

生成式人工智能介绍

人工智能正站在技术革新的前沿,引领我们进入一个充满创造力和可能性的新时代。在人工智能的发展历程中,生成式人工智能代表了一种质的飞跃——它不再局限于分析和识别现有数据,而是能够自主生成与人类创作相媲美新颖内容。这种技术的进步不仅扩大了人工智能的应用范围,也革新了我们的工作、生活和娱乐方式。

生成式人工智能的核心在于其创造性和自主性,它基于深度学习,通过学习大量数据中的模式和结构来生成新的内容。从图像到文本,从音频到代码,生成式人工智能的应用案例层出不穷,不断刷新我们对机器创造力的认知。例如,DALL-E 在图像生成领域所展示的惊人成果,Copilot 在编程领域的辅助能力,ChatGPT 在文本生成和对话交流中的表现,都是AIGC(Artificial Intelligence Generated Content,人工智能生成内容)技术实力的体现。国内相继涌现了一批优秀的 AIGC 大模型,如 CSDN 的“C 知道”、百度的“文心一言”、阿里巴巴的“通义”、360 的“智脑”、字节跳动的“豆包”以及深度求索的 DeepSeek。其中,DeepSeek 的出现尤为引人注目。作为中国新兴的人工智能力量,DeepSeek 的国际影响力非常显著,它不仅打破了美国在 AI 领域的技术垄断,还通过开源模式推动了全球 AI 技术的普惠发展。凭借低成本和高性能的优势,DeepSeek 在全球市场中展现出了强大的竞争力,吸引了众多国际企业和科研机构的关注与合作,其技术创新也促使全球产业链进行了重构,为国际合作带来了新的契机。DeepSeek 的崛起不仅提升了中国在全球 AI 领域的竞争力,还为全球科技合作和产业发展带来了新的机遇与挑战。

随着技术的不断进步,生成式人工智能正在迅速发展,吸引了众多国内外公司和组织的关注。这些参与者的积极投入推动了生成式人工智能技术的创新和应用,使其在艺术创作、编程辅助、个性化服务等多个领域展现出巨大的潜力。

生成式人工智能的发展也伴随着挑战和问题,如版权、伦理和隐私保护等,需要我们在享受技术红利的同时,也对其进行审慎的思考和规范的使用。本项目将带您深入了解生成式人工智能的发展历程、技术原理、应用领域以及面临的挑战,共同探索这一创造性技术是如何重塑我们的世界的。

学习目标

- (1) 掌握生成式人工智能的概念。
- (2) 了解生成式人工智能的发展历程。

- (3) 了解生成式人工智能的常见技术。
- (4) 了解生成式人工智能的典型应用平台。

任务 1.1 生成式人工智能的概念



任务 1.1
学习助手

生成式人工智能是一种能够创建新数据样本的人工智能技术。这些样本可以是文本、图像,也可以是音频、视频等形式,它们既源于训练数据又超越了原始数据的范畴。与传统的人工智能相比,生成式人工智能更侧重于模拟人类的创造力,通过学习数据的潜在结构来生成新颖的内容。了解这些基本概念有助于我们认识生成式人工智能在推动技术创新、提升用户体验和开拓新业务模式方面的巨大潜力。同时,这也为我们提供了必要的背景知识,以便在面对由该技术引发的伦理、隐私和法律等问题时,能够做出更为明智的决策。简言之,掌握生成式人工智能的基础知识对于推动其负责任地发展和应用至关重要。

本任务是掌握生成式人工智能的基本概念,包括其定义、发展历程以及与传统人工智能的区别。

1.1.1 生成式人工智能的定义

生成式人工智能是一种基于算法、模型和规则生成各类内容的技术,这些内容包括但不限于文本、图像、音频、视频和代码等。这类技术能够针对用户需求,依托事先训练好的多模态基础大模型等,利用用户输入的相关资料,生成具有一定逻辑性和连贯性的内容。与传统人工智能不同,生成式人工智能不仅可以处理输入数据,而且能学习和模拟事物内在规律,自主创造出新内容。生成式人工智能生成图像如图 1.1 所示。



图 1.1 生成式人工智能生成图像

生成式人工智能主要基于深度学习和神经网络技术,通过学习大量的数据和模式来生成新的内容。其中,最具代表性的生成式人工智能模型就是生成对抗网络(Generative Adversarial Network, GAN)。GAN 由一个生成器网络和一个判别器网络组成,它们相互竞争并共同进步,以产生逼真的生成样本。

生成式人工智能在多个领域都有应用,如图像生成、文本生成、音频生成等。例如,在图像生成方面,生成式人工智能可以学习大量真实图像数据的分布特征,并生成具有相似特征的新图像;在文本生成方面,生成式人工智能可以学习文本数据的语义和语法结构,并生成具有连贯性和多样性的新文本。

1.1.2 生成式人工智能发展历程

生成式人工智能从早期的概念到现代高度复杂的应用,这一领域经历了几个重要的阶段。生成式人工智能发展历程如图 1.2 所示。

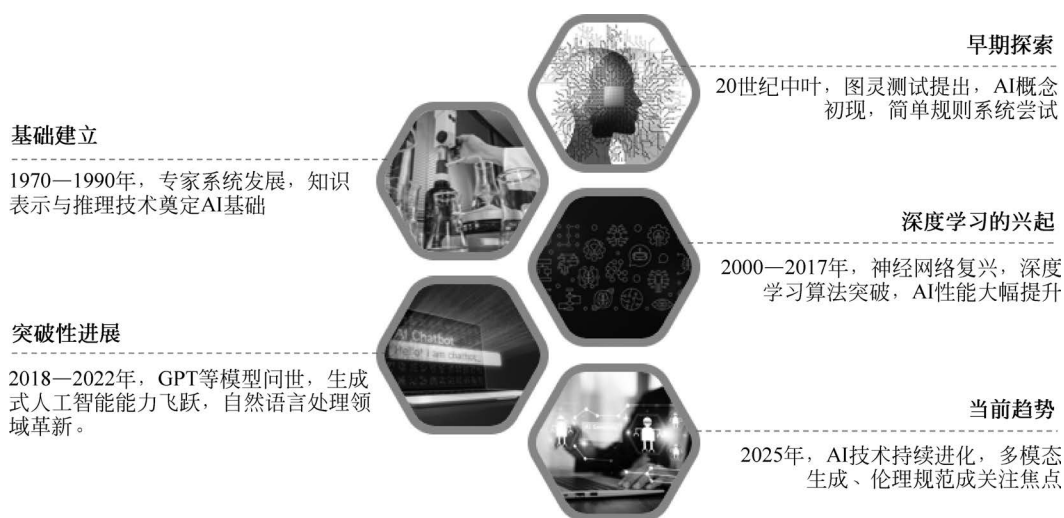


图 1.2 生成式人工智能发展历程

1. 早期探索(20 世纪中叶)

生成式人工智能的起源可以追溯到 20 世纪中叶,艾伦·麦席森·图灵(Alan Mathison Turing)提出了著名的图灵测试,这是评估机器是否具备人类智能的一种方法。

20 世纪 60 年代,约瑟夫·魏岑鲍姆(Joseph Weizenbaum)开发了 ELIZA,这是最早的自然语言处理程序之一,能够模拟心理治疗师的角色。

2. 基础建立(1970—1990 年)

这一时期的人工智能研究主要集中在符号主义和连接主义两种不同的范式上。虽然生成式人工智能不是当时的主流,但是相关技术如神经网络的研究为后来的发展奠定了基础。

3. 深度学习的兴起(2000—2017 年)

21 世纪初,随着计算能力的提升和大数据可用性的增加,深度学习开始兴起。深度学习技术为生成式人工智能提供了强大的工具。

2014 年,伊恩·古德费罗(Ian Goodfellow)等提出了生成对抗网络,这是一种用于生成新数据样本的神经网络架构,极大地推动了生成式人工智能的发展。

2017 年,Google 的研究人员提出了 Transformer 模型,这是一种基于注意力机制的序列到序列模型,极大地提高了序列生成任务的效率和质量。

4. 突破性进展(2018—2022 年)

2018 年,OpenAI 发布了 GPT(Generative Pre-trained Transformer)模型,这是一个预训练语言模型,能够生成高质量的文本。随后的版本 GPT-2、GPT-3 进一步提升了性能。

2022 年年末,OpenAI 推出了 ChatGPT,这是一个对话式的生成模型,能够进行高质量的对话交互,标志着生成式人工智能在文本生成领域取得了显著进展。

随着技术的进步,生成式人工智能被应用于多个领域,如图像生成、视频制作、音乐创

作、药物发现等。

5. 当前发展(2025 年)

当前生成式人工智能的发展趋势如下。

(1) 大规模预训练模型。当前的趋势是构建更大规模的预训练模型,这些模型能够在各种任务上展现出出色的表现,且具有更强的泛化能力。

(2) 多模态生成。多模态生成是研究热点,即模型能够处理多种类型的数据,如文本、图像和音频的组合。

(3) 个性化与定制化。生成式人工智能正越来越多地用于个性化推荐、定制化内容创建等领域。

(4) 伦理与安全。随着技术的发展,人们越来越关注生成式人工智能的伦理问题和社会影响,包括隐私保护、偏见消除等方面。

当前生成式人工智能相关概念如下。

PGC(专业生成内容): 新闻报道、电影、电视节目、专业文章、学术论文等。

UGC(用户生成内容): 社交媒体上的帖子、评论、博客文章、用户上传的视频照片。

AIAGC(人工智能辅助生成内容): 自动校对工具、智能写作助手、PS 中的 AI 滤镜等。

AIGC(人工智能生成内容): ChatGPT 生成的文本、DALL-E 生成的图像、生成对抗网络生成的视频等。

1.1.3 生成式人工智能与传统人工智能

传统人工智能常被称作“基于规则的人工智能”或“判别式人工智能”,其运作依赖于事先设定的规则和庞大的数据集。传统人工智能的核心原理是利用丰富的数据资源对模型进行训练,使其能够辨识并捕捉数据中的关键特征,并依据这些特征执行任务分类或预测。传统人工智能涵盖了一系列机器学习技术,包括决策树、支持向量机(Support Vector Machine, SVM)、逻辑回归等算法。以图像识别为例,如果要训练一个传统人工智能系统来区分猫和狗,就需要向系统提供大量已经标记好的猫狗图片。系统将学习这些图片中猫和狗各自的特征,如猫的尖耳朵或狗的圆鼻子,从而实现对图像的准确分类。

判别式人工智能可以基于已有数据作出判断或分类,主要用于分类或识别数据。生成式人工智能可以创造出新颖的输出,这些输出可能是文本、图像、语音、视频甚至是模拟场景等形式。以下是两种人工智能的主要区别。

1. 主要特点

判别式人工智能专注于执行特定任务,如识别或分类,在预定义的规则下工作,无法自主创造新内容。

生成式人工智能能够创造新事物,如文本、图像、音乐和代码,通过学习数据集中的模式来生成新数据。

2. 学习过程

判别式人工智能通过收集特定数据集、数据预处理、分割、训练模型、模型评估和优化来学习。

生成式人工智能可以收集数据、预处理、创建训练模型、反向传播和优化、调整数据,并最终生成新内容。

3. 功能与应用场景

判别式人工智能主要用于数据分析和预测,如垃圾邮件识别、下国际象棋等。

生成式人工智能用于创造全新的内容,如 GPT-4 生成文本、DeepArt 绘画转换。

4. 学习方法

判别式人工智能通常需要标记数据进行控制式学习。

生成式人工智能可以进行不受控制的学习,不需要标记数据进行训练。

5. 限制

判别式人工智能受限于具体任务,无法创新原创内容。

生成式人工智能生成的内容可能不够一致或准确,细节不受控。

6. 典型应用场景

判别式人工智能的典型应用场景为 Spam Sieve 邮件过滤器、语音助手(Siri、Alexa)、推荐系统(Netflix、亚马逊)、搜索引擎(谷歌)。

生成式人工智能的典型应用场景为 OpenAI 的 GPT-4、DeepArt 绘画转换、创建内容(故事、艺术、音乐)、DeepFake(AI 换脸)。

7. 底层技术

判别式人工智能基于特定算法,如决策树、SVM、逻辑回归等。

生成式人工智能基于大型神经网络模型,如 GPT 系列,能够处理非结构化数据并生成内容。

8. 未来展望

判别式人工智能继续在特定领域内提高效率和准确性。

生成式人工智能将在内容创造、创新和多个行业中发挥更大的作用。

任务 1.2 生成式人工智能的工作流程



生成式人工智能的工作流程至关重要,因为它贯穿了从模型构建到实际应用的整个过程。通过深入了解这个流程,我们可以更好地把握模型设计与训练的关键步骤,如所需算力、数据质量和算法选择等,这对于确保模型的有效性和可靠性是基础性的。此外,工作流程中的每个环节都可能涉及内容安全、个人信息保护、模型安全及知识产权等问题,这些问题如果处理不当可能会导致严重的法律和社会后果。因此,只有全面理解这些流程和技术背景,我们才能够有效地规避风险,确保技术的健康可持续发展。同时,这也有助于我们利用生成式人工智能的潜力来解决实际问题,提高生产力,并在教育、

任务 1.2
学习助手

娱乐、医疗和科研等领域创造出新的可能性。

本任务是掌握大模型构建条件及生成式人工智能的工作流程。

1.2.1 大模型的构建条件

生成式人工智能本质上是一个被用户使用的算法服务,其生命周期包括模型训练、服务上线、内容生成、内容传播四个阶段。该算法服务通常由语言大模型和视觉大模型驱动,离不开算力、数据、算法、生态和人才五个构成条件。因此,要理解和分析生成式人工智能,首先需要对大模型的五个构成条件有所了解。

1. 算力

(1) 训练和运营需求。生成式人工智能需要大量的计算资源来支持其训练和运营。例如,ChatGPT 使用的 Azure 超级计算机包含了 28.5 万个 CPU 核心、1 万个 GPU 和 400Gb/s 的 GPU 间传输带宽。

(2) 算力消耗。在训练阶段,ChatGPT 每天的算力消耗约为 3640P,运行 ChatGPT 需要 7~8 个投资规模 5 亿美元、算力 500P 的数据中心来支撑,每次训练的成本约为 500 万美元。而在运营阶段,仅 GPU 的年投入就需要 7000 万美元。

(3) 高性能硬件。随着算力需求的增长,高性能硬件的研发变得至关重要。英伟达联合多家生成式人工智能企业正共同研发适合此类计算需求的芯片架构和计算引擎。

2. 数据

(1) 数据的重要性。数据被视为知识,是生成式大模型能力提升的关键因素。训练数据集通常包含大量书籍、网络百科全书、论坛和博客等高质量资源。

(2) 数据规模。训练数据集的规模通常接近 8000 亿个分词 token 和上百 TB(数据清洗前)。

(3) 数据质量。为了提升数据集的质量,多数公司会通过外包或众包的方式对数据集进行手动标注。

(4) 数据合规性。数据集来源可能涉及个人信息和国家安全问题,因此需要重视数据合规监管和质量评估。

3. 算法

(1) 技术基础。生成式人工智能基于深度学习技术,涉及统计学、概率论和机器学习等基础知识。

(2) 关键技术。生成式人工智能的主要技术包括 Transformer 神经网络模型、基于人类反馈的强化学习(Reinforcement Learning from Human Feedback, RLHF)、零样本学习(Zero-Shot Learning)和提示学习(Prompt Learning)等。

(3) 未来趋势。未来生成式大模型将朝着支持多模态(图像、文本、语音、视频)的方向发展,具备更广泛的应用能力和“基础设施”式的属性。

4. 生态

(1) 开放 API。谷歌、微软等公司将生成式人工智能成果通过 API 开放给用户调用,并鼓励二次开发。

(2) 迭代生态。大模型借助用户反馈进行优化,依托 GitHub 等开源社区促进版本更新,形成了良好的双向迭代生态。

(3) 产业化进程。大模型有效解决了应用碎片化等问题,降低了应用门槛,促进了人工智能技术的产业化进程。

5. 人才

(1) 跨学科要求。生成式人工智能需要多学科背景的专业人士,包括风险投资、人工智能、航空航天和自动驾驶等领域的专家。

(2) 团队组成。以 OpenAI 为例,核心团队由来自全球顶尖高校和知名企业的 90 位成员组成。

(3) 资金投入。微软在 2019 年向 OpenAI 投资 10 亿美元,并在 2023 年 1 月追加了 100 亿美元的投资,用于研究团队的扩充和技术基础设施的建设。

1.2.2 生成式人工智能的工作流程

生成式人工智能的基本原理是使用概率模型或神经网络模型,将已有数据的结构和规律学习到模型中,并基于这些结构和规律生成新的数据。生成式人工智能的工作流程如下。

1. 采集数据

与传统人工智能项目一样,生成式人工智能工作的第一步是采集数据。对于 GPT 模型来说,其数据由大量文本组成。例如,GPT-4 就是通过互联网上数 GB 的文本进行训练的。

2. 预处理数据

数据预处理即对数据进行清洗,并将其转换为模型可以理解的格式。

3. 模型训练

用大量预处理后的数据来训练生成式人工智能模型。在训练过程中,生成式人工智能模型会学习输入数据的概率分布和结构,这些数据可以是文本、图像、音频或视频等。

4. 模型选择

模型训练完成后,需要选择合适的模型来生成新的数据。不同类型的数据需要选择不同的模型来生成,如自然语言文本可以使用 RNN 或 LSTM 模型来生成,图像可以使用 GAN、VAE、分布式可逆变换模型或扩散模型来生成,如图 1.3 所示。

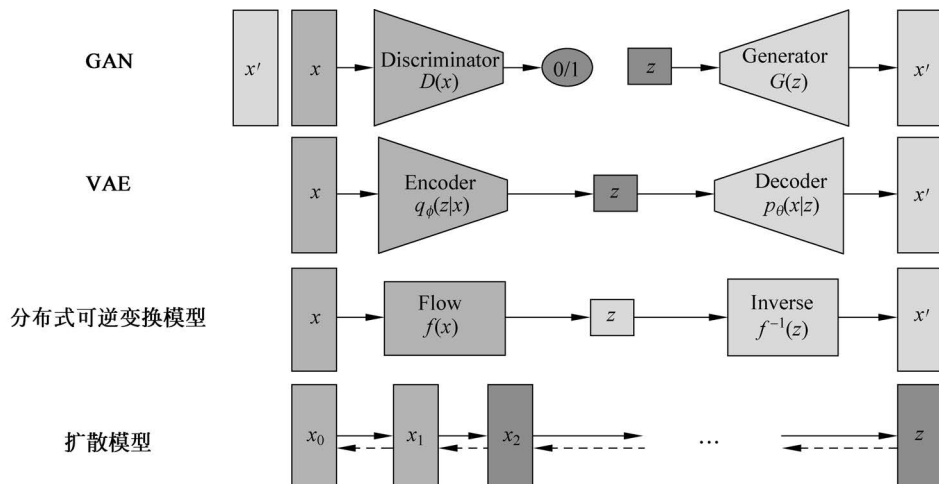


图 1.3 不同类型的生成式模型比较

5. 生成数据

一旦选择好合适的模型,就可以使用该模型来生成新的数据。生成新数据的方式通常是随机采样或条件采样。随机采样是指从模型学习到的数据分布中随机抽样生成新的数据,而条件采样是指在输入一些条件的情况下,从模型学习到的条件分布中采样生成新的数据。

6. 评估生成结果

为达到预期标准,生成的数据需要经过评估。这种评估既可以通过客观度量标准来进行,也可以依赖于人类的主观判断。例如,对于文本,可以检查其语法正确性、逻辑连贯性和意义合理性;对于图像,则可以评价其视觉质量和逼真度。

7. 调整模型

根据评估的结果,可以对模型进行调整和改进,以进一步提高生成数据的质量。例如,增加更多的训练数据,微调模型参数或改进模型架构等。

任务 1.3 生成式人工智能的常见技术



任务 1.3

学习助手

学习生成式人工智能技术,如文本、语音、图像、视频以及多模态生成技术,对于现代社会的发展具有深远的意义。这些技术不仅推动了信息技术领域的创新,还深刻影响着人们的日常生活和社会经济的多个方面。文本生成技术可以自动化创建高质量的书面内容,从文章撰写到代码编写,再到创意写作,极大地提高了内容生产的效率和多样性;语音生成技术则让机器拥有了更加人性化的交流方式,无论是智能助手还是自动化的客户服务,都因之变得更加自然流畅,提升了用户体验;图像生成技术的应用领域广泛,从艺术创作到医学影像分析,再到虚拟现实中的环境构建,都在不断地拓展着人类的视觉边界;视频生成技术更是将这种创造性推向了一个新的高度,通过合成逼真的视频内容,不仅为娱乐和媒体行业注入了新的活力,也促进了教育和培训资源的创新;而多模态生成技术则进一步整合了各种形式的信息处理能力,实现了跨媒体的内容理解和生成,为智能交互系统提供了强大的技术支持,促进了人机交互的自然化和智能化发展。综上所述,掌握并应用这些生成式人工智能技术,不仅可以提升个人或企业的竞争力,还能促进社会整体向更高层次的信息文明迈进。

本任务是掌握生成式人工智能的常见技术,包括文本生成技术、语音生成技术、图像生成技术、视频生成技术、多模态生成技术。

1.3.1 文本生成技术

文本生成的核心是通过训练模型,使其学习文本序列中的上下文关系,然后根据给定的主题生成相关文本,实现了人工智能在文本创作领域的应用。通过引入更多的训练数据、改进模型架构和优化训练算法等方式可增强模型的表现能力。随着这些改进的逐步实施,可以预测未来会有更精确、更流畅、更一致的文本生成结果。

1. 文本生成的核心技术

(1) 循环神经网络(Recurrent Neural Network, RNN)。RNN 是一种处理序列数据的

经典生成式模型,它能够捕捉文本中的长期依赖关系。尽管 RNN 在处理短序列数据时表现良好,但在处理长序列数据时容易遭遇梯度消失或爆炸的问题。为此,长短期记忆网络(LSTM)作为 RNN 的改进版本,通过引入门控机制(遗忘门、输入门和输出门)有效解决了这些问题。

(2) 变换器(Transformer)。与 RNNs 和 LSTM 不同,Transformer 是一种基于自注意力机制的模型,它摒弃了循环结构,采用并行处理输入序列的方式,提高了计算效率。Transformer 在自然语言处理(NLP)任务中表现尤为出色,特别是机器翻译、文本分类等领域。

(3) 生成对抗网络。生成对抗网络由生成器网络和判别器网络组成,通过生成器生成文本并由判别器评估其真实性,从而在训练中不断提升生成质量。生成对抗网络在文本生成中的应用主要体现在增强文本的多样性和流畅性。

(4) 变分自编码器(Variational Autoencoder, VAE)。VAE 是另一种常见的生成式模型,通过编码器将输入数据压缩成潜在表示,然后通过解码器将其重建为输出数据。VAE 在文本生成中的应用主要包括生成多样、合理的文本片段。

2. 文本生成应用场景

(1) 简历润色。通过分析大量高质量简历的数据特征,使用 GAN 或 VAE 等模型可以自动生成简洁、清晰、格式规范的简历内容。用户仅需输入个人基本信息,模型便能快速生成个性化简历。

(2) 面试准备。生成式人工智能能够根据历史面试问题及回答,模拟生成面试官可能提出的问题及回答要点,帮助求职者做好充分准备。这种应用结合 NLP 技术和情感分析,可生成既符合语言规范又包含个人特色的回答。

(3) 歌词生成。歌词创作往往需要灵感和创造性,而生成式人工智能通过学习大量音乐作品的歌词,可捕捉音乐的风格和情感,从而生成新颖、连贯的歌词。RNN 和 LSTM 在这方面尤其适用,因为它们擅长处理文本序列的依赖关系。

(4) 代码生成。随着自动编程的发展,生成式人工智能能够辅助程序员快速生成代码。在编写复杂的函数或程序时,模型通过分析大量的开源代码和 API 文档,可以自动完成代码的填充和优化。这种方式能够大大提高开发效率,尤其对于非专业开发者而言更具实用价值。

(5) 活动计划。在活动策划和管理中,生成式人工智能可以根据历史数据和用户需求,自动生成活动的初步计划,包括时间节点、活动内容、资源分配等。这不仅节省了大量的时间和人力成本,还为活动的策划提供了更加科学合理的方案。

1.3.2 语音生成技术

语音生成技术(Speech Synthesis)是指通过计算机或其他设备,将文本或其他形式的输入转换为人类可理解的语音输出的技术。这种技术在智能助手、导航系统、游戏配音、客户服务系统、新闻播报和语言学习软件等多个领域都有广泛应用。

1. 语音生成的核心技术

语音生成技术主要包括文本到语音(Text-to-Speech, TTS)的合成方法,其中代表性的技术包括 Tacotron、WaveNet 和 FastSpeech 等。

(1) Tacotron。Tacotron 是采用 Seq2Seq 模型结构,结合注意力机制,将文本输入转换

为语音的特征表示,再通过声码器(Vocoder)将特征转换为波形。

(2) WaveNet。WaveNet 是一种基于自回归模型的声码器,能够生成高质量的语音波形,但计算复杂度较高。

(3) FastSpeech。为了克服 Tacotron 生成速度慢的缺点,FastSpeech 采用了非自回归模型结构,大幅提高了生成速度,同时保持了较好的音质。

2. 语音生成应用场景

(1) 智能助手。通过语音生成技术,智能助手可以更加自然地与用户进行交互,提供更加便捷的服务。

(2) 导航系统。在驾驶或步行导航过程中,语音生成技术可以将路线信息转换为语音播报,提高用户体验。

(3) 游戏配音。语音生成技术可以为游戏角色生成逼真的语音,增强游戏的沉浸感和真实感。

(4) 客户服务系统。在自动客服系统中,通过语音生成技术可以实现与用户的语音交互,提供快速、便捷的客户服务。

(5) 新闻播报。语音生成技术可以将新闻内容转换为语音进行播报,方便用户在驾驶或忙碌时获取新闻信息。

(6) 语言学习软件。在语言学习软件中,语音生成技术可以帮助用户模拟真实的语言环境,提高语言学习效果。

1.3.3 图像生成技术

图像生成技术是通过计算机算法生成逼真或创意性的图像。该技术通常依赖于深度学习模型,尤其是 GAN 和 VAE,从噪声或其他输入数据中生成图像。

1. 图像生成的核心技术

图像生成技术主要包括 GAN 及其变种,如 StyleGAN、ProgressiveGAN 等。

(1) GAN。GAN 由生成器(Generator)和鉴别器(Discriminator)两个网络组成。生成器负责生成图像,鉴别器负责区分生成的图像和真实图像,两者相互对抗,共同提高生成图像的质量。

(2) StyleGAN。在 GAN 的基础上,StyleGAN 引入了风格向量(Style Vector)的概念,通过控制风格向量可以生成具有不同风格的图像。

(3) ProgressiveGAN。ProgressiveGAN 采用渐进式增长的策略,从低分辨率到高分辨率逐步训练生成器和判别器,提高了生成图像的质量和稳定性。

2. 图像生成应用场景

(1) 艺术创作。图像生成技术可以生成新的艺术作品,如绘画、插图和设计。技术细节包括 GANs 和 VAEs,能够学习艺术风格并生成新的艺术作品。

(2) 虚拟现实(Virtual Reality, VR)。图像生成技术可以创建逼真的三维场景。技术细节包括实时渲染技术和三维建模算法,可以创建沉浸式的虚拟环境。

(3) 广告设计。图像生成技术可以生成新的广告图像和布局。技术细节包括图像识别和图像合成算法,可以确保广告内容吸引目标受众。