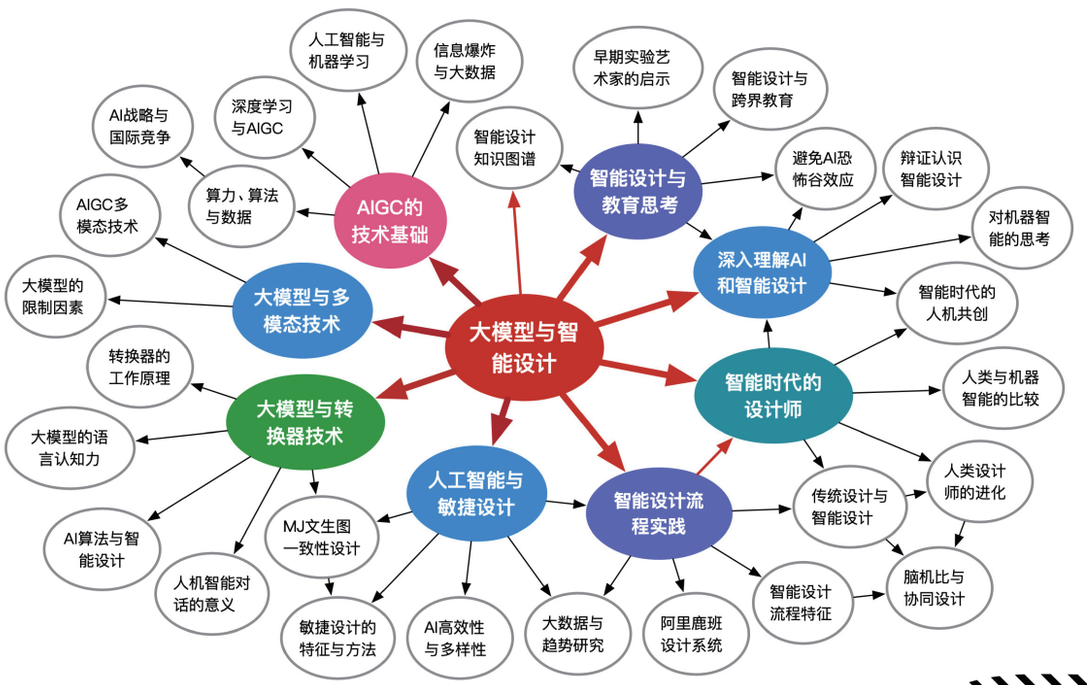
第3章

大模型与智能设计

AIGC 实现了人工智能从感知理解世界到生成创造世界的跃迁，大模型的特征与深度学习算法是智能设计的基础。本章系统阐述了大数据知识和算法、算力与数据的重要价值，同时介绍机器学习与深度学习算法，重点为大模型的技术原理与技术构架，包括转换器、注意力机制与大模型的语言认知能力等。本章还阐释智能设计中的人机关系及各自的优势，并提出人机协同智能设计的方法与工作流程。下面是本章的思维导图，供读者参考。



3.1 AIGC的技术基础

3.1.1 信息爆炸与大数据

为什么生成式人工智能能够成为一个突然增长的行业？其答案就在于当代的信息数据爆炸。当下各种计算设备和应用平台，如智能手机、平板计算机、可穿戴设备、XR 设备、云服务系统和社交网络等的激增，特别是以 5G 为代表的高速移动网络的出现，使今天的互联网已经发展成为一个巨大的数据生成器。人们无论是工作、学习、购物、旅游或是居家生活，每时每刻都在贡献着数据（图 3-1）。在这个由信息构成的“地球村”，我们每人每天都要接触大量的数据和信息，新闻、广告、电子邮件、微信、微商、公众号、微博、直播、短信、互联网、手机视频、电子书籍和广告无处不在。人们除了睡觉外，几乎随时随地都在吸收大量庞杂的信息。随着智能手机的普及，移动短视频 App，如抖音、快手等快速兴起，已成为一种新兴的媒体形式。全球知名咨询公司麦肯锡在报告中指出：“数据已经渗透到当今每一个行业和业务职能领域，成为重要的生产因素。人们对于海量数据的挖掘和运用，预示着新一波生产率增长和消费者盈余浪潮的到来。”如同石油、天然气和电力是今天世界经济运行不可或缺的“燃料”，数据则是 AIGC 技术的引擎。有了大数据的支持，创建复杂的生成式 AI 模型才有可能。

图 3-1　今天的时代是信息爆炸时代，人们每时每刻都在接收与贡献着数据

今天人们所拥有的信息比人类历史上任何时期都要丰富。2019 年，设计师卡尔·菲施等制作了一部动态短片《你知道吗？》（图 3-2，左）揭示了大数据的真相，随即该短片风靡海外社交网络。该作品显示：截至 2018 年，全世界每分钟有 400 小时的视频传到 YouTube 视频网；全世界每天有

6.92 亿人次在使用推特（Twitter）；2020 年有 300 亿的设备连接到物联网；网飞公司（Netflix）的网络电影和网络电视节目每天播放量超过 1.4 亿小时；在线音乐网站（Spotify）的网络音乐流量每天超过 5.6 千万小时；谷歌用户每天使用的搜索量超过 550 亿次；全世界每分钟诞生 570 个网站；最后，作者总结出 2018 年全球新增的数据量超过 33 ZB（33×10^{21}），即数字 33 后面要加 21 个零，这无疑是一个堪比宇宙星系的天文数字。因此，毫无疑问我们正处于大数据时代。大数据不仅数据量大（Volume）、变化速度快（Velocity）、数据类型多（Variety），同时也有潜在的经济价值（Value）和潜在的信息可用性（Veracity）。大数据是结构复杂、数量庞大、类型众多的数据集合，而全球化和网络化就是大数据时代来临的标志（图 3-2，右上）。

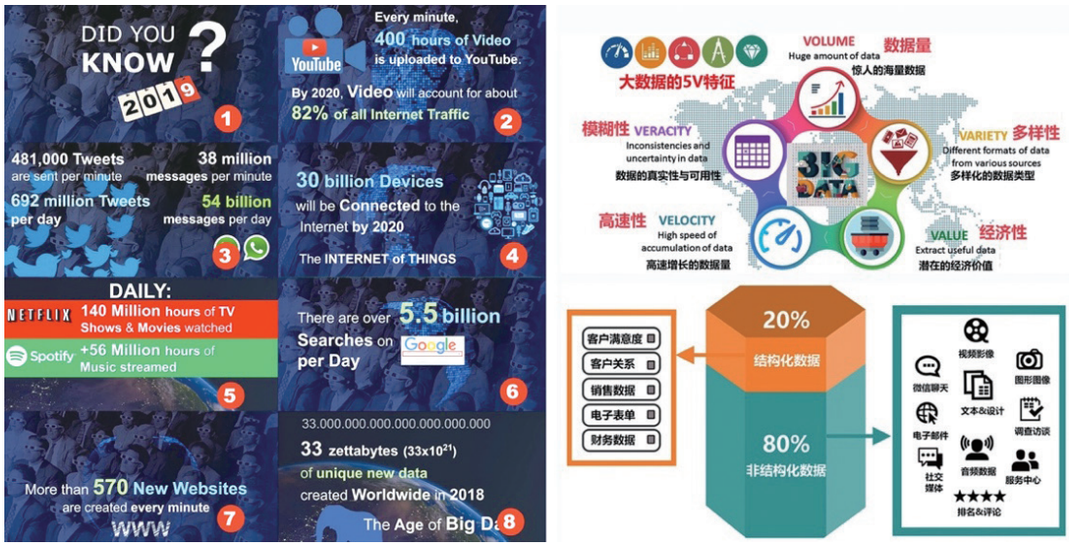


图 3-2　短片《你知道吗？》截图（左）大数据特征（右上）与数据结构类型（右下）

生成式 AI 的快速发展还得益于深度学习（deep learning，DL）算法的出现。该算法能够支持非结构化数据，这使聊天、绘画、照片、动画、视频等数据的智能化成为可能。结构化数据是指具有预定义的数据模型的数据，它的本质是将所有数据标签化、结构化。因此，只要确定标签数据就能读取出来，多数的电子表格、数据报表、财务统计表都属于结构化数据。结构化数据容易被计算机理解和处理。但在日常生活中，结构化数据所占比例很小。非结构化数据是指数据结构不规则或者不完整，没有预定义的数据模型的数据，如图片、音频、视频、文本、网页、电子邮件、社交媒体等，它们往往更难被计算机理解和处理。全世界有 80% 的数据都是非结构化数据，音频、图片、文本、视频这四种载体可以承载世界万物的信息（图 3-2，右下）。虽然人类在理解非结构化数据内容时毫不费劲。但对于早期计算机来说，理解这些内容比登天还难。深度学习在该领域取得了突破性的成就，这不仅使机器视觉研究成为当代人工智能的热门领域，而且也为今天的生成式人工智能的发展奠定了基础。

3.1.2　人工智能与机器学习

近年来 AIGC 的爆发得益于生成式人工智能理论和概念，包括神经网络、生成对抗网络（GAN）、转换器和扩散模型等算法以及机器学习、深度学习、无监督学习和强化学习等理论不断进步和突破。如果说人工智能的目标是让机器能够拥有人类的智力，如探究能力、学习能力以及创造力，那么机

器学习（machine learning，ML）与深度学习就是近 30 年来人工智能所取得的最重要成果。IBM 公司认为："生成式 AI 是一类机器学习技术，它从训练数据中学习。" 机器学习是 AIGC 的基础。图 3-3 不仅形象地说明了 AI 与 ML、DL 及 GAI 的逻辑关系，同时也从时间上说明了从 AI 技术到生成式人工智能的演化历程，即 AI → ML → DL → GAI 的发展历史。

生成式人工智能（GAI）与人工智能（AI）、机器学习（ML）和深度学习（DL）之间的关系

图 3-3　人工智能与 ML、DL 及 GAI（AIGC）之间的逻辑关系

维基百科指出：人工智能的研究历史有着一条从以"推理"为重点到以"知识"为重点，再到以"学习"为重点的清晰自然的发展脉络。机器学习是使用算法来解析数据并从中学习，然后对真实世界中的事件做出决策和预测。机器学习的基础是统计数与概率数，即从数据分析中获得规律并利用规律对未知数据进行预测，从而能够完成人脸识别、语音识别、物体识别、翻译等任务。机器学习技术是一套特定的算法和数学模型，这可以使计算机能够从大量的数据中学习和改进自身的性能。人们可以通过训练模型使计算机能够识别和分类不同的动物特征。例如，在动物分类任务中，人们可以使用大量的动物图像数据集通过机器学习的大模型进行训练。机器学习算法会分析这些图像中的特征，并学会如何区分不同种类的动物。通过提取图像的颜色、纹理、形状等特征，机器学习模型可以学会识别和分类不同的动物，如狗、猫、鸟等。人们通过反复训练和调整模型，就可以提高机器学习的准确性，使其能够在未知图像中识别和分类动物特征。

机器学习常见的方法是使用支持向量机（SVM）算法和 HOG（方向梯度直方图）特征描述符。支持向量机（SVM）通过将数据点映射到高维空间，并找到一个超平面来将不同类别的数据点分开。在分类任务中，SVM 算法可以使用 HOG 特征描述符作为输入，这些描述符可以提取图像的边缘和纹理等特征。人们通过训练 SVM 模型就可以让机器根据图像特征进行分类。HOG 特征描述符通过计算图像中每个像素点的梯度方向和梯度幅值，并将它们汇总为直方图。这个直方图描述了图像中不同像素方向的梯度分布，可以用来表示图像的边缘和纹理等特征。在机器学习中，HOG 特征描述符通常用作输入数据用于训练分类模型，如 SVM 算法。通过使用 SVM 算法和 HOG 特征描述符等技术，机器学习可以从图像数据中学习和识别不同的特征，如图像边缘与纹理等，从而实现图像分类和识别的任务。

3.1.3 深度学习与 AIGC

机器学习过程属于监督学习，即需要人工标记图片特征并训练机器（图 3-4，左上），建立词汇与图片之间的关联性，并在随后的图片检测中完成特征提取（图 3-4，右）。对于机器学习来说，良好的特征提取对最终算法的准确性起了非常关键的作用。系统主要的计算和测试工作都消耗在这部分。但是这部分工作一般都是靠人工完成的，而手工提取特征再打上标签比较费时费力，这不仅需要专业知识，而且很大程度上靠经验和运气。那么机器能不能自动学习图像特征呢？深度学习的出现就为这个问题的解决提出了一种方案。深度学习是机器学习的一个子集，也都是基于神经网络，即通过模仿人类及动物神经网络行为特征来进行分布式信息处理的数学模型。然而，它们的功能有所不同。深度学习对硬件条件要求更高并且需要大量的数据集进行训练，但深度学习能够实现非监督学习的特征提取及分类（图 3-4，左下）的任务。

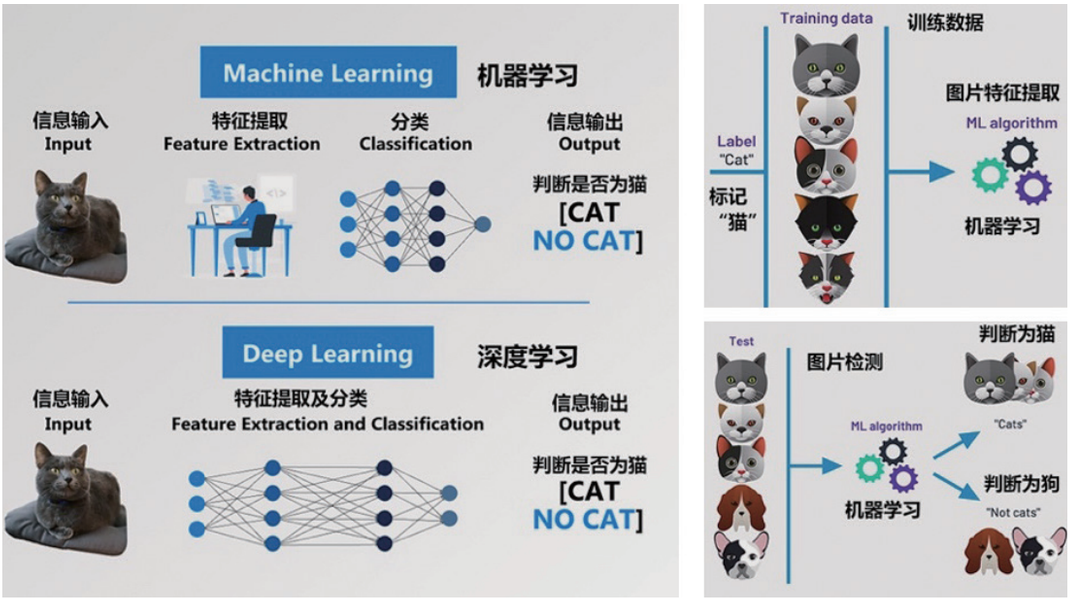


图 3-4　机器学习的原理（左上）流程（右）与深度学习原理（左下）

虽然机器学习与深度学习方法在数据准备和预处理（数据清洗等）方面几乎相同，但机器学习数据集规模小并需要人工特征提取和标记，而深度学习则是多层神经网络结合大数据的逐层信息提取和筛选的过程，并由机器自己完成特征提取。这是生成式人工智能的关键技术之一。多伦多大学教授、计算机学家和心理学家、机器学习领域的泰斗杰弗里·辛顿是反向传播算法和对比散度算法的发明人之一，被誉为“深度学习之父”。2006 年，辛顿等在《科学》上发表了一篇文章，其中有两个主要观点：①多隐藏层的人工神经网络具有优异的特征学习能力，学习得到的特征对数据有更本质的刻画，从而有利于可视化或分类；②深度神经网络在训练上的难度可以通过“逐层初始化”的方法有效克服，该过程可以通过无监督学习实现。这篇文章开启了深度学习在学术界和工业界的浪潮。因此，正是深度学习算法使机器具备强大的学习能力，也使机器学习从技术范畴上升到思想范畴，从而奠定了 AIGC 的基础。表 3-1 是机器学习与深度学习在数据结构、数据规模、学习方法、硬件依赖等方面的对比。

表 3-1　机器学习与深度学习的对比

	机 器 学 习	深 度 学 习
数据结构	结构化数据	结构化数据与非结构化数据
数据规模	可以使用少量的数据做出预测	需要使用大量的训练数据做出预测
学习方法	学习过程划分为较小的步骤，然后将每个步骤的结果合并成一个输出	通过端到端地解决问题来完成学习过程
硬件依赖	可在低端机器上工作。不需要大量的计算能力	依赖于高性能机器，本身就能执行大量的矩阵乘法运算。GPU 也可以有效地优化这些运算
特征提取	需要准确识别特征（人工特征标记）	从数据中习得高级特征，并自行创建新的特征
运行时间	花费几秒到几小时的相对较少时间进行训练	通常需要很长的时间才能完成训练，因为深度学习算法涉及许多层
可解释性	部分算法很容易解释（逻辑回归、简单决策树）而另一些算法比较难解释如 SVM、XGBoost	难以解释，而且常常是不可能的
输出结果	输出通常为数值类型，如评分或分类	输出多种格式如文本、评分或声音、图像

深度学习最大的特点就是可以自主地从数据中抽取数据，其强大的计算能力让它可以处理人类做不到的事情。计算机具有更好的学习和适应能力，这使它成为 AIGC 的基础。深度学习的核心在于模拟大脑的神经网络，其多隐层的人工神经网络具有优异的特征学习能力。今天的深度神经网络一般在 5 层以上，甚至多达 1000 多层，而每层都有具体的分析目标与任务。例如，在人脸识别过程中，有的隐藏层负责识别图像明暗，有的负责识别图像边缘，还有的负责识别人类的五官特征，最后才是组合全部隐藏层数据的输出层（图 3-5）。2012 年机器视觉竞赛 ImageNet 的冠军模型 AlexNet 用了 8 层，2014 年冠军模型 GooleNet 用了 22 层，2015 年的冠军模型 RestNet 用了 152 层，2016 年的冠军模型则多达 1207 层。海量的多层数据计算也使得高性能计算机和图形加速卡（GPU）成为深度学习的“标配”。

算法、算力与数据是 AIGC 能够走上历史舞台的功臣。深度学习的数据规模非常庞大。例如，早期深度学习算法的任务是要求机器从各种动物图片中识别出猫的形象。训练数据高达数万张甚至上亿张图片，由此才能使机器能够区分最小的细节，从而区分出猫与猎豹、黑豹或狐狸等的不同，并将识别错误率降低到最小。但深度学习也会存在一个问题，就是既然绕过了人为提取特征与人为判断规律，就会让深度学习的模型几乎不存在可解释性。深度学习就相当于是一个“黑箱”，虽然我们知道它每次能给出准确或正确的答案，但仍难以解释其原因。随着深度学习的算法与隐层的复杂性不断增加，要想破解这个黑箱就更加困难。因此，深度神经网络的核心仍然是以统计和概率为基础的。人类神经网络可以归纳、演绎、总结现象背后的逻辑与规律，但这点机器依然做不到。机器学习只能通过调用包含大量专业知识和经验的专家系统及海量图片数据库，并根据用户的提示词引导，基于统计和概率来生成足以乱真的艺术作品。

概括地说，深度学习主要使用人工神经网络。神经网络是由神经元（或节点）组成的层次结构。通常包括输入层、隐藏层（可以有多个）和输出层。每个神经元都与前一层和后一层的神经元相连，具有权重和激活函数。深度学习的基本原理如下。

（1）前向传播：主要包括输入层接收原始数据的输入，隐藏层对输入数据进行一系列线性和非线性变换并学习数据的表示，输出层产生模型的最终输出。在前向传播过程中，输入信号通过神经网络传播，每个神经元的激活值经过权重计算和激活函数处理，最终得到模型的预测结果。

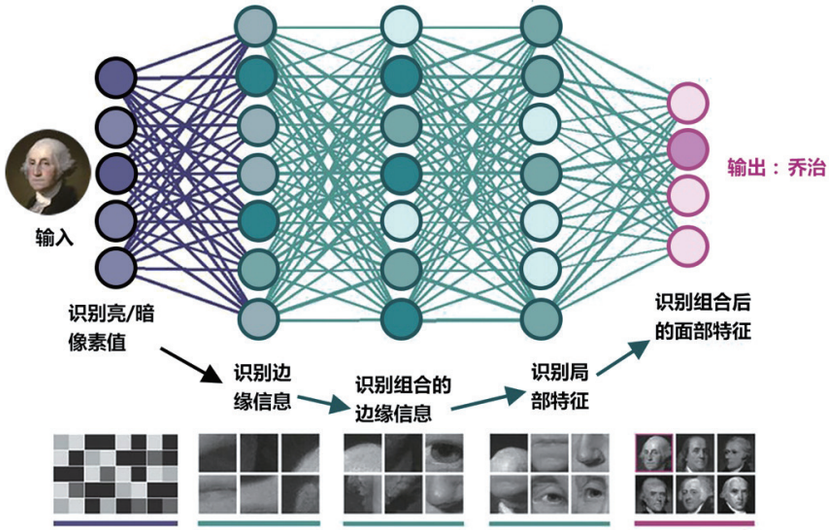


图 3-5　深度神经网络与隐藏层的逐层任务解析

（2）损失函数：损失函数度量模型的输出与实际标签之间的差距。目标是通过调整网络参数（权重）来最小化损失函数。

（3）反向传播：反向传播是通过梯度下降算法来更新神经网络参数，使损失函数最小化。梯度是损失函数对于参数的偏导数，表示了损失函数变化方向和速度。

（4）优化算法：梯度下降是一种优化算法，但有多种变体，如随机梯度下降（SGD）、批量梯度下降和小批量梯度下降等。这些算法通过调整参数以最小化损失函数。

（5）迭代训练：以上步骤构成了一个训练周期。在多个训练周期中，模型通过不断调整参数来逐渐提高性能。

（6）模型评估和调优：使用验证集对模型进行评估，调整模型参数以提高性能。最终，通过测试集对模型进行最终评估。

（7）预测：训练完成后，模型可以用于新数据的预测。

3.1.4　算力、算法与数据

人工智能技术概括来说可以分为 5 个层面（图 3-6）。第 1 层（最底层）是基础层，是人工智能的基础设施建设，包含数据和算力两部分，数据越大，算力越强，则人工智能的解析能力越强。第 2 层为算法层，如卷积神经网络（GAN）、LSTM 序列学习、深度学习等，这些都是机器学习的算法。第 3 层为方向层，如计算机视觉、语音工程和自然语言处理等。此外还有规划决策系统或大数据分析的统计系统，如增强学习等。第 4 层为技术层，如图像识别、语音识别、机器翻译等。第 5 层（最高层）为应用层（为行业的解决方案），如人工智能在金融、医疗、安防、工业、军事、交通、设计和游戏等领域的应用。

在 AIGC 行业中，算法、算力、数据是三个核心概念，它们共同构成了这个领域的基础设施。算力是指计算设备执行算法或处理数据的能力，包括 CPU、GPU、FPGA、ASIC 等。此外还有张

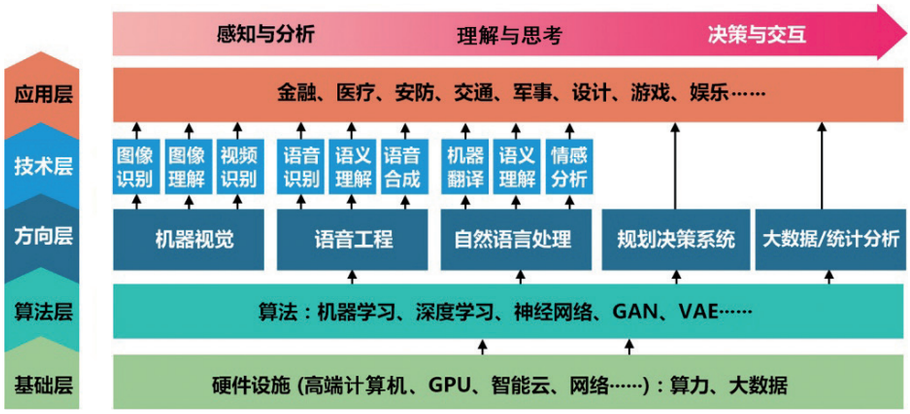


图 3-6　AI 技术的 5 个层面：基础层、算法层、方向层、技术层和应用层

量处理器（tensor processing unit，TPU），Google 开发的高速 ASIC 集成电路，专门用于加速机器学习的硬件等。算力是 AIGC 发展的重要基础设施。例如，据微软 Azure 云服务的高管透露，为了向 AI 初创公司 OpenAI 的深度学习提供支持，微软 2019 年开始为 OpenAI 打造了一台由数万个 A100 GPU 组成的大型 AI 超级计算机，成本或超过 10 亿美元。微软还在 60 多个数据中心总共部署了几十万个可进行智能计算的 GPU（图 3-7，左）。OpenAI 的训练需要处理海量数据以及拥有超大参数规模的 AI 模型，这些需要长期访问强大的云计算服务。为了应对这一挑战，微软必须想方设法将数万个超高速 GPU 卡串联在一起。

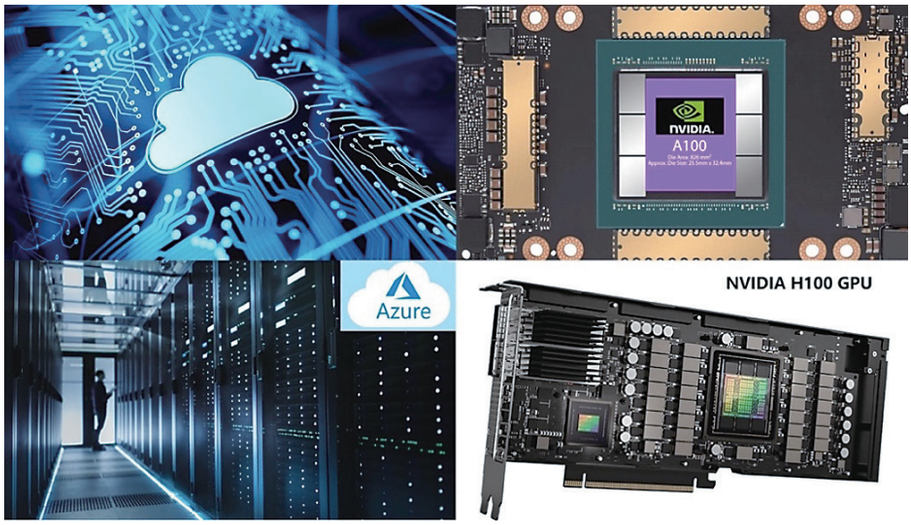


图 3-7　微软 Azure 云服务和数据中心（左）英伟达的 GPU 高速芯片（右）

超大规模的算力虽然耗费金钱，但涉及科技竞争与国家安全。2022 年 8 月，为了保持美国在人工智能领域的技术垄断优势，美国商务部以国家安全为由决定禁止英伟达（NVIDIA）向中国客户销售其两款最先进的、对 AI 技术至关重要的 A100 和较新的 H100 芯片（图 3-7，右）。有了云计算就相当于有了 AI 算力，就可以获得 AI 大模型和生成式 AI 应用的能力，也是中美两国在人工智能领域竞争的关键。中美之间的芯片之争是未来谁能够掌握人工智能领域制高点的关键，更是国运

之争。目前英伟达和 AMD 在高性能计算和人工智能领域具有丰富的产品线和完善的生态系统，外加积累的技术优势和市场地位，预计仍将长期维持 AI 算力芯片领域的龙头地位。与此同时，国内 GPU 研发企业海光信息和寒武纪等也在奋起直追，为国内的 AIGC 发展奠定算力基础。2023 年 8 月，华为推出了 Mate 60 Pro 手机。该机采用了中芯科技自主研发的麒麟 9000S 的 7nm 芯片，许多测试结果显示该手机的速度可以与最新款的 iPhone 5G 手机一样。由此也说明了我国在半导体芯片制造领域的巨大发展潜力。

3.2 大模型与多模态技术

3.2.1 大模型的限制因素

生成式人工智能与设计师高水平的创造力相结合，可以推动互动娱乐以及文化创意产业步入一个新的台阶。生成式人工智能技术与绘画、建筑设计、交互艺术、动画以及电子游戏领域的结合可以为观众带来全新的视觉体验与互动体验，并为教育与服务带来新的生产力，由此推动社会创新发展（图 3-8）。由于 AI 技术的进步和更复杂算法的出现，同时随着质量上升和成本下降，包括 ChatGPT、Midjourney、Dall-E、Stable Diffusion 在内的一系列生成式 AI 应用程序正在迅速普及。越来越多的企业正在将 Al 智能服务纳入业务流程以减少企业人力负担并节省成本。AIGC 具体应用包括 AI 写作、AI 配乐、AI 视频生成、AI 语音合成、AI 绘画、AI 动画、AI 编程等，设计企业可以通过应用编程接口（API）调用这些程序，并运用提示词训练和前缀学习（prefix learning）等工程技术，针对自身的具体需求定向设计产品外观与表现。

图 3-8　AI 插画：人工智能为教育与服务带来新的生产力

与人们的认识不同，大模型并非无所不能的“神器”。到目前为止，机器学习模型仍无法通过其预编程和数据集训练完全自主进行学习和完成创造活动。预训练大语言模型基于其强大的算力与

数据分析能力涌现出了某些表层的“智慧”，但其背后是无数收集源图像和文档的分析师，对数据进行分类的标签员、注释员以及编写代码来训练和测试数据的工程师的努力。在人工智能的幕后，那些看不见的肯尼亚工人对我们最终体验到的 AI 创造性产生了强大的影响（美国《时代》杂志曾发布调查报道称，为了降低 ChatGPT 的危害性，OpenAI 以低廉的价格雇佣肯尼亚工人打标签以训练 AI，由此 ChatGPT 内置的检测器就可以过滤掉暴力、仇恨或种族言论）。训练大型语言模型，如 GPT 系列涉及大量的计算资源和时间。训练时间和迭代周期取决于多个因素，包括模型的规模、训练数据的大小、计算硬件的性能等。

生成式人工智能本身就是一个需要不断改善的迭代原型。例如，目前 AI 大模型并不能回答预训练模型发布日期以后的知识问题（如 ChatGPT 3.5 不能回答 2022 年初知识截止日期以后发生事件）。AI 大模型在开发过程中需要进行不断改进和完善，包括底层模型、算法，以及对数据的分类和标记。对底层模型和算法的改进意味着提高生成式人工智能的学习速度、准确性和效率。此外，设计师可以对机器学习模型中的分步指令进行调整，以使其更好地根据训练数据做出准确的预测。例如，Midjourney 第 6 版推出后，一些人通过输入“故宫、天坛、长城”等国内著名景点进行测试，结果发现虽然画面色彩丰富，主体建筑也非常仿真，但环境呈现则是错误百出（图 3-9，上）。这里除了大模型本身的训练不够外，更多是由一些人为因素造成的，如提示词带有了“林木环绕”“寺庙”“科幻风格”“异域奇幻风景”等不正确的提示，这导致了 AI 生成了错误的图像。笔者用“北京天坛、祈谷坛祈年殿、直径 32.72 米的圆形建筑、鎏金宝顶蓝瓦、三重檐攒尖顶、层层收进、总高 38 米、现实风格”的提示词，就得到了准确的 AI 生成图像（图 3-9，下）。因此，AI 绘图过程中，人工干预是确保系统满足人类需求的关键因素，可以防止 AI 系统出现幻觉或偏见等问题。因此，AI 是一个不断演变和改进的创造工具。设计师需要引导大模型，并通过与 AI 的迭代对话来训练和完善创作过程，这包括从 AI 系统中矫正幻觉并定向启发创意设计。

图 3-9　部分 AI 生成的错误图像（上）和矫正后的准确图像（下）

3.2.2 AIGC 多模态技术

在日常生活中，视觉和语言是最常见且重要的两种模态，而视觉大模型可以构建出人工智能更加强大的环境感知能力，语言大模型则可以学习人类文明的抽象概念以及认知的能力。然而 AIGC 技术如果只能生成单一模态的内容，那么 AIGC 的应用场景将极为有限，不足以推动内容生产方式的革新。深度学习的发展不仅在于推动文本或绘画等静态媒体的智能生成，而且会走向多模态大模型智能创作，如动画、音乐、视频、短片与微电影等动态媒体，同时结合智能编程技术让虚拟世界的融合性创新成为可能，由此能够极大丰富 AIGC 技术可应用的广度。多模态大模型致力于处理不同模态、不同来源、不同任务的数据和信息，从而满足 AIGC 场景下新的创作需求和应用场景，如 VR、AR、数字虚拟人、数字孪生与人机交互。多模态大模型拥有两种能力，一个是寻找到不同模态数据之间的对应关系，例如，将一段文本和与之对应的图片、视频或交互场景联系起来；另一个是实现不同模态数据间的相互转换与生成，例如，根据一张图片生成对应的语言描述，或者根据语言描述创作基于 VR 场景的游戏等。

为了寻找到不同模态数据之间的对应关系，机器学习会将不同模态的原始数据映射到统一或相似的语义空间中，从而实现不同模态信号之间的相互理解与对齐，这一能力最常见的例子就是互联网中使用文字搜索与之相关图片的图文搜索引擎。在此基础上，多模态大模型可以进一步实现文本、图像、音频、视频等数据间的相互转换与生成，这一能力是 AIGC 实现原生创作的关键。AIGC 技术被广泛应用于音频、文本、视觉等不同模态数据，并构成了丰富多样的技术应用。其前沿领域重点为智能化风格迁移、数字内容孪生、数字内容创作和数字内容编辑（图 3-10）。

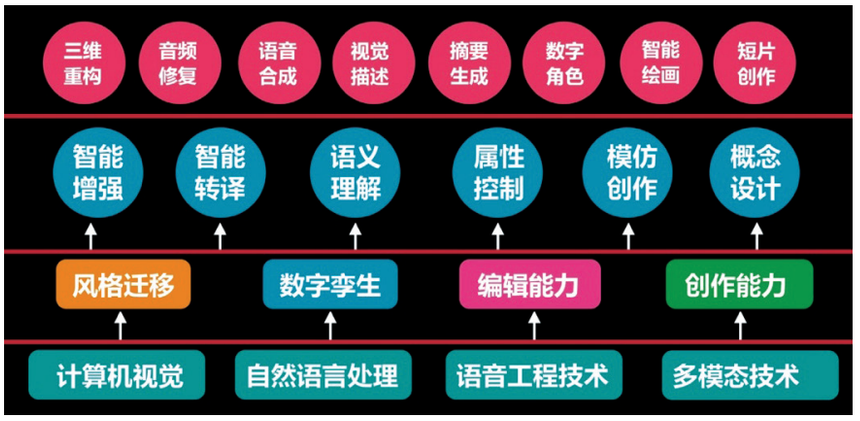


图 3-10　AIGC 多模态技术：风格迁移、数字孪生与内容创作与编辑

2023 年，一系列支持 AIGC 内容创作和多模态应用的通用模型开始向应用场景渗透，助推 AIGC 创作更加多样化的内容。其中，多模态技术能够让 AI 进行文字、图像、声音、语言等不同模态内容融合的机器学习实现文字生成图像、动画、视频以及虚拟人生成等功能，丰富了 AIGC 技术的应用广度。例如，2023 年由美国华裔创办的智能动画工具 Pika Labs 能够生成和编辑 3D 动画、动漫、卡通和电影（图 3-11）。据了解，Pika Labs 已经获得 5500 万美元融资，成为文字生成动画领域的又一家有竞争能力的科技初创企业（另一家为 Runway 公司）。据福布斯报道，Pika Labs 估值目前在 2 亿 ~3 亿美元。除了生成视频，Pika 1.0 还可以实现对现有视频素材中的元素进行修改、更替，例如，更改视频人物的着装，为视频中的“猩猩”戴上墨镜，转换视频的风格，等等。只需要在视

频编辑器中写下提示词，即可生产高质量的视频，或者对视频元素进行编辑和修改。这些创作推动了 AI 多模态技术走向媒体与视频领域。2024 年初，OpenAI 公司推出了其首个文生视频模型 Sora。该模型可以生成长达 1min 的视频，这一技术突破预示着视频内容创作领域的重大变革。虽然截至 2024 年 8 月该技术尚未公开提供公众尝试，但 OpenAI 发布的一系列范例视频已经引发了全球的关注。这些生成视频表明 Sora 模型是迄今为止最令人印象深刻的文生视频工具。

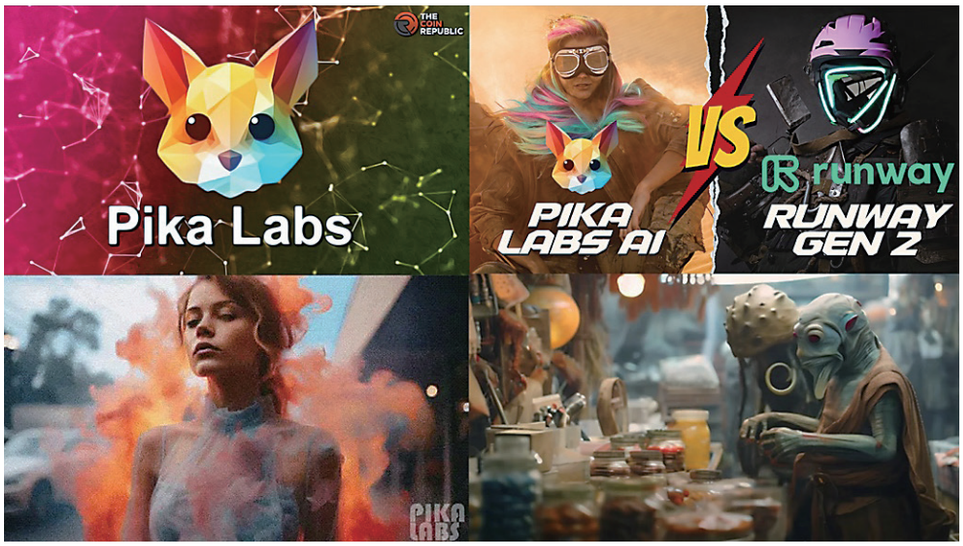


图 3-11　动画领域的智能化创作工具 Pika 和 Runway 网站

3.3　大模型与转换器技术

从 AIGC 的发展路径上，我们可以看出生成式人工智能朝向多模态、多场景和通用性技术的发展趋势。2014 年，生成对抗网络（GAN）的出现标志着 AIGC 开始步入自主生成阶段；2017 年，转换器（transformer）算法使得深度学习摆脱了人工标注数据集的缺陷，大幅减少了模型训练的时间并成为主流模型架构基础。2022 年，扩散模型成为最具影响力的图像生成算法，大语言预训练模型推动 AIGC 应用场景的多元化与智能涌现。该技术使训练大规模数据参数成为可能，帮助 AI 实现机器学习、强化学习和多任务、多语言、多场景的内容生成。在复杂的内容消费场景下，生成算法内容质量得到了进一步提升，能够满足灵活多发、高精度、高质量的内容需求。因此，智能设计充分体现了大模型的技术特点。

3.3.1　转换器的工作原理

大型语言模型（如 GPT 和 Midjourney）之所以具有通用性的特征，是因为它们是在庞大而多样的语料库上进行预训练的，其参数数量可高达千亿以上。这些 LLM 模型使用了转换器架构，可以用来写故事、散文、科研论文、诗歌、回答问题、翻译、聊天，甚至还可以通过大学入学考试或专业资格考试。该架构具有的自注意力机制（self-attention），使得模型能够捕捉长距离的依赖关系和上下文信息。转换器（或称变换器）首次由 Google 研究人员在 2017 年发表的论文《你需要的是

注意力》中提出，并用于解决机器翻译等自然语言处理的任务。转换器摒弃了传统的循环神经网络（RNN）和卷积神经网络（CNN）的方法，而是采用了一种称为自注意力机制的方法，来捕捉文本中的长距离依赖关系，提高了模型的效率和准确性。这些模型通过对大规模文本数据进行预训练，学习语言的语法、语义和上下文关系。这使得它们在多种自然语言处理任务上都能表现出色，因为它们能够理解并生成自然语言的多层次结构。

转换器模型（图 3-12，左）的架构及工作原理并不复杂，它只是一些非常有用的组件的串联，每个组件都有自己的功能。转换器模型由编码器和解码器组成，但其关键因素在于三大核心机制：位置编码（positional encodings）提供词语顺序信息；注意力机制（attention）让模型可以关注关键词语；自注意力机制（self-attention）能够帮助模型学习词语间的依赖关系。这三者相辅相成，使转换器模型得以模拟人类语言处理的方式，达到传统大模型难以企及的效果。转换器有 4 个主要部分：标记化、嵌入、位置编码、多个转换器块以及 Softmax 函数（图 3-12，右）。其中转换器最为复杂，其中许多串联的模块都包含两个部分：注意力模块和前馈组件。从工作流程上看，标记化（tokenization，或分词）是模型开始工作的第一个步骤，该模块由大量标记数据集组成，包括所有单词、标点符号等。Token 可以理解为最小语义单元或“词元”。标记化就是模型获取每个单词、前缀、后缀和标点符号，并将它们发送到库中的已知标记。例如，如果你向 ChatGPT 输入的信息是“写一个故事。”，那么对应的 4 个标记将是 < 写 >、< 一个 >、< 故事 > 和 <。>。一旦输入信息被标记化后，模型就可以将单词转换为数字或向量，这个过程就是嵌入（embedding）。

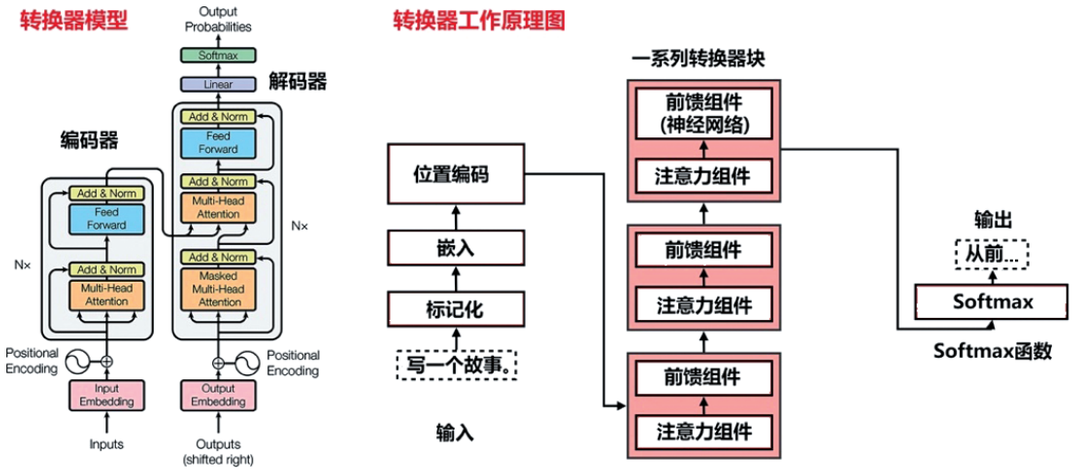


图 3-12　转换器模型（左）与转换器工作原理（右）

从原理上看，为了掌握一个词（如“故事”）的意思，LLM 首先使用大量训练数据并观察“故事”的上下文和邻近词。大模型通过互联网上海量文本可以收集这个数据集，即 LLM 使用数十亿个词汇进行“训练”或调试，由此能够得到与“故事”在训练数据中一起出现的词集，如“讲”“写”“科幻”等。模型处理这个词集时会产生一个向量或数值列表，并根据每个词在训练数据中与“故事”的邻近程度来调整它。这个向量也被称为词嵌入。一个词嵌入可以包含数百个值，每个值表示一个词意的不同方面。就像我们描述一座房子可以用诸如类型、位置、卧室、浴室、楼层等参数，嵌入的值可以定量表示一个词的语言特征。以上过程就是模型生成文本的第一步。位置编码为每个单词添加一个位置向量以跟踪单词的位置，例如，“写（1）”“一个（2）”“故事（3）”“。（4）”，由此可以保证在自然语言处理任务中的词语顺序。

当以上步骤完成后，LLM 的下一步是预测这句话中的下一个单词。这是通过一个非常大的神经网络完成的，其中注意力组件就是帮助模型理解上下文的关键要素之一。注意力机制允许转换器模型在生成输出时，参考输入序列中的所有词语，并判断哪些词对当前步骤更重要、更相关。例如，当转换器要翻译一个英文单词时，它会通过注意力组件快速“扫描”整个英文输入序列，并判断应该翻译成什么中文词汇。如果输入序列中有多个相关词语，该组件会让模型关注最相关的那个，忽略其他不太相关的词语。这种方式与人类大脑的判断非常接近。注意力组件被添加到前馈网络的每个块中。因此，如果你想象一个大型前馈神经网络由几个较小的神经网络块组成，那么每个块都会添加一个注意力组件。

转换器模型中实际使用是多头注意力（multi-head attention）。LLM 使用几种不同的嵌入来修改向量并向其添加上下文。多头注意力帮助语言模型在处理和生成文本时达到更高水平的效率。除了注意力机制外，转换器模型还有一个更强大的机制叫作自注意力机制，其核心思想是允许模型学习词语之间的相关性，也就是词语与词语之间的依赖关系。以句子“我爱吃苹果”为例，通过自注意力，模型会学习到“我”与“爱”有关，“爱”与“吃”有关，“吃”与“苹果”有关。然后在处理时，模型会优先关注这些相关词语，而不是简单按照顺序一个字一个字地翻译。从认知角度来看，自注意力更贴近人类处理语言的方式。因为人类对一个事物的认知不仅在于事物本身，而且会结合事物所处的环境来认识，自注意力机制同样如此。

3.3.2 大模型的语言认知力

通过学习转换器的工作原理，我们可以理解 LLM 能够借助对语言的“理解”而接近与媲美人类智能的本质。语言被认为是人类智能的核心，这是因为语言反映了人类智能的本质和复杂性。语言是一种复杂的符号系统，不仅可以用来表达和传递思想、感情、信息和意图，而且能够进行高效的沟通和交流，包括分享知识，合作完成任务以及建立社会联系等。人类的语言能力远远超越了其他生物的交流方式，也是人类文明能够统治地球的关键因素。语言不仅是一种沟通工具，还是思维和表达的工具。通过语言，人类能够思考、计划和预测未来，并将这些复杂的思维过程转换为可以被族群与社会公众共享和理解的形式。语言使人类能够表达自己的想法、观点和创意。语言还可以使人类能够进行抽象思维，超越直接感官经验。人们可以用语言来讨论抽象的概念、逻辑关系和抽象的数学思维。这种能力有助于人类进行复杂的问题解决和创新。从人类文明角度，语言是文化的重要组成部分，通过语言，人们传递文化价值观、传统知识和社会规范。语言是一种独特而强大的认知工具，它不仅促进了人与人之间的交流，还支持了思维、学习和文化的传承。这些特征使得语言成为人类智能的核心，并在人类发展中发挥着至关重要的作用。历史学家尤瓦尔·赫拉利指出：“每一种人类文化的操作系统都是语言——最开始的时候是文字。我们用语言创造神话和神灵，用语言创造金钱，创造艺术和科学，创造友谊和国家。”

人工智能可以和人类进行语言对话，意味着破解了人类文化的“操作系统”。转换器模型使得 AI 在自然语言理解和生成方面取得了显著突破，也成为生成式人工智能具备“通用性”的基础。这些模型可以在多个任务上进行微调，而无须用户从零开始训练，可以大大提高大模型的工作效率。例如，不同专业的设计师可以针对特定的任务，如建筑设计、室内设计、文案撰写或者角色生成等，通过提示词与参考图等对大模型进行定向训练，由此大模型就可以适应特定领域或任务的数据，进而实现更加专业和定制的性能。高盛集团的一份报告指出：在过去 80 年里，60% 的职业是新兴技术所创造的。因此，对于设计类大学生来说，基于 AIGC 的智能设计不仅意味着挑战，也预示着新

的机会。如果大学生能够在现有的知识体系的基础上，通过跨界学习与反复实践，掌握 AIGC 的设计方法与特征，就可以总结出智能设计的规律性，让自己的技能通过 AIGC 实现跨越式的提升。我们可以通过输入提示词：“一位女设计师正在 AIGC 的协助下进行服装设计。”，借助 AI 绘图再现这个“双赢”的设计场景（图 3-13）。

图 3-13　通过智能绘画表现设计师与 AIGC 一起协同工作的场景

3.3.3　AI 算法与智能设计

复旦大学管理学院院长陆雄文指出：“本轮科技革命的发展，与传统意义上的科技革命有着显著不同。”“它是多赛道并举，相互交融，形成聚能，然后爆发。”基于 AIGC 的智能设计不是针对某一具体艺术设计学科或专业，而是全面影响艺术设计所有的学科和专业，从广告策划到文案写作，从设计草图到高清原稿，从动画脚本到漫画角色设计，从界面设计到后台程序生成，等等。因此，生成式 AI 全面介入设计与创作流程已经成为当前设计学、美术学、建筑学和戏剧影视学等学科都无法回避的现实。AIGC 也成为设计教育新的竞争对手、新的创意资源和工具。AI 不仅带来了创意工具的转换，还带来了艺术与设计观念上的变革。智能设计的核心是人机协同与算法驱动的双重创意系统。与数字媒体时期主要基于人工或基于人工 + 软件的设计模式不同，智能时代的设计流程、设计方法、设计工具和审美法则都在发生重大变化。

最重要的一点是智能设计主要借助大模型算法和机器学习技术，利用大规模数据集进行训练，从中学习并生成具有创意性的作品。大模型能够在设计师的指导下，通过迭代过程不断优化设计方案。这包括对设计参数的调整和重新生成设计的能力，以便使生成的作品逐步接近或满足设计目标。传统艺术设计主要依赖于有限的数据和个体经验，因此很难与之竞争。智能设计系统通常允许与用户进行交互，以便更好地理解用户的需求和偏好。这种交互可以通过用户输入、反馈或调整设计参数来实现。例如，Midjourney 作为目前一款最火的 AIGC 绘图软件之一，虽然在生成图片时让人感觉很惊艳，但是由于大模型的随机性很强，生成角色人物时（如花木兰形象）就面临一个很大的难题：如何保持角色输出的一致性呢？ Midjourney 在 /setting 设置指令下提供的重新混合模式 Remix（图 3-14）就可以帮助解决这个问题。该模式可以修改生成图的提示词，并基于原图进行新的关键

词优化。因此，Remix 模式能帮我们在设计中控制画面一致性，改变你希望改变的地方，例如图片局部的颜色、背景、主题或构图。

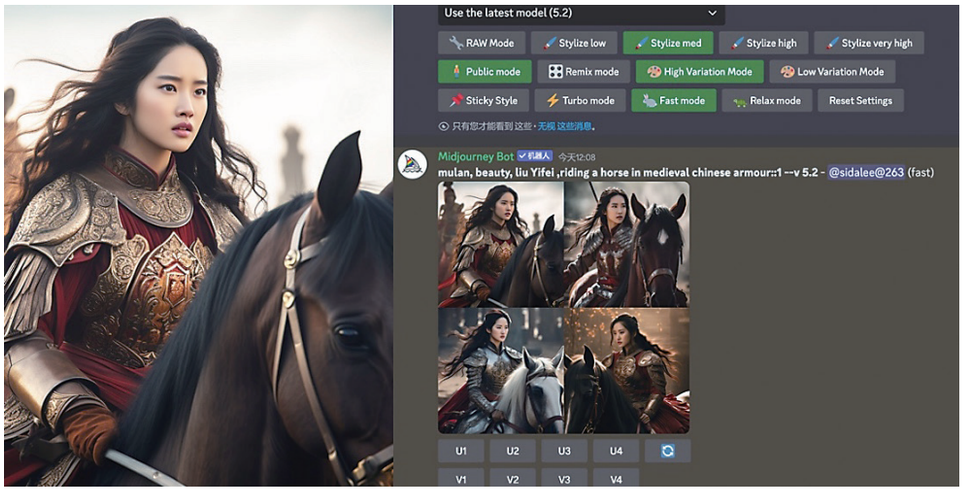


图 3-14　Midjourney 的混合模式使设计保持一致性

例如，在上面的范例中，我们希望在花木兰骑马征战的原稿中，补充加入硝烟战火以及光照、服饰等信息，为此在指令栏中输入 /prefer remix 命令来打开 Remix 重新混合模式，然后单击缩略图下面的 V 变化按钮（变成绿色）。通过修改弹出窗口中的提示词，就可以实现定向设计的效果（图 3-15）。因此，大模型能够通过设计师的反馈来改进设计原型，设计师可以通过语言迭代来优化生成图像。

图 3-15　通过修改合成模式的提示词可实现定向设计

除了混合模式外，几乎所有的 AI 大模型绘画平台都提供了“局部修改”或“局部重绘”的界面与功能。例如，Midjourney 在 4 宫格图下就提供了 Vary Region（局部修改）的按钮。单击该按钮后弹出的“局部重绘”界面中提供了框选和套索工具（图 3-16，右）。如果我们希望修改原稿的部分内容如星巴克杯子，就可以框选该区域，并在输入框中输入修改的提示词：“在杯子上加入‘Merry

Christmas’字样”。单击“确定”后，AI 生成的图片就将输入的英文替换了乱码（图 3-16，左和右）。大模型具有一定程度的人机协同和交互性，能够根据设计师的输入或反馈调整生成的作品。该过程不仅高效便捷，而且可以形成新的创意，由此提升设计师或艺术家的审美和创造力。

图 3-16　通过“局部修改”按钮可以添加或重绘生成画面的部分区域

3.4　人工智能与敏捷设计

3.4.1　AI 高效性与多样性

前文指出：智能设计主要借助大模型算法和机器学习技术，利用大规模数据集进行训练，从中学习并生成具有创意性的作品。因此，AI 具有自动化和高效化的特征。智能设计系统能够在设计过程中执行某些或全部任务，包括自动生成设计概念、优化设计参数，甚至生成最终设计图纸的能力。AI 可以快速地生成大量的艺术作品，而无须人工干预和监督。这使艺术创作的效率大大提高，也同时节省了人力和物力的成本。同时，智能设计也具有多样性特征。AI 大模型可以根据不同的数据和参数，生成不同风格和形式的艺术作品，这使艺术创作的可能性和选择性增加，促进了艺术创作的创新和探索。以“小红帽”原型进行故事角色设计为例，我们借助 AI 就可以得到“森林里的小红帽”“森林里的小红帽与狼”“采蘑菇的小红帽”“骑龙的小红帽”等一系列形象（图 3-17），能够给设计师提供更多的创作灵感与情境想象力。

人机协同创作模式对艺术家和设计师提出了更高的要求。以“小红帽的故事”这个主题进行创作为例，在人机协同智能设计出现以前，传统的创作方式包括选题研究与资料检索，创作团队成员集体“头脑风暴”小组会出方案，策划师提供故事创意及文案脚本，漫画师根据故事设计角色与分镜原稿，计算机师负责出图渲染，设计团队与甲方一起开会研讨设计原型方案，根据甲方意见进行深入修改。以上步骤不仅环节多，周期长，而且往往因为彼此沟通的问题而陷入停顿。随着 GPT4、Dall-E3 等智能设计工具的广泛采用及远程协同设计的普及，设计师的工作效率大幅提升。智能文案、智能出图与协同设计使设计环节大大减少，增强了设计团队与企业的市场竞争力。设计团队能够借助 AI 工具实现“敏捷设计”模式，小步快跑，及时反馈，动态出图，将冗长烦琐的“瀑布流”式的设计改革为更灵活、更紧凑的智能化设计模式。与此同时，我们也必须了解 AI 设计的缺陷与问题。例如，在训练不足的情况下，机器生成的结果常常过于跳跃，如“小红帽给狼开会”这个提示词就会生成一系列令人啼笑皆非的画面（图 3-18），这说明我们对 AI 大模型的理性认识还需要进一步深入。

图 3-17　通过输入不同的提示词可以指导大模型输出不同的主题画面

图 3-18　通过智能绘画表现了小红帽与狼一起开会的场景

3.4.2 人工智能赋能敏捷设计

敏捷设计是一种借鉴软件工程的“敏捷开发”（agile development）模式的新型设计方法，是一种以人为核心、以智能设计平台为辅助，通过大模型原型迭代的产品设计与开发方法。该方法是一种应对快速变化的需求的开发能力，只要在符合价值观和原则的基础上，能让开发团队拥有应对快速变化需求的能力就称为敏捷开发。与传统设计相比，敏捷设计还能更快地对接市场，缩短产品开发周期。AIGC 时代的敏捷设计是一种高度人机协作化的工作方式。设计师会指导 AI 设计平台完成每一次设计原型迭代。这种开发模式可以最大限度地发挥设计师的主观能动性。特别是借助计算机快速文生图的能力，设计师可以缩短与用户沟通反馈的周期，加快提交给用户产品原型的速度（图 3-19）。基于人工智能平台的敏捷设计不仅出图快速，修改方便，而且可以进行“滚雪球式”的产品开发。敏捷设计紧抓用户的刚需与痛点，要求设计师可以从简洁的设计语言开始，发挥 AI 的威力，不要在视觉设计和实现效果上花费过多的精力，及早和持续不断地交付有价值的产品创意让客户满意。

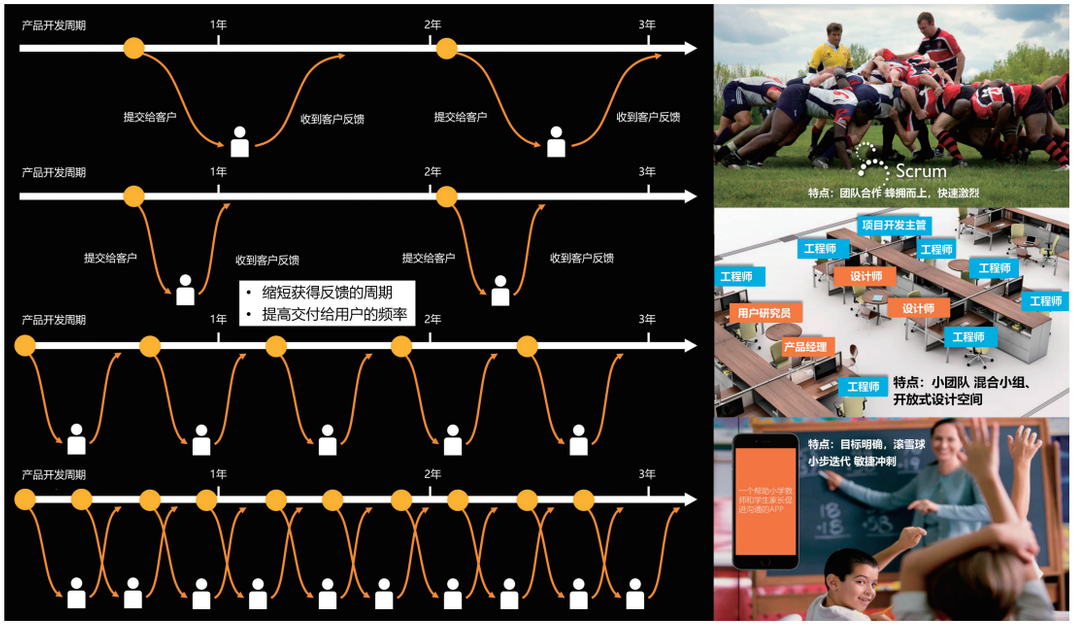


图 3-19　敏捷开发能够缩短与用户沟通反馈的周期，加快产品设计速度

智能科学的快速发展，推动了以人机协同、大模型和敏捷设计为代表的智能设计思想的成熟，这将会逐步替代基于设计思维的设计范式（图 3-20）。设计思维主要是基于观察、调研和用户导向的设计方法，但在时间与效率上无疑落后于当代科技的发展速度。智能大模型不仅能够代替美工师完成海量抠图、构图和多种类型的设计，而且能够借助数据统计与分析改变传统市场与用户的研究方法。随着基于 AIGC 的协同设计方法的普及，预计会有越来越多的企业将智能交互设计平台纳入企业的工作流程。

图 3-20　智能设计是基于人机协同、大模型 / 敏捷设计的流程与方法

3.4.3　大数据与趋势研究

AI 设计通过学习大量数据，有可能发现艺术领域的新趋势和创新点。大模型可以从以往历史设计案例、用户反馈和其他相关数据中筛选合适的资源，为设计师提供更精确和有针对性的设计生成。限于经验、资源、管理与组织结构的不足，小型设计工作室给出的设计方案很难与 AI 设计相抗衡。例如，ChatGPT 3 是由 1750 亿个参数组成的大型神经网络模型。这种大规模的参数量使得该模型能够处理复杂的自然语言理解和图形生成任务。2023 年秋季，OpenAI 公司推出了 GPT 4 Turbo，其参数量高达 1.5 万亿。因此，这些大模型具有更强大的表示能力和学习能力，能够更好地捕捉数据中的复杂模式和关系。更重要的是，大模型服务于全球十几亿的用户，这使得模型在广泛而多样的用户群体中得到验证和应用，从而提高模型的通用性。模型在全球范围内接触到的多元文化和多样化的信息有助于形成更加丰富和全面的设计方案。

由于大模型学习了全球范围内的创意和设计案例，其生成的设计方案更具有创意独特性。这些大模型能够汲取各种设计风格和趋势，形成更具前瞻性和独创性的设计理念。正所谓“三个臭皮匠赛过诸葛亮”，相比设计师单打独斗和冥思苦想的设计，LLM 形成的设计方案不仅更具有通用性，而且还兼有国际视野和创意独特性。我们以咖啡图标设计为例，通过提示 AI 智能设计系统：“请设计一个咖啡店的图标（圆形），徽标中心为咖啡杯子图案，周围的元素有咖啡树叶子的花环，最外侧有‘tulip coffee，since1960’英文字样环绕，简约风格，细节表现，清晰醒目，色彩充满活力。”几分钟后，AI 就可以生成各种创意设计草案（图 3-21），虽然这些图标设计还需要进一步修饰加工，但 LLM 的高效率与多样性可以超越多数设计师的能力。

在上面的咖啡图标设计的案例中，智能设计所具有的实时性、高效性特征是设计团队的“制胜法宝”。众所周知，设计活动本身具有“时间紧、任务急”的特点，许多设计方案都有时间限制或方案提交截止日期。逢年过节，设计师忙企划、赶设计加班加点到深夜更是家常便饭。在这种压力下，能够实时生成设计方案的 AI 系统更是“雪中送炭”，可以成为设计团队的“救星”。因此，智能设计的出现使得设计师能够从容不迫，更有效地安排工作时间与设计进度，从而更高效地推进设计流程。