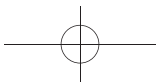


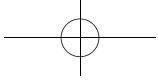
AIGC

让生成式AI成为自己的外脑

成生辉 著

清华大学出版社
北 京





内 容 简 介

本书针对近期较为火热的AIGC技术及其相关话题，介绍AIGC的技术原理、专业知识和应用。

全书共分为九章。第一章介绍AIGC技术的基本概念和发展历程；第二、三章介绍AIGC的基础技术栈和拓展技术栈；第四、五章分别讨论了AIGC技术在文本生成和图像生成两个领域的现状和前景；第六章列举了目前较为热门的AIGC技术应用；第七章描述了AIGC的上、中、下游产业链及未来前景；第八章主要关注AIGC在法律和道德上可能存在的争议与问题；第九章对AIGC技术进行了总结与展望。全书运用可视化的表达方式，对较为复杂的概念进行了生动易懂的阐述。

本书适合AIGC从业人员、相关技术人员以及相关专业的学生参考和学习。

本书封面贴有清华大学出版社防伪标签，无标签者不得销售。

版权所有，侵权必究。举报：010-62782989，beiqinquan@tup.tsinghua.edu.cn。

图书在版编目(CIP)数据

AIGC：让生成式AI成为自己的外脑 / 成生辉著. —北京：清华大学出版社，2023.9
ISBN 978-7-302-64566-5

I. ①A… II. ①成… III. ①人工智能 IV. ①TP18

中国国家版本馆CIP数据核字(2023)第169648号

责任编辑：施 猛 王 欢

封面设计：熊仁丹

版式设计：方加青

责任校对：马遥遥

责任印制：曹婉颖

出版发行：清华大学出版社

网 址：<http://www.tup.com.cn>，<http://www.wqbook.com>

地 址：北京清华大学学研大厦A座 邮 编：100084

社 总 机：010-83470000 邮 购：010-62786544

投稿与读者服务：010-62776969，c-service@tup.tsinghua.edu.cn

质 量 反 馈：010-62772015，zhiliang@tup.tsinghua.edu.cn

印 装 者：三河市铭诚印务有限公司

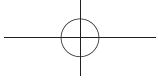
经 销：全国新华书店

开 本：180mm×250mm 印 张：15.75 字 数：242千字

版 次：2023年10月第1版 印 次：2023年10月第1次印刷

定 价：88.00 元

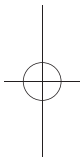
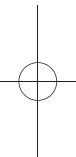
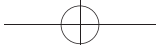
产品编号：102031-01

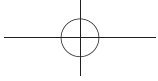


致谢

T H A N K S

感谢刘宇琦、樊宇清、孟怡然、刘铂晗、闫丹、沈闵黑籽、莫晨晨、李翔宇和王俊伟等为本书的出版做出的巨大贡献。特别感谢北京华夏长鸿文化有限公司在策划方面为本书所提供的帮助。





前言

P R E F A C E

AIGC，即生成式人工智能，正推动着人工智能掀开新的一页，以前所未有的速度崛起并席卷全球。无论是学术界还是产业界，都在积极布局AIGC领域，准备迎接一个新时代的到来。

在学术界，研究人员不断探索新的算法和技术，以改进 AIGC 的生成质量和效率。同时，各大高校也推出了相关的课程和研究项目，培养和支持 AIGC 领域的专业人才。

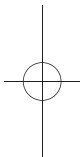
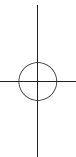
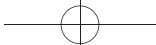
在产业界，许多公司和组织积极应用 AIGC 技术，以提高业务效率和产品质量。例如，新闻媒体、广告公司、内容创作平台等，都在使用 AIGC 技术生成各种类型的内容，以满足不同用户的需求。一些大型科技公司也在积极布局 AIGC 领域，推出各种与 AIGC 相关的工具和平台，以支持 AIGC 应用的开发和部署。

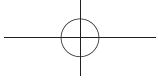
AIGC的发展还面临许多挑战和问题，例如算法可解释性、数据隐私保护等。因此，我们需要持续关注和研究 AIGC，以便更好地挖掘它的潜力，同时避免其潜在的风险和影响。

本书理论结合实践，旨在帮助读者全面掌握 AIGC 的基础知识，为应对未来的技术挑战做好准备。让我们一起探索这个迅速发展的领域，发现其中的无限可能性！

成生辉

2023年6月





目录 CONTENTS

第一章 ✧ AIGC：智造时代来临 / 1

第一节 从UGC、PGC到AIGC / 2

- 一、UGC 首秀登场 / 3
- 二、PGC 持续发力 / 4
- 三、AIGC 引爆热点 / 5

第二节 AIGC 的分类 / 6

- 一、根据生成内容分类 / 7
- 二、根据生成技术分类 / 7
- 三、根据生成目的分类 / 8
- 四、根据生成方式分类 / 10

第三节 AIGC 的发展历史 / 11

- 一、早期探索阶段 / 12
- 二、数据和硬件的提升阶段 / 13
- 三、GAN 技术的引入阶段 / 14
- 四、语言生成模型的兴起阶段 / 16

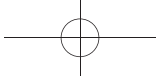
第二章 ✧ AIGC 的基础技术栈 / 19

第一节 识别的技术 / 20

- 一、识别的关键 / 20
- 二、计算机视觉模型 / 21

第二节 理解与输出 / 24

- 一、算法如何理解文本 / 24
- 二、算法如何理解音频 / 27
- 三、算法如何输出文本 / 29



四、算法如何输出音频 / 34

第三节 从输出到可视 / 36

- 一、数据可视化技术 / 36
- 二、数据可视化的建议 / 37
- 三、数据可视化技术与元宇宙 / 40

第四节 从可视到创作 / 41

- 一、跳出传统的 AI 艺术 / 41
- 二、AI 艺术：从理论走向实践 / 43
- 三、艺术的产生到增强 / 45

第三章 ✧ AIGC 的拓展技术栈 / 47

第一节 模块化设计 / 48

- 一、模块化设计的应用 / 48
- 二、模块化设计的优缺点 / 50

第二节 Seq2Seq 模型 / 52

- 一、Seq2Seq 模型的基本原理 / 53
- 二、Seq2Seq 模型的应用 / 53

第三节 多模态模型 / 61

- 一、多模态模型的发展历史 / 62
- 二、多模态模型的应用 / 64
- 三、多模态模型的案例 / 68

第四节 一切的一切：ChatGPT / 72

- 一、ChatGPT 中的 Seq2Seq 思想 / 72
- 二、GPT-4 支持多模态 / 74

第四章 · AIGC 与文本生成 / 75

第一节 AI 文本生成与发展 / 76

- 一、ChatGPT 的出圈 / 80
- 二、国外 AI 文本生成的火热 / 81
- 三、国内 AI 文本生成的发展 / 84

第二节 AI 文本生成的垂直应用 / 87

- 一、专业客服 / 89
- 二、智能搜索 / 91
- 三、智能创作 / 93
- 四、图像也可以用文本表示 / 94
- 五、AI 也是学霸 / 96
- 六、扮演评委不在话下 / 98

第三节 关于 AI 文本生成的一些讨论 / 100

- 一、更好或是更坏 / 100
- 二、监管何去何从 / 102
- 三、隐私愈发重要 / 103

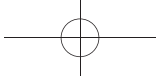
第五章 · AIGC 与图像生成 / 105

第一节 AI 绘画的兴起 / 106

- 一、图像编辑器的魔力 / 106
- 二、神奇的图像生成工具 / 107
- 三、万能的形象生成工具 / 108
- 四、语言处理的奥秘 / 109

第二节 AI 辅助美术设计 / 110

- 一、神通广大的图像处理 / 111



VIII ✧ AIGC：让生成式AI成为自己的外脑

- 二、线条绘制的艺术 / 113
- 三、神奇的风格迁移技术 / 114
- 四、布局设计的底层逻辑 / 115

第三节 AI 辅助海报设计 / 117

- 一、如何使用AI辅助设计海报 / 117
- 二、AI 辅助海报设计的优点与缺点 / 119
- 三、AI 辅助海报设计的未来 / 121

第四节 AI 辅助专业图像生成 / 123

- 一、神奇的算法模型 / 124
- 二、强大的数据集 / 125
- 三、逼真的应用场景 / 128
- 四、未来发展趋势 / 129

第六章 ✧ AIGC 的应用升级 / 131

第一节 内容生成赋能数字化转型 / 132

- 一、用AI 工具打造视听生态 / 132
- 二、多模态交互下的内容创作 / 139
- 三、多模态交互的应用 / 141

第二节 代码生成掀起科技浪潮 / 143

- 一、人人皆可编程 / 144
- 二、AI 生成引发新的内卷 / 147

第三节 策略生成引领游戏革新 / 148

- 一、虚拟形象的生成 / 148
- 二、游戏的开发制作 / 149

第七章 · AIGC 的生态构建 / 151

第一节 上游产业的发展 / 153

- 一、硬件的突破 / 154
- 二、软件的带动 / 155
- 三、数据的支持 / 157
- 四、人才的培养 / 159

第二节 中游产业的带动 / 161

- 一、云计算：算力的提供者 / 161
- 二、大数据：训练的基础 / 162
- 三、机器学习：多线程学习的探索 / 164

第三节 下游产业的繁荣 / 165

- 一、终端产品的打造 / 166
- 二、智能生态的构建 / 166

第八章 · AIGC 的挑战与监管 / 173

第一节 AIGC 的阿喀琉斯之踵 / 174

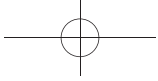
- 一、当前的不足 / 174
- 二、未来的发展 / 176

第二节 法律监管与问题 / 178

- 一、谁是版权所有者 / 178
- 二、技术的 B 面：道德与滥用 / 180
- 三、数据隐私问题 / 181

第三节 产业影响与人员转型 / 183

- 一、AIGC 的产业落地 / 184
- 二、AIGC 对人力资源市场的冲击 / 191



X ✧ AIGC：让生成式AI成为自己的外脑

三、普通员工应如何应对AIGC带来的职业危机 / 192

第九章 ✧ AIGC引领未来的变革 / 197

第一节 探索无限可能性 / 198

一、揭示 AIGC 改变世界的力量 / 199

二、AIGC 的冒险之旅 / 206

第二节 人与 AI 智慧共生 / 211

一、探索 AI 全球治理的未来 / 211

二、掀起商业与社会变革的力量 / 214

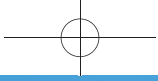
第三节 探索下一代数字世界的无限潜力 / 219

一、元宇宙开发的领军者 / 219

二、革命性的搭档：Web 3.0 / 222

三、一场机遇与争议并存的革命 / 225

参考文献 / 227



第一章 AIGC：智造时代来临

AIGC(AI-generated content)，即生成式人工智能。随着人工智能技术的进步，包括自然语言处理、计算机视觉和机器学习等领域的发展，AI能够创造和生成各种形式的内容，如文章、图像、音频和视频等。

AIGC的应用领域非常广泛。例如，在数字媒体和广告行业，AIGC可以用于生成新闻报道、社交媒体帖子、广告文案和图像等；在虚拟现实、视频游戏和电影制作等领域，AIGC可以用于生成虚拟角色、场景和剧情等。

AIGC的发展和應用引发了一系列讨论，如内容原创性、道德问题以及法律责任等。为此，监管机构正在努力制定相关政策和准则，以确保AIGC生成的内容符合伦理和法律要求。

第一节 从UGC、PGC到AIGC

从UGC到PGC，再到AIGC，内容生成方式逐渐从用户主导转变为专业人士主导，最后引入了人工智能技术。

释义 1.1: UGC

UGC(user generated content)，指由用户主动创作和分享的内容。

在Web 2.0时代，社交媒体平台和在线论坛为用户提供了发布和共享内容的机会。UGC包括用户撰写的文章、发布的照片和视频、发表的评论和社交媒体帖子等。这些内容通常基于用户的个人经验、兴趣和观点^①。

释义 1.2: PGC

PGC(professional generated content)，指由专业人士或团队创作和制作的内容。

与UGC相比，PGC通常具有更高的质量和专业性，这些内容由专业的作家、摄影师、记者、制片人等创建。PGC包括新闻报道、电影、电视剧、音乐专辑、图书等。这些内容经过精心策划、编辑和制作，以满足特定的标准和用户需求。

释义 1.3: AIGC

AIGC(AI-generated content)，指由人工智能生成的内容。

随着人工智能技术的快速发展，AI可以通过机器学习、自然语言处理、计算机视觉和生成对抗网络等技术生成各种类型的内容。AIGC涵盖文字、图像、音频和视频等媒体形式。AIGC可以模仿人类的创造力和风格，生成更符

^① CHEONG H J, MORRISON M A. Consumers' reliance on product information and recommendations found in UGC[J]. Journal of Interactive Advertising, 2008, 8(2): 38-49.

合用户需求的文章、图像、音乐和视频等。

AIGC与UGC和PGC相比，具有一些独特的优势。AIGC可以在短时间内生成大量内容，满足用户日益增长的需求。同时，AIGC可以自动化创作，减少人工创作的成本和时间。然而，AIGC也面临一些挑战，如内容原创性、伦理问题和法律责任等，需要谨慎处理。

如图1.1所示，UGC、PGC和AIGC代表不同阶段和不同参与者在内容生成中的角色变化。从用户驱动到专业驱动，再到引入人工智能技术，这个演变过程反映了技术进步和内容创作模式的变革。

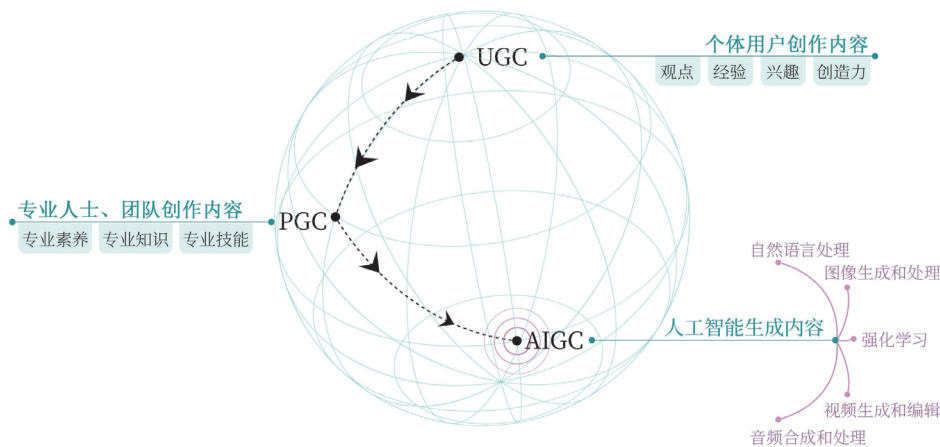


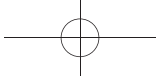
图1.1 从 UGC、PGC 到 AIGC

一、UGC 首秀登场

UGC是由用户主动创作的内容，包括文字、图像、视频、音频等多种形式，它反映了普通用户的观点、经验、兴趣和创造力。用户通常在微博、论坛、视频分享网站等平台上发布和分享自己创作的内容^①。

UGC具有用户主导、多样性、即时性和互动性等特点。用户主导即用户自愿创作内容，表达个人观点和分享体验；多样性是指UGC的内容形式多

^① PUIGSERVER P, SPIEGELMAN B M. Peroxisome proliferator-activated receptor-V coactivator 1a(PGC-1a): transcriptional coactivator and metabolic regulator[J]. Endocrine Reviews, 2003, 24(1): 78-90.



样，反映不同用户的兴趣和创意；即时性是指UGC往往是实时发布的，可以快速反映事件和趋势；互动性是指UGC鼓励用户之间通过互动和分享形成社群。

UGC的优势在于内容丰富多样，能吸引大量用户参与；能形成独特的创意和观点；鼓励用户互动和分享，可促进用户之间建立社交关系，形成社群；不仅能够快速反映当前事件和趋势，还能及时提供信息和观点。

当然，UGC也面临一些挑战。例如，UGC的质量参差不齐，有些内容可能存在不准确、质量低或违法违规的问题，需要设置有效的筛选和管理机制。此外，UGC可能侵犯他人的版权或违反法律法规，需要建立相关法律框架和合规机制。由于UGC的开放性，存在传播虚假信息和谣言的风险，需要用户加强辨别能力，平台完善信息验证机制。

总体来说，UGC作为一种充满活力和创造力的内容形式，不仅给用户带来了表达自我的机会，也给企业、社会 and 平台带来了新的机遇和挑战。在数字时代，UGC将继续发挥重要作用，并对多个领域产生深远影响。

二、PGC 持续发力

PGC与UGC相对，PGC强调专业素养、专业知识和专业技能。

PGC具有专业性、可信度和信息价值等特点。专业性指的是PGC由具备专业素养和知识的人士或机构创作，提供专业的观点、见解和信息。可信度，顾名思义，PGC受到专业人士的监督和审查，通常具有较高的可信度和权威性。信息价值指的是PGC注重提供专业知识和专业技能，以满足用户对专业内容的需求。

PGC的优势在于它由专业人士创作，专业人士具备专业知识和经验，能够提供高质量和可信赖的内容。PGC通常涉及专业领域，提供专业知识的深度剖析和全面覆盖^①。PGC在作者和读者之间形成专业网络，促进专业人士之间的交流和合作。

^① LIN J, WU H, TARR P T, et al. Transcriptional coactivator PGC- 1 α drives the formation of slow-twitch muscle fibres[J]. Nature, 2002, 418(6899): 797-801.

PGC也面临一些挑战，例如，创作PGC需要耗费时间和精力，包括研究、调研、写作和编辑等环节，同时可能需要投入资金和资源。而且PGC的受众通常是特定领域的专业人士或对该领域感兴趣的读者，受众范围相对较小，可能限制内容的传播和影响力。随着行业的不断发展和变化，PGC需要保持更新，以确保内容的准确性和相关性。

综上所述，PGC作为专业人士和机构输出的高质量内容，具有较高的可信度和影响力，对于行业发展、学术研究和商业运作都具有重要作用。然而，PGC也面临着挑战，包括维持更新、受众限制和成本投入等方面。在信息爆炸的时代，PGC的价值仍然不可忽视，它与UGC共同构成了丰富多样的内容生态系统。

三、AIGC引爆热点

AIGC是指通过人工智能技术生成的内容。人工智能可以用于创作文字、图像、音频和视频等多种形式的內容。AIGC的创作过程通常涉及以下技术和方法。

(1) 自然语言处理(natural language processing, NLP)。NLP技术用于处理和生成自然语言文本，包括文本生成、摘要、翻译和对话系统等。

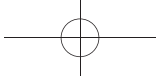
(2) 图像生成和处理。计算机视觉技术可用于生成和处理图像内容，包括图像合成、风格转换和图像增强等。

(3) 音频合成和处理。音频处理技术可用于生成和编辑声音内容，包括语音合成、音乐生成和音频修复等。

(4) 视频生成和编辑。视频生成技术可以利用图像和音频合成方法生成和编辑视频内容，包括场景生成和视频剪辑等。

(5) 强化学习。强化学习算法可用于训练模型，使其生成具有特定目标和约束的内容。

AIGC与UGC和PGC相比，优势更加明显。AIGC不仅可以大大提高内容



6 ✧ AIGC：让生成式AI成为自己的外脑

创作效率，节省时间和人力资源，还可以生成大量内容，并根据用户的偏好和需求进行个性化定制。此外，AIGC可以产生独特和创新的内容，突破传统创作的限制。

AIGC的应用领域非常广泛，包括内容创作、媒体和广告、营销和个性化推荐、虚拟角色和游戏、艺术和创意等。在营销方面，AIGC可以根据用户的兴趣和行为生成个性化的推荐内容，包括产品推荐、新闻推荐和社交媒体内容等。同时，它可以生成虚拟角色的对话和行为，还能设计游戏中的情节和关卡。AIGC也可以生成艺术作品。

总之，AIGC作为一项具有巨大潜力的技术，正在改变内容创作和消费的方式。尽管AIGC面临一些挑战和限制，但随着技术的不断进步和社会适应能力的不断增强，AIGC将在多个领域继续发展，并对创作、传播和消费内容产生深远影响。

第二节 AIGC 的分类

如图1.2所示，AIGC可以根据生成内容、生成技术、生成目的和生成方式进行分类，其中生成技术影响着AIGC的方方面面。

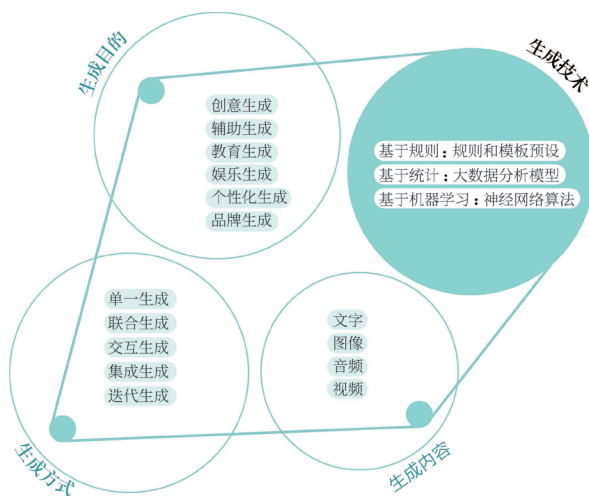


图1.2 AIGC 的分类

一、根据生成内容分类

我们可以根据生成内容类型对AIGC进行分类，具体包括文字、图像、音频和视频等多种形式。

AIGC能够生成各种类型的文字内容，包括文章、新闻报道、故事、对话等。它可以根据给定的主题或写作风格生成与之相符的文字，并且能够模拟不同的语言风格和写作声音。

AIGC可以根据文字描述或简单的指示生成图像内容。它能够生成照片、插图、图表、地图等各种类型的图像，并且可以根据用户需求调整颜色、构图和版式。

AIGC能够生成各种类型的音频内容，包括语音、音乐、声效等。它可以根据指定的语言和情感，生成具有特定语言风格或音乐风格的音频内容。例如，它可以根据文本生成一段具有特定语调的语音^①。

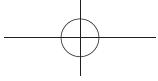
AIGC可以生成各种类型的视频内容，包括短片、动画、广告等。它可以将生成的图像、音频和动画效果组合在一起，生成视频片段；也可以根据脚本或指示，控制视频的主题、情节和节奏。

综上所述，AIGC生成内容是多样化和灵活的，它可以根据用户的需求和输入指示生成各种形式的内容。它的应用领域非常广泛，包括文学创作、广告设计、教育培训等。需要注意的是，AIGC生成的内容基于模型的预测和学习，可能存在一定的主观性和创造性。

二、根据生成技术分类

根据生成技术的不同，AIGC可分为如下几类。

^① CAO Y, LI S, LIU Y, et al. A comprehensive survey of ai-generated content(aigc): A history of generative ai from gan to chatgpt[J]. arXiv preprint arXiv:2303.04226, 2023.



1. 基于规则的生成

基于规则的生成技术使用事先定义的规则和模板来生成内容。规则包括语法、语义和逻辑等方面的规定，它可以确保生成的内容符合特定的要求。例如，在基于规则的文本生成中，首先定义特定句式、词汇选择和句子结构等规则，然后根据这些规则，利用统计模型和概率分布来生成内容^①。它还可以通过分析大量的训练数据，学习数据中的模式和规律，然后根据这些统计信息生成新内容。例如，基于统计技术的语言模型可以根据单词或短语的出现频率，预测下一个单词，从而生成连贯的句子。

2. 基于机器学习的生成

基于机器学习的生成技术使用机器学习算法和模型来生成内容。它通过对大量数据进行学习和训练，建立模型的表示能力，然后使用该模型生成新内容。例如，基于深度学习的生成模型，如循环神经网络(recurrent neural network, RNN)和生成对抗网络(generative adversarial network, GAN)，可以学习生成文本、图像和音频等多种类型的内容。基于强化学习的生成技术，使用强化学习算法和框架来生成内容。它将生成内容视为一个决策过程，通过与环境交互学习生成策略，并不断调整生成过程中的决策和行为，以获得更好的生成效果。例如，在基于强化学习的图像生成中，可以通过不断调整生成器网络的参数，使生成的图像更加逼真和符合要求。

这些生成技术可以单独应用，也可以组合使用，以获得更好的生成效果。不同的生成技术在生成内容的质量、效率和灵活性等方面有所差异，适用于不同的应用场景。因此，用户在选择和应用AIGC时，需要根据具体情况考虑生成技术的类别及特点，以确保生成内容的质量和适用性。

三、根据生成目的分类

根据生成目的的不同，可对AIGC进行分类。例如，在营销领域中，

^① DU H, LI Z, NIYATO D, et al. Enabling AI-Generated Content(AIGC)Services in Wireless Edge Networks[J]. arXiv preprint arXiv:2301.03220, 2023.

AIGC可以生成广告文案、产品说明书等，以满足不同的营销目的；在新闻领域中，AIGC可以生成新闻报道、新闻评论等，以拓展传统媒体的报道范围；在科技领域中，AIGC可以生成科技文章、研究论文等，以推动技术进步；在教育领域中，AIGC可以生成教育材料、测试题等，以提高教学效率。据此进一步细分，可将AIGC分为创意生成、辅助生成、教育生成、娱乐生成、个性化生成和品牌生成。

在创意生成方面，这种类型的AIGC旨在产生具有创意性和独特性的内容。它可以生成艺术作品、诗歌、音乐等具有创意性的内容，旨在启发和激发人们的创造力和想象力^①。

在辅助生成方面，这种类型的AIGC旨在帮助人们完成特定任务或提供帮助。它可以生成各种形式的参考资料、报告、文档等，以满足用户对信息的需求。例如，它可以根据用户提出的问题，生成相关的研究报告或技术文档。

在教育生成方面，这种类型的AIGC可以生成教学材料、教科书、练习题等，以帮助学生学习和理解各种学科知识。它还可以根据学生的个性化需求，生成定制化的学习内容和辅助教材。

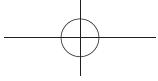
在娱乐生成方面，这种类型的AIGC旨在提供娱乐和娱乐内容。它可以生成电影剧本、游戏情节、角色对话等，以支持电影、游戏和虚拟现实等娱乐产业的创作和开发。它还可以生成幽默段子、趣味小说等，为用户带来娱乐和轻松的体验^②。

在个性化生成方面，这种类型的AIGC旨在根据用户的个性化需求和喜好生成定制化内容。它可以完成个性化推荐、定制化产品设计、个人风格化写作等任务，以满足用户的个性化需求。

在品牌生成方面，这种类型的AIGC旨在帮助品牌和企业进行内容营销和宣传。它可以生成品牌故事、广告语、社交媒体内容等，以提升品牌形象和推广产品或服务。

① WU J, GAN W, CHEN Z, et al. AI-generated content(AIGC): A survey[J]. arXiv preprint arXiv:2304.06632, 2023.

② DEWEY J. Experience and education[C]//The educational forum: vol. 50: 3. [S.l. : s.n.], 1986: 241-252.



这些分类仅为示例，实际上AIGC还可以进一步根据具体需求和应用场景进行功能扩展。不同的生成目的需要不同的生成技术和算法来实现，以确保生成内容符合预期目标并具有相应的质量和效果。在应用AIGC时，用户需要根据具体的生成目的和目标选择适当的分类，并进行相应的调整和优化。

四、根据生成方式分类

根据生成方式的不同，可将AIGC分为单一生成、联合生成、交互生成、集成生成和迭代生成。不同的生成方式有着不同的优势和限制，用户应根据具体情况进行选择。

单一生成是指AIGC专注于生成一种类型的内容。例如，专门生成文本的AIGC，专门生成图像、音频或视频的AIGC，它们会根据指令或模型训练生成相关的内容。

联合生成是指AIGC能够同时生成多种类型的内容，并将它们组合在一起形成更综合的结果。例如，AIGC可以生成一篇文章，并为文章生成配图和相关的音频，从而形成完整的多媒体内容^①。

交互生成是指AIGC能够与用户进行交互，并根据用户的输入和反馈调整和完善生成的内容。通过与用户交互，AIGC可以更好地理解用户的需求和偏好，以生成更符合用户期望的内容。

集成生成是指AIGC能够整合多个不同的模型和算法来生成内容。不同的模型负责不同的任务或内容类型，通过集成它们的生成结果，可以获得更复杂和多样化的内容。例如，AIGC可以同时利用文本生成模型、图像生成模型和音频生成模型来生成一段带有配图和音频的故事。

迭代生成是指AIGC可以根据之前生成的内容进行迭代和改进。它可以通过评估和反馈机制来不断优化生成结果，并根据反馈信息进行调整和学习，

^① BAILEY R, ARMOUR K, KIRK D, et al. The educational benefits claimed for physical education and school sport: an academic review [J]. Research papers in education, 2009, 24(1): 1-27.

以获得更高质量和更符合用户期望的内容。

按生成方式分类是为了描述AIGC生成内容的不同方式和策略。在实际应用中，用户可以根据具体需求和场景选择合适的生成方式，以获得最佳的生成结果。此外，用户可以根据具体需求对这些生成方式进行组合和调整，以满足复杂和多样化的生成要求。

综上所述，AIGC可以根据不同的分类方式进行分组，不同类别的AIGC并不是互相独立的，用户可以依据不同的应用场景和目的进行选择，以提高生成效果，实现更好的应用效果。在未来，随着人工智能技术的不断发展和创新，AIGC将在各种领域得到更广泛的应用。

第三节 AIGC 的发展历史

如图1.3所示，AIGC的发展历史可以追溯到20世纪80年代，当时的研究人员开始探索利用机器学习和自然语言处理等技术生成文本内容。到了20世纪90年代，随着深度学习算法的发展，研究人员开始探索使用神经网络生成文本。但由于硬件性能和数据量的限制，这些技术的应用受到了很大的限制。

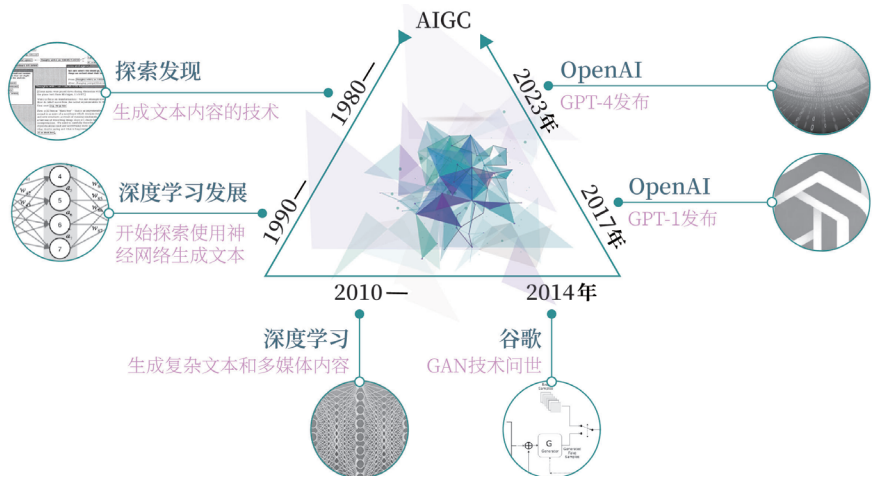
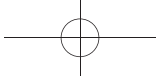


图1.3 AIGC 发展的重要节点



随着互联网的兴起和数据的大量积累，人工智能生成内容技术得以迅速发展。2010年左右，研究人员开始使用深度学习技术生成复杂的文本和多媒体内容。

2014年，谷歌公司发布论文《通过神经网络生成图像》，提出了一种使用神经网络生成图像的方法。这项技术被称为生成对抗网络(generative adversarial networks, GAN)，它可以生成逼真的图像。这是AIGC发展的一个里程碑。

2017年，OpenAI发布了一种新的语言生成模型，称为GPT-1。它使用一种称为“转换器”的神经网络结构，可以生成类似于人类写作的文章。在之后的几年中，GPT-2、GPT-3和GPT-4相继推出，它们的生成效果越来越接近人类写作。

一、早期探索阶段

AIGC的早期探索阶段可以追溯到20世纪80年代，当时的研究人员开始探索利用机器学习和自然语言处理等技术生成文本内容。

在这个时期，人们对于如何使用计算机模拟人类语言的产生和理解充满了好奇。

在20世纪80年代初期，研究人员开始尝试使用基于规则的方法生成文本。这种方法基于语法和句法规则，将用户输入的语言片段转换成规定的文本格式。这种方法存在的问题是需要用户手动编写复杂的规则，并且很难捕捉语言的细微差异和多义性。

随着统计自然语言处理(natural language processing, NLP)技术的兴起，研究人员开始使用概率模型来生成文本。其中较为著名的方法是马尔可夫模型，它是一种基于概率的自然语言处理技术，可以对语言的规律进行建模，但这种方法仍然受到了数据量和计算资源的限制，因为马尔可夫模型需要大量的数据和计算资源来训练^①。

^① CRESWELL A, WHITE T, DUMOULIN V, et al. Generative adversarial networks: An overview[J]. IEEE signal processing magazine, 2018, 35(1): 53-65.

在20世纪90年代，随着深度学习算法的发展，研究人员开始探索使用神经网络生成文本。其中较为著名的方法是循环神经网络(recurrent neural network, RNN)，它可以对序列数据进行建模。这种方法可以学习到语言的长期依赖关系，可以生成更准确的文本，但仍然存在数据量和计算资源的限制。

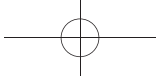
在这个阶段，人们对于AIGC的研究主要集中在文本生成方面。研究人员探索了很多不同的方法和模型，例如基于规则的方法、统计语言模型、神经网络模型等。虽然这些方法存在一定的局限性，但它们为后来的研究提供了很好的启示。

二、数据和硬件的提升阶段

数据和硬件的提升阶段是指从2000年到2010年的这段时间，数据总量的增加以及硬件性能的大幅提升，使得AIGC的研究和应用得以快速发展。本书将从数据和硬件两个方面详细阐述。

在数据方面，随着互联网的普及和数据存储技术的快速发展，人们可以更方便地获取和处理大量的数据。这为AIGC技术的研究提供了更多的数据支持和数据资源。例如，在语言生成方面，人们可以使用互联网上的大量文本数据来训练模型。在图像生成方面，人们可以使用大量的图像数据来训练模型。在数据标注方面，随着人工智能的快速发展，人们可以使用自动化工具对数据进行标注，从而更快速地获取大量标注数据。这些数据的增加和质量提升，极大地推动了AIGC技术的发展和應用。

在硬件方面，随着计算机硬件性能的提升和计算资源的大幅增加，研究人员可以使用更复杂和深层次的神经网络模型。例如，GPU的出现使得神经网络的训练速度大幅提升，研究人员可以更快速地训练更深层次的神经网络模型。云计算和分布式计算的发展，使得研究人员可以利用多台计算机进行并行计算，从而更快速地训练和应用复杂的AIGC模型。这些硬件和计算资源的提升，大大促进了AIGC技术的研究和应用，为人工智能产业的发展带来了新动力。



除此之外，数据和硬件的提升也为AIGC技术带来了新的应用场景。例如，在语音识别和自然语言处理领域，AIGC技术可以帮助人们开发语音助手、聊天机器人等智能应用。在图像识别和计算机视觉领域，AIGC技术可以帮助人们开发自动驾驶、安防监控等智能应用。在生物医学和化学领域，AIGC技术可以帮助人们开发新药和新材料等。

在这个阶段，AIGC技术的研究和应用呈现如下趋势。首先，深度学习成为AIGC技术的主流，这种基于神经网络的学习方式可以自动从大量的数据中学习特征和规律，可以实现端到端的学习和预测；其次，自监督学习成为一个新热点，这种学习方式不需要标注数据，而是从未标注的数据中学习^①，可以大大降低数据标注的成本；最后，AIGC技术开始朝联合学习和跨模态学习方向发展，这种技术可以将多个模态的数据和知识融合起来，从而更好地模拟人类的多模态感知和智能决策能力。

除了技术研究和应用方面的发展，数据和硬件的提升也促进了AIGC产业的发展。越来越多的公司和机构开始投资和研发AIGC技术，同时也出现了一批专门从事AIGC技术研究和应用的公司和机构，如Google、IBM、Facebook、OpenAI等。这些公司和机构在AIGC技术研究和应用方面取得了很多成果，推动了整个行业的发展。

总之，数据和硬件提升阶段是AIGC技术和产业快速发展的时期，数据和硬件的提升促进了AIGC技术的研究和应用，同时也推动了人工智能产业的发展 and 壮大。在未来，随着数据和硬件的不断提升以及人工智能技术的不断进步，AIGC技术研究将会更加深入，拥有更广泛的应用场景。

三、GAN 技术的引入阶段

GAN是一种新型深度学习网络，被誉为“人工智能领域的一个里程碑”。GAN技术的引入阶段是AIGC技术发展的重要阶段之一。GAN的引入，让AIGC技术有了更多的发展可能性和应用场景，成为人工智能研究和应

① WANG K, GOU C, DUAN Y, et al. Generative adversarial net works: introduction and outlook[J]. IEEE/CAA Journal of Automatica Sinica, 2017, 4(4): 588-598.

用领域的重要进展。

释义 1.4：生成式对抗网络

生成式对抗网络(generative adversarial networks, GAN)是一种用于无监督学习的神经网络，由 Ian Goodfellow 于 2014 年开发。

GAN技术的引入阶段为2014年到2016年。GAN是一种基于对抗训练的生成模型，它包含生成器和判别器。生成器通过学习样本数据的分布，生成与真实样本相似的数据。判别器则用于将真实数据与生成数据区分开来^①。生成器和判别器相互对抗，互相提高对方的性能，从而达到生成与真实数据分布相似的数据的目的。

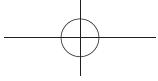
GAN技术的引入，对AIGC技术的发展产生了深远的影响。首先，GAN技术可以生成高质量的数据，例如图像、音频、文本等数据，这使得人们可以更加方便地获取并利用大量的数据。其次，GAN技术的生成模型可以用于数据增强、样本生成和模型预训练等任务，为人工智能应用提供了更多的可能性。最后，GAN技术可以用于图像修复、图像融合等领域，能为人们提供更好的视觉和感官体验。

随着GAN技术的不断发展和完善，越来越多的研究人员开始将其应用到实际场景中。例如，在医学影像处理领域，GAN技术可以用于图像分割、图像配准和医学影像生成等任务。在游戏和虚拟现实领域，GAN技术可以用于游戏场景的自动生成和虚拟人物的生成。在文学创作领域，GAN技术可以用于自动生成小说、诗歌等文学作品^②。

总体来说，GAN技术的引入为AIGC技术的发展注入了新的动力和活力，使得人工智能的应用场景更加丰富和多样化。随着技术的不断完善和推进，GAN技术将会在更多的领域发挥作用，为人们带来更多的惊喜和创新。

① METZ L, POOLE B, PFAU D, et al. Unrolled generative adversarial networks[J]. arXiv preprint arXiv:1611.02163, 2016.

② FLORIDI L, CHIRIATTI M. GPT-3: Its nature, scope, limits, and consequences[J]. Minds and Machines, 2020, 30: 681-694.



四、语言生成模型的兴起阶段

语言生成模型是AIGC技术领域的一项重要技术，从2017年至今，语言生成模型被广泛应用。它可以自动生成人类可读的语言，包括文本、对话、故事等。语言生成模型的兴起，为人工智能在文本处理和自然语言处理领域的应用提供了强大的支持。

早期，人们使用基于规则的方法生成文本，这些规则是由专业的语言学家和领域专家手动编写的。这些方法存在一些明显的问题，如规则的维护和更新成本高、模型的可扩展性差等。因此，人们开始尝试使用机器学习方法生成语言。

随着深度学习技术的发展，基于神经网络生成模型的方法成为主流。最早的基于神经网络生成的模型是循环神经网络(recurrent neural network, RNN)和长短时记忆网络(long short-term memory, LSTM)。RNN和LSTM可以根据已生成的文本来预测下一个单词，从而生成连续的文本。这些模型的优点在于可以处理不定长的输入和输出序列，但存在梯度消失的问题，也容易生成重复的文本。

为了解决这些问题，人们开始使用基于变分自编码器(variational auto-encoder, VAE)和GAN的方法进行语言生成。VAE和GAN都是比较先进的生成模型^①，它们能够生成更加逼真、多样和连贯的文本。VAE和GAN的不同之处在于，VAE通过在隐空间中对输入数据进行编码，并在隐空间中进行插值和解码来生成新的样本；而GAN则是使生成模型和判别模型对抗学习，生成模型产生的样本需要通过判别模型来确定真伪，从而生成更加逼真的样本。

此外，还有一种基于Transformer的语言生成模型，它是目前应用较为广泛的语言生成模型之一。Transformer是由Google在2017年提出的一种新型神经网络结构，它可以处理长序列数据，且效果优于传统的循环神经网络和

^① ELKINS K, CHUN J. Can GPT-3 pass a Writer's turing test?[J]. Journal of Cultural Analytics, 2020, 5(2).

卷积神经网络。基于Transformer的语言生成模型主要是GPT系列模型，包括GPT-1、GPT-2和GPT-3。

GPT系列模型的显著特点及关联词如图1.4所示，它主要具有支持多轮对话、支持多种语言的应答交互、在多领域及应用场景的强可扩展性、根据用户兴趣及使用记录生成智能推荐和自我学习能力等特点。GPT模型通过对大规模语料库进行预训练，可以生成与原始文本相似的连贯且语义合理的文本。GPT-2模型在2019年推出后引起了广泛关注，其生成效果非常出色，甚至可以生成足以欺骗人类的假新闻。GPT-3模型在2020年发布，它拥有了迄今为止最大的参数量，可以自动生成文本、代码、音乐甚至图像。

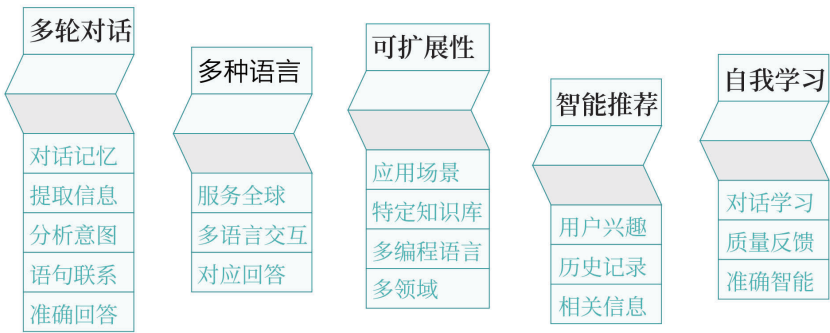


图1.4 GPT 系列模型的显著特点及关联词

语言生成模型还有一些其他应用，如文本摘要、机器翻译、对话系统等。随着自然语言处理技术的不断发展，语言生成模型在人工智能领域的应用前景也越来越广阔。

综上所述，在AIGC的发展历程(见图1.5)中，语言生成模型的兴起对于其技术和产品的进一步完善起到了关键作用。AIGC在语言生成模型的研究和应用方面，不断进行探索和尝试，不断推出更加先进、高效的技术，使得其在人工智能领域中不断保持领先地位。

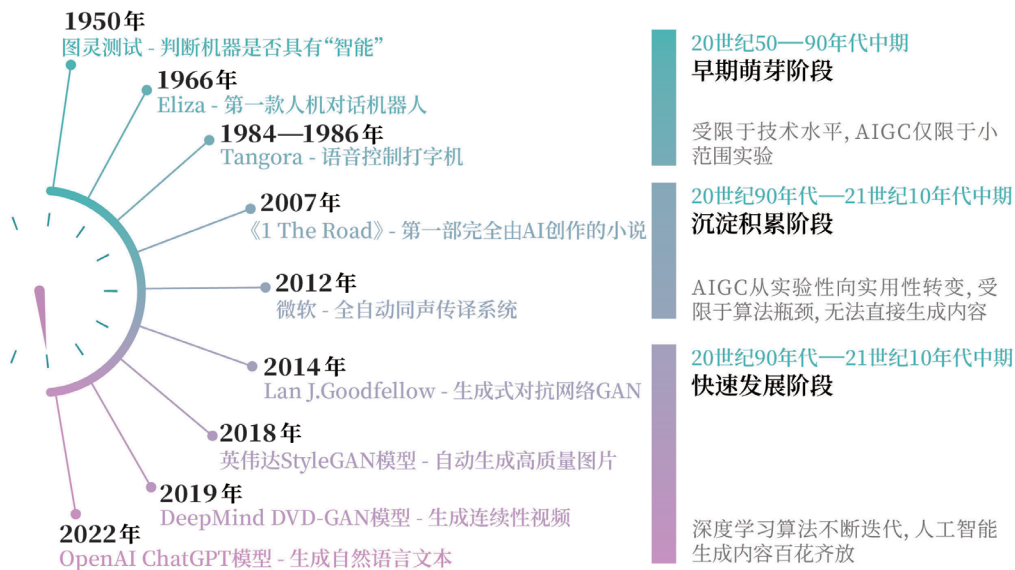
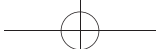


图1.5 AIGC 的发展历程



第二章 AIGC的基础技术栈

AIGC的落地离不开多种技术的结合和发展。本章主要介绍AIGC的基础技术栈。通常来讲，AIGC的技术栈涵盖识别的技术、理解与输出的技术、可视的技术和创作的技术四个部分。本章将按照顺序介绍四种技术，讲述AIGC从识别到创作的全过程。