

AIGC 概述：理论、实践 与发展方向

第 1 章

本章系统构建 AIGC 领域的基础认知框架与能力体系,从 AIGC 的核心定义、数据驱动与创造性两个本质特性切入,梳理人工智能关键技术从博弈探索、问答识别到运动实践再到通用大模型的发展脉络,进而全面展开文本、图片、音视频、多模态等 AIGC 分类及典型应用场景。在此基础上,深入剖析大模型的核心支撑、关键技术与学习生成机制,通过 DeepSeek 等前沿大模型案例理解其对行业生态、市场格局的影响及核心优势,最后结合实践任务,完成从理论认知到工具应用的过渡,全面建立对 AIGC 技术原理、产业应用及社会价值的整体认知,为后续深入探索人工智能技术奠定基础。

学习目标

本章致力于通过 AIGC 理论学习与实践操作的结合,让学生掌握 AIGC 的定义、分类、核心特性及发展历程,理解大模型训练原理、关键技术(蒸馏、推理、RAG)与生成机制;能够识别不同类型 AIGC 内容的应用场景,运用 AIGC 工具完成简单内容创作与信息分析;同时培养对 AIGC 技术的前瞻性视野,树立科学看待其优势与局限性的态度,构建对 AIGC 产业生态的基础认知,形成面向智能时代的技术理解能力与实践应用素养。

【知识目标】

- (1) 掌握 AIGC 的基本定义、核心概念及其发展历程。
- (2) 了解 AIGC 的主要分类及典型应用场景,包括文本、图片、音视频的生成。
- (3) 理解 AIGC 大模型的训练原理、核心基础(数据集)、关键技术(如蒸馏、推理、RAG)及其核心特性。
- (4) 认识当前 AIGC 领域的前沿发展趋势和大模型的重要代表,如 DeepSeek、Qwen、Kimi。

【能力目标】

- (1) 能够识别不同类型 AIGC 内容的生成形式与对应的应用领域。
- (2) 能够初步分析 AIGC 技术在特定信息内容领域的应用潜力。
- (3) 能够理解大模型预训练与微调的基本逻辑,以及不同技术概念(蒸馏、推理、RAG)的区别与作用。

【素养目标】

- (1) 培养对 AIGC 技术发展的兴趣和前瞻性,积极关注新兴 AI 应用。
- (2) 树立正确看待 AI 技术优势与局限性的科学态度,认识其社会与伦理影响。
- (3) 初步构建对 AIGC 产业生态的基础认知,理解前沿大模型(如 DeepSeek)的崛起与影响。

模块 1.1 什么是 AIGC

本模块首先明确 AIGC 的学术定义与技术边界,通过对比决策式 AI 的工作原理,揭示其“从无到有”的内容创造能力。重点阐释 AIGC 的两大本质特性:数据驱动性——依托大规模标注数据与神经网络参数实现模式学习,以及创造性——通过概率分布建模生成具有新颖性的输出内容。模块通过解析 GPT 系列模型参数规模与生成能力的关系,帮助读者建立对 AIGC 技术本质的认知框架,为后续理解技术原理与应用场景提供概念锚点。

知识 1.1.1 AIGC 的定义

AIGC 的核心概念是利用人工智能技术自动生成文本、图像、音频、视频等多种形式的

内容。

1. AIGC 的特点

AIGC 具有数据驱动和创造力两个显著特点。

(1) 数据驱动: AIGC 依赖的大语言模型(large language model,LLM,简称大模型)通过学习海量数据来捕捉和理解其内在模式与生成规则。

(2) 创造力: AIGC 不仅能生成符合逻辑和规范的内容,还具备生成高质量且富有创意内容的能力。

2. AIGC 内容的生成形式

从生成形式看,AIGC 覆盖多类内容产出,如图 1-1 所示。

(1) 文本生成: 支持文生文(如文案续写)、文生代码(如程序片段编写)、文生表格(如结构化数据整理),借由模型对语言逻辑与规则的学习,输出精准或创意性文本。

(2) 图片生成: 包含文生图(如文字描述转视觉画面等)、图生图(如图像二次创作等)、图像修复(如老照片翻新等),让视觉内容突破人工绘制边界。

(3) 视频生成: 可实现文生视频(文字脚本转动态影像)、图生视频(静态图拓展动态故事)、视频生视频(素材剪辑优化),重塑视频创作流程。

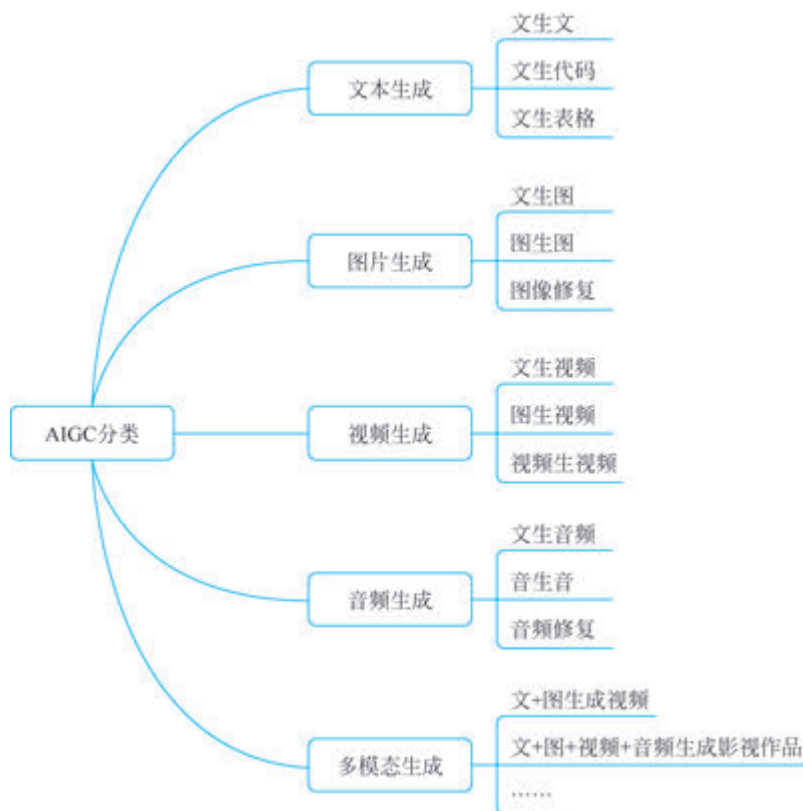


图 1-1 AIGC 的分类

(4) 音频生成：涵盖文生音频(文字转语音、音效)、音生音(音频风格转换)、音频修复(噪声去除),拓展声音创作的可能性。

(5) 多模态生成：融合文、图、视频、音频等,比如文+图生成剧情短片、多素材拼接成影视级作品,探索跨形式内容创作。

3. AIGC 创作的四个发展阶段

AIGC 被业界视为继专业生产内容和用户生产内容之后的新一代内容生产方式,代表了内容创作的进一步智能化演进,内容创作的四个阶段如图 1-2 所示。

(1) PGC(professional generated content)：由专业人员或机构(如媒体机构、出版社、专业工作室)生成的内容,如新闻网站报道等。这些内容通常具有较高的专业性和权威性。

(2) UGC(user generated content)：由普通用户自己创建、生产并分享的内容,如社交媒体帖子、个人博客、短视频等。这些内容通常更具个性化和互动性。

(3) AIUGC(AI-assisted user generated content)：表示人工智能辅助或引导用户生成内容的方式。在这种模式下,AI 作为一种工具,帮助用户更高效地完成内容创作,如 AI 写作助手、智能修图软件等。

(4) AIGC(artificial intelligence generated content)：完全由人工智能独立生成的内容,其创作过程无须或仅需少量人工干预。这是 AI 在内容创作领域深度发展的体现,突破了传统内容生产的边界。

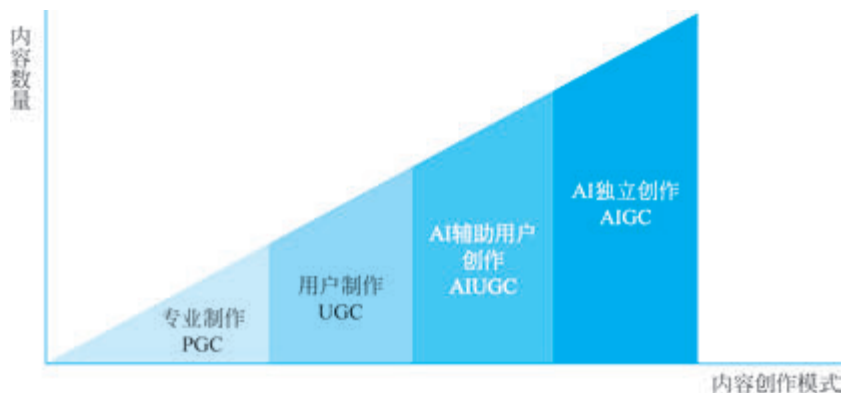


图 1-2 AIGC 内容创作的四个发展阶段

知识 1.1.2 人工智能关键技术的发展历程

人工智能关键技术发展历程主要有以下几个阶段，如图 1-3 所示。

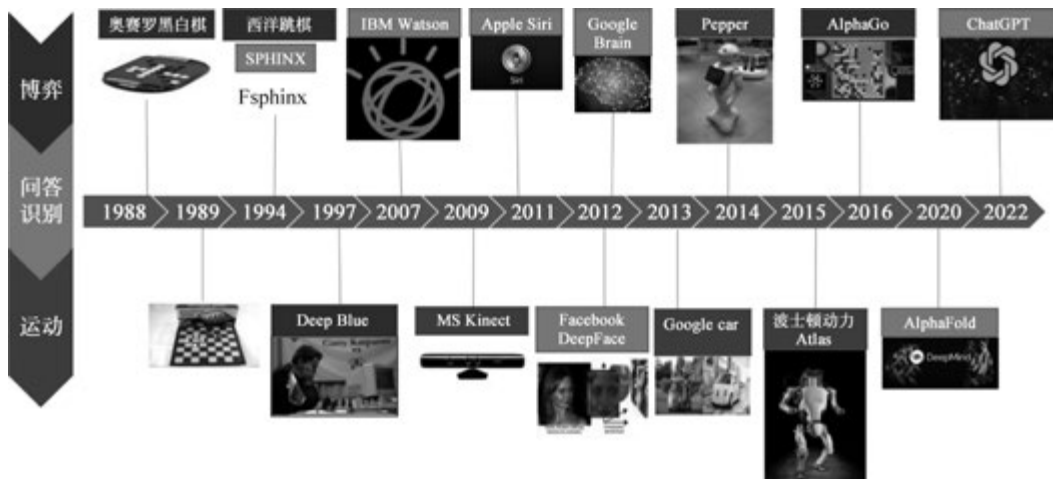


图 1-3 人工智能关键技术发展历程

1. 早期探索阶段

博弈，是 AI 技术早期探索的“试验场”——在规则清晰又充满变化的对弈中，验证算法与算力的潜力：1988 年，奥赛罗黑白棋让机器尝试理解“策略对抗”逻辑，用算法模拟人类棋手思考路径，为复杂博弈场景的 AI 应用打基础；1994 年，西洋跳棋进一步优化搜索算法，让 AI 在规则化博弈里提升决策效率，探索“逻辑推理”的机器实现方式；1997 年，Deep Blue(深蓝)标志性突破！IBM“深蓝”击败国际象棋世界冠军，靠超强算力+深度搜索，证明 AI 在复杂规则博弈中能超越人类，成为 AI 技术“出圈”的里程碑事件。

2. 问答识别阶段

当 AI 走出棋盘，“如何听懂问题、给出答案”成了新挑战。问答识别技术的演进，让机器从“执行简单指令”迈向“理解人类意图”：1994 年，SPHINX 聚焦语音识别基础能力，尝试让机器“听懂”人类语言，突破人机交互的第一道壁垒；2007 年，IBM Watson 在“危险边

缘”知识问答节目中击败人类,整合知识图谱、自然语言处理技术,让 AI 学会“理解开放问题、推理答案”,标志问答从“规则匹配”升级为“知识融合推理”;2011 年,Apple Siri 成为真正走进大众生活的智能助手,把问答识别和日常场景结合(查天气、设提醒),让 AI 从实验室工具变成“生活帮手”,推动技术平民化落地。

3. “运动”阶段

“运动”维度的 AI 技术,让机器从“思维探索”走向“物理实践”——感知环境、自主决策、灵活行动,重塑人机协作模式:2009 年,MS Kinect 通过深度摄像头识别人体动作,让机器“看懂”人类运动,开启智能体感应用(如游戏互动、健身辅助),拓展 AI 在“空间感知”的应用边界;2014 年,Google car(谷歌无人车)融合计算机视觉(识别路况)、路径规划(动态决策)、传感器融合(感知环境),让汽车从“人驾驶”变“机器自主行驶”,探索 AI 在复杂动态场景的落地可能;2015 年,波士顿动力 Atlas: 靠仿生机械结构+运动算法,实现跑跳、越障、团队协作。它证明 AI 不仅能“完成动作”,还能在非结构化环境(如废墟救援)里自主规划、灵活应对,刷新机器人应用的想象空间。

AI 技术早已跳出单一领域,走向“通用技术底座+跨场景应用”的新形态:2012 年 Google Brain 打磨深度学习框架,2013 年 Facebook DeepFace 专注于人脸识别——为 AI 发展提供“算法基础设施”,让技术能快速适配博弈、问答、运动等不同场景;2016 年,AlphaGo 再次突破博弈天花板! 结合深度学习+强化学习,击败围棋世界冠军。围棋规则复杂、变化无规律(没有明确“最优解”),AlphaGo 证明 AI 能在“模糊规则、需要创意决策”的场景里胜出,让技术从“死记硬背规则”进化到“主动创造策略”;2022 年,ChatGPT 引发自然语言处理变革,靠大模型+强化学习,实现开放对话、内容创作,让 AI 从“理解问题”升级为“主动生产内容”,彻底改变人机交互方式;2022 年,AlphaFold 用 AI 破解“蛋白质折叠预测”难题,把技术从“生活应用”推向“科学研究”,展现 AI 跨领域突破的潜力。

知识 1.1.3 AIGC 的核心特性

AIGC 的核心特性、优势、挑战与局限如表 1-1 所示。

表 1-1 AIGC 的核心特性、优势、挑战与局限

核 心 特 性	优 势	挑战与局限
数据驱动	基于大规模数据集训练,能够学习复杂模式并生成高质量、高真实度的内容。例如,通过学习海量文本生成流畅的文章,或基于大量图像生成逼真的画面	对数据质量高度依赖,训练数据中的偏差、偏见或错误可能导致生成内容出现偏差或不准确
多模态生成	支持文本、图像、音频、视频等多种内容形式的生成与转换,能够实现跨模态的创作与融合。例如,根据文字描述生成图片,或将图片转化为视频	跨模态生成效果可能存在不一致性或不同模态之间语义对齐的挑战,导致生成内容在某些方面脱节
创造力和扩展性	能够生成创意性强、独具风格的内容,突破人类创作的局限,并推动 AI 技术在多个领域的创新应用。例如,设计新产品草图或谱写原创乐曲	创造力仍受限于训练数据和模型架构,可能缺乏完全的、自发的原创性,更多是基于对已有数据的重组和变异

续表

核 心 特 性	优 势	挑战与局限
智能交互能力	具备强大的上下文理解能力,能够生成流畅且符合语义、逻辑连贯的内容,支持复杂的对话和交互任务。例如,进行多轮对话并保持语境一致性	在复杂语境或信息模糊不清的情况下,可能出现误解或语义不一致,导致生成的内容不够准确或难以理解
社会与伦理影响	赋能行业应用,显著提升生产效率,降低内容创作门槛,为个人和企业带来新的发展机遇。例如,自动化生成营销文案或辅助设计工作	可能带来虚假信息(如深度伪造)、版权归属问题、伦理道德挑战(如侵犯隐私)以及法律法规争议,需要社会各界共同应对和规范

模块 1.2 AIGC 的分类与应用场景

从回溯技术演进脉络切入,本模块将系统梳理人工智能自 1956 年达特茅斯会议诞生至今的关键发展节点。内容以“技术突破—瓶颈期—范式转移”为叙事主线,重点解析三次 AI 浪潮的驱动因素:符号主义时期的逻辑推理系统,统计学习时期的机器学习算法,以及当前深度学习时期的神经网络革命。通过对比 AlexNet 与 GPT-3 的技术架构差异,揭示 AIGC 作为第三代 AI 技术典型代表的演进必然性,帮助读者理解技术迭代背后的科学逻辑与工程突破。

知识 1.2.1 AI 生成文本——文生文(文章、代码、表格)

文生文是指通过人工智能技术,根据用户给定的文本描述、指令或特定要求,自主生成全新的、连贯且符合语境的文字内容,包括文章、代码、报告甚至表格数据等多种文本形式。

AI 生成的文本广泛应用于文章创作(如新闻稿、小说章节)、编程(如代码生成、错误修复)、智能对话(如客服机器人、虚拟助手)、报告撰写、摘要生成等多种场景。

当前文生文领域的代表性产品包括 OpenAI 的 ChatGPT、百度的文心一言、阿里的通义千问、Kimi 以及智谱清言等,这些模型在理解和生成人类语言方面表现出色,极大地提升了文本生产的效率和质量。

图 1-4 展示了通过 ChatGPT 来辅助老师,总结文章的中心思想以及表达的情感。

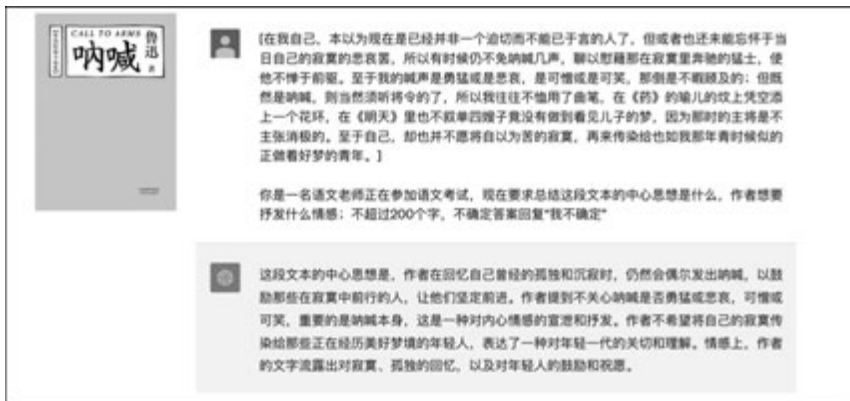


图 1-4 ChatGPT 实现文生文案例

知识 1.2.2 AI 生成图片——文生图

文生图是指利用人工智能技术,根据用户输入的文本描述或关键词,生成与之相符的视觉图像。模型通过理解文本的深层含义,创造出原创的图像作品。

文生图技术在艺术创作(插画、概念艺术)、广告设计(生成产品宣传图)、游戏开发(场景、角色概念图)、室内设计(可视化效果图)、服装设计等领域具有重要的应用价值。

DALL·E3(OpenAI)、Midjourney、Stable Diffusion、Artbreeder、即梦 AI、Kling、Vidu 等是目前文生图领域的领先产品,它们能够生成令人惊叹的、多样化的图像风格和内容。

图 1-5 所示为大模型的文生图能力展示。



图 1-5 文生图案例

知识 1.2.3 AI 生成图片——图生图

图生图是指人工智能基于已有的图像,通过各种算法进行编辑、风格转换、内容扩展或生成新的图像。它不依赖于文本输入,而是直接操作视觉信息。

图生图技术广泛应用于图像增强(提升清晰度)、风格转换(将普通照片转换为油画风格)、背景替换(移除或更换图像背景)、图像修复(修补损坏部分)、图像重绘(AI 重新绘制图像)等。

DALL·E3(OpenAI)、Midjourney、Stable Diffusion、Artbreeder 等同样在图生图领域表现出色,提供了丰富的图像编辑和创作功能。

图 1-6 展示了大模型基于现有图片进行二次创作的能力,图 1-7 展示了大模型图像风格转换的能力。



图 1-6 图生图案例 1

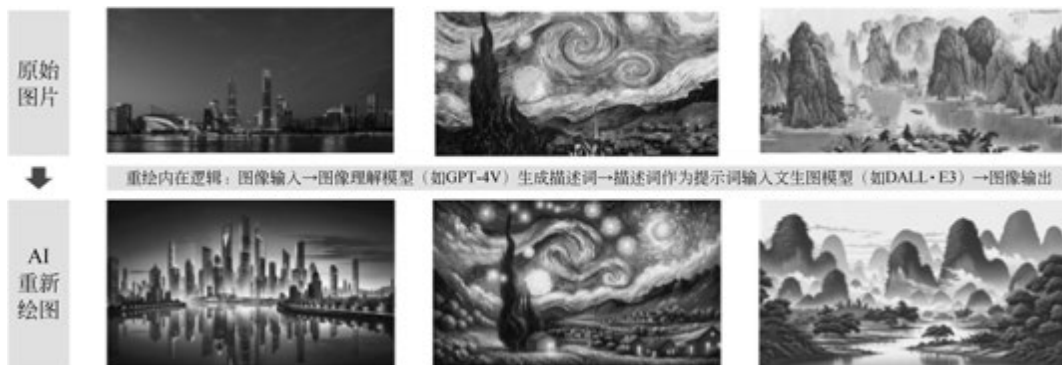


图 1-7 图生图案例 2

知识 1.2.4 AI 生成视频——文生视频

文生视频是指根据用户输入的文本描述,通过人工智能技术生成完整的动态视频内容。用户只需提供文字脚本或创意描述,AI 即可自动生成符合描述的视频,包括场景、人物动作、转场等元素。

文生视频技术适用于快速制作动画短片、营销视频、虚拟人物对话、新闻播报(虚拟主播)、教育内容等,极大地降低了视频制作的门槛和成本。

Kling、Veo、Sora、Hailuo、Wan、Waver 等是当前文生视频领域的热门产品,它们不断提升视频内容的质量和生成效率。

图 1-8 展示了大模型的文生图能力。



图 1-8 文生图案例



熊猫做饭视频

知识 1.2.5 AI 生成视频——图生视频

图生视频是指利用人工智能技术将静态图像、设计元素或现有视频片段转换为动态视频。它不仅支持对静态图像进行编辑和生成,更能赋予图像动画效果、实现流畅的转场,并编织出连贯的故事情节。

图生视频广泛应用于广告制作(将静态产品图动态化)、影视后期(特效制作、场景扩展)、虚拟现实(virtual reality,VR)、增强现实(augmented reality,AR)内容创作以及在线教育课件制作等多个领域。

Kling、GenFlare、PixVerse、Hailuo、Waver、Vidu 等产品在图生视频技术上持续创新,能够生成高质量的动态视觉内容。

图 1-9 展示了大模型的图生视频能力。



图 1-9 大模型图生视频示例



照片动起来视频

知识 1.2.6 AI 生成音乐

生成音乐是指利用人工智能技术,根据用户输入的文本描述(如“欢快、爵士乐”)、情感标签(如“悲伤”)、节奏模式或特定参数设置(如曲速、调性),来创作完整的音乐作品。这包括作曲、编曲、歌词生成、音色合成等多个方面。

AI 生成音乐可用于影视配乐、游戏音效、广告背景音乐、辅助音乐人创作、个性化音乐定制以及虚拟现实场景中的环境音效等。它为非专业人士提供了音乐创作的便捷途径,也为专业人士提供了新的创作工具和灵感。

Suno、Eleven Music、FUZZ、Lyria、海绵音乐、网易天音等是目前 AI 生成音乐领域的知名产品,它们能够根据用户需求创作出各种风格和类型的音乐,甚至能生成带歌词的歌曲。

图 1-10 展示了大模型的音乐创作能力。



图 1-10 大模型音乐创作示例

模块 1.3 AIGC 的基本原理

本模块从技术模态与应用领域两个维度构建 AIGC 的分类框架。技术维度划分为文本生成(如 ChatGPT)、图像生成(Midjourney)、音频生成(Suno AI)、视频生成(Sora)及多模态融合五大类;应用领域覆盖内容创作(广告文案、影视剧本)、教育培训(智能题库、个性化辅导)、科研辅助(文献综述、实验设计)、工业设计(产品原型、工程绘图)等 12 个垂直场景。本模块通过展示 2024 年 AIGC 市场规模数据(全球 217 亿美元,同比增长 115%)及各细分领域占比,帮助读者建立对 AIGC 产业生态的全景认知,理解技术落地的商业价值与社会影响。

知识 1.3.1 AIGC 大模型的核心基础：数据集

AIGC 的强大能力源于其背后复杂而精妙的 AI 大模型。理解这些大模型的核心基础——“数据集”，以及它们如何“学会创造”的过程,对于掌握 AIGC 的本质至关重要。

1. 数据集的含义

数据集可以被形象地比喻为大模型学习的“教材”或“知识库”，包含大模型需要学习和理解的大量信息。数据集的质量和范围直接决定了 AI 模型生成内容的质量、准确性和多样性。图 1-11 展示了在训练不同的大模型时,使用到的数据集的差异。

2. 不同类型的数据来源

大模型在训练时会整合多种类型的数据,包括文本数据(如书籍、文章、网页内容)、图像数据(如图片)、音频数据(如语音、音乐)以及视频数据等。