



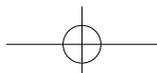
AIGC 概述

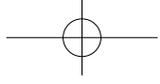
视觉传达设计正处于持续演进的阶段。人工智能与数字工具的广泛应用，使设计手段与创作流程发生了明显变化。图像生成、素材调用、画面调整等环节逐步整合进智能系统，设计已从依赖手绘与现成资源的方式过渡到以指令驱动图像输出的流程。工具结构更加灵活，创作节奏更为高效，设计师在构思图像时，可以在多个方案中快速判断、筛选、微调，从中捕捉最贴近主题的画面方向。随着操作方式变化，设计思维发生转变，从原有的图形堆叠，延展为围绕目标展开的生成式探索。人工智能不替代设计，而是构建出一套新的工作方式，为图像创作打开更多可能。

1.1 AIGC 的定义与发展历程

1.1.1 AIGC 的定义

AIGC (artificial intelligence generated content, 人工智能生成内容) 是基于深度学习、计算机视觉、自然语言处理等技术，让 AI 具备了自主生成文本、图像、视频、音频等内容的能力。它的工作方式不同于 Photoshop 之类的传统设计工具，也不同于设计模板的自动化生成，其通过学习海量数据，掌握设计语言、风格和构图法则，并根据输入的指令自动生成符合语境的视觉作品。换言之，AIGC 不单单是“生产工具”，也是一个能够自主“思考”和“创作”的系统。





1.1.2 AIGC 的发展历程

20 世纪,人们对人工智能的想象多半停留在科幻作品中。《2001 太空漫游》中的 HAL 9000 是冷酷无情的超级智能,《银翼杀手》里的复制人却带着人类无法理解的复杂情感,《黑客帝国》中的矩阵系统则操控着整个世界的数据流(图 1-1)。但在现实中,人工智能的发展并没有一开始就奔向这些“全能 AI”的方向,而是走了一条漫长而曲折的道路。从最初的逻辑推理,到后来让机器学会理解语言、创造艺术、生成图像,AIGC 的发展史是一部充满突破与挑战的科技史诗,每一次技术的跃迁都在不断拓展人类对创造力的认知边界。

1. 早期探索与萌芽(20 世纪 50 年代—20 世纪末)

1956 年,达特茅斯会议(图 1-2)宣布了“人工智能”概念的诞生,但在那个年代,计算机的运算能力极其有限,研究者更多是在纸面上构思 AI 可能的未来。艾伦·图灵(Alan Turing)提出的“图灵测试”试图定义机器是否能思考,而在现实测试中,机器甚至无法正确识别一张图片。尽管如此,科学家们仍然尝试让计算机模仿人类的创造力,哪怕只是简单的文字处理,也足以让当时的研究者兴奋不已。

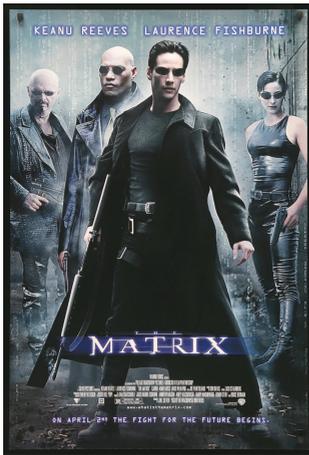


图 1-1 《黑客帝国》海报



图 1-2 达特茅斯会议

1966 年,ELIZA 诞生。这个由麻省理工学院约瑟夫·维森鲍姆(Joseph Weizenbaum)开发的程序(图 1-3),能够模拟心理治疗师与用户进行简单对话。当用户说“我今天很难过”时,ELIZA 会回应:“你为什么会觉得难过?”它的回应方式依靠关键词匹配,利用重复和改写用户输入来构建句子。这种对话形式在当时引起广泛关注。其实 ELIZA 不具备真正的创造力,但它让计算机生成文本的可能性进入了公众视野。

2. 机器学习的兴起与 AIGC 的突破(21 世纪初)

进入 21 世纪,计算机算力持续提升,人工智能研究进入加速发展阶段。2006 年,杰弗里·辛顿(Geoffrey Hinton)(图 1-4)提出深度学习概念,推动 AI 从依赖规则的处理方式,转向基于数据的训练机制。深度学习让 AI 能够识别模式、调整结构、优化结果,图像、语言、音乐等领域的内容生成能力随之拓展,为 AIGC 的发展提供了技术基础。

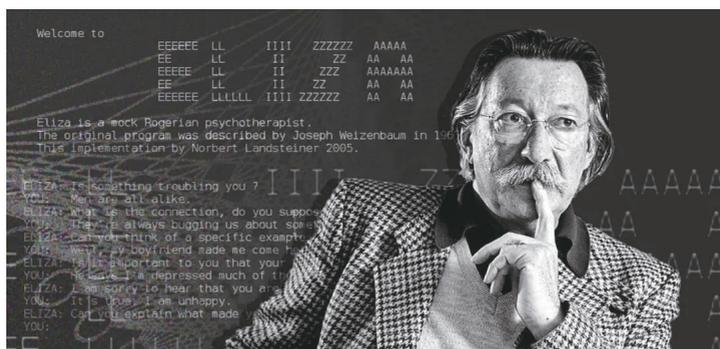


图 1-3 ELIZA 与它的创作者约瑟夫·维森鲍姆

2014 年,蒙特利尔大学的伊恩·古德费罗 (Ian Goodfellow) (图 1-5) 提出了 GAN (生成对抗网络) 模型,彻底改变了 AI 生成内容的方式。GAN 的结构像是一个“对抗性画家”——一部分 AI 负责“画画”(生成内容),另一部分 AI 负责“挑刺”(判别内容质量),两者互相比较劲,让 AI 生成的最终作品越来越接近人类创作。GAN 的引入使 AI 具备了构建完整图像的能力,输出内容逐步呈现出更完整的构图与更连贯的细节。2015 年,DCGAN 引入卷积神经网络结构,提升了图像纹理的表现力。2017 年,WGAN 在训练机制上进行优化,增强了图像生成的稳定性与一致性。2018 年,NVIDIA 推出 StyleGAN,可对图像特征进行调整与组合,支持人物形象的自由变化与风格生成,使 AI 图像创作具备更高的可控性与表达力。



图 1-4 杰弗里·辛顿



图 1-5 伊恩·古德费罗

3. AIGC 进入全民创作时代 (21 世纪 20 年代至今)

2020 年后,AIGC 进入快速发展阶段,AI 生成内容在多个领域得到应用,逐步成为日常创作的重要工具。GPT-3 的发布显著提升了 AI 的文本生成能力,其可用于撰写新闻、故事、诗歌与代码,语言组织更具结构性与连贯性。2021 年,OpenAI 推出 DALL-E (图 1-6),实现基于文本生成图像的功能,为视觉创作提供了新路径。2022 年,Stable Diffusion 开源,使图像生成工具更加普及,用户可通过自然语言控制图像构成,提高了操作效率与输出多样性。2023 年,Runway 发布 Gen-2,支持根据文本生成视频内容,推动视觉叙事向自动

4 AIGC 视觉传达设计

生成模式延展。2024 年，OpenAI 推出 Sora，进一步提升视频内容的生成效果，拓展了 AI 在影视制作中的应用范围。

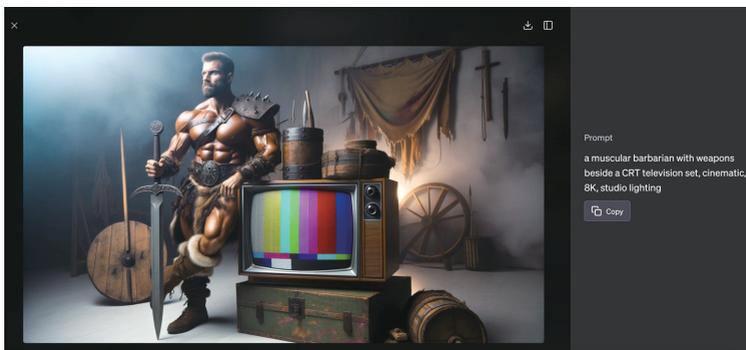


图 1-6 DALL-E 关键词生图

与此同时，中国的 AIGC 技术发展也在持续推进。百度发布文心一言（后更名为“文小言”），支持对话、写作、问答等多种文本任务，并构建出与视觉模型配套的多模态系统。阿里云推出通义千问，强化内容生成的可控性与语义理解能力，适用于写作、文案、营销等多个场景。字节跳动推出豆包，聚焦于创意写作、图文生成与视频辅助脚本，为内容创作者提供轻量化工具。清华大学、复旦大学等高校也陆续推出开源大模型，并在图像、音乐、视频等方向展开实验性应用。这些技术成果共同推动 AIGC 走向日常化、多样化与平台化，为内容创作提供了持续扩展的技术支持。

如今，AIGC 早已不再局限于实验室，而是走进了千家万户，成为创意产业不可或缺的组成部分。Adobe Firefly 让 AI 生成的视觉元素可以无缝融入 Photoshop，让设计师能够更快速地完成创作。Canva AI（图 1-7）使得普通用户也能轻松创建专业级的设计作品。Figma 推出的 AI 辅助设计功能（图 1-8），提高了团队协作的效率。AI 在艺术创作、影视制作、广告营销等多个领域的应用，使得创意生产的方式发生了颠覆性的变化。



图 1-7 Canva AI 功能 Canva Magic Studio

AIGC 的发展为创作带来了新的可能。AI 擅长“模仿”与“组合”，在大规模数据中提取规律，生成具有结构的内容。真正的艺术直觉、文化理解和创意灵感仍由人类设计师

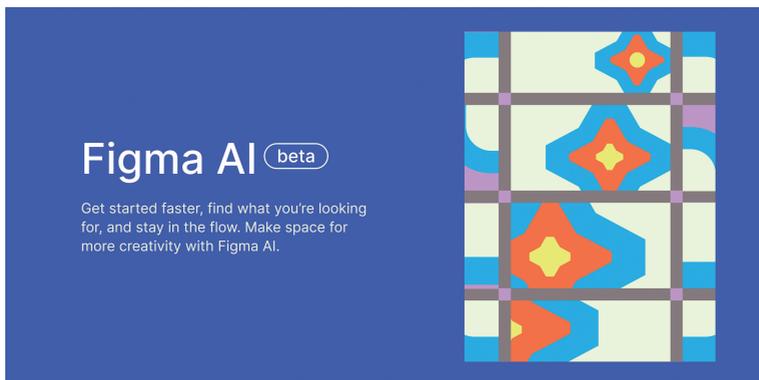


图 1-8 Figma AI 功能

主导。AIGC 在帮助创作者提高效率的同时，也在推动人机协作方式的转变。未来，AI 将逐步发展为智能创作的参与者，深入融入创意行业，协助设计师、艺术家、编剧和导演展开想象力，探索新的表达方式。

1.2 AIGC 图像生成流程

1.2.1 数据准备：学习图像语言，建立视觉知识

AI 的绘画之旅，从一片视觉荒原启程。最初，它像个刚刚接触世界的孩子，眼前充满了混沌：透视是什么？色彩如何搭配？光影能否塑造空间感？这些对人类来说稀松平常的概念，对 AI 来说却完全陌生。如果此时让它拿起“画笔”，它只会在画布上胡乱撒上一堆色块，拼凑出一张混乱无序、毫无逻辑的噪声图。AI 之所以能成为如今的“艺术大师”，正是因为它经历了一场史无前例的视觉训练，在一座巨大的“数字画廊”中，学习如何看、如何理解、如何创造。这座画廊，便是庞大的视觉数据集，它如同一本囊括全球艺术风格、现实场景、奇幻设定的百科全书，为 AI 打造了一个取之不尽的灵感源泉。

在 AI 的世界里，数据就是知识，而 LAION-5B、COCO Dataset、ImageNet、WikiArt 等超大规模数据集，则是它的老师，每一张图片都是一堂无声的课。LAION-5B（图 1-9）作为目前全球最大的开源视觉数据集，几乎是 AI 的“宇宙级画展”——这里有 50 亿张图片，从现实世界的日出日落、乡村城市、森林湖泊，到艺术界的印象派油画、哥特式建筑、未来主义概念设计，应有尽有。AI 的任务，就是在这无穷无尽的图像海洋中，像一名狂热的艺术生一样，疯狂摄取视觉经验，归纳出万物的形态与规则。

但光是看，还不够。AI 不能像一个普通的观众，仅仅停留在“好看”或“不好看”的层面，而是要像一位深度学习的艺术家，拆解画面的结构、分析色彩的分布、揣摩笔触的节奏，甚至推测创作者的意图。为了帮助它完成这一点，每张图片都附带了详细的文本描述，比如“晨曦中的巴黎街头”“战场上挥舞长剑的中世纪骑士”“迷雾笼罩的赛博朋克城市”（图 1-10）。AI 需要像侦探一样，在这些描述与图像之间建立对应关系，推断出“晨曦”意味着柔和的金黄色调，“赛博朋克”代表高饱和度的霓虹色彩，“战场”通常伴随着尘土飞扬和动态姿态。这些知识，最终会成为它创作时的底层逻辑，避免用户在输入“废墟中

6 AIGC 视觉传达设计

的哥特式教堂”时，随意堆砌一个现代化结构建筑，而是能精准调用“哥特式”的尖塔、彩色玻璃、黑暗肃穆的色调，使画面具有风格统一的完整性。（图 1-11）



图 1-9 LAION-5B



图 1-10 MJ 提示词：“晨曦中的巴黎街头”“战场上挥舞长剑的中世纪骑士”“迷雾笼罩的赛博朋克城市”



图 1-11 MJ 提示词：“废墟中的哥特式教堂”

但问题来了，仅仅“看”过亿万张图，AI 就能画出惊艳世界的作品吗？不，绝无可能。如果只会简单地模仿与复制，那它不过是台高级的复印机，而非真正的创造者。AI 还需要理解“画面背后的流程”，学会如何塑造空间、如何构建层次、如何让光影赋予画面情绪。一个人类艺术家在画一幅城市夜景时，路灯的摆放可以用来引导观者的目光；在描绘森林时，树木的错落是在密集与留白之间寻找平衡。这些潜藏在画面之下的视觉原则，是 AI 必须掌握的技能，否则它生成的图像只会是一堆缺乏生命力的像素拼接，而无法真正具备艺术感染力。

为了帮助 AI 进入更深层次的视觉学习，CLIP 学习模型被引入。其训练方式，是让 AI 玩一个“看图说话”的游戏——被喂入海量图片及对应的文字描述，然后要求在图像和文本之间建立匹配关系。

不过，真正的艺术创作，不只是把元素拼在一起，而是要赋予作品风格和灵魂。不同的艺术流派有着独特的表达方式，印象派讲究光影的跳跃，超现实主义偏爱梦幻的错位，赛博朋克擅长塑造冷峻的科技氛围。如果 AI 不了解这些风格，那么它生成的画作就会显得机械而呆板。为了让 AI 能够真正理解风格，研究人员在训练过程中向它提供了大量艺术流派的数据集，并让它学习不同画风的特点。比如，AI 在研究莫奈的作品时，会发现印象派笔触松散、光影流动、色彩层次分明；在接触赛博朋克时，则会看到高饱和度的冷色调、霓虹灯的光晕、充满未来感的建筑形态。这样，当用户输入“印象派风格的星际都市”时，AI 才会将未来横跨在星际间都市的霓虹灯转换成柔和的色块，使整个画面呈现出莫奈式的色彩氛围（图 1-12）。



图 1-12 MJ 提示词：“印象派风格的星际都市”

但即便掌握了风格，AI 还需要解决另一个关键问题：构图。如果画面没有明确的视觉焦点，元素摆放毫无章法，那么即便色彩再丰富，依旧会让人感到混乱不堪。在人类艺术创作中，黄金分割、三分构图、对角线构图等技巧，都是艺术家们在漫长历史中总结出的视觉经验，而 AI 也必须学会这些规则，才能让画面具备平衡感和美感。当 AI 生成“湖面上的独舟”（图 1-13）时，其会利用一定技巧（如黄金分割法），将小船放置在黄金比例线



8 AIGC 视觉传达设计

上,使观者视线自然地落在它身上。在绘制“夜色中的石中剑”(图 1-14)时, AI 会自动调整光影,使月光照在剑刃上,营造出层次感。

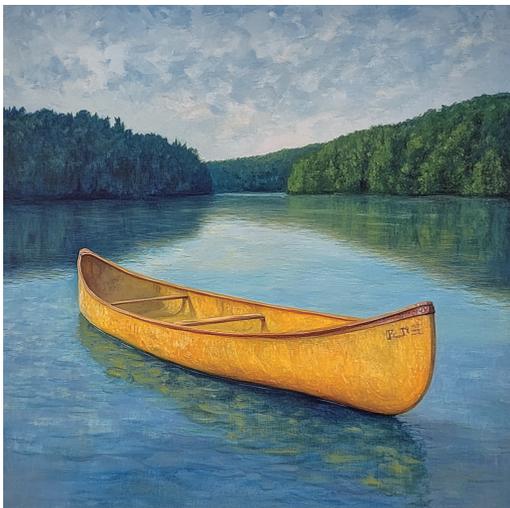


图 1-13 MJ 提示词:“湖面上的独舟”



图 1-14 MJ 提示词:“夜色中的石中剑”

这一切,构成了 AI 视觉学习的完整拼图。AIGC 的数据准备过程,是让它像人类一样,理解这些图像的意义。它必须从数据中学会如何解析画面,如何匹配风格,如何构建合理的视觉关系。

1.2.2 模型解析:拆解文本内容,构建画面逻辑

1. 文本解析:提取关键词,构建视觉要素

AI 生成图像的第一步,是要理解用户的提示词。表面上看,这只是文本输入的过程,但对 AI 来说,这更像是解读一份复杂的创意提案。它要拆解文字,找出画面重点,判断哪些内容是核心,哪些是修饰,并在“脑海”中构建一张“看不见的画布”,提前安排好主体、背景、色彩和风格。

每当用户输入一段描述,比如“身穿紫色魔法披风的少女站在清晨海边的悬崖上”(图 1-15),AI 首先进行“语法分解”,就像解构一幅复杂的画面。在这个过程中,AI 的自然语言处理(NLP)模型会将句子拆分成多个语义单元,并计算它们之间的层级关系。

“少女”是画面核心,决定了主角的身份和存在感。

“紫色魔法披风”定义了服饰特征,可能意味着深紫色、束腰、古朴、复古剪裁。

“海边的悬崖上”影响场景设计,提示地理景观有海浪、悬崖、碎石等。

“清晨”决定光影氛围,意味着清冷色调、光影对比强烈、可能有阴影投射。

这些信息被 AI 赋予不同的权重,确保在最终的生成过程中,重点突出主体,规避所有元素被平均化。比如在“复古电影中的美式乡村田野”这样的描述中,美式乡村田野作为主场景,需要占据画面主体,而“复古电影”作为氛围设定,负责影响色彩和光线效果,而不会改变画面的核心结构(图 1-16)。



图 1-15 MJ 提示词：“身穿紫色魔法披风的少女站在清晨海边的悬崖上”



图 1-16 MJ 提示词：“复古电影中的美式乡村田野”

文本解析不仅仅是“拆解”，AI 还会考虑风格暗示和隐含信息。把图 1-15 的描述改成“身穿紫色魔法披风的少女站在法阵中央”，虽然没有明确指出法阵的材质、颜色、发光效果，但 AI 会根据上下文推测应有的视觉呈现，法阵可能是紫色的能量光环，而非普通的地砖图案（图 1-17）。这样的语义理解，使得 AI 能够在文本描述之外自主填补画面信息，确保生成结果符合人类的视觉预期。

2. 视觉映射：转换语义信息，还原图像细节

文本解析完成后，AI 需要将这些拆解出的视觉元素，映射到具体的画面表现上。这一步，是导演拿到剧本后开始思考场景布置、色调氛围、镜头运用等关键细节的过程。

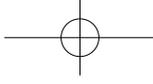
AI 通过语义匹配，在训练数据库中寻找最符合提示词的视觉模式。比如，对于“维多利亚时代的家居”（图 1-18），它不会简单地随机生成一本书和一张桌子，而是从已学习的视觉知识中提取相关特征，确保画面包含：



图 1-17 MJ 提示词：“身穿紫色魔法披风的少女站在法阵中央”



图 1-18 MJ 提示词：“维多利亚时代的家居”



深色木质家具，如雕刻繁复的书柜和扶手椅；
暖色调光源，如烛光、黄铜台灯；
古典建筑装饰，如墙上的复古壁纸或天花板上的水晶吊灯。

这种风格映射，使得 AI 生成的画面准确且保持艺术风格的统一性。再举一个例子，“蒸汽朋克机械师的工作室”（图 1-19）这个提示词，便意味着齿轮、铜管、蒸汽管道、复古仪器等元素的合理搭配，而不会生成包括现代电子设备或极简风格的房间布置。

视觉映射阶段，视觉逻辑同样重要，确保物体与环境的关系符合现实认知。在“漂浮在云端的灰暗城市”（图 1-20）中，AI 需要计算城市的悬浮方式（是否通过磁悬浮、吊索支撑、反重力装置）、光影方向（云层遮挡部分光线，产生柔和的散射光）、视角设定（俯视还是仰视）等问题。这些细节，使得画面在视觉上更具说服力，而不会显得违和或随意。



图 1-19 MJ 提示词：“蒸汽朋克机械师的工作室”

图 1-20 MJ 提示词：“漂浮在云端的灰暗城市”

1.2.3 叙事结构：组织画面节奏，形成故事结构

AI 生成的画面不仅仅是色彩、构图和光影的组合，更是一段故事的瞬间定格。它不只是展示一个场景，而是让观众产生“这里发生了什么？”或者“接下来会发生什么？”的思考。这种未完成的叙事感，让画面变得鲜活，使观看者不仅仅是旁观者，而是故事的参与者。

当输入“沙漠中的旅行者”（图 1-21）后，AI 需要判断这个画面仅仅是一个站在黄沙中的人物，还是一场旅途的片段。他或许正被风沙吞噬，或许刚刚在远方的驿站中短暂停留。如果是“夕阳下的沙漠旅行者”（图 1-22），画面可能会被温暖的金橙色调包裹，影子被拉长，空气中带有热浪的微微扭曲，传递出一天行程即将结束的疲惫感。如果是“风暴中的沙漠旅人”（图 1-23），AI 可能会让狂风卷起飞沙，模糊远方的地平线，旅人低头前行，斗篷被风撕扯，使画面彰显出更强的生存压力。这些细节的变化，决定了画面的故事基调，而不仅仅是对关键词的简单匹配。