

许子祺

——人工交互技术文件检索作业

生成式人工智能：技术嬗变与未来图景

2025 年 6 月 18 日

1 生成式人工智能的核心理论与模型演进

生成式人工智能（Generative Artificial Intelligence, Generative AI）^[1]是人工智能领域的重要分支，致力于通过机器学习模型生成与训练数据相似的全新内容，包括但不限于文本、图像、音频等。与专注于分类和预测的判别式模型不同，生成式模型通过学习数据的概率分布，生成具有创造性和真实性的输出。这种范式转变为人工智能的应用开辟了新的可能性^[2]。

生成式人工智能的发展历程可追溯至多个关键技术突破。2014 年，Goodfellow 等人提出了生成对抗网络（Generative Adversarial Networks, GANs）^[3]，其核心思想是通过生成器（Generator）和判别器（Discriminator）之间的对抗博弈优化模型性能。生成器尝试生成逼真的样本，而判别器则负责区分生成样本与真实样本。两者的优化目标可表示为以下最小-最大博弈问题^[4]：

$$\min_G \max_D V(D, G) = \mathbb{E}_{x \sim p_{\text{data}}(x)} [\log D(x)] + \mathbb{E}_{z \sim p_z(z)} [\log(1 - D(G(z)))]$$

其中， G 表示生成器， D 表示判别器， z 为随机噪声， x 为真实数据。这一机制显著提升了生成样本的质量，尤其在图像生成领域取得了突破性进展。

2017 年，Vaswani 等人提出的 Transformer 架构进一步推动了生成式人工智能的发展。Transformer 通过自注意力机制（Self-Attention）捕捉序列数据中的长距离依赖关系，极大地提升了自然语言处理任务的性能。基于 Transformer 的大型语言模型（Large Language Models, LLMs），如 OpenAI 的 GPT 系列，展现了卓越的文本生成、推理和跨模态能力。例如，GPT-3 通过大规模预训练，能够生成流畅且语义丰富的文本^[5]。

近年来，扩散模型（Diffusion Models）成为生成式人工智能的新热点。扩散模型通过在数据上逐步添加噪声并学习去噪过程，生成高质量的图像输出。Stable Diffusion 等模型在图像生成任务中表现出色，其稳定性优于传统 GANs。扩散模型的数学基础基于马尔可夫链和概率分布的逐步优化，为生成式 AI 提供了新的理论视角^[6]。

2 应用范式、深层挑战与未来展望

生成式人工智能的应用已渗透至多个领域，展现了其强大的潜力。在生命科学中，DeepMind 的 AlphaFold2 通过生成式方法预测蛋白质三维结构，解决了长期困扰生物学的蛋白质折叠问题，其预测精度达到前所未有的水平^[7]。在软件工程领域，生成式 AI 能够自动生成代码片段，显著提高开发效率。例如，基于大型语言模型的代码生成工具可根据自然语言描述生成可执行程序。在文化创意产业，生成式 AI 被用于艺术创作、音乐生成和视频制作，极大地丰富了文化表达形式。

尽管生成式人工智能取得了显著成就，但其发展仍面临诸多挑战。首先，模型幻觉（Hallucination）是一个亟待解决的问题^[8]，即模型可能生成虚假或不符合事实的内容，这在信息敏感领域可能导致严重后果。其次，算法偏见可能因训练数据的偏差而产生，影响模型的公平性和可靠性。此外，生成式模型的高能耗和知识产权归属问题也引发了广泛讨论。这些挑战要求研究人员在技术、伦理和法律层面共同努力。

展望未来，生成式人工智能将朝着多模态和交互式模型的方向发展。多模态模型能够同时处理文本、图像和音频等多种数据形式，提供更丰富的用户体验。交互式模型通过与用户的实时交互，提高生成内容的精准性和可控性。此外，通过模型压缩和优化技术，小型化模型将提升生成式 AI 的可访问性和计

算效率。长远来看，生成式人工智能有望成为通用人工智能（Artificial General Intelligence, AGI）^[9]的重要组成部分，为人类社会带来深远影响。

参考文献

- [1] BANH L, STROBEL G. Generative artificial intelligence[J]. Electronic Markets, 2023, 33(1): 63.
- [2] OOI K B, TAN G W H, AL-EMRAN M, et al. The potential of generative artificial intelligence across disciplines: Perspectives and future directions[J]. Journal of Computer Information Systems, 2025, 65(1): 76-107.
- [3] SAXENA D, CAO J. Generative adversarial networks (GANs) challenges, solutions, and future directions[J]. ACM Computing Surveys (CSUR), 2021, 54(3): 1-42.
- [4] GOODFELLOW I J, POUGET-ABADIE J, MIRZA M, et al. Generative adversarial nets[J]. Advances in neural information processing systems, 2014, 27.
- [5] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[J]. Advances in neural information processing systems, 2017, 30.
- [6] YANG L, ZHANG Z, SONG Y, et al. Diffusion models: A comprehensive survey of methods and applications[J]. ACM Computing Surveys, 2023, 56(4): 1-39.
- [7] BRYANT P, POZZATI G, ELOFSSON A. Improved prediction of protein-protein interactions using AlphaFold2[J]. Nature communications, 2022, 13(1): 1265.
- [8] JI Z, LEE N, FRIESKE R, et al. Survey of hallucination in natural language generation[J]. ACM computing surveys, 2023, 55(12): 1-38.
- [9] GOERTZEL B. Artificial general intelligence: concept, state of the art, and future prospects[J]. Journal of Artificial General Intelligence, 2014, 5(1): 1.