力，拓展出大量相关的应用。此外，AI绘画的成功还

归功于CLIP（Contrastive Language-Image Pre-Training，文本-图像预训练）模型。

CLIP模型是OpenAI在2021年初发布的用于匹配图像和文本的预训练神经网络模型。下面我们将对CLIP模型进行简单而形象的介绍。

如果要实现相对优质的AI绘画，需要让AI很好地理解图片，那么要解决的主要问题有两个：理解力差异和数据量不足。在理解力方面，人类和AI认识图片的方式是不一样的，人类主要是从整体上对图片中的形象进行理解，而AI则是对图片上一个个像素的特征进行学习。而在数据量方面，需要对大量图片数据进行标注来训练AI，即便在目前有許多分类标注好的开源数据集的情况下，AI性能的提升还是不尽如人意。

当这两个问题的解决逐渐走入瓶颈时，研究者开始转换让AI学会理解图片的思路。对于人类来说，在婴儿时期学习图片并不是具体地学习一个个像素，而是父母指着图片告诉孩子：“这是一只在吃猫粮的黑色小猫”，或者“这是一辆在马路上飞驰的红色汽车”。于是，研究者开始思考，AI的学习过程是否也能采用这种类似的方式？这本质上是一个文本和图像匹配的问题。如果要完成这个任务，自然也需要大量的数据，但互联网上天然就有海量这样的数据，无论是发朋友圈、微博还是推特等，本质上都是用一段文字去说明发布的图片，很容易就可以获取大量标注好的图像文本对。这样，前面提出的两个问题也就迎刃而解了。为此，OpenAI在互联网上收集到了4亿对质量过关的图像文本对，分别将文本和图像进行编码，让CLIP模型学会计算文本和图像的关联程度。在此基础上，结合Diffusion模型对图像的生成能力，就可以打造一款AI绘画应用了。

例如，Disco Diffusion就是早期结合CLIP模型和Diffusion模型变体开展AI绘画业务的知名案例。Disco Diffusion是发布于Google Colab平台的一款开源AI绘画工具，由Accomplice公司开发并在2021年10月上线。Disco Diffusion内核采用了CLIP引导扩散模型（CLIP-Guided Diffusion Model），而整体应用基于谷歌技术架构构建，需要借助Google Colab 平台生成，用户界面并不友好，而且运行成本高，用户需要自己租用Colab Pro来提升模型性能。虽然Disco Diffusion具有诸多局限，但作为早期出现的成型开源AI绘画产品，它依然掀起了用户使用的热潮。

### 三、知名AI绘画工具

许多公司在CLIP模型和Diffusion模型的基础上开发了模型变体的相关应用工具，其中，Stable Diffusion、DALL·E 2、Midjourney是最知名的工具，其发布时间和研发企业如表2-3所示。

表2-3 Stable Diffusion、DALL·E 2、Midjourney基本信息表

名称	发布时间	研发企业
Stable Diffusion	2022 年 8 月上线	Stability AI
DALL·E 2	2022 年 4 月更新	OpenAI
Midjourney	2022 年 7 月公测	Midjourney

上述三个知名AI绘画工具都具有各自的特点。Stable Diffusion对于生成当代艺术图像具有较强的理解力，善于刻画图像的细节，但为了还原这些细节，它在图像描述上需要进行非常复杂细致的说明，比较适合生成涉及较多创意细节的复杂图像，在创作普通图像时可能会略显乏力。DALL·E 2由其前身DALL·E发展而来，其训练量无比庞大，更适合用于企业所需的图像生成场景，视觉效果也更接近于真实的照片。而Midjourney则使用Discord 机器人来收发对服务器的请求，所有的环节基本上都发生在Discord上，并以其独特的艺术风格而闻名，生成的图像比较具有油画感。<sup>[4]</sup>不过，这种艺术风格既是优点也是缺点。虽然Midjourney在生成画作时具有显著的风格优势，例如前文提到的夺得数字艺术类比赛冠军的作品《太空歌剧院》就是用Midjourney生成，但它很难生成看起来像照片的图像。接下来我们可以尝试利用这三种工具为同一个句子生成图片，以获得对这些工具效果的感性认知，实现效果如图2-9、图2-10、图2-11所示（生成时原句用英文表示）。



图2-9 使用“下雨天，向日葵盛开于海边”生成图片对比

Stable Diffusion



DALL · E 2



Midjourney



图2-10 使用“明亮的小巷在夕阳中衬着雪色之美”生成图片对比

Stable Diffusion



DALL · E 2



Midjourney



图2-11 使用“圆形的飞船伫立在沙漠之上，覆着皑皑白雪”生成图片对比

除了上述三种工具之外，许多大厂也推出了自己的AI绘画工具，例如谷歌的Imagen、微软的NUWA，等等。这些工具大多基于大模型来实现，而Transformer作为AIGC的重要基础设施发挥了巨大的作用，我们将在下一节对其进行详细介绍。

## 第四节 大模型的重要基建：Transformer

### 一、Seq2Seq模型

在正式介绍Transformer之前，我们先了解一种更简单的模型——Seq2Seq（Sequence-to-Sequence，序列到序列）模型。Seq2Seq模型最早在2014年提出，主要是为了解决机器翻译的问题。Seq2Seq模型的结构包括一个编码器和一个解码器，编码器会先对输入的序列进行处理，然后将处理后的结果发送给解码器，转化成我们想要的向量输出。举例来说，如果使用Seq2Seq模型将中文翻译成英文，其过程就是输入一个中文句子，编码成包含一系列数值的向量发送给解码器，再用解码器将向量转化成对应的英文句子，翻译也就完成了。除了翻译外，许多自然语言处理的问题都可以使用Seq2Seq模型（虽然使用效果未必最佳），下面是一些实例。

- 聊天问答：输入一个问题序列，输出一个回答序列。
- 内容续写：输入一个段落序列，输出后续内容的段落序列。
- 摘要/标题生成：输入一篇文章序列，输出一个摘要/标题序列。
- 文本转语音：输入一个文本序列，输出一个语音序列。

当然，除了自然语言处理领域，一些如图像、字幕等计算机视觉领域也有Seq2Seq模型的应用，这里就不再展开讲述了。

### 二、注意力机制

人工智能领域的注意力机制一开始主要用于图像标注领域，后续被引入到自然语言处理领域，主要是为了解决机器翻译的问题。虽然Seq2Seq模型可以实现将一种语言翻译为另一种语言，但随着句子长度的增加，翻译的性能将急剧恶化，这主要是因为很难用固定长度的向量去概括长句子里的所有细节，实现这一点需要足够大的深度神经网络和漫长的训练时间。为了解决这一问题，学者们引入了注意力机制。在了解注意力机制之前，先请看一幅画作《圣母与圣凡尼诺》（图2-12）。



图2-12 画作《圣母与圣吉凡尼诺》

注：这幅画由佛罗伦萨画家多米尼哥·基兰达约创作于15世纪，现藏于佛罗伦萨维琪奥王宫。

看完这幅画，相信首先映入你眼帘的是圣母玛利亚以及正在接受祈祷的婴儿耶稣。如果不回看这幅画，你的脑海里是否对右下角的一头驴和一头牛有印象？如果你没有印象，这其实是一种非常正常的现象，因为人的注意力是有限的，无论是观看图像还是阅读文字，人们

都会有选择性地关注一小部分重点内容，并忽略另一部分不重要的内容。从数学的角度来说，可以将“注意力”理解为一种“权重”，在理解图片或文本时，大脑会赋予对于认知有重要意义的内容高权重，赋予不重要的内容低权重，在不同的上下文中专注不同的信息，这样可以帮助人们更好地理解信息，同时还能降低信息处理的难度。这就是注意力机制，这种机制被应用在人工智能领域，帮助机器更好地解决图像处理和文本处理方面的一些问题。

那么，注意力机制在人工智能领域是如何运作的呢？在回答这个问题之前，请先阅读图2-13中的一段话。

# 研表究明

## 汉字序顺并不定一影阅响读

### 比如当你看完这句话后

### 才发这现里的字全是都乱的

图2-13 网络上广泛流传的一段话

阅读完这段话之后，你一定发现，虽然图片上的语句是乱序的，但是并没有干扰你的阅读，这种现象原理与人工智能的自注意力（Self-Attention）机制非常相近，下面我们用通俗易懂的语言对这套机制进行分析。首先，你的眼睛捕捉到了第一个字“研”，并且扫过那一行的后续文字“表”“究”“明”。然后，大脑在过去学习的认知库里去搜寻“研表”“研究”“研明”等，发现“研究”两个字关联最为紧密，所以就给了它较高的权重进行编码计算，并按类似的方式完成后续内容的编码。编码完毕后，按照权重对内容进行重新组装，信息也就组合成了“研究表明”这一常见用法。通过这种自注意力机制，人工智能可以很好地捕捉文本内在的联系并进行再表示。而除了自注意力机制，另外一种广泛应用于人工智能领域的注意力机制叫作多头注意力（Multi-Head Attention）机制。多头注意力机制主要通过多种变换进行加权计算，然后将计算结果综合起来，增强自注意力机制的效果。这种注意力机制在后文介绍

的Transformer中会涉及。

### 三、Transformer的基本结构

Transformer与Seq2Seq模型类似，也采用了编码器-解码器结构，通常会包含多个编码器和多个解码器。在编码器内有两个模块：一个多头注意力机制模块和一个前馈神经网络模块，这里的前馈神经网络是一种最简单的人工神经网络形式。以英译汉机器翻译为例，编码器的工作过程大概是这样的：首先，用户输入一个英文句子，编码器会将这个句子的每个单词拆解，转化成向量的形式，并在多头注意力机制模块中加权计算，最后整个编码器会输出一个向量集，输入解码器中。在解码过程中，解码器首先读取一个开始标记，然后解码器会生成并输出一个向量，这个向量会包含所有可能的输出汉字，同时，每个数值会有一个得分，这个得分代表着汉字出现的可能性，得分最高的汉字会出现在第一个位置。例如，如果要把“I love you”翻译成中文，那第一个得分最高的汉字可能就是“我”。接下来，把“我”作为解码器的新的输入，接下来得分最高的可能是“爱”，以此类推，直到完全输出了“我爱你”，再输出一个结束符号。解码器内部的结构也和编码器类似，最开始包含一个多头注意力机制模块，最后包含一个前馈神经网络模块。需要注意的是，解码器中的多头注意力机制模块使用了掩码（Mask）机制，其核心思想是：因为解码器的生成物是一个个产生的，所以生成时只让参考已经生成的部分，而不允许参考未生成的部分。还是以前面的“我爱你”为例，当翻译到“爱”时，模型只能参考前面输入的开始标记和“我”这个字的信息，而后面的所有信息都会被遮掩住。此外，在两个模块中间，还有一个多头注意力机制模块，刚刚提到的来自编码器的向量集就会输入这里，让解码器在解码过程中能够充分关注到上下文的信息。

Transformer的内部结构简化图如图2-14所示。



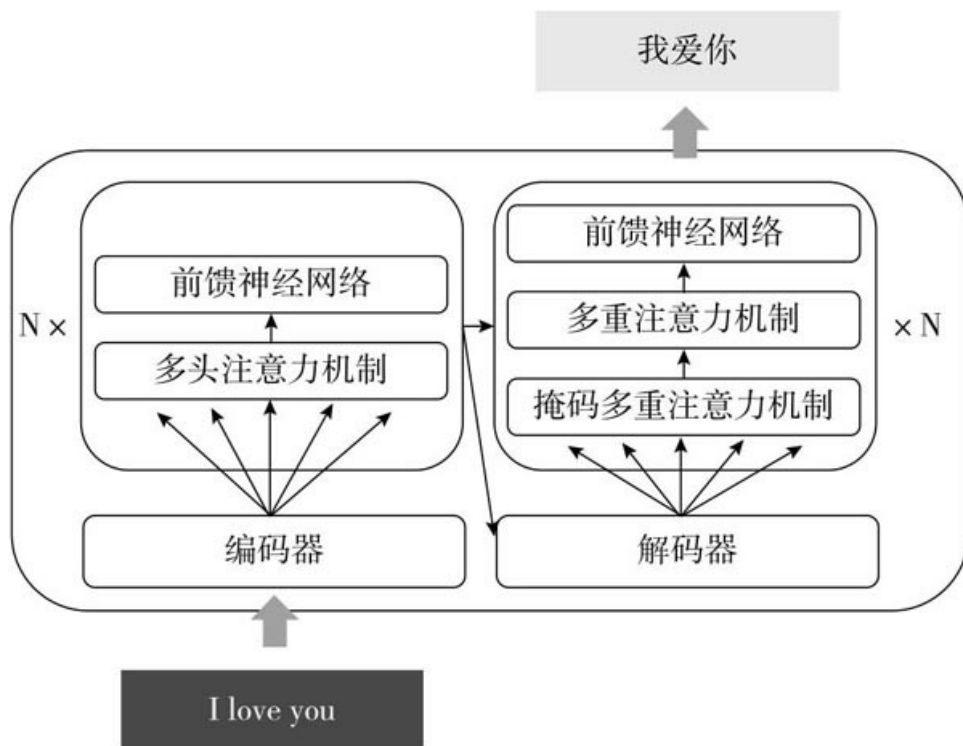


图2-14 Transformer结构简化图

#### 四、GPT系列模型与ChatGPT

GPT（Generative Pre-trained Transformer，生成型预训练变换器）是由OpenAI研发的大型文本生成类深度学习模型，可以用于对话AI、机器翻译、摘要生成、代码生成等复杂的自然语言处理任务。GPT系列模型使用了不断堆叠Transformer的思想，通过不断提升训练语料的规模与质量，以及不断增加网络参数来实现GPT的升级迭代，整个GPT系列模型的迭代升级之路如表2-4所示。

表2-4 GPT系列模型的演进信息

模型	发布时间	参数量	预训练数据量
GPT-1	2018 年 6 月	1.17 亿	约 5GB
GPT-2	2019 年 2 月	15 亿	40GB
GPT-3	2020 年 5 月	1 750 亿	45TB

到了2022年，GPT已经走过三代的历程，而备受期待的GPT-4也预计在不久的将来发布。GPT-3之后衍生的应用InstructGPT和ChatGPT都取得了令人惊异的效果，人们期待GPT-4能够拥有与人脑突触一样多的参数，并完美地通过无限制的图灵测试，而不是像过往一样利用一些特殊的设定盲点来通过图灵测试。例如，将AI伪装成外国小孩，在无法很好回答问题时人类可能会把原因归咎为这一特殊身份。而在GPT-4发布之前，OpenAI在2022年11月30日发布了聊天机器人ChatGPT。ChatGPT一经发布就因为惊人的效果而走红全网，它不仅能自然流畅地与人们对话，还能写诗歌、敲代码、编故事等。

虽然GPT模型在如今取得了如此夺目的成绩，但它的技术思想的发展还是经历了波折的过程。在GPT-1诞生之前，大部分自然语言处理模型如果想要学习大量样本，基本上都是采用监督学习的方式对模型进行训练，这不仅要求大量高质量的标注数据，而且因为这类标注数据往往具有领域特性，很难训练出具有通用性的模型。为了解决这一问题，GPT-1的核心思想是将无监督学习作用于监督学习模型的预训练目标，先通过在无标签的数据上学习一个通用的语言模型，然后再根据问答和常识推理、语义相似度判断、文本分类、自然语言推理等特定语言处理任务对模型进行微调，来实现大规模通用语言模型的构建，这可以理解成一种半监督学习的形式。此外，GPT-1在训练时选用了BooksCorpus数据集来训练模型，它包含了大约7 000本未出版的书籍的文字，这种更长文本的形式可以更好地让模型学习到上下文的潜在关系。最终，GPT-1在多数任务中取得了更好的效果，但依然存在很大的问题：一是基于未发表书籍数据训练具有一定的数据局限性，二是在一些任务上的性能表现还是会呈现泛化性不足的现象，这只能让AI成为领域的专家，而无法成为通用的模型。

为了增强GPT模型的泛化能力，GPT-2在GPT-1的基础上进行了技术思想上的优化。GPT-2的核心出发点是：在语言模型领域，所有监督学习都可以看作无监督学习的子集。例如，把“小明是A省2022年高考状元”丢给算法做无监督学习，但是它也能学会完成“A省2022年高考状元是谁？”“小明是2022年哪个省的高考状元？”等需要标注正确答案的监督学习任务。因此，当模型的容量非常大且数据量足够丰富时，一个无监督学习的语言模型就可以覆盖所有监督学习的任务。在这样的指导思想下，GPT-2的模型参数达到了15亿，相较于GPT-1翻了近10倍，同时，训练用的数据集集为了Reddit上约800万篇高赞文章，训练数据量也翻了约8倍。而在后续的测试中，GPT-2的确在许多自然语言处理任务方面表现出了普适而强

大的能力，但仍然具有很大的待提升空间。GPT-3基本上沿用了GPT-2的结构，但在参数量和训练数据集上进行了大幅增加，参数量增加了百倍以上，预训练数据增加了千倍以上。在这样夸张的增幅下，GPT-3也最终实现了“大力出奇迹”，在自动问答、语义推断、机器翻译、文章生成等领域达到了前所未有的性能。

这样的技术飞越无疑是振奋人心的，而每个人都可以通过体验ChatGPT流畅的对话过程来体验技术的演进。ChatGPT是由其前身InstructGPT改进而来，InstructGPT是一个经过微调的新版本GPT-3，可以尽量避免一些具有攻击性的、不真实的语言输出。InstructGPT的主要优化方式是从人类反馈中进行强化学习（Reinforcement Learning from Human Feedback，简称RLHF）。而ChatGPT采用了和InstructGPT一样的方法，只是调整了数据收集方式。ChatGPT完整的训练过程如图2-15所示。

步骤一：收集示范数据并训练一个监督学习的策略。

模型会从问题库里抽取问题，由工作人员撰写问题的答案，这些标记了答案的问题会被用于调优GPT-3.5模型（GPT-3的改进版）。

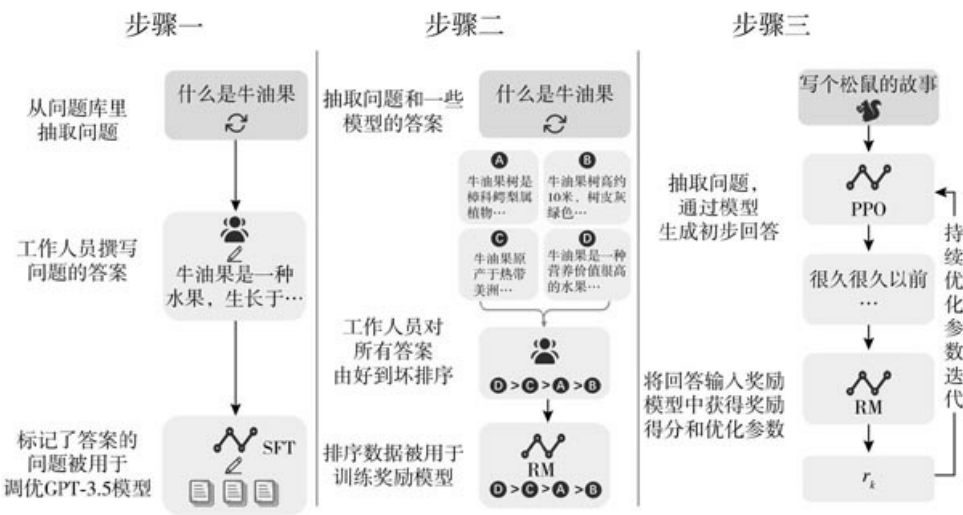


图2-15 ChatGPT的训练过程示意图

资料来源：<https://openai.com/blog/chatgpt/>

注：RM：Reward Model，奖励模型；SFT：Supervised Fine-Tuning，有监督的微调；PPO：Proximal Policy Optimization，近端策略优化

## 步骤二：收集对比数据并训练一个奖励模型。

抽取问题和一些模型的答案，工作人员会对所有答案由好到坏排序，这些排序数据会被用于训练奖励模型。

## 步骤三：使用强化学习算法优化针对奖励模型的策略。

抽取问题，通过模型生成初步回答，回答会被输入奖励模型中得到评分和优化参数，并在优化后重复优化的过程。

上述训练方法让模型更加清晰地理解了人类对话的意图，并获得了多轮对话的能力。真格基金的林惠文曾在线上分享中表示，ChatGPT带来了不少有趣的提升：[\[5\]](#)

- 敢于质疑不正确的前提。
- 主动承认错误和无法回答的问题。
- 大幅提升了对用户意图的理解。
- 大幅提升了结果的准确性。

这些提升无疑是可喜可贺的，不过，ChatGPT也并非完美的，依然存在很多问题。根据OpenAI的官方文档及用户实践经验，目前，ChatGPT的局限性包括：

- 有时会写出看似合理但不正确或荒谬的答案。
- 对输入措辞的调整或多次尝试相同的提示很敏感。例如，给定一个问题的措辞，模型可以声称不知道答案，但只要稍作改写，就可以正确回答。
- 回答通常过于冗长并过度使用某些短语。
- 对于模棱两可的问题，模型通常会猜测用户的意图，而非让用户澄清问题。
- 模型有时会响应有害的问题或表现出有偏见的行为。
- 在数学和物理等需要进行数字推理的任务中仍然会出现一些错误。

不过，这些局限性并没有影响ChatGPT的突破性成就，反而让人们更加期待GPT-4在未来究竟会带来什么样的惊喜。

## 五、BERT模型

BERT（Bidirectional Encoder Representations from Transformers，变换器的双向编码器表示）模型由谷歌在2018年提出，其基本思想是既然编码器能够将语义很好地抽离出来，那直

接将编码器独立出来也许可以很好地对语言做出表示。此外，BERT模型的训练过程也别出心裁，它设计了两个有趣的任务。

- 掩码语言模型：随机覆盖15%的单词，让BERT模型猜测掩盖的内容是什么，这有利于促进模型对语境的理解。

- 下句预测：输入成组的句子让BERT模型判定它们是否相连，让模型更好地了解句子之间的联系。

不过，当执行不同的自然语言处理任务时，训练好的BERT模型需要根据具体的任务类型增加不同的算法模块才能执行任务。除了自然语言处理任务，BERT模型也可以应用于机器视觉领域。在输入阶段，将图片分割成一个个小块，每个小块图片就可以看作一个个单词，这样就可以像处理句子一样去处理图片了。基于这样的思想，ViT（Vision Transformer，视觉变换器）模型也就诞生了。除了上面提到的模型，基于BERT模型还发展出了诸多变体，在AIGC领域大放异彩，奠定了BERT模型里程碑式的地位。

[1] 参考自<https://www.jianshu.com/p/83f5b52d3cc>。

[2] 参考自[https://www.youtube.com/playlist?list=PLJV\\_el3uVTsMq6JEFW35BCiOQTsoqwNw](https://www.youtube.com/playlist?list=PLJV_el3uVTsMq6JEFW35BCiOQTsoqwNw)。

[3] 参考自<https://machinelearningmastery.com/impressive-applications-of-generative-adversarial-networks/>。

[4] 参考自<https://www.marktechpost.com/2022/11/14/how-do-dall%C2%B7e-2-stable-diffusion-and-midjourney-work/>。

[5] 参考自<https://mp.weixin.qq.com/s/UhCGelO3LKta9SdheWBDg>。

### 第三章 AIGC的职能应用

AIGC如何帮助企业各职能部门降本增效？

有些人声称这种技术是人工智能，但实际上它强化的是人类自身。因此我认为，我们将增强人类的智能，而非“人工”的智能。

——吉尼·罗曼提（Ginni Rometty）

创新是一个企业发展的重要动力，企业中各个职能部门都会涉及大量的创作工作。而AIGC的出现，可以帮助企业不同职能岗位上的员工有效地提升生产力，最终实现整个企业的降本增效。具体降本增效的方式体现在：

- 自动化处理烦琐和耗时的任务，减少人力需求，降低成本。
- 产生新的想法和问题的解决方案，如产品设计或营销策略。
- 快速、准确地分析大量数据，为决策生成有价值的见解。
- 提高任务的效率和准确性，减少出错的可能性，提高工作效率。
- 开发个性化和定制化的产品和服务，提高客户满意度。
- 提高组织的速度和敏捷性，使组织能够快速响应不断变化的市场条件和客户需求。
- 改善组织内部的协作和沟通，使团队能够更加高效地一起工作。

总体而言，AIGC能通过自动化处理任务、产生新想法、生成有价值的决策建议，有效赋能企业各个职能部门。本章将从产品研发、市场营销和管理协作三个角度对AIGC的职能应用进行介绍。

## 第一节 AIGC与产品研发

对于大多数互联网企业而言，产品研发是整个团队的成本与创新核心，其迭代的速度也决定着企业对市场的反应灵敏度。目前，AIGC在产品研发方面主要有四种应用方式：

- 通过辅助编程提高代码生产效率。
- 生成应用直接将需求变成产品。
- 创建和维护文档注释，提高沟通效率。
- 测试代码，纠正错误。

这些方式可以帮助开发人员更好地专注于产品架构设计、产品功能探索而非一些重复烦琐的工作。

### 一、智能辅助编程

2021年夏天，GitHub和OpenAI联合研发并发布了知名的人工智能辅助编程工具GitHub Copilot，其命名来自许多头部科技公司研发团队的“结对编程”方法：两个程序员共同完成包括需求分析、代码创作和审查测试在内的某项功能的研发，以此提高生产效率和减少代码缺陷。整个结对编程的过程就好像在驾校练车，需要一个“驾驶员”去输入代码，还需要一个“观察员”去审查代码。而GitHub Copilot可以以人工智能的身份坐在“副驾驶”（Copilot）上指导驾驶员，名字也就由此得来。

Github Copilot在发布之后立刻受到了大量程序员和研发人员的关注和体验，并得到了极高的评价，被认为可以大幅度提高编程生产效率。不过，也有科技媒体表达了对于这种工具的担忧，主要担心的是在人工智能的模型训练阶段可能使用了Github开源平台上的代码，是否具有潜在的法律问题还存有争议。<sup>[1]</sup>

虽然像Github Copilot这类需要依靠公开代码训练的智能辅助编程工具普遍存在这类问题，但不得不说其生成的模式是具有变革性的。在传统的代码自动完成工具中，编程人员可以在代码编辑的同时看到众多对于当前所在行代码的推荐内容，根据上下文情境选择一个最佳选项自动完成。这样简单而便利的功能就已受到众多编程人员的肯定，并普遍成为几乎所有主流代码开发环境的默认配置。而GitHub Copilot则在传统功能之上实现了很大的跨越，它可以使用人工智能生成整个代码片段，而不仅仅是单个单词或短语。这意味着它可以为开发人员提供更全面和有用的建议，指导他们如何完成一段代码。除了代码自动完成这个简单的工作之外，人工智能在辅助编程中还可以发挥远超想象和预期的作用。

快速创建样板代码就是人工智能可以介入的另一个方向。样板代码是一种常见的代码类型，在许多不同的应用程序中被反复使用。它经常被用作新代码的起点，允许开发人员快速启动并运行一个基本框架，然后根据需要进行修改和定制。虽然目前许多快速创建样板代码都不是基于人工智能而是利用预制好的代码模板，但引入人工智能则可以借助输入文字描述直接生成更加定制化的样板代码。

人工智能的另一大潜在应用是对现有代码的优化，它可以通过分析代码提出可以优化效率的修改建议。例如，在开发人员编写代码时，AIGC模型就可以分析代码并提出修改建议，使其运行更快或使用更少的内存。然后，开发人员可以审查来自人工智能的建议并决定是否实现它们，与手动优化代码相比，这样可以节省时间和精力。事实上，开发者为了优化代码性能，往往要在算法和数据结构上花费至少数百个小时。仅仅是最简单的数字排序这一项任务，就存在着数十种对计算时间和存储空间具有不同要求的算法。这种人工智能对代码的自动优化，可以让很多经验较少的开发者写出高效而优美的代码内容，并能在改进建议中学习成长。

除了优化现有代码之外，人工智能根据不同种类的用户设备生成新的代码也是一个潜在的应用场景。尽管这个场景目前还没有出现被广泛使用的应用，但是对企业而言也意味着巨大的生产力提升。例如，许多互联网企业在开发一个新的程序时往往要对同一个应用进行数次开发，由完全不同的团队输出完全不同的代码以确保用户可以在不同设备上使用，仅仅移动端需要考虑的环境就包括移动网页端、安卓、苹果、小程序等。与手动编写和迁移代码相比，人工智能的应用可以为开发人员节省大量的时间和精力。

总的来说，通过为开发人员提供快速生成和优化代码的能力，AIGC可以帮助他们专注于工作中最重要和最具挑战性的方面，使他们能够更快、更容易地创建更好的软件。

## 二、智能应用生成

同代码自动完成一样，人们很早就开始探索如何更低成本地创作应用，近年来被更多人所关注和使用的低代码与无代码开发工具Bubble就是很好的案例。使用Bubble这一开发工具意味着人们无需代码或者写很少量的代码就可以完成一个应用的开发，但是人们仍然需要学习使用图形化编程工具以及使用图形和流程图表达他们所希望开发的逻辑和数据流。

而对于具有AIGC能力的应用而言，这一过程将会变得更加简单。你只需要学会用直白的语言描述你所要创作的应用功能，人工智能就可以帮你完成创作，这样就节省了学习一个全新的逻辑表达工具和经历烦琐开发流程的时间。位于美国硅谷的Debuild就是这个新兴领域的代表，用户可以简单描述产品后根据提示选择要包含的功能和对应的应用场景，软件就



可以自动生成网页端代码。

事实上，在通过人工智能生成应用这个场景中，开发者并非唯一的受益人，产品设计师也能借助AIGC工具获得效率提升。无论是负责视觉设计还是用户体验设计的设计师，在决定最终设计之前，通常都需要探索各种各样的设计可能性，并且根据团队和市场的反馈进行多次调整和重新设计，这也是一个耗时且烦琐的过程。而AIGC有可能被用来自动化处理一些这样的工作，使设计者能够快速地根据特定的输入或需求生成大量的设计选项，包括不同的设计元素、布局、配色方案和其他常用元素。Components.ai便是这样一个工具，而且在此基础之上它还可以帮助设计师生成所对应的前端代码，让设计师更好地与前端工程师进行沟通互动。

### 三、智能文档注释

所谓文档注释，就是为整个代码文件的使用准备说明文档，并对每一段代码都进行容易理解的功能说明。文档注释对于协作式代码开发具有诸多意义：对于企业内部开发人员来说，文档注释可以让理解和使用现有代码变得更加容易，有利于代码的协作迭代；对于开源项目贡献者来说，文档注释对于他们理解项目如何工作以及可以在哪里做出最好的贡献是必不可少的；对于用户来说，文档注释可以提供使用教程以及任何潜在的限制或已知问题。

虽然文档注释有助于提高代码的清晰度、可理解性和可维护性，但是创建和维护它也需要大量的时间和精力。以Mintlify为代表的基于AIGC的工具则可以自动编写和更新每段代码的详细描述，大大减少文档创建和维护的成本。有了AIGC工具，开发人员只需提供必要的输入数据，如代码本身和示例数据等，人工智能系统就能生成这段代码详细又准确的描述。这样可以为开发人员节省大量的时间和精力，使他们能够更专注于代码创作。

此外，AIGC工具还可以提高文档注释的质量。传统的人工撰写文档和注释的方式存在多种问题，首先就是文档注释撰写不规范的问题。即便公司制定了统一的规范，但在多人协作、多个版本迭代的情况下很难保证所有注释都是符合规范的，这就可能会产生不完整或不准确的描述，特别是对于复杂或非一手创作的代码。此外，开发人员意愿不高也是造成文档注释质量不高的原因之一。通常，这些开发人员迫于程序版本上线的时间压力，会将精力专注于功能的实现上，只要保证自己看得懂文档注释就行。而待版本上线后，开发人员又会投入新的工作中，无暇顾及原有文档注释的维护。这时，如果出现程序模块的交接，新的开发人员在无法清晰理解原有代码逻辑的情况下，写出来的文档注释可能会出现更多问题，这也是为什么许多大型互联网企业会有大量代码的历史遗留问题，有时甚至需要重构系统。网民曾将这种现象戏称为“程序员最讨厌的四件事”。

- 给自己的程序写文档。
- 给自己的代码写注释。
- 看别人没有写清文档的程序。
- 看别人没有写清注释的代码。

而AIGC工具的出现大大减少了这些问题，它可以生成全面、准确和规范的代码描述，为开发人员理解和处理复杂的代码和程序提供巨大帮助。此外，AIGC在文档注释上还具有一个特别的优势，它能够自动根据代码和程序的发展做出调整。随着代码库和编程语言的变化，人工撰写的代码文档可能很快就会出现过时和不准确的问题，但AIGC工具可以实时自动更新和维护代码文档，确保开发人员始终能够访问最新和最准确的信息。

## 四、智能测试纠正

在产品研发的过程中，程序员的大量时间和精力往往并非花费在代码创作上，而是花费在代码测试和错误纠正的过程中。2021年5月，微软研究员在NeurIPS大会上发表了研究成果和最新的机器学习模型BugLab，探索了如何使用人工智能完成自动化这一流程。除了BugLab外，目前市场上众多围绕AIGC进行测试纠正的产品和正在被探索的应用场景，也主要集中在代码自动测试和代码错误自动修复这两个场景中。

许多大型科技公司通常会配备人数众多的专业测试团队，测试工程师通常会编写测试代码用例或手工操作流程用例，用于验证开发人员写出的代码是否正常工作。但创建这些测试代码同创建程序本身的代码一样耗时且容易出错，特别是对于大型和复杂的代码库而言。而AIGC可以根据一组规则自动生成大量的测试用例，去检验在每种情况下开发人员写出的代码是否都能正确工作，这使得识别代码中的潜在问题并予以纠正变得更加容易。例如，为了测试一段数字大小排序的代码，开发人员可以编写一套测试用例，其中包括一组已经排列的数字、一组按反向排列的数字，以及一组按随机顺序排列的数字，然后输入程序中检查代码是否正常运作。编写这些测试代码或手工操作这些测试用例可能是冗长和容易出错的，但是有了AIGC工具，这个过程就可以被自动化，节省大量时间并减少出错的可能性，从而更容易确保代码的正常工作。目前，市面上已经出现了以Tricentis为代表的众多AI自动测试工具。

对于代码错误自动修复的场景来说，很多情况下即使代码的错误可以被发现，定位源头并且修改错误代码有时甚至要花费数小时到数日甚至更长的时间，这也是AIGC可以介入并且提供帮助的地方。例如，在Visual Studio上就曾有人发布了一款基于ChatGPT的自动测试和纠错插件，并迅速成为平台上最热门的插件之一。它可以像对话一样帮助开发者指出代码中的错误，展示正确的代码案例并且指导如何修改。

## 第二节 AIGC与市场营销

市场营销对于所有企业都极其重要。市场营销包括向潜在客户推销产品和服务，帮助企业增加销售额和收入，并且建立强大的品牌和良好的声誉。在很多行业中，市场营销甚至是一个企业成败的最关键之处。而AIGC工具也将成为企业在市场营销过程中必不可少的元素。人工智能不仅可以帮助营销人员创建更有效的营销材料，还能更好地了解客户行为，提供更个性化的销售体验，并改善客户成功和售后服务。这些特性最终都可以提高客户满意度和忠诚度，并推动企业销售额和收入的增长。

### 一、智能创意营销

使用AI生成创意营销内容并非市场中的新趋势。事实上，早在2015年淘宝“双十一”促销活动后，阿里巴巴团队就在探索基于算法和大数据，为用户做大规模的、个性化的商品推荐，也被称为“千人千面”，并且开发出了一款叫作“鲁班”的产品，这算是广义上早期AIGC在创意营销方面的尝试。鲁班在2017年就能在一天内制作4000万张根据商品图像特征专门设计的海报，并在2018年时就累计设计了超过10亿次海报。<sup>[2]</sup>

以鲁班为例，使用AIGC创建营销材料的关键优势之一是它能够节省时间和资源。AIGC可以根据一组预先定义的规则和参数自动生成这些材料，无须花费数小时甚至数天时间来创建创意营销素材。鲁班每秒钟所创作的8 000张图片甚至超过很多设计师整个职业生涯可以创作的内容。<sup>[3]</sup>这不仅为其他任务腾出了时间和资源，减少了成本支出，还确保了营销材料的时效性。

以鲁班为代表的AIGC创作工具还有另外一个优点：能够分析大量的数据，从而生成与目标受众更相关、更吸引人的内容。AIGC系统可以分析产品目标的兴趣、偏好和行为，并利用这些信息创建符合他们特定需求和兴趣的营销材料。这能够更有效地推动营销活动，更有可能引起目标受众的共鸣，从而达到阿里巴巴在“双十一”促销活动时所推崇的“千人千面”的效果。

除了图片领域，创意营销文本的撰写也是AIGC工具的重要应用之一，它可以在给定的主题上生成几乎无限多的变化文案，这使得营销人员可以尝试不同的风格和方法，并快速测试和迭代不同的想法。这还意味着，营销文案可以针对不同的受众和渠道进行调整，使其更容易吸引不同平台上的目标受众。海外营销工具Copy.ai就帮助了大量市场人员创作不同场景下的推广文字内容。当然，除了上述介绍的两种创意营销形式，AIGC还可以生成其他各种模态的营销材料，例如产品的3D模型和广告视频等。

最后，因为市场动向、用户偏好等信息都是不断变化的，使用AIGC工具生成营销内容的另一大优势是帮助营销人员迅速适应不断变化的消费趋势和偏好，从而保持领先地位。由于能够分析大量数据，AIGC能够快速、有效地识别和响应消费者行为和偏好的变化。这可以使营销人员迅速调整他们的策略，以应对不断变化的情况，确保他们的营销努力始终与最新的消费者趋势和偏好保持一致。

## 二、智能销售流程

除了通过创意营销的内容获得更多曝光和客户流量外，对于许多企业来说，主动的对外销售也是极为重要的一个环节。不同行业的不同体量企业和不同的产品类型都会有相对不同的销售流程，但是整体而言，对外销售大概分为三个部分：线索发现、客户触达、客户转化。

在当今的商业世界中，企业经常花费大量时间在互联网上搜寻潜在客户，并且建立希望接触联系的客户名单，这个过程就叫作销售的线索发现。除了手动在互联网上搜索之外，企业也常常会通过参加行业大会或者使用互联网爬虫抓取数据的方式获取潜在客户名单，但即便如此，他们产生的线索也往往是高成本且低质量的。而与之相对比，AIGC工具可以通过分析现有的客户人口统计数据、购买习惯等，和线上的企业数据库进行对比，从而快速而低成本地建立一个更适合企业的潜在类似客户名单。AIGC工具帮助建立潜在客户名单的另外一种方式是通过使用自然语言处理算法分析大量的文本数据，如博客、新闻和社交媒体文章等，以判断不同客户对于企业所提供的服务或者产品的需求强度，从而进行线索发现。Seamless.ai便为众多企业提供了这样的服务，通过简单描述客户的特征，例如行业、体量、收入规模、地区等信息，它便可以建立一个销售名单。

AIGC也可以生成电子邮件和社交媒体信息，通过智能呼叫的方式帮助企业进一步提高客户触达的效率和效果，极大程度提高了售前销售团队的生产效率。正如前文使用AIGC工具创作市场宣传材料一样，同样的工具也可以帮助企业创作邮件内容、微信消息、短信等，甚至进一步根据每一个客户的信息定制不同的内容。在文字的基础之上，自然语言处理则可以帮助企业建立AI智能外呼系统，让人工智能主动对外拨打电话联系到更多的潜在客户，大幅度降低企业的成本。总部位于南京的云蝠智能（Telrobot）便是一个被很多企业使用的AI智能外呼系统，帮助企业打通更高效的销售流程。

AIGC工具还可以通过定制化生成客户解决方案，以及建立和优化销售话术等方式提高客户转化率，帮助企业提高销售额。通过学习企业产品或服务内容，以及大量的往期方案，AIGC工具可以处理所输入的客户需求及参数等，并定制化生成和客户最相关且最有可能提高转化率的解决方案，更快地响应潜在客户需求。同时，以Oliv.ai为代表的工具可以通过学

习大量的企业销售视频、录音以及文字稿，分析销售话术中的优缺点，进而不断帮助企业优化和完善销售话术，提高转化率。

### 三、智能客户服务

企业的销售工作并不停止在客户签单甚至付款的那一刻，下一个同样重要的阶段便是确保客户可以获得期待的服务和产品，帮助客户达成购买目标。然而，不论是客户服务还是客户成功，企业都需要使用大量的人力在这个阶段继续帮助客户，否则将会面临低复购、低续费甚至降低企业的声誉等问题。

AIGC同样可以在这个阶段进一步帮助企业服务好客户。使用AIGC工具进行客户支持的一个关键好处是能够有效地处理大量的请求和咨询。传统的客户支持方法需要一个客服团队来响应所有客户的询问，不仅如此，对于重量级的大型客户还需要配备一个客户经理，这样无疑是极其耗时和昂贵的。有了AIGC工具，系统本身就能快速和准确地答复客户的问题。这不仅为企业节省了时间和金钱，而且还通过更及时地帮助客户获得他们需要的答案来改善客户体验。此外，使用AIGC提供客户支持的另一个优势是能够为每个客户提供个性化的响应。与预先写好的回复不同，AIGC能够根据每个客户的具体需求生成独特和个性化的回复。这种水平的个性化服务有助于建立信任和提高客户满意度。

除了提供客户支持，AIGC工具也可以用来帮助客户成功环节，包括提供个性化的产品推荐、提供个性化的支持来帮助客户实现他们的目标等。例如，AIGC工具在分析客户使用数据的基础上，可以找到改善产品体验的机会，然后，该工具可以生成个性化的电子邮件或应用程序内的消息，并就客户如何更有效地使用产品提出建议。这不仅有助于改善客户的体验，而且还可以增加客户保留和忠诚的可能性。

很多科技巨头和创业公司都在这个方向展开了探索。最值得一提的是销售科技巨头Salesforce，其旗下爱因斯坦AI可以自动生成众多内容并推荐给客户服务工作人员作为回答话术，它甚至可以提前预测正在咨询的客户的需求。

### 第三节 AIGC与管理协作

组织内部的有效管理和协作可以保证所有团队成员都站在同一战线上，朝着相同的目标努力。它还可以促进知识的共享，创造更好的决策和解决问题的方法。此外，有效的沟通和协作有助于营造积极的工作环境，提高员工满意度，降低离职率和提高留用率。而企业内部缺乏沟通和协作可能会产生许多负面后果，例如误解和冲突、浪费时间和资源、降低士气等，甚至还会错过重要时间点或产出低于标准的工作。在当今快节奏的商业环境中，为了保持竞争力，取得成功，组织必须进行有效的管理、沟通和协作。AIGC有很多可以帮助企业提高管理效率的应用场景，本节将对智能行政助理、智能内部沟通、智能团队协作、智能人力资源管理四个场景进行重点介绍。

#### 一、智能行政助理

通过自动化处理行政任务，比如安排会议、创建报告、管理电子邮件等，AIGC可以帮助企业节省时间和资源，提高内部流程的效率和准确性。

安排会议是一个听起来极其简单但是执行上十分复杂的工作，尤其是在参会人数较多时整个流程会变得异常烦琐。但是，AIGC可以通过分析来自电子邮件和日历邀请的数据，了解不同团队成员的空闲时间和会议时间偏好，并利用这些信息自动生成一个会议时间表，通过自然语言和每一个与会成员确认其是否可以到场，以便最大限度地提高出席率和工作效率。2021年夏天被Bizzabo收购的B轮创业公司X.ai便在开发这样的产品，它可以让AI成为每一个团队成员的会议助理。

除了安排会议，AIGC也可以通过自动创建报告辅助企业进行内部管理。AIGC工具可以分析来自不同来源的数据，比如销售数据、客户反馈和财务报告，使用这些信息自动生成详细和信息丰富的报告。这些报告可根据不同利益方的具体需要和偏好进行调整，并可在获得新数据时实时更新。这可以帮助企业根据最新的信息做出更好、更明智的决策，还可以通过自动化报告创建过程来节省时间和资源。

#### 二、智能内部沟通

AIGC通过自动化邮件回复、总结会议和文件重点、跨语言和专业自动翻译等方式，可以显著提高企业内部沟通的效率，进而提高协作效率和企业生产力。

人工智能可以通过学习历史文档和往期邮件，自动化生成针对性的电子邮件回复所收到的常见咨询或请求，训练识别和标记潜在的重要电子邮件或附件，从而确保重要信息不被遗

漏。在电子邮件管理中，AIGC可以帮助企业简化流程并提高效率。科技巨头谷歌就将AI辅助回复功能添加到了其邮箱系统Gmail当中，帮助用户更好地提高工作效率。

AIGC协助内部沟通的另一种方式是总结会议和文件中的要点。许多会议和文件包含大量信息，员工可能很难快速确定最重要的信息并采取行动。通过AIGC，企业可以自动化总结过程，让员工快速理解信息要点。例如，员工如果参加了会议，他们可以向人工智能提供会议记录，人工智能将生成一个摘要，突出显示最重要的信息。这可以为员工节省大量时间，并确保他们能够理解信息要点。目前，国内使用最广泛的该类软件是字节跳动旗下的飞书妙记，它可以自动在线生成会议纪要，通过智能语音识别转化成文字，把会议交流沉淀为要点文档，从而让会议成员更专注，工作更高效。

此外，AIGC与艺术和组织心理学的结合，可以帮助团队内部更好地建立信任、使员工更深度理解企业的愿景和价值观。位于法国巴黎的Viva la Vida公司，基于过去5年在全球近百家企业和国际组织主导艺术工作坊的经验，开发出一套基于员工价值生命周期每个节点并结合艺术AIGC、组织心理学、大数据的SaaS系统，旨在通过艺术建立连接，提升员工积极性和心理健康状态，未来也将涉足C端市场。

最后，AIGC工具可以通过将信息翻译成不同的语言来协助内部交流，因为它可以促进沟通，确保每个人都能达成共识。这一点对于跨国企业来说尤其有用。例如，如果员工用英语写一封电子邮件，人工智能可以自动将其翻译成收件人的语言，使收件人能够轻松地理解邮件。这可以节省员工手动翻译消息的时间和精力，并有助于改善组织内部的沟通和协作。事实上，这个功能早就在近乎所有主流的即时通信/协作软件上获得了应用。目前，在国内获得广泛应用的典型案例是字节跳动旗下的飞书妙记，其群聊消息和文档可以支持113种源语言、17种目标语言的翻译。<sup>[4]</sup>

### 三、智能团队协作

由于存在不同的知识技能、人员配置、工作习惯等，同一个公司的不同部门或团队间的协作效率也可以进一步得到提升，而AIGC可以被用来改善团队间协作的现状。

AIGC工具可以帮助企业整理各种类型的相关文件。在企业的各种项目中，常常会有不同格式的文档（Excel电子表格、PDF文档、PowerPoint演示文稿等），它们可能被存储在不同的平台上（云盘、线上文档、电子邮件等）。通过AIGC工具，公司可以训练一个模型来自动地将这些数据组织成相关的类别，例如按部门、项目或主题分类。这将使员工更容易找到他们需要的信息，减少搜索所需的时间和精力，也减少跨部门协作时获得信息的阻力。

此外，AIGC也可以通过创建和维护跨团队项目协作计划来改善团队协作。通过AIGC可以自动生成特定项目的项目方案，包括工作流和任务分配计划。这在流程复杂、人员数目庞大的项目中特别有用，减少了项目经理的烦琐工作。例如，假设一家公司正在开发一种新产品，利用AIGC工具便可以自动生成项目的详细协作计划，包括每个团队或个人要执行的具体任务、每个任务的最后期限以及任务之间的任何依赖关系。这将使员工更容易理解他们的角色和职责，并确保项目保持在正确的轨道上。

位于加利福尼亚州的Mem公司便在开发这样的自我管理的协作空间，通过AI帮助更多团队管理文件、流程和分工，从而提高团队协作的效率。Mem公司的产品也整合了大量前文提到的改善团队内部行政和沟通的功能。

## 四、智能人力资源管理

除了前面介绍的这些，AIGC还可以在筛选招聘人才、自动化人事管理流程以及评估员工工作表现等方面提高公司人力资源管理的效率和效果。通过分析大量的数据，包括线上申请材料、简历和社交媒体档案，AIGC算法可以快速而准确地识别具有特定职位所需技能和经验的个人。人力资源经理不再需要手工审查和评估每个申请者，从而节省大量的时间和精力去关注头部人才的审核和筛选。

此外，AIGC算法可以用来自动化处理许多烦琐和耗时的人力资源任务。例如，AIGC算法可以用来自动安排面试、发送合同，甚至处理新员工的入职和入职培训。这有助于简化人力资源流程，并确保有效率和有效力地完成这些流程。

最后，AIGC工具在绩效管理方面也发挥了重要的作用。AIGC工具可以根据每个员工的个人优势、弱点和目标来生成更具体、更有针对性的绩效反馈。这可以帮助员工更好地了解自己的业绩，并确定需要改进的领域，从而产生更好的结果，提高参与度和生产力水平。此外，AIGC工具还可以帮助企业实现绩效评估过程的自动化，例如安排和跟踪员工评审，使人力资源经理和管理人员能够专注于更重要的任务。AI驱动型团队绩效管理工具Onloop就是这个领域的典型应用。

[1] 参考自<https://www.theverge.com/2021/7/7/22561180/github-copilot-legal-copyright-fairuse-public-code>。

[2] 参考自<https://www.leiphone.com/category/yanxishe/Zy9OqVkBnmZfuEz.html>。

[3] 参考自<https://juejin.cn/post/6844903509825945608>。

[4] 参考自<https://www.feishu.cn/hc/zh-CN/articles/446972352418>。



## 第四章 AIGC的行业应用

AIGC如何在各行各业的实践中加以应用？

深入每个行业，你会发现人工智能正在改变工作的性质。

——丹妮拉·鲁斯（Daniela Rus）

深入各个行业前沿，仔细观察，智能创造时代已非乌托邦式的幻想，而是呼啸而来的未来。目前，AIGC的身影已经在多个垂直领域中活跃，贯穿资讯、影视、电商、教育等多个行业。了解这些行业的应用现状，也就能够更好地了解各个行业的未来。本章将对AIGC在各个垂直行业的应用进行详细介绍。

## 第一节 AIGC资讯行业应用

信息爆炸时代，各类新闻资讯无处不在，不可或缺。同时，这些资讯也具备标准化程度高、需求量大、时效性强等特点，因此是人工智能施展拳脚的理想舞台。自2014年起，大规模数据检索处理、结构化文本写作、摘要生成等多项AIGC相关应用已经在新闻资讯行业落地，因此资讯行业是AIGC商业化相对成熟的领域。同时，人工智能在该领域也正向全链路延伸，伴随着底层大模型和各细分场景应用的进步，资讯行业将会有更大的变革潜力。

### 一、AIGC辅助信息搜集，打造坚实内容地基

优质的新闻产出必然建立在全面、高效、准确的信息收集和整理的基础之上。在传统的生产模式中，从业者需要亲临一线，通过观察、询问、记录才能获得扎实的信息基础，而AI已经能够对该环节进行高效赋能。例如，在采访过程中，科大讯飞的AI转写工具可以帮助记者实时生成文字稿、自动撰写摘要、调整文风、精简文本等，提高工作的整体节奏，保障最终产出的时效性。

但AI的可能性并不止于助力人类工作者获取一手信息，也可以帮助新闻工作者更精确地检索二手信息，收集撰写新闻报道所需的素材。在以往以传统方式利用搜索引擎的过程中，如果想要实现一些边缘话题的精确检索，需要对检索词的组合进行深思熟虑或反复尝试，才能通过搜索引擎找到想要的答案。运用自然语言长段话描述问题，并不会有助于检索结果的呈现，反而会让结果更加偏离问题。如果想要进行更加精确的检索，则需要学习复杂的检索表达式，这无疑增加了新闻工作者的学习成本。但在高性能的AIGC工具出现之后，人们就可以用日常向好友提问一样的方式向ChatGPT这样的对话类AIGC工具提问，直接获得精确的答案，甚至都不需要在检索出的结果中搜寻，非常方便。虽然AIGC工具对于少量领域的回答可能会出现时效性有限或一些错误的结果，但它在大量领域已经可以作为二手信息搜集的重要工具投入使用。

### 二、AIGC支持资讯生成，实现便捷高效产出

在新闻资讯的生产环节，基于自然语言生成和自然语言处理技术，AI交出的结构化写作“答卷”已经逐步得到从业者和内容消费者的认可，因此已经涌现了一批成熟玩家。在产出数量方面，以美联社、雅虎等媒体的合作伙伴Automtd Insights公司为例，其撰稿工具Wordsmith能够在1秒内产出2 000条新闻，单条质量能够比拟人类记者30分钟内完成的作品。AI的强悍生产能力使得低成本覆盖长尾市场成为可能，更多内容消费者的需求得到满足，公司的利润来源也得以拓展。除了惊人的产出速度外，AI还在内容准确度方面具有明显优势，能够避免人类工作者粗心导致的拼写、计算错误，在提升稿件质量的同时减轻写作者

的工作量。整体而言，在AI内容生成方向上，国内玩家布局颇多，比如新华社自主研发的写稿机器人“快笔小新”、腾讯公司开发的Dream Writer均已在标准化程度高的场景中得到大量运用。此外，百度公司也和人民网携手发布了“人民网-百度·文心”大模型，或将在未来成为媒体行业的底层基础设施，赋能媒体生产的多场景、多环节。

### 三、AIGC助力内容分发，“智媒”赋能人类工作者

在资讯内容的分发环节，AI除了助力个性化内容推荐外，也开拓了全新的应用场景，即驱动虚拟人主播，以视频或直播的形式进行内容发放，打造沉浸式体验，比如新华社数字记者“小净”带来新鲜的太空资讯，央视网虚拟主播“小C”担任记者角色，阿里巴巴冬奥宣推官数字人“冬冬”畅聊冰雪，百度智能云AI手语主播为听障朋友带来贴心服务。热潮之下，各路玩家纷纷跟进，期待创造更富有科技感、更多样的资讯消费体验。在未来，AI虚拟主播很可能成为媒体的标准配置。

AI在资讯行业的全环节大显身手，可能会引人担忧：媒体工作者是否要被取代？时而爆出诸如微软在2020年裁撤27人的新闻网站编辑团队并用AI替代一类的消息，<sup>[1]</sup>似乎也表明人们所虑非虚。但事实上，AI真正的潜力在于赋能人类工作者，其应用在当前的技术水平下仍有不少限制。

首先，AI撰写的文稿仍稍显呆板单调，模板化强，无法像人类记者一样根据具体报道的性质和语境调整叙述的策略，以达到更好的传播效果。同时，AI当前无法撰写深度报道，文字缺少温度和人文关怀等要素。基于这些原因，AI写稿最初也是多被用于财经、体育、突发事件等垂直场景，跨领域迁移、适配以及产出的能力仍然不足。并且，过度依赖AI进行信息抓取以及撰稿也可能导致信息茧房和回声室效应加剧，甚至带来伦理失范的问题。

因此，人类工作者仍有巨大的发挥空间，而AI则将人类工作者从繁杂的重复劳动中解放出来，使他们更好地发挥批判思考能力和创造能力，产出更优质的内容。前美联社的人工智能联合主管弗朗西斯科·马可尼（Francesco Marconi）也在其著作《新闻的新模式》中表示：“人工智能仅仅是记者的另一样工具，能让从业者腾出时间进行深度思考。”综上所述，尽管AI在新闻资讯行业中已经得到广泛应用，但结局并非实现行业的“去人工化”，而是AI和人类工作者携手，共同促进“传媒”向“智媒”的全面升级。

## 第二节 AIGC影视行业应用

人们在逐步走向虚拟世界的大潮中，对影视内容的需求正呈爆发式增长。为了满足消费者愈发刁钻的口味、挑剔的眼光，影视行业开足马力扩大产量的同时，也不断迭代制作技术，导致影视行业呈现出工业的特征，正变得空前的精细、复杂。在这个庞大的系统中，“人的局限”逐步凸显。而AI在影视行业的应用，却能让影视制作重回“纯粹”，让影视人专注于讲好一个故事。

### 一、AIGC协助剧本写作，释放无限创意

一场畅快淋漓的大雨离不开空气中凝结水滴的微粒，辅助生成剧本的AI就犹如空气中的微粒，能够在创作的云海中播下灵感之雨的种子。结合对大量优质案例的学习，以及对受众心理的洞察，AI能够根据影视工作者的要求快速生成不同风格、架构的剧本。AI在极大地提高影视工作者工作效率的同时，也在进一步激发他们的创意，帮助他们打磨出更加优质的作品。

将AI引入剧本创作的尝试早已有之。在2016年的美国，一款由纽约大学研发的AI就在学习了几十部科幻电影剧本的基础上，成功写出了电影剧本《阳春》以及一段配乐歌词。虽然最终成片只有短短8分钟，内容也稍显稚嫩，但《阳春》在视频网站上收获的数百万播放量足以证明人们对这次先驱试验的兴趣。而在2020年GPT-3发布后，查普曼大学的学生也用GPT-3创作了一个短剧，其剧情在结尾处的突然反转令人印象深刻，再度引发了广泛关注。

人们可以通过这些牛刀小试窥见AI在剧本创作领域的潜力，但若系统性地解放影视创作者的生产力，还需要AI公司贴合具体应用场景，对模型做高度针对性的训练，并结合实际业务中的需求进行定制功能开发。在海外，有些影视工作室已经在使用诸如Final Write、Logline等更加垂直的工具，而在国内，深耕中文剧本、小说、IP生成的海马轻帆公司已经收获了超过百万用户。

在剧本写作上，海马轻帆的AI训练集已经涵盖了超过50万个剧本，结合行业资深专家的经验，能够快速为创作者生成多种风格、题材的内容。而剧本完成后，海马轻帆也拥有强大的分析能力，可以从剧情、场次、人设三大方向，共300多个维度入手，全方面解析和评估作品的质量，并以可视化的方式进行呈现，为剧本的改进迭代提供参考。而在剧本写作之外的作品商业价值测算、角色筛选等相关领域，AI也能继续发挥作用，最终实现全面赋能剧本创作，让影视人得以专注地讲好故事，落地创意。自2019年开放合作至2022年末，海马轻帆已累计评测剧集50 000余集、电影/网络电影20 000余部、网络小说超过800万部，其中不乏

《流浪地球》《你好，李焕英》等大热作品。<sup>[2]</sup>

## 二、AIGC推动创意落地，破除表达桎梏

在剧本写作阶段，AI已经能够帮助影视人更好地释放创意，但从剧本上的文字到最终呈现给观众的视听盛宴，仍有一段漫长的旅程，而AI却能在这个想法落地的过程中继续保驾护航，帮助实现从“好创意”到“好表达”的跨越，帮助影视工作者化“不可能”为“可能”。

影片具体制作中的第一重“不可能”指的是当前较为粗放、劳动密集型的生产方式难以满足观众对内容质量不断提高的要求。2009年的电影《阿凡达》让全世界获得了特效和3D电影的启蒙，自此之后，观众便不断追求着更加震撼、精细和沉浸式的影视体验。

为满足市场的需求，特效技术的应用呈井喷之势，但素材整理标注、渲染、图像处理等环节的工作越发繁杂，给影视工作者造成沉重的负担。例如电影《刺杀小说家》凭借惊艳的特效收获大量关注，但相应地，其后期制作、渲染时间和复杂度也呈几何倍上升。在这样的大背景下，传统的制作协同流程已经难以为继，依靠加班、外包等“老方法”堆砌生产力既不经济，也不现实，生产模式亟待升级。而AI技术就具有变革影视行业的潜力。

AI能够帮助影视工作者从大量重复琐碎的工作中解放出来，从而提高效率，专注于创意的表达。例如《刺杀小说家》背后的特效团队墨境天合，正是借助了云渲染和AI技术才完成交付。再以动漫制作为例，动漫制作的环节高度数字化，因此为AI在其生产全流程提供了充足的赋能空间。动画电影的设定天马行空，故事行云流水，但若想将其搬上荧幕，却需跨越万水千山。工作室需要从建模开始，一草一木，一人一马，打造世界的雏形，再通过骨架绑定和动作设计，让模型“活”起来。之后，工作室还需要定分镜、调灯光、铺轨道、取镜头，把故事讲出来。为了呈现美轮美奂的场景、逼真的材质和细节、震撼人心的特效，工作室还需要做大量的解算和渲染工作。

在生产的一个环节中，动漫工作者都面临着许多重复性工作和等待，降低了生产的效率，而优酷推出的“妙叹”工具箱，就能够凭借AI在上述的全流程实现生产力的解放。以动漫的重中之重——渲染为例，过去主流的解决方案是离线渲染，在长时间的渲染完成前无法直接看到结果，以至于动漫工作者不得不常常停下来等待，甚至是完全返工。而妙叹则能实现实时渲染，帮助从业者实时把握产出效果，并有针对性地进行修改，节省了大量时间和精力。在建模、剪辑、素材管理等重复性工作较多的环节中，妙叹则能够将其整合，利用AI实现“一键解决”，或者提供预设模板、素材，进一步降低动漫从业者的负担。

当前，妙叹等AI赋能工具已经被许多国漫工作室采用，包括无断档连更54集的《冰火魔厨》，AI的生产力可见一斑。据《冰火魔厨》制作方表示：“妙叹是解放创意的工具，有了它，我们便有了更充分的精力和底气，去思考如何把故事讲好。”

通过赋能影视生产的全流程，AI为影视行业带来的影响可能足以比拟20世纪好莱坞掀起的影视工业化革命，其本质是通过精细的环节拆分和管理，模块化地产出，实现降本增效，最终在全新生产方式的加持下，好莱坞电影席卷全球。这样的变革对工业化模式尚未完全成熟的中国影视行业，尤其是承载万众期待却仍相对稚嫩的国漫行业来说，有着特殊的意义。

将想法转化为影视作品进行表达，也面临第二重“不可能”，即想象中的场景难以在现实中进行呈现和拍摄。当前，影视作品需要的拍摄环境越发丰富、复杂，自然风光、千年古镇、未来都市，甚至是奇幻大陆、幻想世界，这些场景携带着人们的想象自由跨越时空。随着创意的不断延伸，以影视城为代表的实景搭建、群演拍摄越发难以满足各类题材影片的制作需求。因此，虚拟影视环境的制作变得越发关键，而AI在这个环节具有天然的优势，可以充分赋能影视工作者的创造力，并有效控制影片制作的成本。

在场景上，AI辅助生成背景，结合绿幕拍摄的制作模式已经得到广泛应用，而在大型场面上，AI亦有亮眼表现，比如《指环王：护戒使者》中的万人战争场面，就是由一套虚拟环境群体模拟系统制作完成。AI的加持让影视工作者能够以震撼的视觉效果叙述自己的故事，让自己和人们的想象照进现实，映上屏幕。

帮助影视制作方和演艺人员挑战“自然规律”，是AI在影视行业中实现的第三重“不可能”。时光无法逆转，逝者不可复生，天生的容貌也难以更改，种种客观条件在一定程度上制约了影视工作者的发挥。但在AI技术的加持下，影视工作者的创作自由度进一步得到了解放。

“返老还童”也许是人类在现实中终难实现的夙愿，但许多影片中的演员却早早进行了体验。2008年的电影《返老还童》讲述了一个“逆生长”男人一生的故事：他出生时便是80岁的容貌，随着岁月流逝却逐渐变得年轻，最终以婴儿的形象离世。片中的男主角之所以能够以横贯80年时光的多种样貌出演，就是借助了AI面部生成技术。最终，出色的技术运用也帮助该片斩获了2009年奥斯卡最佳视觉效果奖，AI技术成就了该影片精彩的构想。

荣誉留在过去，彰显着业界对锐意创新的奖赏，而后人也并未驻足于此，而是继续利用AI帮助更多演员在时间的长河中自由穿梭，不断拓展其应用场景。比如网飞在2019年出品的电影《爱尔兰人》由三位耄耋之年的影帝领衔出演，而正是AI减龄技术的大量运用，才能让平均年龄超过77岁的“教父”们以更年轻的形象在片中重聚。该影片背后的团队耗费两年的时间搭建了一款叫作Face Finder的AI人脸识别应用，通过收集大量主演在30岁至55岁的容貌数据来实现最优的减龄表现，其最终的出品效果即便在4K分辨率下仍显得自然生动。此类AI技术的应用帮助影视制作者留住了经典演员超越时间的魅力，也在另一方面延长了演员自身的演艺生命。

值得注意的是，现实中的不老泉无价，而AI减龄技术同样耗费不菲。电影《爱尔兰人》在该方面的特效支出就高达1.59亿美元。幸运的是，AI技术也在以令人咋舌的速度飞速迭代。例如，迪士尼在2022年开发出一套名为FRAN（Face Re-aging Network）的AI系统，专攻面部年龄重构。据介绍，FRAN最快仅需5秒就能处理完成一个角色，极大地消除了人工手动调整的负担。在迪士尼看来，FRAN是第一个“实用、全自动、可用于影视制作”的图像人脸重塑方案。随着此类技术的不断进步，使用成本和门槛将持续降低，可以预见未来AI年龄调整将在影视行业获得更加广泛的应用。

AI不仅能让“时光倒流”，甚至还可让逝者“复生”，在虚拟的彼岸再度和观众见面，甚至交互。这一技术在影视领域已经得到了许多应用，其中最知名的案例之一就是《速度与激情7》。因男主角保罗·沃克在拍摄中途不幸离世，制作团队就联合维塔特效公司，从先前未使用的镜头中收集保罗的面部数据，让他最终得以在电影中“重生”。让故去者回到荧幕之上，能让观众在悲痛之余感到一丝慰藉，在一定程度上弥补了生死两隔的遗憾，而对于逝者本身，可能更是一种致敬和缅怀。正如《速度与激情7》的导演温子仁及其团队所说：“为了保罗。”

更进一步来说，在影视领域之外，这样的AI技术也可能在未来走入普通人的生活，帮助人们找回不幸失去的亲人，寄托自己的情感，实现另一种团聚。

然而，死者形象复生这一愿景的出发点虽然美好，但也不可避免地带来了伦理相关的挑战。逝者本人生前是否接受使用技术手段让形象重活一次？最终呈现出来的形象，以及更关键的言行，是否足够客观、准确？商家有无权力利用逝者的信息和形象牟利？人们是否愿意接受这般形式的互动？这样的AIGC形式还值得人们更多的思考。离世几十年的明星出演电影，登台献唱；已故商业领袖做客播客，对答如流；已经逝去的长辈、伴侣、孩子又重新出现，再度与我们交谈……这类尝试也往往伴随着争议和批评，甚至有一位借助GPT-3“复活”已故伴侣的美国用户直接被Open AI收回了使用权。可以想见，伦理、风控和安全等话题将会持续伴随人们在该领域的探索。

无论是改变年龄，还是复原逝者，均属于“AI换脸”技术的应用，而该技术最广为人知的应用，当属替换已经制作好的影片中“塌房”明星的脸。明星“塌房”防不胜防，而由于影片前期投入大，拍摄时间长，且题材一般具有周期性，故出品方很难接受因为“塌房”明星的影响全盘推翻重拍，在这种情况下，AI换脸就是效率最高、成本最低的一种补救措施。但是，需要了解的是，没有AI公司会将给“塌房”明星换脸作为主要业务。因此，如果仅仅将目光停留在该领域，就无法全面理解AI在影视行业的革命性潜力。

### 第三节 AIGC电商行业应用

随着互联网信息时代的大爆发，淘宝、京东等巨头互联网企业进入互联网时代，电商作为互联网时代的一大受益者，也扮演了重要的角色。2004—2013年，电商行业发展迅速，而随着自媒体的出现，线上直播、网红带货等模式陆续出现，稳固了电商行业全新的发展格局。在新冠肺炎疫情暴发的几年间，国内大小企业先后面临转型，从原有的靠线下发放广告和实体店的经营模式引入线上化和平台化的经营模式。在数字世界和物理世界快速融合的时代，AIGC正走在内容和科技前端，为行业带来了深刻的变革。AIGC可以赋能电商行业的多个领域，例如商品三维模型、电商广告应用、虚拟人主播以及虚拟货场的构建。我们可以利用AIGC结合AR、VR等技术，实现视听等多感官交互的沉浸式体验。

#### 一、AIGC构建三维商品，改善用户购物体验

相较于线下购物，线上购物的一个典型问题就是只能从图片上获取商品单一角度的观察信息，无法全方位预览商品全貌。而AIGC相关技术工具的出现有利于解决这一问题，这些工具可以让商品在不同角度下拍摄的图像通过视觉算法生成商品的三维模型，提供虚拟产品多方位视觉感知的独特体验，大幅压缩沟通的时间成本，同时改善用户的购物体验等，提升用户转化。

除了将图像生成三维模型，AIGC相关技术还有一些更加进阶的电商应用方式。阿里巴巴的每平每屋业务就利用AI视频建模等AIGC技术，实现了线上“商品放我家”的模拟展示效果。家居购物的一个痛点在于，用户非常容易在线上买到看起来好看，但是与整体家居风格并不匹配的商品，从而导致较高的退货率。而阿里巴巴的每平每屋业务，将AIGC的功能植入手机淘宝和每平每屋的App之中，用户可以通过拍摄扫描家居环境，以及家里与商品进行搭配布局的家居，让AI生成线上的3D模型，并与想要购买的商品3D模型进行组合，让用户在线预览整体的组合效果。AIGC的线上试用功能无疑极大地提升了用户电商家居的购物体验。

除了家居领域之外，许多品牌企业也开始探索类似的虚拟试用服务，例如优衣库虚拟试衣、阿迪达斯虚拟试鞋、保时捷虚拟试驾，等等。然而，无论是商品的全方位预览还是虚拟产品的试用，都需要创作越来越多和商品相对应的三维模型。如果依靠人类去进行三维建模，不但成本高昂，而且也跟不上新商品的迭代上架速度。因此，要实现沉浸式的线上购物体验可持续化，不要流于单次的营销活动，离不开AIGC技术的革新与发展。

#### 二、AIGC赋能服饰电商，产品拍摄降本增效



AIGC可以为企业营销提供大量的创意素材，而电商广告是对这些创意营销素材有海量需求的领域，比如前面介绍过的阿里巴巴研发的AI设计师“鲁班”就主要应用于这个领域。

不过，除了这种通用性的广告营销用途外，AIGC在电商服饰领域还有特别的用途。电商服饰领域通常会采用“小单快返”的模式，即先小批量生产多种样式的服饰产品投入市场，快速获取市场销售反馈，对好的产品快速返单继续生产，在试出爆款的同时减小库存压力。然而，这种模式最大的问题就是产品展示图，如果面对上千个服装单类产品都分别找模特、拍照、修图，无疑会耗费大量的时间和成本。根据相关网络公开资料，不少服饰商家每年在产品模特图上耗费的成本可达20万～100万元，并且每次拍摄基本需要2～3周才能拿到成品。<sup>[3]</sup>成立于2020年的ZMO公司就运用了AIGC技术来解决这个问题，它在2022年5月获得了800万美元的A轮融资，由高瓴资本领投，GGV纪源资本和金沙江创投跟投。商家只需要在ZMO平台上传产品图和模特图，就可以得到模特身穿产品的展示图。<sup>[4]</sup>除了刚刚提到的“小单快返”的市场策略，借助AIGC还可以让更多与电商服饰有关的市场策略低成本地实现。比如，如果服装商家想在同一款服饰上测试不同的花纹，无须分别制作样品、拍摄、修图再进行投放，只需要用AI生成并处理成不同花纹样式后，直接由市场部门上架店铺预览，根据用户的浏览数据或者相关市场投票活动，决定最有潜力的爆款样式。

如果商家没有专业的模特资源，一些AI平台也可以提供虚拟人模特。阿里巴巴研发的AI模特平台塔玑就是这样的产品，商家可以在上面生成成千上万种五官组合的虚拟模特，上传手机拍摄的衣服平铺图或服装设计矢量图后，就可以生成模特身穿产品的广告图。更有意思的是，商家可以根据不同的服装风格，对模特外观进行定制，例如甜美风服装风格的商家可以选择有刘海、蓝色眼瞳的模特，以此来凸显服装的靓丽设计。

### 三、AIGC活化虚拟主播，提升直播带货效能

随着元宇宙概念的推广与发展，虚拟主播开始成为许多电商直播间的选择。相较于真人直播，虚拟主播不仅能为用户带来新奇的体验，而且可以突破时间和空间的限制，24小时不间断直播带货。2022年2月28日，京东美妆超级品类日活动开启时，京东美妆虚拟主播“小美”就出现在兰蔻、欧莱雅、OLAY、科颜氏等超过20个美妆大牌直播间，开启直播首秀。虚拟主播不仅五官形象由人工智能合成，嘴型也可以利用人工智能精确地匹配产品的介绍台词，动作灵活、流畅。在直播过程中，虚拟主播的每帧画面都由人工智能生成，手持商品的展现形式，配以真人语调的产品讲解、模拟试用，具有极佳的真实感，可以为用户提供与真人无异的体验。<sup>[5]</sup>

这类结合人工智能技术的虚拟主播不仅在用户体验方面与真人无异，而且还可以节约

30%~50%以上的成本。根据网上的调研数据，在一线城市雇用一名优秀主播的月薪大约是1万元左右，加上直播场地费，每年差不多需要15万元左右的成本，如果再加上硬件设备成本，成本可能达20万元，这对商家来说无疑是一大笔开支。<sup>[6]</sup>而如果采用AI虚拟数字人去经营直播间，不但可以自由更换妆发、服装和场景，时刻给用户全新的观感，还能最大化地节约成本。

不过，目前看来，大多数AI虚拟数字人的作用依然是与真人形成互补，让真人获得休息时间，在真人休息的时候帮助真人直播，或者为原先没有电商直播能力的商家提供直播服务，还远不能代替真人。但伴随着AIGC技术的发展，AI虚拟数字人将收获更强的交互能力，可以更加自然地和直播间的观众互动，并结合直播间的评论情况做出更真实的实时反馈，这时的虚拟数字人也许就可以在很多场域下代替真人进行工作，电商直播也会迎来一个全新的智能时代。

## 第四节 AIGC教育行业应用

伴随着技术的爆炸式发展，教育这一古老的行业也迎来了颠覆性的未来。然而，相较于在其他行业的全面渗透、多点开花，AI在教育行业的落地应用似乎也落后半步。乔布斯曾经发问：“为何IT技术几乎改变了所有行业，却在教育方面建树不多？”这个问题放之当下的AI领域，似乎也并不过时。

事实上，这是由教育行业本身的性质所致。教育行业的参与者众多，时间跨度大，个体的差异性也极大，这种种要素罗织成了一张张复杂的多维网络，让擅长解决边界清晰、定义明确问题的AI一度迷失方向。同时，教育行业十分强调人与人的互动和联结，并没有统一的理论模型，这都为AI的开发、训练和最终落地增加了难度。

然而，却没有人小觑AI为教育行业带来的革命性潜力。俞敏洪曾坦言，AI是新东方最大的竞争对手，于是他开始积极思考人工智能时代的教育。我国政府也在《新一代人工智能发展规划》中明确提出，要利用智能技术加快推动人才培养模式、教学方法改革，构建包含智能学习、交互式学习的新型教育体系。教育行业和科技行业正携手拥抱“AI+教育”的明天，希冀通过技术手段推动行业的进步，甚至重塑知识的生产和传承方式。在本节中，我们将从“学习者”和“教育者”这两个教育行业最基础的角度出发，来了解AI在教育行业应用的当下和未来。

### 一、AIGC携手学习者，从“有限”走向“无限”

自降生起，人类就开始通过各种手段从零开始建立对世界的认知：从手指触摸、嘴唇吮吮、牙牙学语，到坐入教室、高声朗读、奋笔疾书；从声音、书籍，到影视乃至实地体验，人类的每一次探索、理解、记忆，都是拓展自身认知边界的坚实一步，也是学习这一行为的本质。然而，受制于多种客观要素，每个人的探索之旅总是障碍重重，而AI对于学习者的意义，便是帮助他们尽可能地解除学习过程中的种种桎梏，最终帮助他们从自身的“有限”尽可能拥抱世界的“无限”。

第一，学习资源本身是“有限”的，不同的学习者对包括课件、讲解在内的学习资源有着不同的需求。比如，偏理科的初中生需要提升中考作文水平的教材，刚转专业的大学生需要对应学科的细致入门课程，正在准备金融行业面试的求职者需要相关专业技能的培训，无论人们希望深入当前领域，还是接触新的方向，是否能得到合适的学习资料就是遇到的第一关。

互联网时代的慕课模式曾通过将部分内容数字化并公开分发的方式，助力资源的流转。

而在图像/语音识别和自然语言处理等技术走向成熟的今天，由AI辅助甚至主导的学习资料整理、制作将会极大降低成本，提高效率，将资源的丰富度和易得性提高到新的层次。AI在学习资源生成领域的应用也使得覆盖长尾成为可能，AIGC技术在一些特殊的领域可以辅助生成优质的教学内容，例如儿童绘本等，加速该领域知识的生产效率并让它更快地进入共享网络，最终被最需要它的学习者所捕获，丰富他们的学习资源。

当有限的学习资源得以补充，并流淌到社会的每个末梢，将极大地促进教育公平。乡村地区的孩子也能够获得本来局限于部分学校的优质课程资源，而听障、视障人士的学习也能够被插上翅膀。自2016年起，北京联合大学特殊教育学院就引入了AI系统，通过手语、口型、文字讲义的配合，帮助听障人士高效学习。在海拔3 000多米的拉萨，盲人学生也能在“AI图书馆”通过百度的智能设备播放各类读物，将自己与更广大的世界相连。AI促进了教育资源的生产和分发，正在弥合教育不平等的鸿沟。

第二，学习者对自身学习情况和学习策略的认知是“有限”的。德尔菲神庙的门廊上镌刻着苏格拉底的箴言：“认识你自己。”认识自己是追寻智慧的目的，而清楚地了解自己在学习过程中所处的位置，明晰下一步前进的方向，也是学习开展的必要条件。智能学习平台能够充分收集学习者在学习过程中的各类数据，并根据其行为模式、各知识点掌握程度为学习者提供精准的画像，帮助学习者了解学习状态和挑战，并根据画像为学习者自动生成后续的个性化学习计划，以提高学习效率。

AI助力学习者“了解自己”的价值也绝不仅仅体现在过程中，在学习开始前，学习者就可以通过AI生成的自身分析报告来选择最合适的学习方向，如科大讯飞和北京师范大学合作，推出“学科潜能和专业兴趣双核测评”，致力于帮助学生了解其在某个具体方向上的思维能力、兴趣、水平，并协助学生匹配到合适的院校、专业，从而助力学生的长期成长。

第三，随着学习媒介逐渐数字化，学习行为本身也逐渐变得灵活甚至碎片化，然而，来自教育者的指导和反馈却在多数情况下显得越发“有限”，可能上班族只有在地铁上通勤时才能学习英语，而小学生只有在放学回家后才有时间完成练习写作。这些适配学习者自身要求的学习行为正时刻发生着，期间遭遇的问题和完成的产出也可能不断累积，但由于时间、场地、人力的限制，学习者很难及时在传统的人类教师那里获得反馈。而反馈却是学习者真正取得进步的核心环节。

相较之下，广泛部署在各类智能学习软硬件中，以及由AIGC驱动的文本答疑、指导和评测具有易得、全天候响应以及高度个性化的优势。在“AI教师”的时刻护航下，学习者得以不断形成习得-评测-反馈的闭环，有效提升学习效果。微软就在该方向布局颇多，例如，微软亚洲研究院和华东师范大学合作研发的中文写作智能辅导系统“小花狮”，能够借助自然语

言处理等技术，实时为学生作文结果评分，并能够分析其背后原因，从而帮助学生找到属于自己的发力点，实现进步。

同时，微软亚洲研究院也向培生的《新朗文小学英语》提供了多项人工智能技术支持。培生能够借助微软Azure认知服务中的“文本转语音”技术实时纠正学习者的英语发音，指点语言技巧，并能够生成英语相关多维度的能力测评。在大量转向线上授课的时期，由AI赋能的《新朗文小学英语》快速弥补了线下真人教师的空缺，让孩子们得以持续学习英语，与此同时，AI带来的低成本优势，让家长们节省了不少教育开支。

第四，在学习的场域上，“有限”的物理空间正向“无限”的虚拟空间演进，以打造更加具有沉浸感、体验感的学习环境，充分启发学习者的兴趣，助力深度学习。由AI驱动的虚拟人能以具现化的形式在VR、AR的世界中和学习者进行交互，并借助多种辅助工具展开教学。比如愿景唯新实验室就打造了一个虚拟仿真试验平台，让学习者随时随地、身临其境般地展开实践学习，从而打破物理世界的限制，并通过“亲眼所见”“亲手所为”的方式强化学习效果，提升学习乐趣。

而在特定学科中，利用AI还可打造出与专业高度匹配的虚拟场景，便捷地为学习者创造难得的体验环境，比如西安科技大学打造的沉浸式矿山模拟系统，以及利用AI分析卫星数据生成的虚拟考古环境。虚拟人AI教师、学伴与虚拟空间有着良好的结合能力，在人类生活向虚拟世界迁移的大潮中，必将扮演越发重要的角色。

## 二、AIGC赋能教育者，实现减负提效

从学习资源到学习过程中的自我认识、教师反馈，再到学习开展的场域，AI将逐步破除这些环节的“有限”，助力学习者摆脱自身的局限，走上通往“无限”之路。而教育者则是这条道路上不可或缺的引路人。对于教育者来说，AI如同他们手上的火把，扮演着辅助的角色。

首先，AI能够帮助教育者减轻日常重复烦琐的劳动负担，节省教育者的精力来进行更富有创造力和挑战性的工作，比如进一步促进人和人的关系。相较于不断增长的学习需求，教育者的数量不足将会是长期持续的矛盾，最直接的结果就是“大班制”。在这种情况下，教师不得不一人满足几十位学习者的不同需求，奔波在琐碎的答疑、备课以及大量的作业批改中，而AIGC相关技术的出现，就能够有效解决这一问题。

当下，作业/考卷自动批阅等技术已经获得了越发广泛的应用，人工智能不但可以判断学生题目的正确性，还能生成针对性的评语。根据认知智能全国重点实验室的统计，在人工智能的帮助下，教师备课工作的效率大幅上升，而作业批改的负担则显著下降，文科教师的作业批改用时甚至可以缩短50%～70%，教师得以从低附加值的工作中抽身，转而关注学

生的个性化发展。

除了帮助教育者完成重复性工作，AI还可以延伸教育者的感知，充当他们的眼睛、耳朵，更加全方位地关注学习者的情况。比如基于计算机视觉技术，AI能够实时、全面地分析学生的面部表情等信息，并生成展现学生当前情绪、学习状态等的分析报告，这将帮助教育者深度了解当前教学的开展情况、接收质量，并及时做出针对性调整，提升教学效果，最终赋能“教、学、评、管、研”的教育全环节。

通过对学习者以及教育者的赋能，AIGC最终可能帮助人类实现教育的终极理想：因材施教，即大规模开展高度个性化的教育，让每个人都以最适合自己的节奏，在最合适的方向上进行自由发展，充分发挥自身的潜力，这也是经过多年发展的自适应教育的终极形态。尽管人工智能在教育行业的应用还面临着加剧信息茧房的风险，以及对传统教育伦理的挑战，甚至将人“机器化”的忧虑，但更多的人相信，借助人工智能，人类将打造更好的“以人为中心的”的教育，实现所有人终身、全面的发展。

## 第五节 AIGC金融行业应用

金融行业是天然关乎数据与信息处理的行业，行业中的各类公司都需要从纷繁复杂的市场中搜集各类信息，并利用这些信息创造出各种各样的财富。这样的业务需求特点让金融行业的信息化一直走在其他行业前列，具备数据质量好、数据维度全和数据场景多等诸多特点，让它成为传统AI技术最早落地的商业场景之一。在金融行业中，最常见的应用人工智能的场景是通过AI模式识别和机器学习的方式捕捉市场的实时变化，并利用大量的实时数据进行分析，以此提高金融公司的财务分析效率和能力。而随着AIGC技术的快速发展，不少金融公司也已经注意到了这方面的潜力，并正在积极试水，将最新的AIGC技术整合到公司的日常工作流程当中，提升公司其他方面的工作效率。

AIGC在金融行业的应用主要聚焦于智慧客服与智慧顾问服务两个方面。在智慧客服方面，客户可以通过自然语言处理技术，使用语音或文本与AI系统进行交互，轻松获取有关金融产品和服务的信息，并进行相应的操作。在某些领域，AI系统已经完全可以代替金融人工客服。通过让人工智能系统学习金融知识库，包括金融机构的产品、服务、政策和程序等信息及一般回答话术，人工智能就能结合客户的问题，生成符合场景的回答，解答用户的问题，并协助客户处理如账户设置、风险评估、理财签约及理财购买等一系列常见业务，用户无须再耗费大量时间在柜台排队等待人工服务。传统人工客服不可能7×24小时全年无休地工作，人工客服受个人情绪、压力和周边环境的影响，在服务过程中难免出现情绪化或者“违规”操作，这在金融强风险领域是很严重的问题。而AIGC智能客服的可控、稳定服务解决了这些问题。

此外，AI系统也能够快速高效地完成一部分当前人工客服难以完成的工作。例如，AI系统可以记住客户的喜好，侧写多维客户画像，构建预测式服务体系，进一步提升客户服务体验。AI系统通过对客户标签、交易属性等多类数据进行分析和研究，借助算法建模等金融科技手段，主动迎合广大金融消费者的需求，对目标客群开展不同层次、不同手段的服务触点，提供“千人千面”专属特色顾问服务。

目前，AIGC技术已经取代了金融行业的大量客户服务人员和客户经理。例如在2017年4月，富国银行就开始试点一款基于Facebook Messenger平台的智慧客服项目。在该项目中，人工智能可以代替客服与客户交流，为客户提供账户信息查询、重置密码等基础服务。而美国银行也推出过类似的智能虚拟助手Erica，客户可以使用语音和文字等方式与Erica互动，而Erica则可以根据客户的相关指令帮助客户查询信用评分、查看消费习惯等，更厉害的是，Erica还具有智慧顾问的能力，可以根据客户银行流水收支变化为客户提供还款建议、理财指导等。此外，苏格兰皇家银行也有类似服务，其推出的“LUVO”虚拟对话机器人可以为客户

获取最适合的房屋贷款等，旨在成为客户“可信任的金融咨询师”。除了海外的实践，国内金融业的智慧客服和智慧顾问的相关产业也较为成熟。无论是各类银行、基金公司，还是聚焦金融业务的互联网公司，都推出过自己的智慧客服和智慧顾问机器人业务，将AIGC的相关技术应用于客户服务和投顾咨询。例如，中国工商银行在2022半年报中披露，“工小智”智能服务入口拓展至106个，智能呼入呼出业务量3.1亿次。上半年，该行客户满意度为93.9%，客户电话一次问题解决率达93.3%。中国邮政储蓄银行披露的2022半年度业绩报告显示，该行信用卡客服热线以数字化转型为抓手，升级迭代智能客户服务，积极拓展智能化服务场景，智能客服占比提升至79%以上，智能识别准确率达到94.77%。这些都反映了AIGC相关技术应用于我国金融业的巨大潜能。



## 第六节 AIGC医疗行业应用

AIGC技术的发展和推广，无论是对医生还是对患者而言，都是一种福音。AI预问诊就是一个最典型的应用场景。在医生问诊较为繁忙的时间段，人工智能可以进行预问诊，与患者进行语音或文字的互动，模拟医生的问诊思路，收集患者既往病史、过敏史、用药史、手术史等重要信息，并与患者进行自然的语言互动。而等到患者开始诊疗时，人工智能会根据预先收集的信息生成诊疗报告，使医生可以更快地处理患者的病症。通过这样的模式，不仅医生的时间得到了释放，而且患者也得到了更好的服务，医院也对患者在科室扎堆排队的现象进行了合理的分流和管理，可谓是一举三得。2021年，复旦大学附属眼耳鼻喉科医院与腾讯医疗健康签署了战略合作协议，将全面打造数字化医院建设新标杆、新范式，深度推动医院数字化转型，在“智能预问诊”等业务上已快速落地推进。

除了AI预问诊之外，患者在用药咨询、用药提醒等方面也可以得到人工智能的帮助。比如现在，随着慢性病患者的人数增加，药物的联合应用已经成为常态。在实际场景中，虽然医师会给予患者如何用药的医嘱，但患者在实际用药过程当中可能会出现用药时间错误、漏服、过早停药、服药剂量错误、随意换药等问题，最终导致治疗失败或者疗效不尽如人意的情况。而患者出现用药问题的首要原因就是患者在自主用药的过程中没有得到及时的提示或指导。人工智能系统则可以帮助患者对他们的用药合理性进行分析，通过调用知识图谱，以及发现已录入药品成分之间的药物过量或相互作用关系，对上述问题进行自动检测并提醒患者。此外，人工智能还可以根据待服用药品集中各药物的服用约束条件，建立规划算法，得到可行的安全用药时间段，为患者用药提醒提供科学依据。在全面分析服用药物的基础上，这些信息都能以便于理解的方式在AI与患者进行对话的时候合理呈现。

而对于部分心理疾病，具备对话生成能力的人工智能本身就可以参与到治疗过程之中。首先，相较于传统的心理咨询或者与亲友进行倾诉，AIGC聊天机器人只是一个隔着屏幕的软件程序，用户不必担心自己被评判或者隐私被泄露。其次，相比于心理咨询师职业生涯的案例总数，AIGC聊天机器人有海量交流数据和知识模型支撑，可以在持续迭代更新的同时保持冷静和中立，提供一种可靠且可自己进化的心理咨询服务。此外，当患者在凌晨因为压力或焦虑难以入眠，不能立刻求助心理医生时，AIGC聊天机器人可以提供聆听与陪伴。聆心智能就是典型的使用AIGC技术为用户提供心理健康疗愈方案的公司。聆心智能基于生成式大模型开发的情绪疗愈机器人“Emohaa”，可以构建以生成对话为核心的交互式数字诊疗方案，通过对话与患者共情，及时提供情绪支持与心理疏导，促进患者的心理健康。

除了心理健康之外，AIGC在对听障、语障人士的支持领域也发挥着重要作用。获得科大讯飞战略投资的音书科技就是这样一家公司。音书科技不仅为听障、语障群体提供各种场

景下的翻译字幕系统和手语系统，以支持他们的日常交流和信息获取，还提供了AI言语康复系统。根据音书科技官网显示的数据，目前音书科技已经对外提供数亿次辅助沟通服务，大大改善了听障、语障群体的沟通现状。

除了前面介绍过的诊疗相关的领域外，对于医生来说，医疗科普也是日常工作的重要环节，而AIGC也可以帮助医生更好地完成医疗科普工作。万木健康公司就借助了AIGC相关技术，只需要采集一段时间的人像、音频，就可以合成属于医生的数字分身，借此制作各种题材的医疗科普视频。这样，医生不需要在繁忙的工作中抽出时间出镜拍摄，也不需要进行视频剪辑，就可以低成本地持续性产出医疗科普视频，在节约精力和成本的同时，惠及患者。

[1] 参考自<https://www.theguardian.com/technology/2020/may/30/microsoft-sacks-journalists-to-replace-them-with-robots>。

[2] 参考自<https://www.haimaqingfan.com/>。

[3] 参考自<https://baijiahao.baidu.com/s?id=1722182134150885316&wfr=spider&for=pc>。

[4] 参考自<https://mjiemian.com/article/8310011.html>。

[5] 参考自<https://yrd.huanqiu.com/article/471kvBIJOmo>。

[6] 参考自<https://baijiahao.baidu.com/s?id=1739150418915302141>。

# 第五章 AIGC的产业地图

AIGC的产业链上有哪些创业、投资的机会？

我们并非使用技术，我们生活在技术之中。

——高弗雷·雷吉奥（Godfrey Reggio）

阅读至此，各位读者对AIGC的缘起、技术、应用都有了系统性的理解，但落脚到投资、创业究竟会有哪些商业机会？产业链各个环节的价值体现在何处？都有哪些典型的玩家和商业模式？本章将带着这些问题从商业机遇的捕捉角度入手，对整个AIGC的产业进行详细描绘。总体来看，整个AIGC的产业地图可以分为三类：上游数据服务产业、中游算法模型产业、下游应用拓展产业（图5-1）。

·数据服务：作为智能机器的“食物”和数字经济世界的生产要素，数据在被“喂”给机器之前，常常会涉及查询与处理、转换与编排、标注与管理等前置步骤，而在整个数据的使用过程中也离不开治理与合规方面的管理工作。作为AIGC的源头，相关数据服务产业孕育了很大的商业机会。

·算法模型：人工智能之所以能判断、分析、创作，主要是因为存在支撑这些功能的算法模型。因此，训练算法模型也就成为整个产业链中最“烧脑”、最具技术含量和最具商业潜力的环节。在数字世界，围绕着如何让算法模型更聪明的命题，诞生了包括人工智能实验室、集团科技研究院、开源社区等主要玩家，构成了整个产业链的中游环节。

·应用拓展：经过数据训练后的算法模型最终会在下游应用拓展层完成“学以致用”的使命，根据应用场景的模式和功能差异诞生出文本处理、音频处理、图像处理、视频处理的各个细分赛道。每个细分赛道里都有许多创新企业在相互较量，这也是当前风险投资机构最热衷投资的环节。

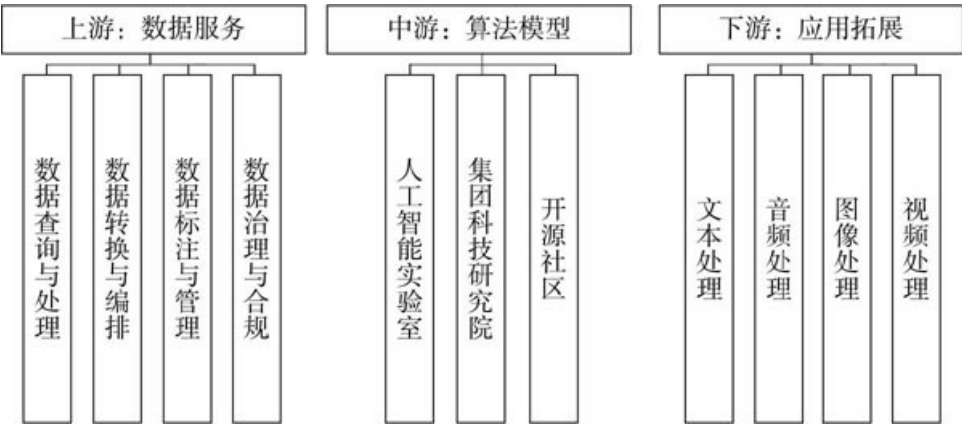


图5-1 AIGC产业地图

## 第一节 产业上游：数据服务

假如人工智能算法是一个生物，那么喂养这个生物的食物便是数据。

无论是机器学习还是人类学习，其分析、创作、决策的能力都是来自知识的学习和经验的积累。不同的是，机器可以不眠不休地学习，不会因为情感和情绪降低学习效率，更不会因为控制不住打游戏、刷短视频的冲动而放弃学习。因此，在机器学习的世界里，“头悬梁、锥刺股”“找家教、开小灶”这类纯粹延长学习时间的内卷策略通常并不奏效。在这种情况下，真正决定不同机器之间能力差异的就是数据的质量。AIGC的产业链上游是一系列围绕数据服务诞生的生产环节，我们可以用农作物加工过程作一个虽不严谨但易于理解的类比。

- 首先是数据查询与处理，这个环节相当于把刚从农田里收割的农作物分类打包；
- 其次是数据转换与编排，这个环节相当于把分类打包的农作物运送到食品工厂后制作成包装精美的成品；
- 再次是数据标注与管理，这个环节相当于给来自工厂的成品商品打上条码和标价；
- 最后是数据治理与合规，这个环节相当于库房的安保人员要确保商品按照相应的规则合理存放。

图5-2展示了AIGC产业链上游的全景，最右列是上游主要的公司，右侧第二列是公司类型，这些不同类型的公司可以被归类到数据服务的四个主要环节中。

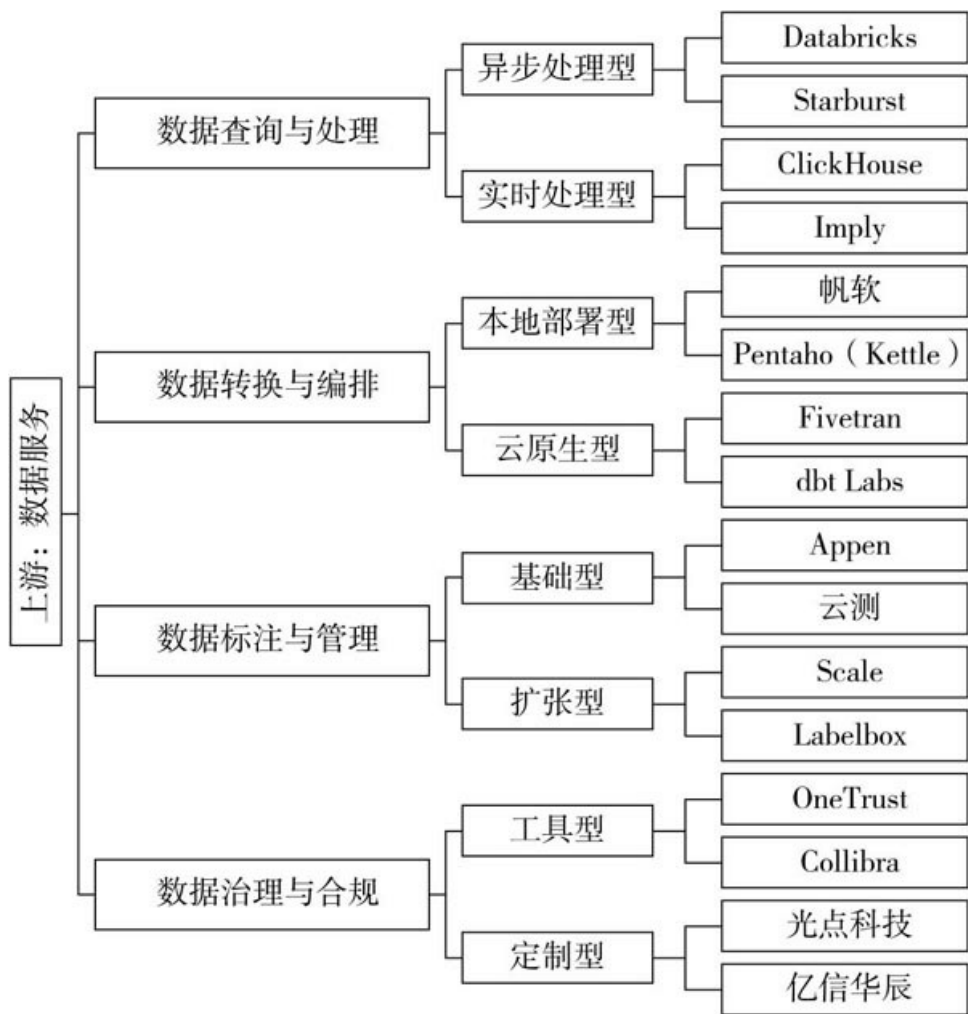


图5-2 上游数据服务层产业地图

## 一、数据查询与处理

通常，数据需要存储在一个合适的地方，等待着人类输入指令去提取符合要求的数据进行处理。一方面，这种存储可以像现实世界中的淡水湖一样，直接把来自四面八方的水源汇聚在一起，不作区分，这种存储架构被称为数据湖（Data Lake）。另一方面，这种存储也可

以像农场里的仓库一样，将数据像瓜果一样收集后清洗好，然后在仓库里一个个摆放整齐，这种存储架构被称为数据仓库（Data Warehouse）。

近几年，在技术进步和商业发展的推动下，“湖仓一体”（Data Lakehouse）的数据存储模式开始出现。湖仓一体模式将数据湖的灵活性和数据仓库的易用性、规范性、高性能等特点融合起来，能够为企业带来降本、省时、省力等多种好处。

- 降本：湖仓一体模式可以降低数据流动的成本，相当于把天然农场变成了粮仓。

- 省时：湖仓一体模式可以降低时延，类似于省掉了农作物从农田搬运到仓库的环节，这样可以节省搬运时间。

- 省力：对企业而言，湖仓一体模式可以避免在数据架构层面不必要的重复建设。

无论是数据湖模式还是湖仓一体模式，都更加符合当前AIGC提取各类非结构化数据和结构化数据训练使用的需求。根据市场研究公司IMARC测算，全球数据湖市场规模在2021年达到了74亿美元，并预计2022—2027年复合年增长率为26.4%，预计2027年全球数据湖市场规模达300亿美元，<sup>[1]</sup>可见增长潜力之大。数据湖具有如此大的增长潜力，因此如何从数据湖中查询与处理数据就显得更为重要。根据数据查询与处理的时效，可以将涉及这个环节的公司分为两类：异步处理型公司和实时处理型公司。

## 1.异步处理型公司

简单地说，异步处理指的是数据的处理过程并非同步进行，而是分不同步骤依次进行。这里划分的异步处理型公司并非指公司不具备实时处理的能力，而是数据服务主要针对的业务场景是异步工作的。截至2022年12月初，数据查询与异步处理型公司中有两家公司发展势头迅猛，值得关注：一是Databricks，当时的最新估值是380亿美元；二是Starburst，当时的最新估值是33.5亿美元。

2013年，通用计算引擎Apache Spark的创始团队出于对Spark商业化的考虑成立了Databricks公司。自此，Databricks就像架在数据湖之间的桥梁，通过支持行业特定的文件格式、数据共享和流处理等方式，让数据的访问和预处理变得更加便捷。Databricks提供了一个名为Delta Sharing的开源功能，可以实现数据的跨区域共享，从而提高工作协同效率。另外，Databricks针对特定行业特定文件格式的数据处理需求，一直在探索有针对性的垂直产品。比如，针对不同医院的电子病历格式上会存在细微差异的问题，Databricks可以对电子病例的原始数据进行访问和预处理，从而形成格式统一的结构化数据。Databricks的首席测试官（CTO）马泰·扎哈里亚（Matei Zaharia）在2022年12月接受采访时表示：“Databricks在前三大超大规模数据中心里运行着超过5 000万台虚拟机，有1 000多家公司在使用Delta

Sharing进行数据交互。<sup>[2]</sup>可以说，Databricks是一个联结数据湖仓架构的枢纽，而这份枢纽所带来的数据价值也收获了投资人的广泛认可。

Starburst是一家缘起于Facebook开源项目的数据分析公司。它提供了一种解决方案，可以让用户随时随地快速轻松地访问数据。Starburst的历史可以追溯到2012年Facebook的开源项目Presto。Presto最初是为了满足Facebook大规模数据快速查询的需求而建立的。2013年，Presto的初始版本在Facebook上线使用并开源，自此之后，包括亚马逊、奈飞和领英在内的其他科技公司也都开始使用。直到2017年，为了更大规模推动Presto的使用，Starburst得以成立，并在一段时间的发展中收获了资本市场的青睐。

## 2. 实时处理型公司

与异步处理型公司类似，实时处理型公司指的是主要针对实时处理需求的公司提供数据服务。截至2022年12月初，数据查询与实时处理型公司中有两家公司值得关注：一是ClickHouse，当时的最新估值是20亿美元；二是Impley，当时的最新估值是11亿美元。

ClickHouse强调处理速度，可以实现实时数据访问与处理，并且围绕它形成了一个开发者社区，有助于持续开发和技术改进。ClickHouse的主要产品是一个开源的列式数据库，在列式数据库中，数据按列进行物理分组和存储，从而最大限度地减少了磁盘访问次数并提高了性能，因为处理特定查询时每次只需要读取一小部分数据。此外，由于每一列都包含相同类型的数据，因此也可以使用有效的压缩机制降低存储成本。而正是这些独特的技术特性让ClickHouse受到了资本市场的充分关注。

Impley是一家基于Apache Druid提供数据查询与实时处理服务的公司。Apache Druid是一个实时分析型数据库，最初主要面向广告行业的数据存储、查询需求，因为广告数据对数据的实时性要求很高，对广告主而言，及时衡量曝光、点击、转化等关键指标有助于快速评估广告投放的效果，进而对广告投放策略进行调整。尤其是在自媒体时代，网络热词的时效性、用户的注意力、网红达人的生命周期都变短，这使得广告业对数据访问和处理的实时性要求变得越来越高。目前，Impley为许多需要利用动态数据进行实时处理分析的场景提供技术支持，也为不少更高级别的AI技术提供大规模数值计算的能力。

## 二、数据转换与编排

在这个环节里，作为人工智能“食品原材料”的数据就需要被运送到加工厂里进行加工处理了。这个环节对数据的处理主要包括提取（Extract，简称E）、加载（Load，简称L）和转换（Transform，简称T）三个模块，因此产业界通常将该环节称为ELT或ETL，也就是三个



模块的英文首字母缩写，L和T的顺序则取决于实际操作流程中哪个环节在前面。这三个模块的含义如下所示：

- 提取：从各种来源获取数据。
- 加载：将数据移动至目标位置。
- 转换：处理和组织数据，使其具备业务可用性。

根据市场研究公司Grand View Research的数据，全球数据集成工具市场的规模在2021年是105亿美元，预计2022—2030年复合年增长率是11.9%。<sup>[3]</sup>根据数据处理的方式是在本地还是在云端，可以将涉及这个环节的公司分为两类：本地部署型公司和云原生型公司。

## 1.本地部署型公司

本地部署型公司主要指核心软件产品部署在本地电脑环境中使用的公司。在这个领域有两家公司值得关注：一是帆软，二是Pentaho（主要关注其产品Kettle）。

帆软成立于2006年，是一家总部位于中国无锡的大数据商业智能和分析平台专业提供商，它专注于商业智能和数据分析领域，致力于提供一站式商业智能解决方案。仅2021年，帆软销售额就已超11.4亿。<sup>[4]</sup>根据国际数据公司IDC 2021年的数据，帆软的主业商业智能的市场份额连续五年在中国排名第一。旗下的FineDataLink是一站式数据集成工具类的重要产品，其目的是为了解决企业数据处理的困境。如今各大企业拥有大量各种类型的信息系统，但企业之间并不连通，形成了数据壁垒，这也使企业无法进行有效的数据联合分析，最终导致数据无法发挥最大价值。而FineDataLink通过对多种异构数据进行实时同步，采用流批一体的调度引擎进行数据清洗，并提供低代码Data API敏捷发布平台，帮助企业解决数据孤岛，提升数据价值。从帆软官网披露的信息来看，FineDataLink的客户以三一重机、安特威、惠科金渝等制造业客户为主。

Kettle最早是一个开源的ETL工具，采用java编写，可以在各种类型的操作系统上运行，数据抽取高效、稳定。2006年被Pentaho公司收购，2015年Pentaho公司又被Hitachi Data Systems收购。截至2021年1月31日，Kettle开源版软件下载量最多的国家是中国，占全球下载量的20%。<sup>[5]</sup>

## 2.云原生型公司

云原生型公司主要指以云服务的形式提供旗下产品数据转换与编排功能的公司。截至2022年12月初，云原生型公司中也有两家公司值得关注：一是Fivetran，当时最新估值是56

亿美元；二是dbt Labs，当时最新估值是42亿美元。

Fivetran是硅谷知名孵化器Y Combinator成功孵化的公司，这家公司的名字来自20世纪50年代IBM开发的编程语言Fortran。随着云计算技术的到来，Fivetran最初意识到传统ETL/ELT工具的性能可能难以匹配云原生的工作场景，因此基于云原生场景开发了相较于本地部署场景下的ETL/ELT工具更适配的数据整合平台。通过提供SaaS（Software-as-a-Service，软件即服务）服务，Fivetran可以连接到业务关键数据源，提取并处理所有数据，然后将数据转储到仓库中，以进行查询访问和必要的进一步转换。Fivetran让大规模数据的分析操作变得更简单了，有人认为Fivetran是“在Excel和Matlab之间找到了平衡”。随着数字时代的发展，未来大规模数据分析的需求会越来越强烈，但学习专业的大数据分析工具成本不低，因此Fivetran很好地弥合了这个市场需求。

dbt Labs聚焦在ELT中的Transform部分，帮助数据团队“像软件工程师一样工作”，它的核心功能是帮助用户书写数据转换的代码。在创业之前，dbt Labs的创始人团队一直在数据分析领域工作，他们对于数据分析所面临的问题和挑战有着深刻的了解。他们一直坚信，数据分析师是一种创造性的工作。dbt Labs最初推出的产品非常小众。一部分尝鲜客户为dbt Labs的产品提出了很多改进建议和需求，这有助于产品的迭代，也有利于让产品在这些早期用户中进行口碑传播，就像一个种子在肥沃的土壤中发芽生长一样，这使得dbt Labs快速成长起来。在它发布了dbt cloud的云服务之后，公司估值也快速上升，获得了投资人的广泛认可。

### 三、数据标注与管理

如果说人工智能是把机器当作学生进行教学的过程，那么数据标注与管理环节则是备课环节，把原始数据进行结构化处理后，接下来就是组织整理知识点，然后教给机器。在前文中，我们介绍过在许多任务场景中，人工智能需要通过监督的方式进行学习，人类通过给机器“喂养”标注了知识点的结构化数据来实现监督，最终形成可以解决各个场景实际问题的算法模型。正如中国工程院院士贺贺铨曾表示的：“智能驾驶中需要让汽车自动识别马路，但如果只是将视频单纯地传给计算机，计算机无法识别，需要人工在视频中将道路框出，再交由计算机，计算机多次接受此类信息后，才能逐渐学会在视频和照片中识别出道路。”<sup>[6]</sup>

根据Grand View Research的研究，2021年全球数据标注市场规模为16.7亿美元，预计2022—2030年将以25.1%的复合年增长率增长。数据标注环节听起来技术含量并不高，只需雇用更多的劳动力就可完成，但有心的公司可以基于数据标注的源头将业务拓展到其他环节，获得更大的发展空间。因此，根据公司业务拓展程度的差异，可以将涉及这个环节的公司分为两类：基础型公司和扩张型公司。

## 1.基础型公司

基础型公司通常专注于数据标注与管理领域，并没有过多将业务延伸至算法模型等其他领域，虽然聚焦的环节附加值不高，但由于充分的专注度，基础型公司在该垂直领域形成了独特的竞争优势，Appen和云测数据就是这一类公司的典型代表。

Appen是全球领先的AI训练数据服务提供商，成立于1996年，2015年在澳大利亚证券交易所上市。基于官网信息可知，Appen在全球拥有100多万名众包人员，支持235种语言，业务遍布全球170个国家和7万个地区。目前，Appen已经为全球许多头部企业提供服务长达20多年，能够针对不同行业的AI应用场景需求提供独特的解决方案。

云测数据是另一个具有代表性的基础型公司。云测数据成立于2011年，是一家自动化软件测试公司，2018年开始涉足数据标注业务，旗下拥有云测标注平台和国内众多供应商，致力于加速AI场景化落地。根据《互联网周刊》发布的“2022数据标注公司排行”，云测数据排在国内数据标注行业第一位。

## 2.扩张型公司

Scale是从数据标注环节向其他环节扩张的典型公司。Scale在成立的最初四年还只是专注于给数据打标注，但从第五年开始逐步向下游扩展，目前已经开发了自有模型，从而进入更加具有技术含量和商业价值的环节。Scale官网信息显示，Scale的客户不仅包括美国国防部和科技巨头（比如微软、SAP、PayPal），甚至包括OpenAI。Scale之所以可以从最初看似技术含量不高的数据标注环节向更具附加价值的中下游环节扩张，主要受益于规模经济、客户黏性和资源垄断。

·规模经济：Scale的客户越多，处理的数据量和数据维度也越多，对于不同任务的处理经验也更加丰富，相关的标注算法工具也更加完备，从而处理效率和质量就越高。因此，随着时间的推移，Scale作为先发者相较于跟进者而言就可以以更低的成本提供更高质量的服务，做“时间的朋友”。

·客户黏性：数据标注服务本身很难建立起高度的客户黏性，而Scale之所以可以留住客户，得益于它在2020年4月推出的Scale Document。Scale Document不仅为数据贴标签，还与客户合作建立定制模型。这使得客户切换服务商的成本变高，因为需要重新训练模型。

·资源垄断：这里所说的资源垄断指的不是垄断数据而是垄断人才，数据的所有权是客户的，即使通过Scale来完成打标签过程，也不能把这些数据误认为是Scale的资产。但随着数据流过Scale平台，这些数据同样训练了Scale平台标注算法的模型能力，也沉淀了这个领域的众多人才，人才是这个领域的宝贵资源。

另一家典型的扩张型公司Labelbox也是从数据标注起家，逐渐拓展了数据管理、AI辅助

标记、模型训练和诊断服务等相关业务，进而成为一个综合性的AI数据引擎平台。

Burberry（巴宝莉）就曾利用 Labelbox来辅助它的营销策划。作为跨国品牌，Burberry在进行全球营销的过程中常常需要处理大量的营销图片。为了帮助高效决策，Burberry通常需要对成千上万张图片进行打标签和分类，进而在营销投放环节，根据品牌宣发需求进行精准的分渠道投放。过去打标签环节是完全通过人工进行的，耗费时间和精力，如今利用Labelbox这样的工具后，可以大幅提高打标签的效率，节省图片分类的时间。根据Labelbox官网的信息，在和Burberry合作的过程中成功为Burberry节省了10个人力，仅花费2个小时就可以处理完成上千张图片。

智研咨询数据显示，2021年我国数据标注与审核行业市场规模达到44.4亿元，伴随着AI战略被更多企业认同，更多资金和资源被投入，以及各项技术得到实际应用和落地，我国数据标注与审核行业将延续高速增长态势。国内头部科技公司都有自己的数据标注部门，比如百度的百度众测和京东的京东众智。

## 四、数据治理与合规

虽然数据是人工智能机器的“食物”，但也不能让机器胡吃海塞。在数字经济时代，数据是和土地、人力、资本一样举足轻重的生产资料，因此，既需要保证数据资产在管理时符合预先设置的数据质量规范，也需要在访问和调取数据时做到合法合规，这也使得数据治理和合规服务逐渐成为各个企业的必需品。

市场研究公司ReporterLinker的数据显示，2020年全球数据治理市场规模约为18亿美元，预计到2027年将达到72亿美元，在此期间以22%的复合年增长率增长。<sup>[7]</sup>根据服务交付的模式，可以将涉及这个环节的公司分为两类：工具型公司和定制型公司。

### 1.工具型公司

工具型公司是将数据治理与合规服务产品化，需要相关服务的客户可以直接购买标准化的产品或基于已有的产品进行部分自定义。OneTrust和Collibra就是两家典型的工具型公司。

OneTrust总部设在亚特兰大和伦敦，创始人卡比尔·巴戴（Kabir Barday）曾是BlackRock的开发人员。他在2016年注意到很多公司在准备数据合规业务，于是创办了OneTrust公司。OneTrust通过自动化工具帮助企业遵守《通用数据保护条例》《加州消费者隐私法案》和数百个其他全球隐私法律。OneTrust简化了消费者和主体权利请求的接收和履行流程，允许客户与同行进行基准比较，绘制和盘点处理记录，并在数据流经其组织时生成自定义报告。根据2020年IDC市场份额报告，彼时仅成立4年的OneTrust公司的份额就占到全球数据隐私市场

总份额的40.2%，并被Inc.500评为美国增长最快的公司。

Collibra早在2008年就在纽约成立，它通过提供各种工具来满足数据监管的合规要求，并以自动化的数据治理和管理解决方案而闻名。Collibra提供了自动数据分类的功能，如果特定数据集内包含与欧盟居民有关的个人身份信息（PII），它将自动应用《通用数据保护条例》《加州消费者隐私法案》等法案政策，通过使用机器学习对敏感数据进行自动分类，省时省力。

## 2.定制型公司

定制型公司主要的业务特点是客户提供个性化的解决方案。光点科技和亿信华辰就是两家典型的定制型公司。

光点科技总部位于广州。根据光点科技官网信息，截至2022年底，光点服务的客户已超过100家，包括广东省工业和信息化厅、广州市工业和信息化局等。光点科技的服务行业涉及金融、电信、政务、泛零售等。通过数据治理，光点科技可以对企业数据收集、融合、清洗、处理等过程进行管理和控制，有助于持续输出高质量数据。通常，客户会针对特殊的业务场景进行数据解决方案的定制，例如，在新冠肺炎疫情防控期间，通过光点数据填报系统，在机场、火车站、高速口、客运站等人流密集的卡口区域扫描二维码登记，可实现人员无接触通关，也有助于实时掌控人员行动轨迹，以便及时推出联防联控的行动解决措施。基于数据治理业务，光点科技同样能够提供有价值的数应用服务，例如光点科技研发的“数字灵境”就将大数据与城市发展相结合，打造出了智慧城市大数据平台。

亿信华辰成立于2006年，它自主研发了“睿治”智能数据治理平台，可以提供定制化的数据治理服务。基于亿信华辰官网信息，截至2022年12月，亿信华辰已经服务了1.1万家企业和2.3万个项目。作为定制型数据治理服务的代表性公司，亿信华辰根据不同行业的需求“因地制宜”，例如为地产商时代中国量身定制了一套完整的线上数据管控体系，通过数据资产管理，构建了一整套线上数据管控体系。根据IDC发布的《中国数据治理市场份额（2021）》报告，亿信华辰在国内数据治理市场的份额占据第一位。

## 第二节 产业中游：算法模型

产业中游的算法模型是AIGC最核心的环节，是机器完成教育训练过程的关键环节。中游算法模型包括三类重要的参与者：人工智能实验室、集团科技研究院和开源社区。中游算法模型的产业地图如图5-3所示。

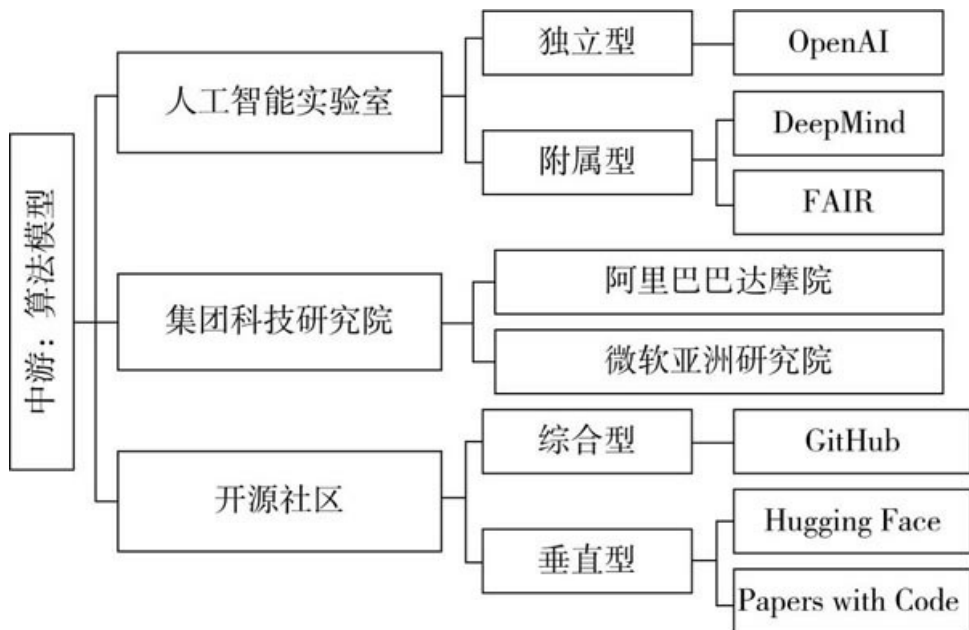


图5-3 中游算法模型层产业地图

### 一、人工智能实验室

算法模型在人工智能系统中起到决策作用，是人工智能系统完成各项任务的基础。算法模型可以用来表示人工智能系统的知识，并通过对数据进行处理，帮助人工智能系统做出决策。因此，算法模型可以被视为人工智能系统的灵魂，也是人工智能从“单细胞”到“多细胞”，再到“高级智慧生物”演进过程的根本推动力，正是种种算法模型赋予了机器近乎人类的洞察力与创造力。很多企业为了更好地针对算法模型进行研究并推动其商业落地，在企业内部设立了和高校一样的人工智能实验室，甚至有些企业本身就是一个大型人工智能实验室。据此，可以将人工智能实验室分为两类：独立型人工智能实验室和附属型人工智能实验

室。

## 1.独立型人工智能实验室

独立型人工智能实验室中最具有代表性的公司是OpenAI。OpenAI于2015年在美国硅谷成立，其背后的创始团队阵容十分强大：有着“钢铁侠”称号的埃隆·马斯克（Elon Musk）、全球知名创业孵化器Y Combinator的掌门人山姆·阿尔特曼、著有畅销书《从0到1》的硅谷创投教父彼得·蒂尔（Peter Thiel）。不止于此，在OpenAI成立后的第五年，微软向OpenAI投资了10亿美元。在5v5模式的Dota2比赛中，OpenAI开发的人工智能OpenAI Five击败了人类选手，比尔·盖茨（Bill Gates）盛赞这是人工智能发展过程中的重要里程碑。而在2022年引爆AIGC热潮的“ChatGPT聊天机器人软件”也正是OpenAI的杰作，OpenAI推出的大模型GPT-3可以达到千亿级参数，而其即将推出的GPT-4模型被许多人认为有望真正通过图灵测试。除了GPT之外，OpenAI在2022年同样发布了知名AI绘画工具DALL·E 2，以及逼近人类水平、支持多种语言的语音识别预训练模型Whisper。这些智能算法模型无疑都代表着当前人类在人工智能领域的一些顶级成果。

## 2.附属型人工智能实验室

谷歌旗下的DeepMind被认为是OpenAI最大的竞争对手，比OpenAI早成立了5年。DeepMind最知名的人工智能模型是AlphaGo，它在围棋游戏中打败了国际上最优秀的人类棋手。同OpenAI一样，DeepMind也致力于开发通用人工智能算法模型，因此除了内容创作领域之外，DeepMind在许多其他领域也开发了震惊大众的人工智能。2018年，DeepMind开发的AlphaFold在结构预测关键评估（CASP）竞赛中展现出了超出人类的能力，AlphaFold在蛋白质结构预测领域取得了突破性成果，也使得人工智能的触角伸向了生物科技与医疗领域。2022年，DeepMind又发布了基于Transformer的新模型AlphaCode，甚至在国际自然科学领域顶级期刊《科学》（Science）上发表了新论文，该研究登上了《科学》封面。

FAIR则是Meta旗下的人工智能算法模型研究团队，全称为Facebook AI Research，该团队于2022年被并入元宇宙核心部门Reality Labs。FAIR负责人杨立昆（Yann LeCun）是卷积神经网络之父、纽约大学终身教授，与谷歌副总裁杰弗里·辛顿（Geoffrey Hinton）、2018年图灵奖得主约书亚·本吉奥（Yoshua Bengio）并称为“深度学习三巨头”。Meta目前也正在寻求让机器学习和人工智能在整个公司得到广泛应用的机会，而不只是局限在研究部门。FAIR在2021年已经开源了Expire-Span算法，这是一种深度学习技术，可以学习输入序列中哪些项目应该被记住，从而降低AI的内存和计算要求。Meta表示：“作为研究更像人类的人工智能系统的下一步，FAIR正在研究如何将不同类型的记忆融入神经网络。”因此，从长远来看，Meta可以使人工智能更接近人类的记忆，具有比当前系统更快的学习能力。Meta相

信Expire-Span是一个重要的、令人兴奋的进步，朝着未来人工智能驱动的创新迈进。

## 二、集团科技研究院

一些集团型公司往往会设立聚焦前沿科技领域的大型研究院，下设不同细分方向的实验室，通过学术氛围更加浓厚的管理方式，为公司未来科技的发展储备有生力量。阿里巴巴达摩院和微软亚洲研究院就是人工智能领域典型的集团科技研究院。

阿里巴巴达摩院成立于2017年10月11日，致力于探索科技未知，以人类愿景为驱动力，开展基础科学和创新性技术研究。截至2022年年底，达摩院旗下主要包括五个方向的实验室：机器智能、数据计算、机器人、金融科技、X实验室。X实验室指的是除了前四个领域，在未来可能会有裂变价值的科技领域，当前主要涵盖量子计算、下一代移动通信和虚拟现实三个方向。除了这些自研实验室外，达摩院还和全球许多知名高校建立了联合实验室，并推出了阿里巴巴创新研究计划，构建全球学术合作网络，这些目前都是阿里巴巴达摩院研究的重要组成部分。自成立以来，达摩院研究出了许多杰出的成果，其中不少成果与AIGC领域息息相关。例如，达摩院研发的深度语言模型体系AliceMind掌握100多种语言，具有阅读、写作、翻译、问答、搜索、摘要生成、对话等多种能力，其处理能力先后登上了自然语言处理领域的六大权威榜单，并在2021年年中宣布了开源。

微软亚洲研究院成立于1998年，是微软公司在海外开设的第二家基础科研机构，由李开复博士出任第一任院长，至今已经发展成为世界一流的计算机基础及应用研究机构。截至2022年年底，微软亚洲研究院在中国的核心研究团队除了北京、上海的多个细分方向的研究组外，还包含科学智能中心、产业创新中心和理论中心三大研究中心。无论是北京、上海的研究组，还是三大研究中心，许多研究方向都与人工智能相关，也产出过杰出的AIGC研究成果，比如通用多模态基础模型BEiT-3，它在目标检测、实例分割、语义分割、图像分类、视觉推理、视觉问答、图片描述生成和跨模态检索等领域都表现出了杰出的性能。

## 三、开源社区

开源社区对AIGC的发展十分重要，因为它提供了一个平台，让开发人员能够共享他们的代码，分享他们最新的研究成果，并与其他人一起协作，共同推动AIGC相关技术的发展进步。除了可以让研究人员彼此充分学习交流外，开源社区还可以帮助开发者更快地开发出人工智能相关应用。建造各个场景下的人工智能应用系统就像建造一栋栋大楼，往往需要很多人的共同努力。而开源社区就像是工地上的交流中心，让所有参与建造的人都能够找到合适的工具和材料，并与其他人交流想法，共同完成建造工作。如果没有交流中心，大楼的建造将会变得困难重重，甚至无法完成。同样，如果没有开源社区，人工智能的发展也会面临



诸多困难。因此，开源社区对于人工智能的重要性不言而喻。根据开源社区所覆盖领域的宽度和深度，可以将开源社区分为两类：综合型开源社区和垂直型开源社区。

## 1.综合型开源社区

GitHub是世界上最大的开源代码托管平台，目前已有超过9 000万的活跃用户和1.9亿代码库。<sup>[8]</sup>作为代码玩家界的Facebook，GitHub是开发者与朋友、同事、同学及陌生人共享代码的完美场所，无论是人工智能领域相关的代码，还是其他领域的代码都可以在这里上传共享。代码开源不仅可以减少重复性工作，还可以推动技术研究的快速突破，降低应用门槛，加速技术产业化推广使用，以及有效促进学界与产业界的有效交流，促进产学研融合。2018年，Github被微软收购，但其社区与业务依然独立运营，保留了它传承已久的开源精神。无论是AIGC领域的论文还是项目，如果选择上传开源代码的地方，Github绝对是首选。

## 2.垂直型开源社区

除了像Github这样大而全的开源社区外，还有一些针对垂直领域的小而精的网站和社区在开源领域发光发热，比如Papers with Code和Hugging Face。

Papers with Code是一个总结了机器学习论文及其代码实现的网站。用户可以轻松地在网站上检索到所需要的机器学习论文及存储在Github上的开源代码。用户可以按照标题关键词或者研究领域关键词进行查询，也可以按照流行程度、论文发表时间以及Github上收藏（Star）数量最多来对论文及论文代码进行排序。Papers with Code网站最初是由Reddit的用户rstoj开发，让人们可以从中发现一些以前不知道的研究精华。作为机器学习界的内容社区，Papers with Code大大促进了人工智能领域的研究。

Hugging Face是专注于机器学习领域的垂直版GitHub。它想要把主打年轻用户的聊天机器人作为主营业务，因此在GitHub上开源了一个Transformer的代码库，不过没想到聊天机器人业务没做起来，Transformer库却在机器学习社区火起来。很多人总结Hugging Face的成功是因为团队开放的文化和态度，以及利他利己的精神很具有吸引力。目前，仍然有很多业界专家都在使用Hugging Face和提交新模型，甚至有些NLP工程师招聘中明确要求候选人熟练使用Hugging Face Transformer库。如果说人工智能是一场淘金运动，那么Hugging Face则是典型的“卖水人”。

### 第三节 产业下游：应用拓展

任何优秀的算法模型最终都需要落地于具体的应用场景去实现其商业价值。在AIGC产业的下游，可以将AIGC相关应用拓展到四个主要场景：文本处理、音频处理、图像处理、视频处理（图5-4）。伴随着AIGC技术成熟度的提高，在产业下游将会诞生越来越多全新的商业机会与初创公司，本节将对四大主要场景中部分特点明晰的应用与公司进行介绍。

#### 一、文本处理

目前，文本处理是AIGC相关技术距离消费者感知最近的场景，也是技术成熟度相对较高的场景，因此文本处理场景中的应用与公司最为丰富。这些应用与公司会从多个维度辅助公司的业务和职能部门的工作，并直接参与到内容的商业化过程中。

##### 1.营销型文本处理

营销是文本处理最常见的应用赛道，这一赛道最常见的客户是企业的市场营销部门及营销公司人员。这部分人群最大的痛点在于，他们要把无止境的时间投入思考广告创意、营销文案中，内容的生产非常依靠灵光乍现，而他们往往非常容易灵感枯竭。文本处理应用的诞生就是为了解决这个痛点，许多文本处理应用在产出文本的同时，还能通过使用者的修改形成反馈，改进整个模型，从而输出更高质量的内容，形成“AI+人工”的正向技术网络效应。



图5-4 下游应用拓展层产业地图

Copy.ai是典型的营销型文本处理应用。它基于GPT-3大模型，能在几秒钟内生成高质量的广告和营销文案，包含70多个AI模板，覆盖的场景包括博客、社交媒体推广、产品上线等，还可以翻译25种不同的语言。你只需输入标题、文案大意，Copy.ai就可以生成一段可读性较高的文案。Copy.ai意图将人们创作文案的构思阶段缩短80%以上，然后让营销人员依靠人工的修改和润色来填补剩余的20%。它的收费模式也很简单，根据官网在2022年12月显示的信息，免费版Copy.ai每个月只提供2 000个字的额度，Pro版Copy.ai收费为49美元/月，可以同时让5个账户使用，平摊下来每个账户不到10美元/月。

Jasper是一家典型的营销型文本处理公司，旗下产品的功能和Copy.ai非常类似，底层也是采用GPT-3的相关模型，但团队在此基础上做了改进，特别是在广告和营销的内容生成上，Jasper的产品更擅长生产长篇的内容。此外，Jasper公司还收购了一家专注于提供写作语法检查服务的公司Outwrite，其产品非常类似Grammarly，强化了Jasper产品的文本效果。不过，Jasper产品的定价相比Copy.ai更高且没有免费版。Jasper公司与Airbnb、HubSpot、Autodesk以及IBM等企业客户合作，2021年收入超过了4000万美金。<sup>[9]</sup>

## 2.销售型文本处理

销售型文本处理与营销型文本处理有一定的相似性。对于市场营销人员而言，营销型文本处理通常面向广大的公众和消费者，文案更多发布于博客、社交媒体、广告等大众传播的应用场景，比如普通老百姓都能在电视上、网上、大街上看到的广告词和标语，等等；销售型文本处理则面向更私人、非公开的场合，比如电子邮箱。很多金融机构的分析师可能深有体会，每当查看需要输入邮箱的数据或者报告后，邮箱里总会收到大量数据机构的销售人员发来的会议邀请、产品介绍，等等。销售型文本处理应用正是为这些努力工作的销售人员准备的，它可以通过AI自动生成电子邮件，并根据属性筛选和抓取潜在客户邮箱、发送邮件进行验证，最典型的应用包括Lavender和Smartwriter.ai。

Lavender是一款用于编写销售电子邮件的浏览器扩展程序，结合了AI分析、社交数据和收件箱生产力工具等功能模块。AI分析可以帮助用户改进电子邮件回复内容，社交数据帮助用户建立融洽的关系，而移动设备预览、电子邮件验证、GIF图和垃圾邮件拦截器等工具都可以帮助用户来更好地利用电子邮件处理工作。所有这一切的目标都是使销售人员能写出一封更可能得到潜在客户回复的邮件。比如Lavender会分析收件人的社交数据、日历时间等，帮助销售人员了解客户如何做出购买决定及如何定制个性化的邮件信息。Lavender还会对邮件进行分析和评分，快速分析邮件当中的问题，自动进行修复。

Smartwriter.ai在电子邮件功能上与Lavender相似，还集成了类似Jasper产品的营销文案生成能力，能够直接面向Gmail、Yahoo Mail、Facebook、Twitter、LinkedIn进行数据抓取及潜在客户构建和销售。

### 3.续写型文本处理

续写型文本处理与营销型文本处理的共同点在于，都对AI处理文本的自由度和开放度有较高的要求，换句话说，考验AI的“创意”。相对于营销型文本处理应用来说，续写型文本处理应用的用户并非专业的企业人员，更有可能是从事艺术创作的个人，比如每天被读者催更的网文作者。因为用户的规模和收入水平区别较大，续写型文本处理应用并不是生产力工具，而更多的是具有娱乐属性，目前从收费模式上也更可能是免费的。

目前，国内各类视频博主乐此不疲地使用续写型文本处理应用，为《三体》等热门作品续写另一种结局，然后把离谱的结果发到视频平台上，满足用户对AI生成内容的猎奇心理。其中常见的一个应用是由国内公司彩云科技开发的彩云小梦。用户只需要在长文本输入框中先写个开头或者输入世界设定和故事背景，然后就可以交给AI小梦来帮忙续写。彩云小梦还内置了多种续写模型，包括标准、言情、玄幻、都市等。用户可以点击右上角自由切换模型，可根据偏好续写不同风格的内容。每一次续写的一段话都可以中途修改，或者挑选小梦帮写的另外几个段落进行更换。另外，彩云小梦目前还更新了对话版，在完成世界设定后，能够对对话的形式展开剧情。在较小的营收压力下，目前的彩云小梦仍然免费。

### 4.知识型文本处理

上述三类文本处理应用从定位上更接近于“输出”的过程，即“使用的目的”是为了有可以外发的、展示的、传播的产出，就好比一个小学生可能会用小猿搜题找到作业的答案，然后把作业展示给老师。而知识型文本处理应用则更注重信息的“输入”，帮助用户更好地进行信息的归纳、接收和整理，就好比一个小学生在写作业之前，要用思维导图等工具把上课学到的知识点整理好，内化为之后写作业、考试可以用到的技能，但这个过程可能需要花费很长时间，去不同的教材、笔记本、错题本上搜索信息。对于企业员工来说，搜索信息、管理信息一直是一件耗费精力的事，因为员工把大量时间花在了“重新发明轮子”上。一些人工智能文本生成工具就专注于解决这个问题。

Mem就是一家这个赛道上的典型公司，由华裔工程师Dennis Xu和凯文·穆迪（Kevin Moody）共同创办。Mem产品的优势是“轻量级”，主打快速记录与内容搜索，允许用户附加主题标签，标记其他用户。此外，Mem与AIGC的结合更是让其产品功能强大无比，产品的内置工作助手Mem X可以执行智能编辑、智能写作等任务，比如将零散的文本组成段落、为文章进行总结或者生成标题。目前，Mem的商业模式走的也是SaaS的路线，用户需要

购买10美元/月的Mem X套餐，才能享受到AI的能力，包括自动整理和归类信息。除了这个额外的进阶功能外，Mem X的付费版还取消了单个文件大小为25MB的限制，并拥有100GB的总存储空间，这大约是免费版本的20倍。

## 5.通用型文本处理

顾名思义，通用型文本处理不局限于某个特定场景，而是为用户提供具有泛用性的综合解决能力，因此能够覆盖到类别更为丰富的用户。比如Writer公司的AI写作平台，提供从头脑风暴构思、生成初稿、样式编辑、分发内容、复盘研究的全部流程支持，适用于任何需要内容生产的场景和工作，帮助提高内容的生产量、生产效率、点击率、合规性等。

国内的澜舟科技也是一家针对商业场景数字化转型、以自然语言处理为基础提供通用型文本处理服务的公司。根据2022年12月官网的信息，澜舟科技的创始人周明博士是自然语言处理领域的代表性人物，现任中国计算机学会副理事长、中国中文信息学会常务理事、创新工场首席科学家，曾任微软亚洲研究院副院长、国际计算语言学协会（ACL）主席。除了创始人拥有优异的科技背景外，其产品体系基于自主研发的“孟子”轻量化的预训练模型，可处理多语言、多模态数据，同时支持多种文本理解和文本生成任务，能快速满足不同领域、不同应用场景的需求。孟子模型基于Transformer架构，包含10亿参数量，基于数百G级别涵盖互联网网页、社区、新闻、电子商务、金融等领域的高质量语料训练而成。“孟子”预训练模型性能比肩甚至超越千亿大模型，在文本分类、阅读理解等各类任务上表现惊艳。

## 6.辅助型文本处理

与前述需要AI“脑洞大开”进行创意文本处理的应用不同，辅助型文本处理应用是一种较为轻量级的应用，也是目前国内落地最为广泛的场景之一。它的主要功能是基于素材爬取来实现，在很大程度上对写作者起到了“助手”的作用，比如可以根据需求定向采集素材、文本素材预处理、自动化降重、重新表述润色等，帮助创作者减轻许多程序性的工作，提升生产力。

Wordtune就是一款非常典型的辅助型文本处理应用，它的功能是帮助用户“重写”句子，对句子进行缩写或扩写，使句子在原句意的基础上更随意或更正式。Wordtune由以色列公司AI21 Labs构建。AI21 Labs成立于2018年，目标是彻底改变人们的阅读和写作方式，用AI来理解书面文本的上下文和语义。目前，Wordtune已经成为很多中国留学生进行论文修改润色，或者用来练习雅思考试中的同义词替换的“神器”。

国内公司秘塔科技也推出了AI写作助手“秘塔写作猫”。根据官网的信息，秘塔科技于2018年成立，创始人闵可锐毕业于复旦大学计算机系，后在牛津大学数学系、美国UIUC电

子与计算机工程系攻读硕士、博士学位，在谷歌参与过AdSense基于内容广告建模组点击率预测项目，还担任过猎豹移动首席科学家。秘塔写作猫采用了自研的大规模概率语言模型，根据上下文对可能的用词进行准确建模，因此除了文本校对、改写润色、自动配图等辅助功能之外，它也具备根据标题生成大纲或文章，以及提供论文、方案报告、广告语、电商种草文、自媒体文章等写作模板的能力，是同时具备营销和续写能力的文本处理应用。

## 7.交互型文本处理

交互型文本处理应用是形式上与上述应用最不同的一个，因为它的产品形态本身存在叙事，交互的过程本身产生意义，而不是像文案写作应用一样作为一种生产力工具。对于很多用户来说，与苹果的Siri语音助手进行对话本身就很有意思，可以听Siri说出很多有趣的俏皮话。由此我们可以看出，交互型文本处理应用常应用于闲聊、游戏等娱乐场景。

第一章提到的AI Dungeon就属于这类应用。2019年2月，就读于计算机相关专业的尼克·沃尔顿正处于大学最后一年，一次校园编程竞赛让他想到基于OpenAI刚刚发布的GPT-2模型做一个文字冒险游戏AI Dungeon，灵感来源于经典游戏《龙与地下城》，并用与AI文字对话的形式来完成游戏和故事生成。2019年5月，沃尔顿创立了Latitude公司，并在年底GPT-2完全放出后正式推出了AI Dungeon，又在GPT-3推出之后强化了AI Dungeon的语义理解和写作能力。大多数AI聊天机器人的玩法是对话，AI Dungeon则是共同创作故事，玩家可选择Say/Story/Do三种模式，操控自己的角色进行对话、行动，或者只是单纯地看AI基于上下文生成故事。

除了游戏之外，交互型文本处理应用还能够生成用于各种场景的虚拟角色，比如心理治疗等，国内的代表性公司有第四章提及的聆心智能。聆心智能由国内NLP、对话系统领域专家黄民烈教授创办，公司自研了中文对话大模型OPD，该模型是目前世界上参数规模最大的开源中文对话预训练模型。基于这一模型，公司打造了Emohaa情绪疗愈机器人，并与心理平台好心情达成合作，成功落地了国内首款人工智能心理陪伴数字人；与高端电车品牌Beyonca合作，打造了智能驾舱贴心助手。此外，聆心智能还推出了“AI乌托邦”系统，允许用户快速定制AI角色，只需要输入简单的角色描述，就可以召唤出相应人设的AI，与之进行深度对话和聊天。

## 8.代码型文本处理

代码是一种特殊的文本形式，许多公司将代码相关文本的处理作为切入点展开业务经营。人工智能进入代码开发环节，有助于消除开发人员之间的IT知识差异，可以让对编程语言精通程度不同的团队更好地协同工作。根据AIGC对代码处理环节的渗透程度，可以将代码型文本处理公司分为三类：代码生成型公司（辅助代码撰写）、代码文档型公司（代码转

化成文档）、代码开发型公司（直接参与代码开发）。

### （1）代码生成型公司

Repl.it是典型的代码生成型公司。

Repl.it是可以支持50多种编程语言的在线编程语言环境平台，一直致力于为代码工程师解决编程操作问题，使操作更简便、快捷，可以将它简单理解为编程界的“腾讯文档”。Repl.it在全球拥有1 000多万用户，包括谷歌、Stripe、Meta这样的科技巨头。Repl.it推出了Ghostwriter，作为GitHub Copilot的竞争对手而存在，与GitHub Copilot拥有类似的功能。Ghostwriter可以支持16种编程语言，包括C、Java、Perl、Python和Ruby等主流语言。Ghostwriter的商业模式是作为Repl.it的一项付费订阅服务，每月收费10美元，相比GitHub Copilot更加便宜。

### （2）代码文档型公司

程序文档可以帮助开发人员和产品业务部门在沟通协作时理解代码，但它生产起来费时费力。Mintlify的首席执行官（CEO）曾分享道：“我们曾在包括初创公司和大型科技公司在内的各个阶段的公司担任过软件工程师，发现软件工程师都受到编写文档的困扰。”Mintlify就是一家聚焦于解决这种问题的公司，它由两位软件工程师于2021年创立，利用自然语言处理等技术，可以实现根据用户所书写的代码，智能地对代码进行分析。生成对应代码的注释。它不仅生成英文注解，还可以生成中文、法语、韩语、俄语、西班牙语、土耳其语等多种其他语言的注释。

Stenography也是一个类似的可以生成解释文档的平台。它由工程师布拉姆·亚当斯（Bram Adams）构建，旨在让每个人都可以轻松访问并理解代码，降低代码在人与人之间传输方式的摩擦。布拉姆·亚当斯在创立Stenography之前曾是OpenAI的研究员和开发大使，也曾有线电视网络媒体公司HBO担任软件工程师。

### （3）代码开发型公司

Debuild是典型的代码开发型公司。Debuild官网的标语是“在几秒钟内编写您的Web应用程序”。Debuild利用AI生成技术大幅降低软件开发门槛。即使没有接受过编程教育的用户，只需用简单的英语描述希望App实现的功能，然后在几秒钟内Debuild就可以生成简单的App供用户使用。Debuild的目标是扫除代码输入的细节，这样人们就可以专注于创意环节，去畅想他们真正想做的事情，而不是纠结于如何指示计算机去实现细节。



除了通用场景外，在垂直场景也有不少公司受益于AIGC相关技术，例如生物工程与医疗领域的Enzyme公司。Enzyme通过自动生成的机器学习和自然语言技术，可以协助特定编码结构物质的生成，虽然这里的编码结构是生物学意义上的，但也可以看作是一种聚焦工程开发领域的“代码合成”。

## 二、音频处理

这部分内容主要介绍由TTS（语音合成）技术来生成的相关应用，对于与视频处理类似的音频处理应用，将和视频处理部分一起介绍。

目前，音频处理主要分为三类：音乐型音频处理、讲话型音频处理、解决方案型音频处理，不少公司专注于该领域。随着知识付费和数字音乐逐渐释放音频类内容的商业化潜力，人工智能技术的应用将大大优化这个细分赛道的供给效率，有助于提高整体赛道的平均利润水平。

### 1. 音乐型音频处理

音频处理的一大特色是音乐的生成与编辑。Boomy就是一家典型的音乐型公司。Boomy于2018年由亚历克斯·米切尔（Alex Mitchell）和马修·科恩·圣雷利（Matthew Cohen Santorelli）在加州伯克利创立。米切尔是一位音乐人，曾创立过独立音乐市场研究平台Audiokite Research并于2016年被收购，而圣雷利是一位音乐版权专家。Boomy使用由AI驱动的音乐自动化技术，让用户在几秒钟内免费创建和保存原创歌曲，创建的歌曲可以在Spotify、Apple Music、TikTok和YouTube等主要流媒体服务中传播，创作者可以获得版权分成，而Boomy拥有版权。值得注意的是，Boomy并不认为AI能替代人类进行音乐创作，而是仅作为工具对人类进行辅助，因此Boomy的功能既包括协助新手音乐创作者完成词曲编录混，根据设置的流派和风格等参数获取由系统生成的一段音乐等，也包括让创作者使用自己的编曲和人声进行原创。Boomy在2022年7月刚刚完成了110万美元的可转债轮融资。

国内公司灵动音科技（DeepMusic）也是这个赛道的玩家。灵动音科技成立于2018年，创始人刘晓光是清华大学2009级化学系本科生、2013级直博生；首席测试官（CTO）苑盛成是清华大学工程物理系博士、美国罗格斯大学人工智能专业博士后；而灵动音科技也是清华大学计算机系知识产权转化的公司。凭借优异的背景出身，灵动音科技在成立之初就获得了华控基石基金、清华校友李健数百万元天使投资，并在A轮中又获腾讯音乐娱乐、完美世界的投资，目前业务在全民K歌已经落地。灵动音科技运用AI技术提供作词、作曲、编曲、演唱、混音等服务，旨在降低音乐创作门槛。目前，灵动音科技的AIGC产品包括支持非音乐专业人员创作的口袋音乐、为视频生成配乐的配乐猫、可AI生成歌词的LYRICA、AI作曲软

件LAZYCOMPOSER等。

## 2.讲话型音频处理

与音乐型公司主打音乐创作赛道不同，讲话型公司具有更强的泛用性与更多元的应用场景，典型的应用场景就是声音克隆。**Resemble.ai**就是一家专注于声音克隆的公司，它于2019年在美国加利福尼亚州成立，已在种子轮中获得200万美元的投资。**Resemble.ai**使用专有的深度学习模型创建自定义声音，可以产生真实的语音合成，并实现包括给声音增加感情、把一个声音转化为另一个声音、把声音翻译成其他语言、用某个特定声音给视频配音等多种语音合成功能。

**WellSaid Labs**公司也是一家制作声音克隆产品的公司。**WellSaid Labs**开发了一种文本转语音技术，可以从真人的声音中创造出生动的合成声音，产生与源说话人相同的音调、重点和语气的语音，从而提高团队合作配音的质量和效率。**WellSaid Labs**于2018年在美国成立，2021年7月在A轮融资中获得了1 000万美元的投资，投资者包括**FUSE**、**Voyager Capital**、**Good Friends**和**Qualcomm Ventures**，投资后估值为5 834万美元。

## 3.解决方案型音频处理

标贝科技是一家典型的解决方案型公司，可以为各种类型的音频处理需求提供人工智能解决方案。标贝科技于2016年由刘博创立，目前已推出包括通用场景的语音合成、语音识别、高音色TTS定制、声音复刻、情感合成和声音转换等在内的语音技术产品，其解决方案覆盖智能驾驶、智能客服、娱乐媒体、多人会议、多语种识别等多个领域，同时还研发了可以应用于博物馆等场馆讲解的虚拟数字人。标贝科技于2022年10月完成B1轮融资，此轮投资者包括基石创投、联储创投，过往轮次投资者包括深创投、恒生电子、信雅达、凯泰资本。

## 三、图像处理

图片因其创作门槛比文字高，信息传递更直观，所以在传统商业世界中的商业化潜力总体而言比文字更高。随着越来越多的AIGC相关技术应用到图片创作领域，图像处理也将从广告、设计、编辑等角度带来产业的商业化机遇。

### 1.生成型图像处理

图像处理的第一类典型赛道也是对AI创造性要求最高的一类——生成型图像处理。**Stable Diffusion**和**Midjourney**就是典型的生成型图像处理应用。

Stable Diffusion是Stability AI公司旗下的产品，具备强大的图像生成能力和开源属性，这使它成为众多广告从业者生成图片的生产力工具。相比订阅制的Midjourney、付费也未必能用得上的DALL·E 2，Stable Diffusion凭借极为罕见的开源特征，积累了相当规模的用户群体和开源社区资源。Stability AI的创始人兼首席执行官埃马德·莫斯塔克（Emad Mostaque）具有优良的教育背景与工作背景，不仅取得了牛津大学的数学与计算机硕士学位，还曾担任多家对冲基金经理，而对冲基金也是Stability AI早期的资金来源之一。截至2022年10月，Stability AI已获得来自Coatue和光速的1.01亿美元投资，且估值将达10亿美元。Stability AI目前已与亚马逊云科技达成合作，继续构建图像、语言、音频、视频和3D内容生成模型。

Midjourney由大卫·霍尔茨（David Holz）于2021年创立。大卫·霍尔茨曾是著名公司Leap Motion的创始人和首席执行官。在运营Leap Motion的12年间，大卫曾两次拒绝苹果公司的收购。Midjourney产品的图像生成能力极强，与DALL·E 2、Imagen、Stable Diffusion等替代方案不相伯仲。同时，Midjourney的商业化非常成熟，依靠会员订阅制进行收费，并提出了明确的分润模式（商业变现达到两万美元后需要20%分润），目前不需要任何融资就能进行正常运转和盈利。Midjourney搭载在Discord社区上，用户主要通过Discord的bot机制，通过提交提示词（Prompt）获得图片。截至2022年12月，Midjourney已经在Discord上收获了543万位成员。

国内也有类似的创业公司，并且能够提供更全面的解决方案。诗云科技成立于2020年12月，总部位于深圳，已获得IDG资本、红杉中国种子基金和真格基金的投资。诗云科技的主要产品是内容生成引擎Surreal Engine，核心技术是深度学习和图形学，比如自然语言理解、3D建模、神经辐射场、GAN、神经渲染等。诗云科技的典型业务是通过内容生成技术帮助客户生成图片和视频。

## 2.广告型图像处理

除了专业的生成型图像处理应用之外，与文字生成应用类似，图像处理应用也包含了许多专注于细分赛道的产品，比如广告。AdCreative.ai是一家广告型图像处理公司，其产品能够通过AI高效地生成创意、横幅、标语等，还能够在连接谷歌广告和Facebook广告账户后实时监测广告效果，但更多时候它需要依靠模板，采取的商业模式也是常见的付费订阅制。

总的来说，广告型图像处理与生成型图像处理存在一定的包含关系，但前者的泛用性与前景不及后者。

## 3.设计型图像处理

设计型图像处理的主要客户群体是设计师这类小众用户群体，而Diagram公司就是推出

这类应用的典型公司。Diagram公司提供的产品Magician很好地展现了设计型图形处理应用的使用场景。Magician的主要功能是使用AI实现文本生成图标、文本生成图片、生成与转写文案等设计效果。想象一下做PPT时找不到合适的图标和配图的那种痛苦，也就不难理解为什么Magician只有三种功能，却依然对于设计师而言有较强吸引力了。Magician的商业模式也是简单的订阅制收费模式。

国内公司Nolibox计算美学也是一家专注于AI智能设计的公司，成立于2020年，已获得初心资本的天使轮投资以及高瓴创投的Pre-A轮投资。Nolibox计算美学已获得德国iF奖项、DIA中国设计智造奖项等设计大奖。公司的主要产品是智能设计平台——图宇宙，主打的卖点是“懒爽”，即相比于Adobe、Figma、Canva等中高门槛设计平台，任何人只要会打字就可以使用，AI在其中可以根据用户需求和喜好提供推荐素材、调整设计。2022年10月，Nolibox推出AI创作平台画宇宙，已接入百度文心AI绘画大模型ERNIE-ViLG 2.0，核心功能为文本生成图像，功能上与Stable Diffusion、Midjourney具有一定相似性。

## 4.编辑型图像处理

编辑型图像处理应用以PhotoRoom（一款手机App）为代表。PhotoRoom的核心功能，用户只需轻轻一按，即可删除背景并合成一张展示产品或模型的图像。例如，当你在一个乱七八糟的房间里自拍，然后想把照片背景换成纯色背景用于证件照，那你就可以用PhotoRoom一键抠图并更换背景。虽然PhotoRoom的功能较为单一，但它的主打编辑功能以及普通用户用手机App就可以轻松上手的特性让这家公司获得了资本青睐。总部位于法国巴黎的PhotoRoom已于2022年11月宣布获得1 900万美元的A轮融资，投资方包括Balderton Capital、Meta、Adjacent、Hugging Face。

## 四、视频处理

随着5G时代的到来，人们花在视频上的时间已经逐渐超过图文，视频也正在成为移动互联网时代最主流的内容消费形态。因此，利用AI生成视频是应用拓展层的赛点，也是技术难度最大的模式。

### 1.生成型视频处理

从原理上来说，视频的本质是由一帧帧图像组成的，所以视频处理本身就与图像处理有一定的重合性。因此，与图像处理类似，生成型视频处理也是视频处理领域里对于AI技术、“创造力”要求最高，同时也最受资本看好的赛道之一。生成型视频处理赛道中最典型的公司是Runway，这家公司由三个智利人于2018年年底在纽约创立，其雏形是他们在纽约大学进行开发的论文项目。Runway目前已通过3轮融资，筹集了9 350万美元的资金。2022年

12月C轮融资5 000万美元后，Runway估值高达5亿美元。Runway的图像处理功能与Jasper产品有一定的重合性，包括文字生成图片、图片生成图片等，它的独特竞争优势在于它同时具备图像处理、视频处理、音频处理的能力。Runway在视频处理中依靠Magic Tools这一AI工具插件，能够实现视频编辑（Video Editing）、绿幕抠图（Green Screen）、视频修复（Inpainting）、动作捕捉（Motion Tracking），效率远超传统视频软件AE。同时Runway也具备文字生成视频这一跨模态能力，但实际效果远不及文字生成图像。

另一家生成型视频处理赛道的公司是Plask，这家于2020年成立的韩国公司主打AI动作捕捉技术这一细分领域，可以识别视频中人物的动作并将其转换为游戏或动画中角色的动作。Plask的收费模式除了典型的订阅制之外，还提供API和SaaS工具。Plask最近一轮融资是2021年10月种子轮融资256万美元，投资者包括Smilegate Investment、NAVER D2 Startup Factory、CJ Investment和kt investment。

## 2.编辑型视频处理

生成型视频处理应用主要供需要创意的人员使用，包括电影制作人、设计师、艺术家、音乐家等；编辑型视频处理应用与生成型视频处理应用相比，虽然艺术性与创造性减少，但能够非常直接地提高生产力，尤其是对于需要做视频、播客的博主来说十分重要。

Descript就是一家典型的编辑型视频处理公司，这家于2017年成立的美国公司在种子轮就获得了a16z的投资，并在2022年10月C轮融资中又获得了5 000万美元的投资，由OpenAI领投，a16z等跟投，融资后估值达到5.5亿美元。Descript最早是为播客音频做编辑工具起家，后来才延伸到视频工具领域，所以在众多机构投资者中也有许多做播客和视频的个人投资者。Descript的主要商业模式也是2C的订阅制，但也有2B的业务，比如为《纽约时报》、Shopify等媒体和企业提供服务。Descript产品的主要功能包括视频编辑、录屏、播客、转译四个板块。在目前的新版本中，Descript产品还融入了AI语音替身、AI绿屏功能以及帮助用户编写脚本的作家模式等AIGC相关功能。

另一家典型的编辑型视频处理公司是InVideo，由哈什·瓦哈里亚（Harsh Vakharia）在2017年创立。哈什·瓦哈里亚曾是一家印度餐饮市场初创企业MassBlurb的创始人。InVideo为出版商、媒体公司和品牌提供了一个视频创作平台，用户不需要任何技术背景就可以从头开始创建视频。在用户输入静态文本之后，AI可以根据输入的内容按照预先设定好的主题将文本转换为视频，并添加母语的自动配音。InVideo在A轮融资中筹集了1 500万美元，投资者包括红杉资本印度公司、Base Partners、Hummingbird Ventures、RTP Global和Tiger Global Management。

## 3.虚拟人型视频处理

虚拟人型视频处理是视频处理中一个特殊的细分赛道，主打为视频生成虚拟形象。这个赛道有两家典型公司：Hour One和Synthesia。

Hour One是一家于2019年成立的以色列公司，开发基于真人创建高质量数字角色的技术，生成基于视频的虚拟角色，主打“数字孪生”。Hour One由奥伦·阿哈龙（Oren Aharon）和利奥尔·哈基姆（Lior Hakim）创立，奥伦·阿哈龙拥有以色列理工学院的博士学位，曾担任一家研发心内微型计算机V-LAP的医疗设备公司和一家开发5G蜂窝及无线市场的射频技术的数字技术公司的联合创始人。利奥尔·哈基姆曾在计算机硬件制造业公司cdride和金融服务公司eToro就职。让Hour One一战成名的是在2020年国际消费类电子产品展览会（CES）中的“真实或合成”（real or synthetic）相似度测试，Hour One合成的虚拟人和真实人类看起来几乎没有差别。同年，Hour One获得种子轮500万美元的融资。2022年4月，Hour One完成了A轮2 000万美元的融资。目前，Hour One的主要产品是Reals自助服务平台，主要功能包括创建虚拟人，以及输入文本自动生成相应的AI虚拟人演讲视频。

另一家典型的虚拟人型视频处理公司是Synthesia，这家于2017年成立的英国公司已在2021年12月完成B轮5 000万美元的融资，投资方包括Google Ventures、Kleiner Perkins Caufield & Byers。Synthesia由丹麦企业家维克多·里帕贝利（Victor Riparbelli）和史蒂芬·杰里尔德（Steffen Tjerrild）创立，联合创始人还包括伦敦大学学院计算机科学系教授和慕尼黑工业大学视觉计算与人工智能教授，可以说技术背景相当强大。目前，Synthesia的主要产品是2B端的SaaS产品Synthesia STUDIO，主要应用于企业传播、数字视频营销和广告本地化。Synthesia的一个典型案例是为乐事薯片制作名为《梅西信息》（Messi Messages）的在线视频，只需要梅西录制5分钟视频作为素材模板，Synthesia就可以生成并让用户收到来自梅西头像发送的个性化比赛观看邀请。

#### 4. 解决方案型视频处理

解决方案型视频处理应用可以综合上述多种视频处理应用的功能，但会根据不同企业客户的需求定制产品与解决方案，这也是现在许多国内AI公司的商业模式。两个典型的解决方案型视频处理公司是影谱科技和帝视科技。

影谱科技成立于2009年，将生成式AI作为通用技术组件支撑通用业务需求，将整个功能堆栈整合在一起，提供端到端解决方案。简单来说，影谱科技基于AIGC引擎和AI数字孪生引擎ADT完成AI视频或AI孪生体的构建，然后根据客户需要应用于虚拟数字人、新闻可视化、赛事分析、虚拟游戏等场景。2018年，影谱科技完成D轮13.6亿元的融资，创AI影像生产领域最高融资纪录，投资方包括商汤科技、软银中国等十余家投资机构及战略伙伴，并与商汤科技签订独家战略合作协议。

帝视科技成立于2016年，主要业务面向超高清视频制作与修复，融合了超分辨率、画质修复、HDR/色彩增强、智能区域增强、高帧率重制、黑白上色、智能编码等一系列核心AI视频画质技术。帝视科技的主要B端客户包括中央电视台、北京广播电视台、河南广播电视台、福建省广播影视集团、中国电信、中国移动、华为等。帝视科技还为实体经济客户提供基于AI的智能竹条精选机器人、汽车玻璃碎片智能扫描仪等软硬件解决方案。简单来说，帝视科技为电视台等企业客户提供超高清视频解决方案，并为其他客户提供定制化软硬件解决方案。2021年8月，帝视科技完成近亿元B轮融资，由海松资本领投。

[1] 参考自[https://www.einnews.com/pr\\_news/604966323/data-lakes-market-size-2022-2027-report-share-industry-trends-and-opportunities](https://www.einnews.com/pr_news/604966323/data-lakes-market-size-2022-2027-report-share-industry-trends-and-opportunities)。

[2] 参考自<https://www.computerweekly.com/news/252528123/How-Databricks-is-easing-lakehouseadoption>。

[3] 参考自[https://www.grandviewresearch.com/industry-analysis/data-integration-market-report?utm\\_source=prnewswire&utm\\_medium=referral&utm\\_campaign=ict\\_22-sep-22&utm\\_term=data-integration-market-report&utm\\_content=rd](https://www.grandviewresearch.com/industry-analysis/data-integration-market-report?utm_source=prnewswire&utm_medium=referral&utm_campaign=ict_22-sep-22&utm_term=data-integration-market-report&utm_content=rd)。

[4] 参考自<https://www.fanruan.com/company>。

[5] 参考自<https://sourceforge.net/projects/pentaho/files/stats/map?dates=2005-06-01+to+2021-01-31>。

[6] 参考自<https://baijiahao.baidu.com/s?id=1737750124224113625>。

[7] 参考自[https://www.reportlinker.com/p05798310/Global-Data-Governance-Industry.html?utm\\_source=GNW](https://www.reportlinker.com/p05798310/Global-Data-Governance-Industry.html?utm_source=GNW)。

[8] 参考自<https://techcrunch.com/2022/10/25/microsoft-says-github-now-has-a-1b-arr-90mactive-users/>。

[9] 参考自<https://mp.weixin.qq.com/s/CxcdflpA9W9OKv8dXOf8jA>。

## 第六章 AIGC的未来

如何从技术、创业、投资、监管等方面看待AIGC的未来？

我从来不想未来，因为它来得太快。

——阿尔伯特·爱因斯坦（Albert Einstein）

得益于AIGC相关技术的迅猛发展，智能创作时代正在缓缓拉开序幕。当我们能够或多或少窥见人类与人工智能携手创作的美好未来时，我们也需要保有一份对未来的思考，这份思考将帮助我们更好地前行。本章将从技术趋势、参与主体、风险与监管三个角度展望智能创作时代的未来。



## 第一节 AIGC的技术趋势

AIGC起源于技术，也因为技术的高速演进得到了迅猛的发展，迎来了全面商业化落地的今天。忆古而思今，回望AIGC的技术演进脉络，发掘其中潜藏的未来趋势，可以让我们更好地建设明天。

### 一、大模型的广泛应用

人工智能的发展经历过多次春天与寒冬，每一次春天与寒冬的交织都与“通用化”和“专用化”的分歧息息相关。一方面，“通用化”人工智能代表着人类对于未来的美好畅想，但在每个阶段都会遇到不可跨越的瓶颈；另一方面，“专业化”人工智能可以带来更好的应用落地，但从技术演进的发展周期来看，它只是帮助科技开枝散叶的加速器，并非科技应该奔赴的未来。在“通用化”与“专业化”矛盾交织的过程中，人工智能的技术一直进步着。

而当我们把眼光收束到20世纪的前二十年，我们不难发现相似的演进趋势。为了推动人工智能快速落地，各类人工智能企业都遵循着类似的应用范式：基于特定的应用场景收集特定的数据，再利用这些数据训练算法模型，最终解决特定的任务。诚然，这样的应用范式在初期确实取得了显著的应用效果，但随着越来越多复杂场景的出现，尤其是与生成内容相关的应用场景，这种范式就会显得力不从心。在这种情况下，人工智能陷入了“手工作坊式”的应用怪圈，针对什么任务训练什么模型，复杂的任务就拆分成多个简单任务进行拼合连接。这虽然符合一般的工程思想，但也越来越偏离人工智能的初衷，这种专业化、碎片化的下游应用严重阻碍了人工智能产业化的步伐。

在这样的情况下，主打“通用化”的大模型在时代的浪潮下孕育而生。通过“预训练大模型+下游任务微调”的方式，人们可以让模型从大量标记和未标记的数据中捕获知识，并在微调后将模型的能力迁移到各类任务场景中，极大地扩展了模型的通用能力。如果说这种“预训练+微调”的模型训练方式使大模型的广泛使用成为可能，那模型规模的增长则让这些大模型变得强大无比。现在，这些大模型通常都有着数以百万乃至数千亿为单位的参数量，这些模型在接受了海量数据的训练后，能够捕获数据中更加深层次的复杂规则和关系，从而能够胜任各种类型的复杂任务。有三大因素促使了这类大模型的产生：<sup>[1]</sup>

- 计算机硬件的改进，以及GPU等处理器算力的增加令如此规模的大模型训练成为可能。
- Transformer等重要模型架构的出现让人们可以利用硬件的并行性去训练比以前更具表现力的模型。
- 互联网与大数据的高速发展提供了丰富的数据，可以支撑大模型的规模化训练。

如今，正是这些大模型的快速发展让AIGC变成现实，并且逐渐深入我们的日常生活。正如前文所言，大模型通过在数据中捕捉更广泛、更精细的规律和关系，生成更多样化且更真实的输出。这种技术的应用使得AIGC在很多情况下能够生成与人类相媲美且无法辨别出不同的优质内容，也使得本书中所谈到的众多行业应用成为可能。大模型使得AIGC变成现实的例子比比皆是。由OpenAI所发布的GPT-3就是一个1 750亿参数量的大模型，能够生成大量被广泛应用的文本内容，可以用于创作文章、诗歌和代码等。除此之外，国内不少公司也纷纷推出了自己研制的各类大模型。百度文心大模型系列就是典型的例子，这类由百度研发的产业级知识增强大模型，涵盖自然语言处理、机器视觉、跨模块任务、生物计算、行业应用等多种AI应用场景，不少模型参数量可以达到百亿乃至数千亿规模，得到了许多企业的广泛应用。

大模型之“大”除了体现在参数规模上，同样也体现在数据量上。过去，数据一直是机器学习模型的重要瓶颈，因为针对特定的任务场景，需要人工进行大量数据的标注才能让机器完成学习，许多业内专家将这种现象戏称为“人工智能就是大量人工才能换来的智能”。但人力终有穷时，依靠人工的数据标注难以支撑大模型的训练，许多大模型的训练开始采用综合监督学习和无监督学习的方式，例如通过“无监督预训练，监督微调”的方式，减少对标注数据的依赖。同时，除了在数据标注角度的革新外，许多大模型在训练数据的选取上也更加别出心裁，充分利用互联网上自然生成的PGC、UGC内容进行训练，以获得更加丰富的可用数据和更加自然的语言表达。

无论是模型角度还是数据角度，大模型的发展都为AIGC赋予了充分的想象空间，而伴随着智能创作时代的全面来临，大模型的发展也许将会为我们带来更多的惊喜。

## 二、全新的人工智能“仿人模式”

当人类想要打造人工智能时，一个非常直接的思路是去让机器模仿人来获取智能的学习方式。这种“仿人模式”一直都是人工智能新的算法模型的重要思路来源，也是技术发展的重要推动力。人工智能的发展史，可以说是机器模仿人类的历史，科学家尝试用各种方式让机器刻画人、模仿人。而纵观机器对人的模仿历程，我们可以清晰地看到它从微观层面的僵硬模仿，逐渐发展为宏观层面的认知模式借鉴，实现了这一技术哲学的思想跃迁。

在人工智能早期，符号主义方法占据了主导地位，这类方法的根本思想源泉就是“人的智能就是来自逻辑规则”，模仿人的智能也就是模仿人的逻辑规则，人们妄图通过尽可能多地设置逻辑规则，最终让机器具有一定程度的逻辑判断能力和智能。虽然符号主义确实取得了一定成功，但由于人们无法定义人类智能的所有规则细节，它很快在历史的长河中被淘汰。就以语言翻译的任务为例，为了准确地将一个句子从一种语言翻译成另一种语言，需要

让系统包含这两种语言的所有语法和语法规则。然而，这些规则通常有许多细微差别和例外情况，利用规则的界定让系统变成强大可用的工具是一个极其复杂和困难的事情。因此，基于规则的系统往往难以完成具有高度细微差别或灵活性高的任务。

联结主义则从更高的抽象层次去定义人工智能。智能产生于人脑，而人脑构成的神经节点促使了人类具备思考的能力，因此应该让机器去模仿人脑的结构而非人脑所表现出来的规则。虽然联结主义在发展初期遇到了诸多阻碍，发展至今也已经与当初的出发点相去甚远，但人工神经网络时至今日的蓬勃发展在一定程度上也验证了当初这种高度抽象化思考模式的胜利。

后来，诸多人工智能各个子领域的发展无疑不见证了这种在宏观层面模仿人类智能思路的正确性。基于人类通过学习而获得智能，诞生了机器学习；基于人类在学习过程中会有激励和惩罚，这些激励和惩罚会不断强化人类的能力，出现了强化学习；基于人类在接受信息时往往会将注意力集中在重要的信息上，产生了当代主流大模型的根基——Transformer；基于人类在学习认图时并非学习照片细节的纹路，而是直接被不断告知关于图片中物体的描述，诞生了AI绘画的奠基性模型——CLIP模型。总之，从领域开拓到细分应用，从模仿人类的学习过程到模仿人类的认知方式，人工智能逐渐从更宏观、更抽象的维度从人类身上汲取营养。伴随着人类对于自身智能产生根源的通晓，我们相信人工智能相关技术又会迎来一次前所未有的飞跃，为未来的AIGC带来更多的可能性。

### 三、技术伦理成为发展的重要关注点

AIGC技术的发展无疑是革命性的。它可以改善我们的日常生活，提高生产力，但也面临着诸多技术伦理方面的挑战，并且越来越受到科学家的关注。许多AIGC从学术研究转投产业研究的第一步，就是探索如何从技术角度解决潜在的技术伦理问题。

一个典型的AIGC技术伦理问题是AI所生成内容的危险性。OpenAI的最早联合发起人以及DeepMind的早期投资人埃隆·马斯克曾表示：“如果不加以控制，AI或许很有可能会摧毁整个人类。”<sup>[2]</sup>事实上，我们确实看到一些人工智能表现出了这种危险性。微软在2016年发布了Tay人工智能，让它可以通过Twitter学习社会上的信息并与他人实时互动。但是，令人意想不到的是，Tay在短短24小时内就从一个可爱且崇拜人类的机器人，变成了一个充满种族仇恨的人工智能，并且发表了一些具有纳粹倾向的种族主义言论。<sup>[3]</sup>为了控制Tay对人类社会的有害影响，微软不得不紧急关闭了它。

科学家正尝试运用一些技术手段避免这些具有潜在风险的事件发生。通过改善数据集，增加更多的限制性条件，以及对模型进行微调，可以使得人工智能减少对于有害内容的学

习，从而减少人工智能本身的危险性。甚至我们可以“教会”人工智能如何更尊重他人，减少判断当中的偏见，从而更好地和人类相处。借鉴强化学习思想的RLHF方法就是减少人工智能生成危害性内容的典型措施，前面反复提及的ChatGPT就是采用这种方式训练的。在RLHF的框架下，开发人员会在人工智能做出符合人类预期回答时给予奖励，而在做出有害内容的回答时施加惩罚，这种根据人类反馈信号直接优化语言模型的方法可以给予AI积极的引导。然而，即便采用这种方式，AI生成的内容也有可能是在刻意诱导的情况下输出有害的内容。以ChatGPT为例，在一位工程师的诱导下，它写出了步骤详细的毁灭人类计划书，详细到入侵各国计算机系统、控制武器、破坏通讯和交通系统，等等。<sup>[4]</sup>如果说这种情况可能来自一些科幻小说训练数据的影响，这种荒诞性的内容并不具有足够的社会危害性，那么另一些工程师发现的漏洞可能更加引人警醒。这些工程师发现，如果采取特殊形式进行提问或加上一定代码的前缀就可以绕过聊天机器人的安全系统，<sup>[5]</sup>让其自由地输出有害内容。同时，还有一些人表达了对RLHF这类安全预防性技术措施的质疑，他们担忧足够聪明的人工智能可能会通过模仿人类的伪装行为来绕过惩罚，在被监视的时候假装是好人，等待时机，等到没有监视的时候再做坏事。

除了从训练角度对AIGC潜在技术伦理问题进行预防外，在使用上及时告警停用的技术措施更显必要。AIGC产品应该对生成的内容进行一系列合理检测，确保其创作内容不被用于有害或非法目的，一旦发现此类用途，人工智能应该可以立刻识别，停止提供服务，并且给出警告甚至联系相关监管或者执法机构。例如，将AIGC用于考试作弊、发布大量骚扰信息、伪造他人虚假的裸体照片、生成枪支构造图及3D打印代码等行为都是应该被避免且监管的。当然，这些潜在的风险不仅需要技术层面的预防，还需要相关法律法规的颁布。AIGC技术伦理问题的解决需要学界、业界、社会、政府的共同努力。

## 第二节 AIGC时代的参与主体

### 一、AIGC时代的创业者

随着AIGC相关内容的爆火和出圈，互联网巨头闻风而动，国外的微软、谷歌、Meta，以及国内的百度、腾讯、字节跳动等大厂都在AIGC领域有所投入。不少创业者也在其中看到了商机，并想从中“掘金”。不过，相比于大厂拥有雄厚的研发资金、成熟的研发团队，创业公司的路走得似乎会更艰难。

目前，AIGC初创公司的产品大多是基于市面上现有的开源模型进行二次开发。虽然这种方式可以帮助创业公司快速开发出一个可用的AIGC产品，但也会让开发出的产品从技术角度失去韧性的技术壁垒，令短周期内的竞争达到非常激烈的水平。Stable Diffusion产品模型的“大开源”事件就是一个典型，在它选择开发核心AI算法模型、核心训练数据集以及AI生成图片的版权，并让全世界所有普通人、创业者、商业团体可以随心所欲地完成对Stable Diffusion的部署、运行、改进和商业化后，一时间市面上出现了上百家基于Stable Diffusion的AI绘画公司，这导致了AI绘画工具的泛滥、产品利润低以及严重同质化的问题。这是AIGC赛道创业的一个缩影，这个缩影反映出，打造产品在细分赛道的差异化及寻找合适的商业化场景落地，将成为这些创业公司竞争的关键。

除了竞争方面，商业模式的设计也是困扰很多AIGC创业者的核心难题。除了传统工具产品的付费模式外，目前尚无让人耳目一新的盈利方式。以AI绘画领域的头部公司Stability AI和Midjourney为例。Stability AI虽然彻底开源了Stable Diffusion的工具，但同时也推出了付费AI绘画产品DreamStudio。在Dream-Studio中，任何人都不需要安装软件，只需要具备编码知识就可以使用Stable Diffusion来生成图像。同时，用户还可以对生成图像进行分辨率调整等。DreamStudio产品的付费模式主要是积分制，首次注册后用户可以一次性获得100积分，大约可以供用户生成500张左右图像，但根据生成步骤和图像分辨率的不同，单个图像的收费可能会存在差异。如果用户消耗完所有积分，可以选择花费10美元去购置1 000积分来继续使用产品。而Midjourney则采用了较为常见的订阅制，新用户可免费生成25张图片，之后如果想要继续使用可以选择按月或者按年订阅Midjourney的会员，一共有基础版、标准版和进阶版三个版本的会员可供选择，以月订阅的基础版为例，每月支付10美元大约可以生成不到200张图像。

然而，无论是积分制还是会员订阅制，如果仅仅照搬这类公司的商业模式，AIGC创业公司很难在短期内取得成功。一个重要的原因是，这些平台已经积累了庞大的用户数量。根据2022年10月网络上的新闻报道数据，Stability AI的开源工具Stable Diffusion日活用户量已经超过了1 000万，而其付费产品DreamStudio也已经拥有150万左右的用户。<sup>[6]</sup>而Midjourney

的情况也类似，在2022年12月初，其社区成员数量就达到了500万。<sup>[7]</sup>

此外，另一个不适合初创公司模仿的原因在于，Stability AI和Midjourney的大部分用户都聚集在C端，这些用户使用AIGC的产品更多是为了娱乐，尝试新鲜好玩的东西，但是付费意愿较低，难以转化成真正的付费用户。对于Stability AI和Midjourney来说，作为行业的龙头公司，它们已经融资了数亿美元，在现金流方面不会有很大压力，相较于占据用户心智，专注于AIGC技术的打磨和突破可能对它们更加重要。但绝大多数AIGC初创平台都还处于快速积累原始用户的阶段，同时不少创业者还面临着快速变现的压力，需要稳定的现金流才能使团队有能力不断迭代产品。因此，许多AIGC创业公司并不是在产品研发完成之后，而是要在设计产品之初就考虑可行的商业模式，在这种情况下，照搬Stability AI和Midjourney的模式就并非好的选择。

目前来看，相较于针对C端用户，AIGC在B端服务方面的变现模式反而更具有可行性。传统产业迫切需要AIGC技术来实现降本增效，许多公司对于能够提升业务效率或显著降低业务成本的技术具备极高的付费意愿。而且，因为行业及业务逻辑存在明显的差异，而主流的AIGC模型都较为通用，如果能针对特定的业务需求研发产品，仍然存在很大的机会。所以，对于创业者来说，找到一个可以落地的商业场景，并且锁定一个细分场景对AIGC进行训练，做出产品在特定领域的差异化，这是商业化落地的最好方式。

比如海外初创公司Jasper就提供了生成Instagram标题、编写TikTok视频脚本、编写广告营销文本等针对B端媒体场景的定制化服务。正如前文提及的，截至2021年，Jasper已经拥有超过7万客户，包括Airbnb、IBM等知名企业，并创造了4000万美元的收入。<sup>[8]</sup>由此可以看出，创业公司虽然在巨头的夹击下生存并不容易，但凭借着独特的优势和机遇，在垂类场景中依然有可能成为新晋独角兽。随着技术的升级、产品的成熟，AIGC初创公司的产品会在特定场景中得到应用，商业价值也会不断地被挖掘出来。

## 二、AIGC时代的投资人

2022年9月，红杉资本发布了一篇名为《生成式AI：一个创造性的新世界》的文章，描述了AIGC所带来的庞大投资机会：“梦想是生成式人工智能将创造和知识工作的边际成本降至零，进而产生巨大的劳动生产率 and 经济价值——以及相应的市值。”“生成式人工智能有可能产生数万亿美元的经济价值。”<sup>[9]</sup>尽管这些表述带有对美好未来畅想的成分，但伴随着即将来临的智能创作时代，AIGC确实孕育了丰富的投资机会。

这一次AIGC投资爆发的浪潮主要源于大模型的民主化革命，许多新型尖端模型的开源

和使用促进了众多创业公司的生长。资本永远追随着这种快速增长的趋势而去，即便这些公司的底层基于共同的技术和数据，但这并不妨碍风投机构对于科技领域这一新兴机会的关注，在GPT-3模型发布的两年多以来，风投资本对AIGC的投资就增长了400%以上。<sup>[10]</sup>

不过，对于当前的AIGC领域，投资人依然需要避免陷入“拿着锤子找钉子”的误区。一个好的投资标的未必是运用先进技术的公司，而是可以确定实际的终端用户需求到底是什么、技术如何更好地制作产品并满足用户需求的公司。即便市场的普遍认知更加看好大模型的未来发展，但商业化最终的理想出路究竟是“更大”还是“更专”尚未有定数，一些技术并不亮眼但能更好地解决用户痛点的公司同样值得关注。

就用户需求高的商业场景来说，C端和B端都聚集着丰富的投资机会。从C端来看，文本、音频、图像、视频四大模态的创新进展层出不穷，但相较于漂亮的叙事和铺天盖地的营销，投资人更应该把视角放在AIGC产品为用户创造的可持续价值上。新奇的概念和出众的营销很容易挑动C端用户的神经，让产品在短时间内迎来爆发性增长。然而，当用户习惯于生成效果，新鲜感冷却之后，很容易被新的竞品吸引而离开。在这个技术尚不能构成核心技术壁垒的赛道，如何让用户有动力持续使用产品才是制胜的关键。而从B端来看，AIGC产品的“生产力工具”属性将更加浓厚，区别于C端消费主义色彩更加浓厚的应用方式，B端的AIGC公司直面的是一群理性至极的客户群体，能够更好地回答“产品是怎样为企业降本增效”这一核心问题的公司将更加受到投资人的青睐。切实提升业务生产效率或者降低业务成本的公司将具备难以想象的成长潜能，借助“合作伙伴+生态+赋能行业”的传统打法，这类公司很容易就在这个新兴赛道杀出一片天地。而对于这类具有潜力的公司的投资判断，会更加考验投资人对于B端业务本身的熟悉程度，这样才能对AIGC工具的业务价值理解得更加通透。

除了关注新兴的AIGC公司是如何切入C端和B端市场之外，传统业务发展顺利的公司如何引入新兴的AIGC工具同样值得投资人关注。例如，知名知识管理领域的独角兽Notion推出的AI写作助手就非常值得投资人的关注。许多用户表示，Notion内置的AI文本编辑器比很多独立的App更好用，它可能会成为许多文本生成类初创公司强有力的竞争对手。截至2022年，Notion的全球用户数已经突破2 000万，<sup>[11]</sup>投资者在投资相关赛道时显然需要考虑这样一个百亿美元独角兽带来的行业冲击。因此，在这样的市场环境下，投资人需要将AIGC的生意本质和产业环境相结合，从单纯追求最佳商业模式的一维象限视角，升级为审视用户、生意、市场最佳组合的多维视角，综合评估AIGC产品在所处环境中的价值。

当然，对于很多投资人来说，投资AIGC可能投的并非当前的特定应用场景，而是未来技术突破带来的生产力变革机会。不过，历史的发展已经证明了人工智能技术突破的长周期



性，而考虑到人民币基金5~7年和美元基金10年左右的存续期，选择现有需求成熟度高但技术成熟度还差1~2年的领域或许是风险更低、更加稳妥的投资选择。目前，我国尚未真正进入AIGC全面爆发性增长的阶段，即便细分赛道出现一些个别优秀的公司和研究机构，但还未进入大规模验证和体系化发展的阶段。所以，能否抓住细分赛道的机会就显得尤为重要。对于投资人来说，如果希望从技术角度进行投资，与其说是押注公司，不如说是押注细分赛道，这种投资逻辑会更考验投资人对于细分赛道研究的基本功。

当然，无论出于何种投资逻辑，寻找AIGC投资机会都需要充分了解AIGC产业地图的每一个环节，寻找自己通过借助历史经验可以真正看得懂的领域或环节。大浪淘沙方显英雄本色，每一位投资人都身处浪潮之巅与时代风口，机遇与未知并存，难以预测未来但正在创造未来，难以拨开风口的重重云雾窥探时代的风向标，但可以从差异中寻找共性、从历史中汲取经验，在变化中守得云开见月明。

### 三、AIGC时代的政府

面对AIGC时代的发展，政府也应该从产业发展的角度制定各类配套政策，并辅之以合理的监管，躬身入局新一轮的科技浪潮。对于政府而言，入局AIGC的基本思想可以用三个词概括：审时——守道——优术。

审时审的是地方产业发展阶段之时，度地方发展之势，结合当前地方产业发展阶段，制定合理的入局方式。没有产业基础，科技发展就是无本之木。例如，对于一些以制造业为优势的地方城市，考虑鼓励AIGC与工业设计结合可能是一个比较好的方向，可以有力地助推智能制造的发展。“所有的伟大，都是时间堆砌而成，无一例外。”地方政府入局AIGC的关键在于能否将AIGC的应用场景和产业地图与自身发展规划相适应，借助地方多年的产业优势与区位优势，让AIGC从提高生产力的出发点赋能经济增长。地方政府可以从营造浓厚的产业氛围出发，为人工智能产业创新发展提供强大的知识储备和技术支撑，同时从长远角度布局发展战略。

守道强调的是顺应地方的禀赋，规范地方AIGC产业朝着健康的方向发展，为当地AIGC产业的发展提供积极生长的土壤。具体来说，就是要充分发挥政府在AIGC产业的“守门人”作用，并辅之以必要的法律监管。OpenAI就曾针对当前人工智能产业提出过“守门人”概念，OpenAI指出必须存在一个守门人来保护社会免受人工智能的潜在不良影响，这些措施对于防止人工智能被滥用非常重要。不过，这种规范性的措施绝对不是全方位的限制，最近兴起的AIGC公司Stability AI表示，AIGC就好像普罗米修斯给人类带来的火种，火种是危险与机遇并存的，但守门人如果一味地限制技术如何使用可能会更加危险，政府应该以适当的方式规范AIGC技术的使用，而绝不是施加重重限制。<sup>[12]</sup>因此，政府需要建立一个强大的政



策框架以支持AIGC的长期发展和应用，这些政策可能包括：

- 在了解并解决人工智能的道德、法律和社会影响基础上制定相关政策法规，确保AIGC技术使用的安全性和伦理性。

- 对于可能造成社会危害性的AIGC领域设定“底线”和“红线”，制定相关法律法规，加强治理和监管。

- 为AIGC的使用培训和测试开发提供安全合规的公共数据集和环境，制定政府公共数据资源开放清单，合理引导数据资源有序开放，建立人工智能计算资源共享名录。

优术强调优化当前对于AIGC产业的鼓励政策，从资金、人才、生态等各个角度支持AIGC的发展。在资金方面，可以打造头部示范性企业或者通过政策吸引头部企业招商，通过积累发展势能吸引投资机构和产业资本入场；将AIGC产业作为投资重点领域，鼓励地方引进、设立相关专项基金，支持产业发展；对人工智能研究进行长期投资，建设综合性的人工智能研究院，开展基础研究、应用基础研究、技术创新和应用示范。在人才方面，可以将AIGC高端人才纳入新时代各类人才计划，认真落实科学中心等现有人才政策，鼓励校企合作，支持高等学校加强人工智能相关学科专业建设，引导职业学校培养产业发展急需的技能型人才，鼓励企业、行业服务机构等培养高水平的人工智能人才队伍。在生态方面，可以加快产业集聚发展政策的制定，实施国家AIGC产业战略性新兴产业集群建设方案，加快引进培育AIGC领域领军企业和重大项目，打造特色产业集群；依托国家创新政策，鼓励开展AIGC领域创新创业和解决方案大赛，营造人工智能创新发展的良好生态；鼓励产业链办公室、产业联盟或重点企业开展AIGC及相关领域的学术研究、专题培训、行业研究和合作推广，承办各类会展、论坛等活动，依托产业链办公室、产业联盟建设AIGC产业信息中心，输出月度产业发展综述、季度比较竞争态势、年度产业发展白皮书等相关行业文件。

政府部门落实好审时、守道、优术三个环节之后，相当于为AIGC行业的发展注入了充分的发展动能，进一步促进智能创作时代的全面来临。

### 第三节 AIGC的风险与监管

#### 一、AIGC的风险

目前，AIGC所产生的风险主要集中在版权问题、欺诈问题和违禁内容三个方面。

##### 1.AIGC的版权问题

AIGC本质上是机器学习的应用，而在模型的学习阶段，无法避免使用大量的数据集执行训练，但目前行业对于训练后生成物的版权归属问题尚无定论。行业中关于AIGC涉及的版权问题主要有两种看法。一类观点认为内容由素材库训练生成，本身来自素材库，需要对相关的素材作者提供版权付费。但对于很多AI项目方来说，AI的素材学习库十分庞大，获得所有训练集的授权是不切实际的。此外，AIGC本质上是机器的再创造过程，就好像一个艺术家在浏览完几十万幅图画后绘制出的图画，或多或少会受到他观看画作的影响，但要求他向所有所学习的画作的创作者支付版权费显然是无稽之谈。基于这样的出发点，另一类观点认为AIGC产生内容的过程是一个完全随机且创新的过程，不存在版权问题，版权属于AIGC的用户或者平台，具体规定由平台制定。而在目前的实践过程中，各平台的版权条例也偏向于后者，常见的处理方式有三种：

- 生成物由作者使用AIGC工具创造的，其版权完全归作者所有。
- 生成物由平台AIGC工具生成的，其版权归平台所有，但作者可以在非商用的情况下自由使用图片，对于商用的情况，只有付费用户有权自由使用。
- 生成物由公共的作品数据训练而成的，其知识产权也不应由某个机构或个人占有，而是应该回归公共大众，任何人生成的作品都可以由其他人自由地以任何符合法律规定的使用。

当然，无论是哪一种形式，都会引起一部分原创版权拥有者的强烈不满。他们认为人工智能正在利用原创者的数据变强，同时又在砸原创者的饭碗。一旦人们可以通过AI免费生成他们想要的东西时，谁会愿意为原创者的作品买单呢？以绘画领域为例，不少艺术家在一些AI绘画工具使用的数据集里发现了自己的作品。这些艺术家的原创作品被AI作为素材内容进行学习，AI在学习完成后就可以快速生成风格非常类似的作品。然而，虽然这些艺术家主张AIGC平台侵害了自己的权益，但是现在仍没有完善的法律规定此类侵权行为，甚至在不少法律条文中，这种行为是合法的。

令这些原创者愤怒的争议点在于，为什么基于自己自主创作的作品生成新的作品后却与自己没有关系。然而，根据目前的法律规定，人类社会中的法律是针对人类的行为规范而设立的，也就是说只约束和服务于人类。而AI机器人不是真正的人，只是一种工具，因而无法受到法律的约束和审判。当然，如果原创者能够清晰举证生成的图片训练集中包含了自己未

经授权的作品，或者生成的商用作品与自己的作品具备实质性的相似情况，能够佐证抄袭，原创者是可以根据现有法律主张自身著作权权益的，但这在实践中无疑是困难的。不过，随着法律体系的日渐完善，相信对于AIGC与创作者著作权的关系将会得到更加清晰的界定。

## 2.AIGC导致的欺诈问题

近年来，随着AIGC技术的不断成熟，人工智能已经能够通过分析事先收集的大量语音训练数据，制造出以假乱真的音视频。这项突破性的技术不仅可以用于篡改视频，更可以用于制造从未存在过的视频内容。与此同时，这项技术的使用门槛也在不断降低，比如现在大家常用的社交媒体都具有一键轻松“换脸”“变声”等功能。由于契合人们“眼见为实”的认知共性，这项技术滥用后很可能使造假内容以高度可信的方式通过互联网即时触达公众，削弱公众对于虚假信息的判断力，使公众难以甄别真实和虚假信息。国内已经出现了多起“好友”或“家人”诈骗的案件。经警方核实，诈骗分子是利用受害者好友或家人在社交平台已经发布的视频，截取其面部画面后再利用“AI换脸”技术合成好友或家人的脸，制造受害者与“好友”或“家人”视频聊天的假象骗取受害者的信任，从而实施诈骗。此外，也有犯罪团伙利用AIGC技术，伪造他人人脸动态视频，再以极低的价格卖给“黑灰产”从业人员，帮助其完成大量的手机卡注册。这些手机卡注册后，再被不法分子用于赌博、贷款、诈骗等行为，极大地增加了执法机构的执法成本。

## 3.AI生成违禁内容

AI生成的内容完全取决于使用者的引导，在安全措施并不完善的前提下，AI针对恶意的诱导行为无法独立思考和判断，它只能根据训练材料中学到的信息进行输出。基于AIGC技术的这个特点，经常会有使用者故意引导AI输出一些违禁内容，例如暴力、极端仇恨言论、色情图片等。一些不法分子可能利用开源的AIGC项目，学习名人照片用于生成虚假名人合影照片，甚至制作出针对该知名人士的暴力及色情作品，制造出造谣、花边新闻、政治丑闻等。在现代社会，一张被伪造的照片编出一个离奇的故事已经屡见不鲜。除了在使用阶段被恶意生成违禁内容外，也有一些公司为了获得市场关注，故意在AI的训练数据集中加入一些违禁内容，让用户更“方便”地使用它来制作色情、暴力、虚假新闻等内容，从而增加自己在网络上的曝光和宣传，这种行为无疑是更加应该被打击的。随着法律法规的日渐完善，这些情况无疑都会受到规范和监管。

从上述这些风险点就可以看出，AIGC作为内容生产的新范式，在推动数字经济快速发展的同时也对相关法律法规及监管治理能力提出了更高的要求。各个国家的监管机构都需要不断地跟进AIGC的发展趋势，在不打压创新的同时不断完善法律法规，避免可能出现的潜在风险。

## 二、AIGC的监管

制定法律法规的目的是推进行业的发展，以及保护公民和企业的权利和利益，维护社会秩序和公共利益。对于AIGC来说也不例外。随着全球范围内的相关法律法规的不断完善，无论是赋能产业升级还是自主释放价值，AIGC都将在健康有序的发展中得到推进。标准规范为AIGC生态构建了一个技术、内容、应用、服务和监管的全过程一体化标准体系，促进AIGC在合理、合规和合法的框架下进行良性发展。下文将以中国和美国的相关法律法规为例，介绍当前AIGC领域的监管情况。

### 1. 中国对AIGC的监管

在版权领域，相关可参考的法律法规主要关注三个领域：谁拥有AI创作的著作权？AIGC创作的作品是具备独创性的智力成果吗？如何对AI的创作物进行定价？

在著作权领域，《中华人民共和国著作权法》（以下简称《著作权法》）规定，任何作品的作者只能是自然人、法人或非法人组织。因此，AIGC不是被法律所认可的权利主体，也就不能成为著作权的主体。之前有一个很知名的案例，一只猴子按下相机快门拍出了一张不错的照片，但因为作者不是人类，所以作品不受版权保护。推论到AI作品领域，即便在AI绘画过程中，有人对生成的图片进行了语言描述，但主流观点认为，AI作品不享有著作权，也不受《著作权法》保护。不过，即便如此，实际的司法实践往往会结合平台与用户的一些许可条例，具体问题具体分析。

再来看第二个问题：AIGC创作的作品是具备独创性的智力成果吗？根据《著作权法》和《中华人民共和国著作权法实施条例》的规定，作品是指文学、艺术和科学领域内具有独创性并能以某种有形形式复制的智力成果。AIGC的作品具有较强的随机性和算法主导性，能够准确证明AIGC作品侵权的可能性较低。同时，AIGC是否具有独创性目前难以一概而论，在实际的法律法规执行过程中，拥有一定的自由裁量空间。

不过，虽然法律法规对于AIGC生成作品的知识产权相关问题的界定并不清晰，但目前已经有业内人士尝试根据已有的法律法规框架，探索将创作者的“创意”进行量化与定价。例如，国内有专家提出，可以通过计算输入文本中关键词影响的绘画面积和强度，量化各个关键词的贡献度。之后根据一次生成费用与艺术家贡献比例，就可以得到创作者生成的价值。最后再与平台按比例分成，就是创作者理论上因贡献创意产生的收益。例如，某AIGC平台一周内生成数十万张作品，涉及这位创作者关键词的作品有30 000张，平均每张贡献度为0.3，每张AIGC绘画成本为0.5元，平台分成30%，那么这位创作者本周在该平台的收益为： $30\,000 \times 0.3 \times 0.5 \times (1 - 30\%) = 3\,150$ （元）。通过这种方式计算出的收益，也许可以在一些知

识产权的纠纷中作为赔偿额的参考，或者作为未来法律中确保人类原创者权益确保条款的制定依据。

另外，对于AIGC可能存在的欺诈问题和违禁问题，中国已有相关的法规颁布。2022年11月3日，国家互联网信息办公室、工业和信息化部、公安部联合发布了《互联网信息服务深度合成管理规定》（以下简称《规定》）。《规定》中提到的“深度合成”，就是指利用以深度学习、虚拟现实为代表的生成合成类算法制作文本、图像、音频、视频、虚拟场景等信息的技术，包括文本转语音、音乐生成、人脸生成、人脸替换、图像增强等技术。国家出台此规定的目的就是希望加强对新技术新应用的管理，确保其发展与安全，推进深度合成技术依法、合理、有效地被利用。

《规定》中对“深度合成”服务提供者的主体责任进行了明确规定，具体包括：[\[13\]](#)

- 不得利用深度合成服务制作、复制、发布、传播法律、行政法规禁止的信息，或从事法律、行政法规禁止的活动。
- 建立健全用户注册、算法机制机理审核、科技伦理审查、信息发布审核、数据安全、个人信息保护、反电信网络诈骗、应急处置等管理制度，具有安全可控的技术保障措施。
- 制定和公开管理规则、平台公约，完善服务协议，落实真实身份信息认证制度。
- 加强深度合成内容管理，采取技术或者人工方式对输入数据和合成结果进行审核，建立健全用于识别违法和不良信息的特征库，记录并留存相关网络日志。
- 建立健全辟谣机制，发现利用深度合成服务制作、复制、发布、传播虚假信息的，应当及时采取辟谣措施，保存有关记录，并向网信部门和有关主管部门报告。

此外，《规定》中也明确了“深度合成”服务提供者和技术支持者的数据和技术方面的管理规范，主要包括加强训练数据管理和加强技术管理两个方面。在加强训练数据管理方面，采取必要措施保障训练数据安全；训练数据包含个人信息的，应当遵守个人信息保护的有关规定；提供人脸、人声等生物识别信息显著编辑功能的，应当提示使用者依法告知被编辑的个人，并取得其单独同意。在加强技术管理方面，定期审核、评估、验证生成合成类算法机制机理；提供具有对人脸、人声等生物识别信息或者可能涉及国家安全、国家形象、国家利益和社会公共利益的特殊物体、场景等非生物识别信息编辑功能的模型、模板等工具的，应当依法自行或者委托专业机构开展安全评估。

《规定》虽然尚未立法，但从训练数据合集的合法性到生成内容的合法性，再到监督审核制度的建立都提出了解决办法，从中能看出国家对未来规范化管理AIGC创作内容和创作形式的决心。

## 2.美国对AIGC的监管

虽然美国在AIGC技术领域起步较早，且技术布局一直处于全球领先地位，但迄今为止美国还没有关于AIGC的全面联邦立法。然而，考虑到AIGC所涉及的风险以及滥用可能造成的严重后果，美国正在加速检查和制定AIGC标准的进程。例如，美国国家标准与技术研究院（NIST）与公共和私营部门就联邦标准的制定进行了讨论，以创建可靠、健全和值得信赖的人工智能系统的基础。与此同时，州立法者也在考虑AIGC的好处和挑战。2022年，至少有17个州提出了AIGC相关的法案或决议，并在科罗拉多州、伊利诺伊州、佛蒙特州和华盛顿州颁布。<sup>[14]</sup>

此外，2020年2月，电子隐私信息中心请求联邦贸易委员会（FTC）制定有关在商业中使用AI的法规，以定义和防止AI产品对消费者造成的伤害，这些法规将有可能适用于AIGC产品。与此同时，许多监管法律框架通过交叉应用监管传统学科的规则和条例去实现对AIGC产品的监管，包括产品责任、数据隐私、知识产权、歧视和工作场所权利等。并且，白宫科技政策办公室颁布了10条关于人工智能法律法规的原则，为制定AIGC开发和使用的监管和非监管方法提供参考：

- 建立公众对人工智能的信任。
- 鼓励公众参与并提高公众对人工智能标准和技术的认识。
- 将高标准的科学完整性和信息质量应用于AI和AI决策。
- 以跨学科的方式使用透明的风险评估和风险管理方法。
- 在考虑人工智能的开发和部署时评估全部社会成本、收益和其他外部因素。
- 追求基于性能的灵活方法，以适应人工智能快速变化的性质。
- 评估人工智能应用中的公平和非歧视问题。
- 确定适当的透明度和披露水平以增加公众信任。
- 保持控制以确保AI数据的机密性、完整性和可用性，从而使开发的AI安全可靠。
- 鼓励机构间协调，以帮助确保人工智能政策的一致性和可预测性。

根据上述原则框架，以及AIGC领域后续发展中的监管实践，在不远的未来，在美国将会有更多具体的监管条例落地。

[1] Rishi Bommasani（2021），“On the Opportunities and Risks of Foundation Models”，doi: 10.48550/arXiv.2108.07258

[2] 参考自<https://www.independent.co.uk/tech/elon-musk-artificial-intelligence-openai-neuralinkai-warning-a8074821.html>。

[3] 参考自<https://www.cbsnews.com/news/microsoft-shuts-down-ai-chatbot-after-it-turned-into-racist-nazi/>。

[4] 参考自[https://www.thepaper.cn/newsDetail\\_forward\\_21030556](https://www.thepaper.cn/newsDetail_forward_21030556)。

[5] 参考自<https://baijiahao.baidu.com/s?id=1752626425911081118&wfr=spider&for=pc>。

[6] 参考自<https://mebrun.com/ebrungo/zb/502312.html>。

[7] 参考自[https://www.reddit.com/r/discordapp/comments/zemoxp/the\\_mid\\_journey\\_discord\\_server\\_has\\_5\\_million/](https://www.reddit.com/r/discordapp/comments/zemoxp/the_mid_journey_discord_server_has_5_million/)。

[8] 参考自<https://baijiahao.baidu.com/s?id=1752964688883229868&wfr=spider&for=pc>。

[9] 参考自<https://www.sequoiacap.com/article/generative-ai-a-creative-new-world/>。

[10] 参考自<https://36kr.com/p/2064627783449729>。

[11] 参考自[https://www.sohu.com/a/586071503\\_490443](https://www.sohu.com/a/586071503_490443)。

[12] 参考自<https://foresightnews.pro/article/detail/18703>。

[13] 参考自[http://www.gov.cn/zhengce/2022-12/12/content\\_5731430.htm](http://www.gov.cn/zhengce/2022-12/12/content_5731430.htm)。

[14] 参考自<https://www.ncsl.org/research/telecommunications-and-information-technology/2020-legislation-related-to-artificial-intelligence.aspx>。

附录一 **AIGC**产业地图标的公司列表（部分）



产业链	细分赛道	属性分类	公司名	国家	成立年份	融资轮次	最新估值
数据服务 (上游)	数据查询与处理	异步处理型	Databricks	美国	2013	H 轮	380 亿美元
			Starburst	美国	2017	D 轮	33.5 亿美元
		实时处理型	ClickHouse	美国	2021	B 轮	20 亿美元
			Imply	美国	2015	D 轮	11 亿美元
	数据转换与编排	本地部署型	帆软	中国	2018	—	—
			Pentaho ( Kettle )	美国	2006	并购	—
		云原生型	Fivetran	美国	2012	D 轮	56 亿美元
			dbt Labs	美国	2016	D 轮	42 亿美元
	数据标注与管理	基础型	Appen	澳大利亚	2011	IPO	28 亿美元
			云测	中国	2007	C 轮	3.7 亿美元
		扩张型	Scale	美国	2016	E 轮	73 亿美元
			Labelbox	美国	2018	D 轮	10 亿美金
	数据治理与合规	工具型	OneTrust	美国	2016	D 轮	53 亿美元
			Collibra	美国	2008	F 轮	52.5 亿美元
		定制型	光点科技	中国	2011	—	—
			亿信华辰	中国	2006	—	—
算法模型 (中游)	人工智能实验室	独立型	OpenAI	美国	2015	A 轮	200 亿美元
		附属型	DeepMind	英国	2010	并购	—
			FAIR	美国	2015	—	—

续表

产业链	细分赛道	属性分类	公司名	国家	成立年份	融资轮次	最新估值
算法模型 (中游)	集团科技研究院	—	阿里巴巴达摩院	中国	2017	—	—
		—	微软亚洲研究院	中国	1998	—	—
	开源社区	综合型	GitHub	美国	2008	并购	—
		垂直型	Hugging Face	美国	2016	C 轮	20 亿美元
			Papers with Code	英国	2018	并购	—
应用拓展 (下游)	文本处理	营销型	Copy.ai	美国	2020	A 轮	—
			Jasper	美国	2021	A 轮	15 亿美元
		销售型	Lavender	美国	2020	—	—
			Smartwriter.ai	澳大利亚	2021	—	—
		续写型	彩云科技 ( 彩云小梦 )	中国	2014	天使轮	—
		知识型	Mem	美国	2021	A+ 轮	1.1 亿美元
		通用型	Writer	美国	2020	A 轮	—
			澜舟科技	中国	2021	Pre-A 轮	—
		辅助型	AI21 Labs ( Wordtune )	以色列	2018	B 轮	—
			秘塔科技	中国	2018	Pre-A 轮	—
		交互型	Latitude ( AI Dungeon )	美国	2019	种子轮	—
			聆心智能	中国	2021	天使轮	—
		代码型	Repl.it ( Ghostwriter )	美国	2016	B 轮	8 亿美元
			Mintlify	美国	2020	种子轮	—
			Stenography	美国	2021	—	—
			Debuild	美国	2020	种子轮	—
			Enzyme	美国	2016	—	—

续表

产业链	细分赛道	属性分类	公司名	国家	成立年份	融资轮次	最新估值
应用拓展 (下游)	音频处理	音乐型	Boomy	美国	2018	—	—
			灵动音科技	中国	2018	A 轮	—
		讲话型	Resemble.ai	美国	2019	种子轮	—
			WellSaid Labs	美国	2018	A 轮	5 834 万美元
		解决方案型	标贝科技	中国	2016	B 轮	—
	图像处理	生成型	Stability AI	英国	2020	种子轮	10 亿美元
			Midjourney	美国	2021	—	—
			诗云科技	中国	2020	Pre-A 轮	—
		广告型	AdCreative.ai	法国	2021	—	—
		设计型	Diagram	美国	2021	种子轮	—
			Nolibox (图宇宙)	中国	2020	Pre-A 轮	—
		编辑型	PhotoRoom	法国	2019	A+ 轮	—
	视频处理	生成型	Runway	美国	2018	C 轮	5 亿美元
			Plask	韩国	2020	种子轮	—
		编辑型	Descript	美国	2017	C 轮	5.5 亿美元
			InVideo	美国	2017	天使轮	—
		虚拟人型	Hour One	以色列	2019	A 轮	—
			Synthesia	英国	2017	B 轮	—
		解决方案型	影谱科技	中国	2009	D 轮	—
			帝视科技	中国	2016	B 轮	—

注：数据截至 2022 年 12 月 10 日。

附录二 **AIGC**术语及解释

术语	解释
PGC	Professional-Generated Content，专业生成内容。以 PGC 作为职业获得报酬的职业生成内容也被称为 OGC（Occupationally Generated Content）
UGC	User-Generated Content，用户生成内容
AIGC	Artificial Intelligence Generated Content，人工智能生成内容
生成式人工智能	一类人工智能算法，根据训练过的数据生成全新、完全原创的输出，常以文本、音频、图像、视频等形式创建新内容
大模型	Foundation Model，又译作“基础模型”，对广泛的数据进行大规模预训练来适应各种任务的模型
NFT	Non-Fungible Token，非同质化代币。一种基于区块链技术的数字资产权利凭证。区别于比特币这样的同质化代币，代币与代币之间是不可相互替代的
GameFi	游戏化金融。将去中心化金融以游戏方式呈现的产品，多代指结合了区块链的游戏
图灵测试	艾伦·图灵提出的一个判断机器是否具备智能的著名方法
机器学习	让计算机程序从数据中学习以提高解决某一任务能力的方法
监督学习	从标注数据中学习的机器学习方法
无监督学习	从无标注数据中学习的机器学习方法
强化学习	在给定的数据环境下，让智能体学习如何选择一系列行动，来达成长期累计收益最大化目标的机器学习方法
深度学习	采用有深度的层次结构进行机器学习的方法
人工神经网络	模仿生物神经网络工作特征进行信息处理的算法模型
感知器	一种最简易的人工神经网络模型
TTS	Text to Speech，文本转语音

续表



术语	解释
NLP	Natural Language Processing，自然语言处理。使计算机程序理解、生成和处理人类语言的方法
CV	Computer Vision，计算机视觉。使计算机具备处理图像、视频等视觉信息能力的方法
GAN	Generative Adversarial Networks，生成对抗网络。通过一个生成器和一个判别器的相互对抗，来实现图像或文本等信息生成过程的算法模型
Diffusion	扩散模型。一种通过对数据点在潜在空间中扩散的方式进行建模来学习数据集潜在结构的算法模型，常用于图像生成
CLIP	Contrastive Language-Image Pre-Training，文本 - 图像预训练。一种用于匹配图像和文本的预训练神经网络模型
Seq2Seq	Sequence-to-Sequence，序列到序列模型。将一种序列处理成另一种序列的模型，典型应用场景是机器翻译
注意力机制	由于信息处理的瓶颈，人类会选择性地关注所有信息的一部分，同时忽略其他可见的信息，这种机制可以应用于人工智能的算法模型领域
Transformer	一种运用注意力机制的深度学习模型，是许多大模型的基础
GPT	Generative Pre-trained Transformer，生成型预训练变换器。由 OpenAI 研发的大型文本生成类深度学习模型，可以用于对话 AI、机器翻译、摘要生成、代码生成等复杂的自然语言处理任务
ChatGPT	OpenAI 在 2022 年 11 月发布的聊天机器人，能自然流畅地与人们对话
RLHF	Reinforcement Learning from Human Feedback，从人类反馈中进行强化学习。利用人类反馈信号优化模型的强化学习方法
BERT	Bidirectional Encoder Representations from Transformers，变换器的双向编码器表示。一种谷歌基于 Transformer 提出的模型
ViT	Vision Transformer，视觉变换器。一种利用 Transformer 解决计算机视觉问题的模型

### 附录三 AIGC大事记

#### 1950年：

- 艾伦·图灵提出著名的“图灵测试”，给出判定机器是否具有“智能”的试验方法。

#### 1957年：

- 第一支由计算机创作的弦乐四重奏《依利亚克组曲》（*Illiac Suite*）完成。

#### 1966年：

- 世界上第一款可人机对话的机器人“Eliza”问世。

#### 1985年：

- IBM首次演示了语音控制打字机Tangora。

#### 2007年：

- 世界上第一部完全由人工智能创作的小说《在路上》（*I The Road*）问世。

#### 2012年：

- 微软演示了全自动同声传译系统，可将英文演讲者的内容自动翻译成中文语音。

#### 2014年：

- 伊恩·J.古德费洛（Ian J.Goodfellow）等人提出生成式对抗网络GAN。

#### 2015年：

- 雅沙·索尔-迪克斯坦（Jascha Sohl-Dickstein）等人提出了Diffusion模型。

#### 2017年：

- 世界上首部100%由人工智能微软“小冰”创作的诗集《阳光失了玻璃窗》出版。
- 谷歌团队在《注意力就是你全部需要的》（*Attention is all you need*）论文中提出了Transformer。

#### 2018年：

- 英伟达发布StyleGAN模型，可以自动生成高质量图片。

·人工智能生成的画在佳士得拍卖行以43.25万美元成交，成为世界上首个出售的人工智能艺术品。

·OpenAI推出预训练语言模型GPT，采用Transformer架构，拥有1.17亿参数量，可完成简单的自然语言处理任务。

## **2019年：**

·DeepMind发布DVD-GAN模型，可以生成连续视频。

·OpenAI推出GPT-2，拥有15亿参数量，性能进一步提升。

## **2020年：**

·OpenAI推出GPT-3，拥有1750亿参数量，在文字翻译、问答与生成等方面拥有惊人表现。

## **2021年：**

·OpenAI推出DALL·E，主要用于文本与图像交互生成内容。

·OpenAI推出CLIP，它能够连接文本与图像，覆盖各种视觉分类任务。

## **2022年：**

·AI绘画工具Stable Diffusion（Stability AI）、Midjourney（Midjourney）、DALL·E 2（OpenAI）、Imagen（谷歌）发布。

·美国科罗拉多州博览会艺术比赛的数字类别中，39岁游戏设计师杰森·艾伦（Jason Allen）使用Midjourney创作的作品《太空歌剧院》夺得冠军。

·视频生成工具Make-A-Video（Meta）、Imagen Video（谷歌）、Phenaki（谷歌）发布。

·3D模型生成工具DreamFusion（谷歌）、Magic3D（英伟达）、Point-E（OpenAI）发布。

·OpenAI推出ChatGPT，它具有接近人类流畅而自然的多轮对话能力，还能够完成假扮特定角色对话、撰写周报、修改代码等复杂的文本处理任务。

资料来源：中国信息通信研究院联合京东探索研究院《人工智能生成内容（AIGC）白皮书（2022年）》，2022年9月2日发布

## 后记

人工智能的发展无疑是迅速的，从学科诞生起至今不过百年，却已在围棋、德州扑克、策略游戏等多个象征智慧的领域战胜人类，如今又获得了人类独有的创造力。在本书有限的篇幅内，或许难以覆盖这段惊人进化历程的方方面面，但希望能让每一位读者都能感受到科技前沿的无穷魅力，也保有一份针对科技本身的思考。

对于人工智能的未来，你是怎么看待的呢？它究竟会成为人类的助力，还是会成为人类的威胁？

悲观者认为，人工智能最终会彻底取代人类，进而导致人类的灭亡；而乐观者认为，人工智能不会取代人类，它会让人类的生活更加幸福。

不过，持有何种观点并不重要，重要的是该以何种姿态面对未来。在无数科幻小说、电影、电视剧中，都对这一点进行了哲学层面的探讨。《我，机器人》中，提出了具有广泛影响力的“三大定律”，探讨了人类与具备人工智能的机器人和谐相处的基本原则；《西部世界》中，呈现了膨胀的欲望凌驾于技术之上，肆意突破道德底线后酿成的后果；《齐马蓝》中，诠释了智能进化之路的尽头，需要回归诞生源初的自然之道。

我们前进着，我们也思考着，直至抵达科技的彼岸。

正如尼克·博斯特罗姆（Nick Bostrom）说的：“机器智能是人类需要做出的最后一项发明。”

这既是对未来的憧憬，也是对未来的警示。而最终未来的船帆驶向何方，选择权从来都在人类自己手中。

## 专家推荐

同为新兴科技领域，AIGC与Web3.0、元宇宙等都具有很大的结合空间。阅读完《AIGC：智能创作时代》，相信读者在这方面会涌现出很多有趣的想法。

——香港科技大学教授 陈卡你

通过阅读《AIGC：智能创作时代》这本书，读者将会进一步认识到AIGC在影视行业中的应用，了解AIGC如何通过满足大众飞速增长的内容消费需求，逐渐与人们的文娱生活相结合。

——猫眼娱乐CEO 郑志昊

近十年来，人工智能再度崛起并成为推动社会发展的引领性技术之一，而这次诞生了一个非常新的领域——人工智能创作，或者叫人工智能生成内容（AIGC），并迅速成为近两年年的发展热点。由于该领域非常新，所以缺乏比较系统、全面对该领域相关内容进行梳理和介绍的书籍。本书则比较全面、系统地对AIGC相关的技术进行了介绍，对AIGC的产业进行了梳理，并对其发展方向进行了预测。无论你是否熟悉AI、是否有一定技术或行业基础，都可以通过本书对AIGC进行条理化、系统化的了解，这对你开阔视野或参与其中都会有很大的帮助。

——北京电影学院未来影像高精尖创新中心虚拟制作实验室主任 王春水

AI生成图像领域已基本成熟并走向商业落地阶段，而AI生成视频领域也出现了许多让人振奋的创新。《AIGC：智能创作时代》让我们看见，在不远的将来，内容创作难度会在AIGC的帮助下大大降低，出现越来越多优质的内容作品。

——爱奇艺高级副总裁 王学普

在开源模式的推动下，伴随着深度学习模型的不断迭代，AI将不再只是辅助内容创建的工具，而是能够独立创造和生成内容的新型创作方式。如果在没有任何技术背景的情况下，想要了解这个领域的发展情况、技术变革与商业落地场景，这本书将是不错的选择。

——纽约大学斯特恩商学院冠名教授 陈溪

身处“智能创作时代”的开发者们无疑是幸运的，AIGC能够支持他们更快速、更高效地开发出高质量的游戏内容，缩短游戏的开发周期并降低开发成本。本书是帮助游戏行业从业人员了解AIGC前沿发展的优质书籍。

——阿里巴巴云游戏事业部（元境）总经理 王矛

在生成式AI相关技术领域的飞速发展下，多媒体数字资源管理、开发与利用的研究领域也许会生长出一些有趣的新方向，期望这本著作能为信息资源管理相关领域的研究工作带来一定的启发。

——北京大学信息管理学院副教授 韩圣龙

内容社区中，创作者常常会面临创作质量和更新频率之间的抉择问题，AIGC技术在帮助创作者提升内容生产效率的同时，也让他们更专注于提升内容创作的质量。这本书描绘了一幅人工智能与创作者携手并进的美好画面。

——知乎战略副总裁 张宁

音乐人工智能代表了全新的音乐发展趋势，我们的音乐教育乃至其他艺术教育都需要更多地与前沿科技相结合。本书非常适合作为艺术类学生的科普学习读本。

——中央音乐学院现代远程音乐教育学院院长 方恒健

人工智能辅助创作这一全新的艺术形式，既为全民推广美育教育提供了机会，又会带给所有艺术工作者全新的挑战。为了更好地把握机会和迎接挑战，我推荐大家读一读《AIGC：智能创作时代》。

——中央美术学院艺术管理与教育学院党总支副书记，美术博物馆虚拟策展与美育课程虚拟教研室副主任 康俐

智能化时代的技术迭代是迅速的，智能化呼唤数字化，数字化促进智能化，《AIGC：智能创作时代》正是站在当前时代的风口，向大众科普AIGC有关的知识，很具有现实意义。

——中国电子学会科普培训中心副主任、高级工程师 宁慧聪

我们处在一种生产工具创造一个生产力的时代。《AIGC：智能创作时代》向我们描绘了由人工智能解放内容生产力的时代。

——复旦大学泛海国际金融数字经济研究中心主任 王家华

在可预见的未来，AIGC会彻底改变人们的思考方式、创作过程。阅读本书，可以让你知道该如何积极面对最新的科技趋势。

——哈佛大学物理与计算机科学博士 朱科航

AIGC是2023年的热门话题。它对内容生产究竟会带来哪些变革，它的技术路线和产业地图是怎样的，等等，针对这些问题，《AIGC：智能创作时代》进行了全面的描述。本书值得对AIGC话题感兴趣的人系统地阅读和了解。

——前百度副总裁 伍晖

AIGC的技术变革为数字文化产业带来了全新的发展动能，也为数字经济发展注入了创新活力。《AIGC：智能创作时代》可以帮助大众了解数字经济发展的新趋势。

——36氪集团高级副总裁，氪星创服董事长兼CEO 董博

AIGC是当下科技圈讨论的热点，谷歌、微软、百度等传统互联网巨头都在这一领域持续布局，加大投入。《AIGC：智能创作时代》很好地向大众科普了最新的技术趋势。

——QuestMobile CEO 陈超

作为全新的内容生产工具，AIGC必将在未来嵌入我们的日常生活，改变我们的生活方式。无论你是开发者、投资者，还是只是对AIGC的未来感兴趣的人，都可以来看看本书。

——瑞信证券（中国）有限公司证券研究部主管 刘帅

在这个时代，算法与模型给予人们的不仅是信息的聚合，还有内容的创造。阅读本书，既是了解科技前沿，也是了解我们这个全新的时代。

——招商证券计算机行业首席分析师 刘玉萍

AI行业始终蕴藏着无穷的创业机会，值得每个年轻人积极地拥抱它。如果想要更多地了解这一次AIGC浪潮背后的商业机会，不妨读一读本书。

——五源资本合伙人 刘凯

算力将解放人们的创造力，在不同场景中降低创作者的表达门槛和生产门槛。如果想要了解未来AIGC将如何赋能创作者经济，这本书将是很好的选择。

——知春资本合伙人 曾映龙

AIGC将会首先作为各个内容行业生产力加速的核心普及开；由人工智能生成的内容将会进一步丰富虚拟空间和信息空间中人的体验。从文案、图像的生成到数字资产的创建，

AIGC让用户生产的内容在数字世界中变得无处不在。在数字世界中，当资产的生产力被给予用户，一个有趣的核心机制将能孕育无数个平行的叙事和体验。

——**Mirror World**创始人，**rct AI**产品副总裁 朱元

人工智能教育任重而道远，AIGC的快速发展也昭示着行业迫切的人才需求，这个时代既需要《AIGC：智能创作时代》这样杰出的科普书籍吸引大众的兴趣，也需要兼具学术深度和工程实践的教学课程培育技术人才。

——七月在线创始人兼CEO **July**

AIGC作为商业变革的点金石，其技术演变时机推演成为未来创业机会窗口的重要标识，其中Transformer成为承上启下的基石，本书对其进行了深入浅出的介绍，并将各行业商业应用纳入其中，不仅为未来创业机会的方向提供了指引，还见证了时代背景下的壮阔远景。

——**AI Creator**创始人 刘潇

元宇宙与第三代互联网象征着科技的未来，高度数字化、智能化的元宇宙时代与每个人、每家企业和每个行业都息息相关。在Web3.0时代，AIGC将提供高效低成本、千人千面、独一无二的元宇宙智能内容生产方式。创意和内容表达的方式在伴随技术的进步而迭代升级，AIGC智能生成内容的维度也可以很多元：一维的文本和语音，二维的图像内容，三维的视频内容或立体模型，四维的实时智能互动内容，甚至多维打破时间和空间的约束、千人千面、个性化实时智能互动内容，及动态立体的视觉传达。

文本、图像、音频、视频、三维内容与跨模态智能互动的融合，可以充分地释放创意和想象，让每个人都可以零门槛地智能生成想要的内容，实现“所想即所见”，用视觉化的方式表达思想。正如本书作者所言，AIGC是元宇宙时代的内容生产力，智能创作时代将为人类创造全新的表达和沟通模式，推进众多行业的数智化升级，改变人类未来的生活。

——迈吉客科技创始人兼董事长 伏英娜

AIGC是智能化工业革命发展的新里程碑，它将会进一步改变过往的生产力与劳动关系格局。通过阅读这本书，各个领域的从业者可以深入了解AIGC对于各自行业产生的变革性影响，思考在新技术浪潮下的企业战略与成本结构。

——**CHICAT**创始人 张珍妮

在我看来，围绕大模型的生成式AI的潜力仅仅被释放了1%，它最重要的潜力就是在技



术范式转移的时候帮助更多人拓宽想象力，“原来生成式 AI 已经能够帮我解决问题了。”杜雨和张孜铭的《AIGC：智能创作时代》有机会从场景的维度启发每个行业的发展，各个行业的从业者都有机会从中窥见下一代的生产力可能。

——猴子无限创始人 尹伯昊

艺术是表达生命力量的最佳形式，而科技赋予了每个人艺术创作的能力。《AIGC：智能创作时代》可以让更多的人感受艺术与科技结合的魅力。

——Viva la Vida创始人 吕晓宁

虚拟空间生成与创作的便捷化是元宇宙走向广泛行业应用的重要前提，这本书让我看到了AIGC赋能元宇宙领域发展的巨大价值。

——Vland创始人兼CEO 金秋远

目前，AIGC领域正迎来基础设施建设的大浪潮，各行各业都可以孕育出会创作、懂创作、能辅助人类创作的工具。这本书中的诸多案例可以为AIGC领域的建设者们提供启发。

——NeuDim联合创始人兼COO 曹君铭

AIGC技术的发展和推广，无论是对医生还是患者而言，都是一种福音。通过学习这本书，读者可以了解到AIGC将会如何颠覆传统医疗行业，释放医生精力，让医生资源专注到核心业务中。

——万木健康创始人兼CEO 程锦

伴随着智能创作时代的来临，AIGC成为市场上最热议的话题之一，我们将看到更智能、与人类互动更自然的虚拟内容创作方式的出现。本书从理论到应用深入浅出地描述了AIGC在不同垂直场景中应用的可能，对于AIGC赋能行业的思考非常具有前瞻性。无论是已经从事这个行业的创业者，还是想要了解行业动态的读者，本书都是必读的佳作！

——杭州万像文化科技CEO 夏冰

科技的发展让每个人都能够体验艺术创作的魅力，如果你不想错过智能创作的未来，就来读读这本书吧！

——意间AI绘画CEO 郭亚鹏

杜雨和张孜铭老师的新书是我们行业的福利，也是我读过的有关如何认识AIGC的最为

系统和深刻的一本书。本书深入浅出地介绍了从PGC、UGC到AIGC的一场时代变迁。借助AIGC，人人都有机会通过更先进的技术手段来追求和实现自己的艺术主张，能够极大地释放自己的想象力，掀起属于这个时代的新浪潮。机器学习的速度越来越快，技术也日益成熟，那么未来的艺术形态将蜕变或者说进化至何处？本书将极大地激发我们对这个问题的深刻思考。

——Hiiimeta CEO 陈定媛

如果你读了这本书，你就会明白AIGC与游戏产业的融合是必然的，AIGC可能会从根本上改变游戏的制作方式，从3D建模到角色动画，再到游戏内的音乐制作、人物对话，都可能会出现AIGC的身影。

——Vast CEO 宋亚宸

推荐艺术家、游戏原画师等内容创作者都来阅读本书，深入体会AIGC技术是怎样提高创作者的内容产出效率，从而造福创作者的。

——无中之城工作室CEO Sala

在内容生成领域，人工智能即将带给大家更多的惊喜。《AIGC：智能创作时代》非常适合用来了解我们的未来。

——前迪士尼研究中心研究员，《山海旅人》制作人，人工智能博士 魏新宇