生成式人工智能缘起、发展与现状

**摘要**：本文回顾了生成式人工智能的发展历程，从人工智能的早期探索到知识工程阶段，再到经历的两次“冬天”，直至机器学习与神经网络的复兴以及大数据和深度学习的兴起。接着，详细分析了生成式人工智能在多个领域的应用现状。最后，文章讨论了生成式人工智能在广泛应用中面临的风险。本文旨在展示生成式人工智能的巨大潜力，同时引起对其潜在问题的重视，推动其健康发展和安全应用。

**关键词**：生成式人工智能；机器学习；ChatGPT；深度学习；潜在风险

**作者**：0214845 软件214班 彭慧

随着技术的迅速进步，人工智能已经成为当今世界经济和社会发展的重要推动力。生成式人工智能（Generative Artificial Intelligence）作为人工智能的一个重要分支，基于机器学习的生成模型的构建，其核心在于能够从已有的数据中学习，并创造性地生成新的数据[1]，这一能力使得生成式人工智能在多个领域展现出巨大的潜力和应用前景。在自然语言处理领域，生成式人工智能已经能够生成高质量的文本，并在对话系统、机器翻译、文本摘要等方面得到广泛应用。例如，OpenAI公司于2020年提出的第三代文本生成模型(GPT-3)[2]，以其强大的模型能力和小样本学习能力，入选了2021年MIT Technology Review的“全球十大突破性技术”。其发布的ChatGPT自然语言处理模型，自公开发布以来在5天内就吸引了超过百万用户，在一个月内拥有了超过5000万活跃用户，引发了生成式人工智能的研究热潮[3]。本文在研读和分析一系列文章的基础上，对生成式人工智能的缘起、发展与现状进行文献综述。

1. 生成式人工智能的缘起

在人工智能的探索历程中，我们一直追求的目标是让机器像人类一样思考、交流。人工智能发展初期，科学家尝试使用符号逻辑和知识推理来实现这一目标，但在处理复杂自然语言时遇到了挑战。随着时间的推移，深度学习技术的崛起使研究人员开始尝试利用大规模数据和自监督学习来训练模型，以期提升机器对语言的理解能力。2017年谷歌公司提出Transformer（Vas wani et al.,2017）模型，促使自然语言处理模型的参数量得到大幅扩展[4]，它的问世标志着自然语言处理领域的重大突破，为我们打开了通往更广阔智能世界的大门。

然而，传统的人工智能（AI）主要依赖预定义规则和算法进行任务处理，其局限性在于缺乏灵活性和适应性，难以应对复杂和动态的环境变化。例如，传统人工智能中的判别式模型虽然在一些任务上表现出色，但它们往往只能进行分类或回归等判别任务，无法生成新的数据。这种局限性在自然语言处理、图像处理等需要创造性和生成新数据的领域尤为明显。正是由于这些不足，我们引入了生成式人工智能。

生成式人工智能通过建模联合概率分布，可以利用无标记数据进行预训练，生成高质量的新内容，具有更好的泛化能力和创造性，适用于更广泛的应用场景。其引入不仅弥补了传统模型的局限，还为我们带来了在各个领域中创造性地生成和处理数据的新能力，从而推动了人工智能技术的进一步发展和应用。

1. 生成式人工智能的发展

至今为止，人工智能的发展跌宕起伏，因为技术应用、学术报告的负面影响 和市场危机等在上世纪七十年代和九十年代经历了两次低谷，以及三次高潮

2.1 起源与早期探索（20世纪50年代-70年代）

1950年，英国数学家阿兰·图灵提出了著名的“图灵测试”，用于判断机器是否具有人类智能[5]。这一概念成为人工智能领域的基础之一。1956年，在达特茅斯会议上，约翰·麦卡锡、马文·明斯基、克劳德·香农和纳撒尼尔·罗切斯特等科学家提出了“人工智能”的概念，标志着人工智能作为一个独立研究领域的诞生。这一时期的研究主要集中在符号逻辑和知识推理上，开发了如逻辑推理、问题解决等早期AI程序。著名的项目包括纽厄尔和西蒙的逻辑理论家（Logic Theorist）和GPS（General Problem Solver）。尽管这些早期项目展示了AI的潜力，但由于计算能力和数据资源的限制，实际应用进展有限。

2.2 知识工程与专家系统（20世纪70年代-80年代）

1970年代，人工智能研究进入“知识工程”阶段，专家系统成为研究热点。这些系统通过编码专家知识来解决特定领域的问题。

例如，MYCIN是一个用于医学诊断的专家系统，它能够根据输入的症状和实验结果，提供诊断和治疗建议。DENDRAL则用于化学分析，通过分析质谱数据推断分子结构。专家系统在某些领域取得了显著成功，展示了AI在专业领域的应用潜力。然而，这些系统的局限性也逐渐显现出来，它们通常难以扩展和维护，且在面对新的或复杂的问题时表现不佳。

2.3 人工智能的冬天（20世纪70年代末-90年代初）

人工智能在20世纪70年代末至90年代初经历了两次“冬天”，这段时期被称为AI的黑暗时代。第一次AI冬天始于20世纪70年代中期，其主要原因是人工智能领域的专家系统并未能兑现早期的巨大期望。虽然专家系统在模拟专业知识方面取得了一定进展，但它们的能力受限，无法处理复杂的现实世界问题，导致了技术的局限性和商业应用的失败。这一情况使得资金投入和研究兴趣逐渐减弱，进而引发了AI研究的第一次低谷期。

第二次AI冬天发生在1980年代末期，其主要原因是在过度乐观的预期下，技术进展未能达到商业应用的期望水平。在这一时期，许多人工智能项目由于成本高昂、效果不佳而被终止，这导致了资金投入的减少和公众对人工智能的兴趣下降。加上一些突出的失败案例，如日本第五代计算机项目未能实现预期的成果，使得人工智能研究陷入停滞状态，被认为是一项“死亡的科学”。

这两次“冬天”时期对人工智能领域造成了深远的影响，但也促使人们反思和重新评估人工智能的发展路径。随着技术的进步和新的研究方法的出现，如深度学习和大数据的应用，人工智能逐渐走出低谷，迎来了新的发展机遇。

2.4 机器学习与神经网络的复兴(（20世纪80年代末-90年代）)

1980年代末，神经网络重新受到关注，尤其是反向传播算法的提出，使得训练多层感知机成为可能。反向传播算法的发明解决了神经网络训练中的关键问题，大大提高了其性能。与此同时，机器学习开始崭露头角，研究重点从基于规则的系统转向统计学习方法和数据驱动模型。支持向量机（SVM）、决策树、贝叶斯网络等方法逐渐被广泛应用。

2.5 大数据与深度学习的兴起（21世纪初至今）

21世纪初至今，大数据与深度学习的兴起对人工智能领域产生了巨大影响。随着互联网的发展，大规模数据的积累以及计算能力的提升为机器学习和深度学习算法提供了丰富的资源和基础设施，开启了大数据时代。深度学习算法的突破性进展在2012年由AlexNet在ImageNet图像识别挑战赛中取得胜利开始，其在图像、语音和自然语言处理等领域表现出色，推动了人工智能应用的快速发展。2017年，谷歌提出的Transformer模型进一步推动了自然语言处理技术的发展，其衍生模型如BERT、GPT等在多种NLP任务上展现了优异性能，标志着自然语言处理领域迈入了大模型时代。2018年6月，美国OpenAI公司提出了基于Transformer的预训练语言模型（Generative Pre-trained Transformer）GPT-1。GPT-1（Radford et al.,2018）基于自回归理念，采用12个Transformer解码器，构建从左向右单向预测的语言模型，参数 量达1.17亿。GPT-1的构建首先基于大型语料库进 行无监督的预训练，然后通过有监督的微调技术为下游自然语言处理任务提供解决方案[4]。

1. 生成式人工智能现状与潜在风险

生成式人工智能发展速度惊人、迭代速度呈现指数级发展，被认为是迈向通用人工智能（Artificial General Intelligence，AGI）的重要阶段性成果[7]。2023年，全球范围内，科技巨头积极参与基础大模型竞赛，形成了所谓的“百模大战”，表现为参数体量增大、模态覆盖面增多、垂直领域任务应用增多等特点。代表性模型包括OpenAI的ChatGPT和GPT-4、微软的Bing Chat和Orca、谷歌的Bard和PaLM2、Anthropic的Claude、META的LLaMA和ImageBind、Stability AI的Stable Diffusion以及Amazon的Titan等。这些模型在对多模态数据的理解与生成能力方面取得显著进展，围绕文本、音视频、电光信号等多类型模态数据展开综合研究。同时，微软于9月发布了综述，将多模态基础大模型的主要研究方向归纳为视觉理解模型、视觉生成模型、统一视觉模型、基础大模型加持的多模态大模型和多模态智能体等五个方向。OpenAI也宣布了ChatGPT多模态能力解禁，允许用户使用GPT-3.5或GPT-4模型上传图像进行对话。这些新模型的出现带来了前所未有的能力，包括在开放世界视觉理解、视觉描述、多模态知识、常识、场景文本理解、文档推理、编码、时间推理、抽象推理、情感理解等多个领域和任务上的极强竞争力。

尽管生成式人工智能技术在各个领域展现出了巨大的潜力，但其广泛应用也存在一定风险。高度逼真的生成内容可能被用于传播虚假信息和制造假新闻，威胁社会稳定。生成的内容可能泄露个人隐私，引发隐私泄露问题[8]。训练数据的偏见或不平衡可能导致生成内容反映这些偏见，引发种族和性别歧视等问题。此外，生成的内容可能侵犯知识产权和版权[9]，导致法律纠纷。最后，生成式人工智能技术可能面临黑客攻击、安全漏洞和技术滥用的风险，威胁到系统和数据的安全性。因此，尽管生成式人工智能技术带来了巨大的潜力，但必须谨慎管理其应用，以克服潜在的风险和挑战。

1. 评述

生成式人工智能技术在当前已经取得了巨大的进展，为各个领域带来了新的应用和可能性。从早期的符号逻辑和专家系统到现在基于大规模数据的生成模型，人工智能的发展历程经历了多个阶段。尤其是近年来，随着深度学习和大数据技术的兴起，生成式人工智能技术的应用领域不断拓展，涵盖了自然语言处理、计算机视觉、音乐和艺术创作等多个领域。尽管这一技术有着巨大的潜力，但高度逼真的生成内容可能会被滥用，导致虚假信息的传播和隐私泄露等问题。政府与社会都需要对生成式人工智能技术的应用进行谨慎管理，充分认识到其潜在风险，并通过合作共同努力，推动这一技术的健康发展，实现其更广泛的社会价值。

**参考文献：**

1. 张熙,杨小汕,徐常胜.ChatGPT及生成式人工智能现状及未来发展方向[J].中国科学基金,2023,37(05):743-750.
2. Brown TB,Mann B,Ryder N,et al.L .anguage models are few-shot learners. Advances in Neural Information Processing Systems,2020:1877-1901.
3. Jackson S. OpenAI executives say releasing ChatGPT for publicuse was a last resort after running into multiple hurdlles-andthey're shocked by its popularity.[2023-01-25]/[2023-06-29].
4. 卢宇,余京蕾,陈鹏鹤,等.生成式人工智能的教育应用与展望——以ChatGPT系统为例[J].中国远程教育,2023,43(04):24-31+51.
5. 李冠华. 人工智能生成技术方案的可专利性研究[D].青海师范大学,2024.DOI:10.27778/d.cnki.gqhzy.2023.000597.
6. [1]邓志文.文生视频模型Sora的时间性结构分析——对生成式人工智能的现象学思考[J/OL].编辑之友:1-7[2024-05-30].
7. 王亚珅,方勇,江昊,等.2023年生成式人工智能技术主要发展动向分析[J].无人系统技术,2024,7(02):101-112.
8. 李莉.ChatGPT：生成式人工智能应用风险及对策研究[J].湖北经济学院学报(人文社会科学版),2024,21(02):98-102.
9. 陈永伟.超越ChatGPT：生成式AI的机遇、风险与挑战[J].山东大学学报(哲学社会科学版),2023(03):127-143.