

生成式 AI 赋能算网融合发展

Generative AI Empowers Development of Computing-network Integration

孟利超, 李晶晶 (电子科技大学, 四川 成都 611731)

Meng Lichao, Li Jingjing (University of Electronic Science and Technology of China, Chengdu 611731, China)

摘要:

以 ChatGPT 为代表的生成式人工智能应用引起信息产业新浪潮,也催生了对算力、通信、安全等的新挑战,而以云计算、边缘计算、通信网络技术为支撑的算网融合代表了未来通算一体的发展趋势。生成式 AI 与算网融合的双向赋能,有望克服算力与通信瓶颈,实现生成式 AI 应用更加广泛的推广,同时有望推动算网融合架构向智能化方向发展。分析了生成式 AI 与算网融合的概念、现状及趋势,思考了生成式 AI 应用对算网融合的诉求以及生成式 AI 技术对算网融合架构的赋能作用。

Abstract:

The applications of generative artificial intelligence represented by ChatGPT cause a new wave in the information industry, and at the same time give birth to new challenges to computing power, communication, security, etc. Computing and network integration supported by cloud computing, edge computing, and communication network technology represents the development trend of the integration of communication and computing in the future. The two-way empowerment of generative AI and computing network integration is expected to overcome the bottleneck of computing power and communication, realize the wider promotion of generative AI applications, and at the same time promote the development of computing network integration architecture in the direction of intelligence. It analyzes the concept, research status, and development trend of generative AI and computing-network integration, and considers the demands of generative AI applications for computing-network integration, as well as the empowering role of generative AI technology for computing-network fusion architecture.

Keywords:

Generative AI; Computing and network integration; Two-way empowerment; Communication computing integration

引用格式: 孟利超, 李晶晶. 生成式 AI 赋能算网融合发展[J]. 邮电设计技术, 2023(7): 8-13.

关键词:

生成式 AI; 算网融合; 双向赋能; 通算一体化

doi: 10.12045/j.issn.1007-3043.2023.07.002

文章编号: 1007-3043(2023)07-0008-06

中图分类号: TP183

文献标识码: A

开放科学(资源服务)标识码(OSID):



0 引言

近期,以 ChatGPT 为代表的生成式人工智能(AI)技术得到迅猛发展并掀起新的人工智能浪潮,引起了社会各界的浓厚兴趣。ChatGPT、DALL-E-2^[1]和 Co-dex^[2]等生成式 AI 技术属于人工智能生成内容(AIGC)的范畴,涵盖了通过 AI 模型生成自然语言、图像、音乐等数字化内容,其研究领域及经典模型间的关系如图

1 所示。生成式 AI 能够利用大型人工智能算法根据用户需求以更快速度和更低成本协助或取代人类创造丰富的高质量内容,因而正在引领和重塑内容创作和知识表示等研究应用领域。

大量生成式 AI 应用涌现的同时,其依托海量数据和参数支撑的模型特性和面向用户服务实时响应的应用特性对算力、网络等资源提出了新挑战。生成式 AI 应用催生了对算网融合的需求,通过算力与通信网的深度融合,有望实现云边协同、算网一体,从而能够从算力、通信、安全等方面为生成式 AI 应用的推广赋

收稿日期: 2023-05-16

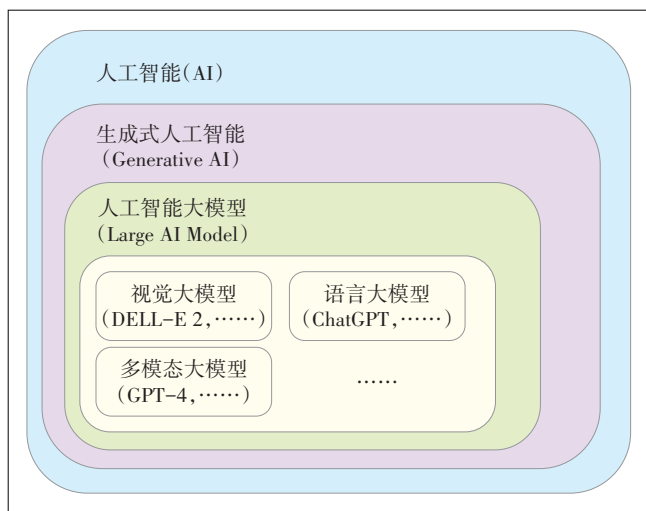


图1 生成式AI研究领域及经典模型间的关系

能。另一方面,生成式AI技术的迅猛发展也为算网融合架构的内生智能化提供了契机。在算网融合架构中整合生成式AI技术,有望推动算网融合关键技术向智能化方向发展,在资源管理、网络优化、服务提供等方面实现更加高效、稳定和优越的性能表现。通过二者的双向良性互动,将有望助推通信行业与人工智能研究领域的深度融合。

本文首先分析了生成式AI与算网融合的基本概念、研究现状及发展趋势。其次阐述了生成式AI对算网融合的需求和算网融合在克服算力、通信、安全瓶颈方面的可行研究路径。随后,从资源管理、网络优化、服务提供3个方面阐述了生成式AI技术对算网融合关键技术发展的赋能作用。

1 生成式AI发展现状及趋势

1.1 生成式AI基本概念

近年来,AIGC在计算机科学界之外受到了广泛关注,整个社会对各类生成式AI产品表现出浓厚兴趣。AIGC指的是使用生成式AI技术生成的内容,其基于生成式AI算法训练得到参数量庞大的大型人工智能模型,能够根据用户输入或需求,以更快速度和更低成本,协助或取代人类创造丰富的个性化高质量内容。例如,ChatGPT是OpenAI开发的用于构建会话人工智能系统的语言模型,可以有效地理解并以有意义的方式响应人类语言输入。DALL-E-2是OpenAI开发的一种最先进的生成式人工智能模型,能够在几分钟内从文本描述中生成独特的高质量图像。

从技术上讲,AIGC指的是给出能够帮助或指导模

型完成任务的人工指令,并利用生成式AI算法生成满足指令的内容。这一过程通常由2个阶段组成:第一,提取和理解用户意图信息;第二,根据提取的意图信息生成所需的内容。近期取得的核心进步归功于在更大的数据集上训练更加复杂的生成式模型,使用更大的基础模型架构以及广泛计算资源的支持。此外,研究人员也探索了将新技术与生成式AI算法相结合的方法。例如,ChatGPT能够利用来自人类反馈的强化学习技术来确定给定指令的最佳响应,从而随着时间的推移提高模型的可靠性和准确性^[3]。结合这些进步,生成式AI技术取得了重大进展,在不久的将来,也将成为机器学习的一个重要研究领域^[4]。

1.2 生成式AI研究现状及发展趋势

生成式模型在人工智能领域有着悠久的历史,伴随着隐马尔可夫模型和高斯混合模型的发展,可以追溯到20世纪50年代。在深度学习出现后,生成式模型的性能得到显著的提高。在自然语言处理(NLP)中,传统的句子生成方法不能有效地适应长句,因而递归神经网络(RNNs)被引入到语言建模任务中对相对较长的依赖性进行建模。随后出现了长短时记忆(LSTM)和门控循环单元(GRU),能够利用门控机制在训练过程中控制记忆。同时,在计算机视觉领域,2014年生成对抗网络(Generative Adversarial Networks, GANs)首次被提出,在各种应用中取得了可喜的成果。随后,变分自编码器(VAEs)和扩散生成模型等其他方法也被开发出来,用于进行细粒度控制,提高生成高质量图像的能力。在这个阶段,生成式AI算法得到迅速的发展,并为大型模型训练奠定了基础。2017年,Vaswani等人提出了具有并行训练能力的transformer模型。2019年,DeepMind发布了能够生成连续视频的DVD-GAN模型。2021年,OpenAI推出了支持从文本生成图像的DALL-E。

此后,借助预训练的大型人工智能模型,AIGC的能力得到了极大的增强,被大规模使用。此类大型模型能够从大量数据中进行学习,从而更好地捕捉语言的细微差别,更加准确和健壮。以生成式AI为支撑的人工智能内容生成范式也为元宇宙^[5]和数字孪生^[6]等新型应用提供了广阔前景。生成式AI的通用架构如图2所示。

在未来,生成式AI有望在绿色架构、分布式和可扩展性算法、安全性和隐私保护等方面得到进一步发展。

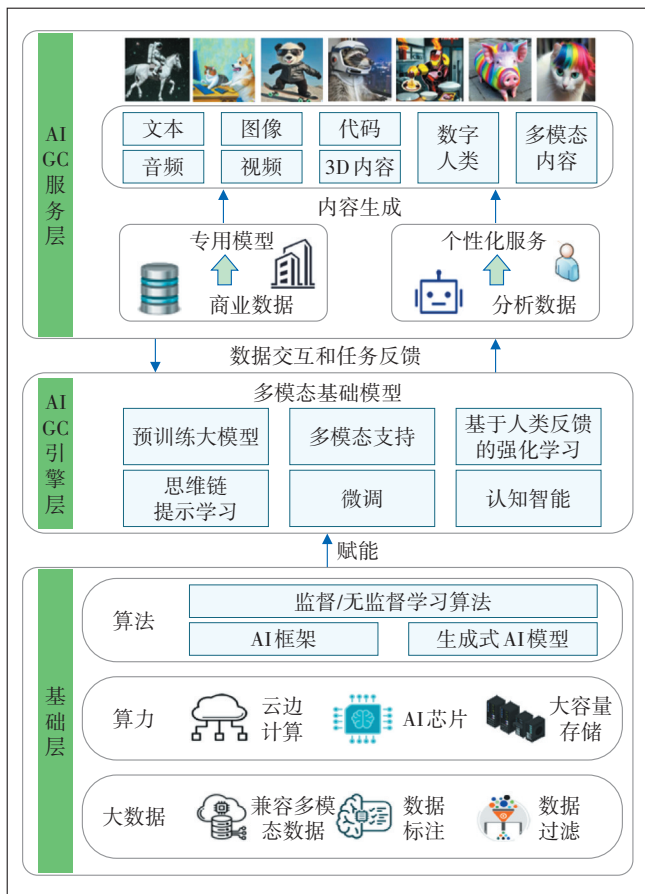


图2 生成式 AI 的通用架构

2 算网融合发展现状及趋势

2.1 算网融合基本概念

生成式 AI 应用服务催生了对新一代通信网络基础设施和关键技术进行革新的机遇与挑战^[7], 其中实现算力与网络的一体化融合发展有望实现“云”“网”“边”“端”的算力通信高效协同, 满足新型人工智能大模型的算网需求。

算网融合指的是计算和网络技术在互联网时代的融合和相互依存的发展趋势, 可将计算和网络资源相结合, 提供更强大、高效和灵活的计算和通信能力。算网融合以网络通信基础设施和异构的计算设施为基础, 利用软件定义网络、网络功能虚拟化、云计算、边缘计算等技术实现数据、计算、网络等多种异构资源的统一感知、编排、调度、管控, 从而实现云边协同、边边协同和通算一体, 为算网融合上层应用提供支撑^[8]。面向服务框架的算网融合分层架构如图 3 所示。

2.2 算网融合研究现状及发展趋势

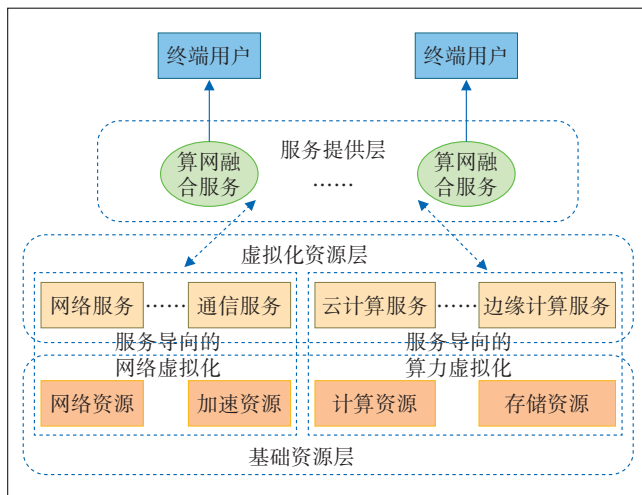


图3 面向服务框架的算网融合分层架构

算网融合是国家战略、行业发展、技术演进共同驱动下的必然发展趋势。在战略驱动方面, 我国在“十四五”期间提出了“东数西算”战略布局, 并在国务院印发的《“十四五”数字经济发展规划》中明确提出云网协同与算网融合发展的要求。在行业驱动方面, 智慧城市、智慧医疗、智能制造等应用服务需求和数字经济下大连接、高算力、强安全的迫切诉求亟需算网融合发展为其提供支撑。在技术驱动方面, 出现了面向边缘侧算力提升的边缘计算、分布式计算、云边协同计算等较为成熟的技术方案, 在通信网络演进中 SDN、NFV 等技术日趋成熟, 这些技术的发展演进为算网融合提供了技术支撑。

在算网融合的研究中, 文献[8]从算网设施、算网平台、算网应用、算网安全 4 个层级阐述了算网融合的背景、内涵和现状。文献[9]从算力发展面临的算力、存储、功耗、带宽挑战切入, 指出算网融合需要网络域、计算域协同创新。文献[10]研究了云网融合的体系架构和关键技术。文献[11]从智能物联网视角研究了“云网边缘端业”的端到端架构和云网集成应用框架。

算网融合发展的未来趋势体现在智能化、弹性化、边缘化方向。随着人工智能和机器学习的发展, 算网融合将更加智能化和自动化, 自动化网络管理和资源优化算法将减少人为干预, 提高网络性能和效率。同时, 未来算网融合也将更加弹性灵活, 资源的动态分配和任务卸载将成为常态, 使计算和网络能够根据需求实时调整和优化。此外, 边缘侧算力的重要性持续凸显, 边缘设备的计算和存储能力将得到提

升,边缘计算将成为处理实时数据和低延迟应用的关键。在推动云网边端智能协同、实现智能云网资源融合、建立智能内生机制方面,生成式AI技术将有望充分发挥其赋能作用。

3 生成式AI应用服务催生算网融合新需求

生成式AI应用服务表现出了对强大算力、敏捷通信、低时延实时响应、可扩展性与隐私保护的迫切需求^[12],算网融合和云边协同技术方案是解决上述挑战的潜在研究方向。

3.1 生成式AI应用服务特点

生成式AI已在多个领域得到广泛应用。总的来说,生成式AI应用服务具有实时交互、个性化体验、多模态支持、自适应学习、多领域应用、高可用性和可扩展性等特点。

a) 实时交互。用户可以提出问题、请求或指令, AI系统会立即回应并提供相应的结果或操作。

b) 个性化体验。通过学习和理解用户的行为模式和喜好,为每个用户提供定制化的建议、推荐或解决方案。

c) 多模态支持。除了文本外,还可以接受语音、图像、手势等多种输入形式,并能够理解和处理这些输入。

d) 自适应学习。可以从与用户的交互中学习,不断优化自身模型算法,提供更准确、智能的响应和服务。

e) 多领域应用。能应用于多个领域,例如可用于语音助手、智能客服、虚拟导游、智能家居控制等多种场景。

f) 高可用性和可扩展性。可以处理大量的用户请求并提供快速响应,同时可以根据需求进行水平扩展,以应对高峰时段或大规模用户的需求。

生成式AI的实时交互、高可用性和可扩展性特点对通信网络的带宽、时延提出了更高的要求,多模态支持、自适应学习与多领域应用的特征对算力尤其是边缘端的算力提出了更大的需求。此外,在为用户提供个性化体验的同时还需要实现用户数据隐私保护,因而催生了对克服算力瓶颈、保障实时通信、实现隐私保护的新需求。这些需求有望通过算网融合架构和相关关键技术的支撑得到满足。

3.2 生成式AI应用服务对算网融合的需求

算网深度融合能够为大带宽低时延业务服务和

各类型生成式AI应用服务提供算力、时延、带宽、可靠性、安全性等方面的保障,以满足未来人工智能应用的发展需求^[13]。

针对算力瓶颈问题,依托算网融合架构实现更加绿色节能的高算力供给,以更低的能耗成本实现可扩展的模型训练和推理是可行的解决途径。在高速网络方面,算网融合有望提供高速、低延迟的网络连接,通过快速获取和交换数据加速计算过程。在分布式计算方面,算网融合可以将计算资源分布在多个节点上,通过并行执行任务提高整体的计算能力,并通过负载均衡和任务调度等技术优化计算资源的利用效率。同时,依赖于云计算平台,能够获得弹性的计算资源,可根据需要动态调整计算能力。此外,图形处理器(GPU)、张量处理器(TPU)等加速器与AI芯片等专用硬件的研发,也能够加速生成式AI应用的训练和推理过程,从而克服算力瓶颈问题。

针对实时通信问题,算网融合的研究有望满足低时延高带宽实时通信的需求。具体来说,算网融合能够通过升级网络基础设施来提供更高带宽和更低延迟。同时,5G网络的部署和边缘计算的发展为生成式AI应用的实时通信需求提供有力支持。5G网络能够支持大规模数据传输和实时通信,边缘计算能够将计算资源推向网络边缘,使生成式AI模型更接近终端用户。此外,针对算网融合中的网络优化和协议改进的研究也有助于满足实时通信需求。例如,网络流量管理、数据压缩、带宽分配和拥塞控制等技术可以提高网络性能,减少通信时延。

针对隐私保护问题,算网融合有望通过采用数据加密、身份验证、安全协议和隐私保护技术手段,满足生成式AI应用对安全和隐私保护的需求。其一,可以采用数据加密技术保护通信过程中传输的数据,采用身份验证机制确保通信各方可信,防止恶意攻击者入侵。其二,采用安全协议和通信通道来确保生成式AI应用与其他系统间的安全通信。其三,算网融合架构下可以融入数据脱敏、差分隐私、同态加密等隐私保护技术,在生成式AI应用处理数据时减少对个人身份和敏感信息的暴露,同时保持数据的可用性和有效性。

总的来说,如图4所示,算网融合架构能够为各类型生成式AI应用服务提供基础架构设施保障,在保护用户数据隐私的同时克服算力和通信资源瓶颈问题,为生成式AI应用服务的推广赋能。

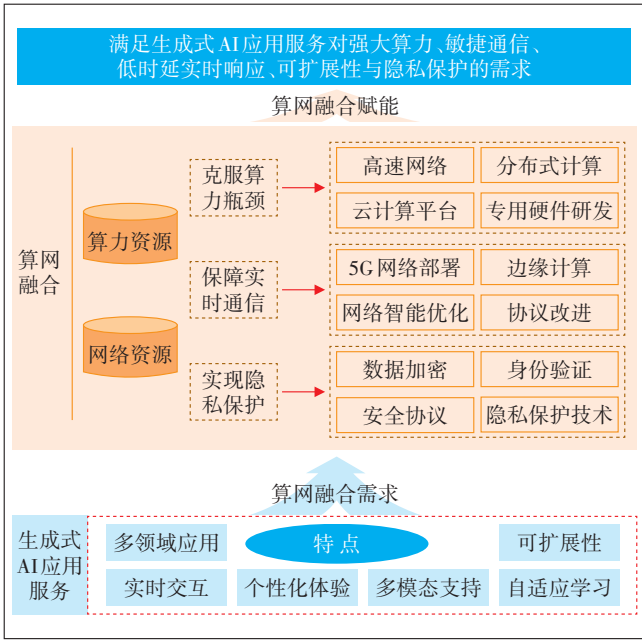


图4 算网融合赋能生成式 AI 应用的整体架构

4 生成式 AI 技术赋能算网融合关键技术发展

如前所述,算网融合有望为生成式 AI 应用服务提供广泛支撑,除此之外,生成式 AI 技术也能为算网融

合关键技术的发展赋能^[14],在资源管理、网络优化、服务提供等方面构建智能化、自主化解决方案,从而实现生成式 AI 技术与算网融合架构的双向赋能,推动信息社会的进一步发展^[15]。生成式 AI 技术赋能算网融合关键技术整体架构如图 5 所示。

4.1 生成式 AI 赋能资源管理

在资源管理方面,生成式 AI 的赋能作用体现在网络资源分配优化、动态网络调度、故障和安全管理、资源优化建议、自主决策和自适应网络等方面。

a) 网络资源分配优化。生成式 AI 可以分析网络拓扑、流量负载和用户需求等多个因素,智能分配网络资源,包括带宽、计算资源、存储资源等,优化资源的分配方式,提高网络的性能和效率。

b) 动态网络调度。生成式 AI 可以通过学习网络的历史数据和实时信息,预测网络流量和负载的变化趋势,自主地调整网络资源的分配和路由策略,从而实现动态的网络调度。

c) 故障和安全管理。生成式 AI 可以监测和分析网络的故障和安全事件,快速识别异常行为,预测潜在的故障和安全威胁,并提供实时的故障修复和安全防护建议,保障网络的稳定性和安全性。

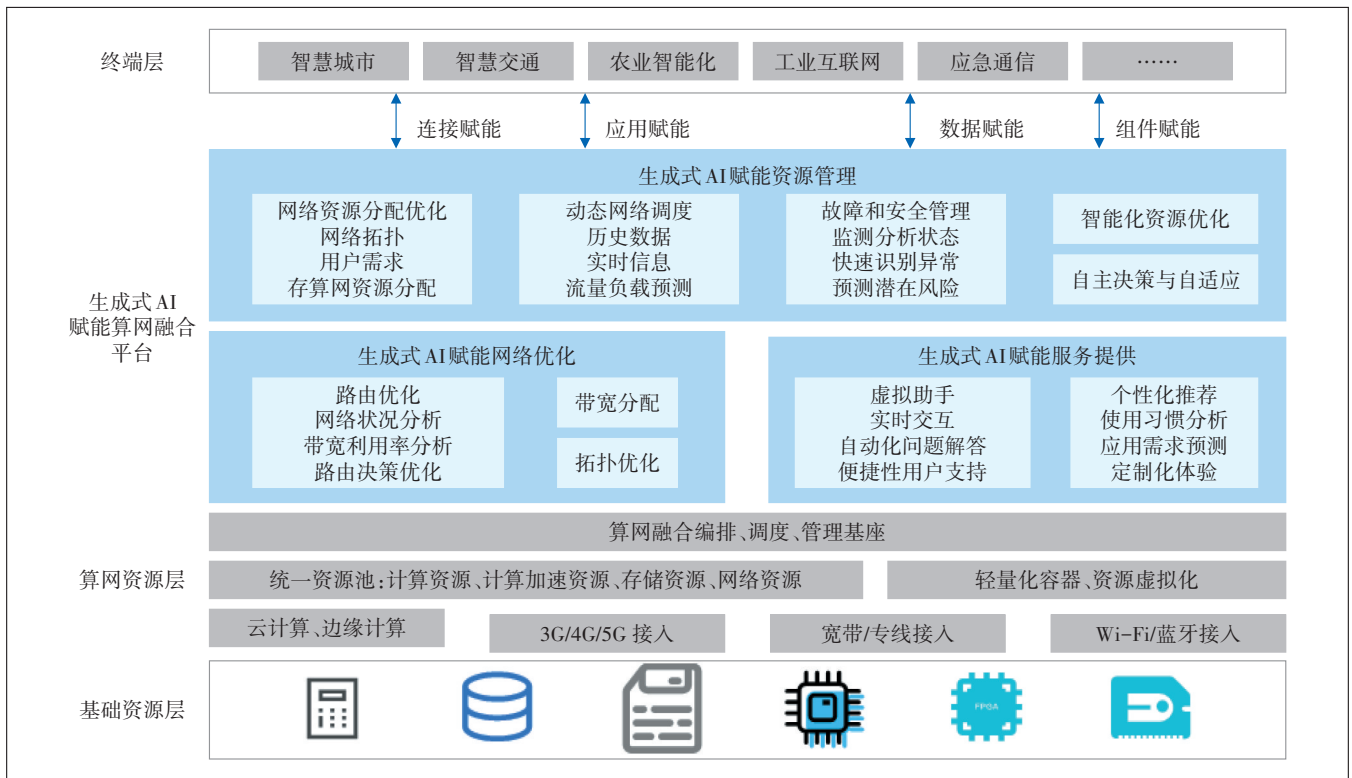


图5 生成式 AI 技术赋能算网融合关键技术整体架构

d) 资源优化建议。生成式AI可以分析网络数据、性能指标和用户行为,发现潜在的瓶颈和优化机会,并生成相应的建议和策略,帮助网络管理员和决策者做出更好的资源管理决策。

e) 自主决策和自适应网络。生成式AI结合强化学习方法,能够通过与环境的交互来优化网络资源管理策略,从而实现网络的自主决策和自适应。

4.2 生成式AI赋能网络优化

在网络优化方面,生成式AI技术能够为路由优化、带宽分配、网络拓扑优化等提供智能化支撑。

a) 路由优化。生成式AI可以基于历史数据和实时监测,分析网络状况,包括延迟、带宽利用率等指标,预测最佳路由选择,优化网络路由决策。

b) 带宽分配。生成式AI可以根据实时的流量需求和网络状态,智能分配带宽资源,提高网络的带宽利用率和吞吐量。

c) 网络拓扑优化。生成式AI可以分析网络拓扑结构,识别潜在瓶颈和优化机会,生成优化建议,如增加连接、优化网络布局等,以改善网络的性能和可扩展性。

4.3 生成式AI赋能服务提供

在服务提供方面,生成式AI技术能够为虚拟助手、个性化推荐等服务赋能,为用户提供更智能、个性化的服务体验。

a) 虚拟助手。生成式AI可以用于构建虚拟助手,提供实时的交互和支持。通过生成式AI模型的对话生成能力,可以与用户进行自然语言对话、回答问题、提供建议和执行一系列任务,从而增强用户与服务提供者之间的互动和便捷性。

b) 个性化推荐。生成式AI技术可以根据用户的偏好和历史行为,生成个性化的推荐结果。在计算与通信网络的深度融合中,生成式AI可以分析用户的网络使用习惯、应用需求等信息,推荐适合用户的服务和内容,提供定制化的体验。

5 结束语

目前生成式AI应用服务在各个行业中引起了广泛兴趣,一方面生成式AI应用的进一步发展依赖算网融合克服其面临的算力、通信、安全瓶颈,另一方面生成式AI技术的应用有望推动算网融合向着内生智能化方向发展。本文分析了生成式AI与算网融合的发展现状,并进一步指出了二者双向赋能的发展趋势与

可行方案。二者双向赋能,将有望推动计算与通信行业把握新一轮人工智能技术浪潮的先机,进而催生新需求、新技术、新应用和新业态。

参考文献:

- [1] RAMESH A, PAVLOV M, GOH G, et al. Zero-shot text-to-image generation [C]//Proceedings of the 38th International Conference on Machine Learning. Virtual Event: PMLR, 2021: 8821-8831.
- [2] CHEN M, TWOREK J, JUN H, et al. Evaluating large language models trained on code [DB/OL]. [2023-04-02]. <https://arxiv.org/abs/2107.03374>.
- [3] OUYANG L, WU J, JIANG X, et al. Training language models to follow instructions with human feedback [C]//36th Conference on Neural Information Processing Systems (NeurIPS 2022). New Orleans, United States: Curran Associates, Inc., 2022: 27730-27744.
- [4] 黄实, 张文帝, 杨雅清. ChatGPT拉开AI新时代帷幕[J]. 中国电信业, 2023(3): 42-45.
- [5] WANG Y T, SU Z, ZHANG N, et al. A survey on metaverse: fundamentals, security, and privacy [J]. IEEE Communications Surveys & Tutorials, 2023, 25(1): 319-352.
- [6] WANG Y T, SU Z, GUO S L, et al. A survey on digital twins: architecture, enabling technologies, security and privacy, and future prospects [DB/OL]. [2023-04-02]. <https://arxiv.org/abs/2301.13350>.
- [7] 张嗣宏, 张健. 以ChatGPT为代表的生成式AI对通信行业的影响和应对思考[J]. 电信科学, 2023, 39(5): 67-75.
- [8] 穆域博, 韩淑君, 柴瑶琳, 等. 算网融合的现状和发展趋势[J]. 信息技术, 2022, 16(2): 27-33.
- [9] 赵先明. 算网融合定义未来[J]. 通信技术, 2022, 55(6): 720-726.
- [10] 乔爱锋. 云网融合体系架构及关键技术研究[J]. 邮电设计技术, 2022(6): 14-18.
- [11] 张磊, 黄洪波, 原全新. 5G AIoT赋能云网融合创新发展实践及应用[J]. 邮电设计技术, 2023(1): 21-26.
- [12] 兰祝刚, 任然. ChatGPT对运营商影响及发展建议[J]. 中国电信业, 2023(4): 28-31.
- [13] 胡滨雨, 郭敏杰. 新一轮AI爆发下电信运营商的挑战和机遇[J]. 中国电信业, 2023(4): 23-27.
- [14] 冯养杰. 5G通信技术与人工智能的融合与发展前景探析[J]. 电子元件与信息技术, 2023, 7(3): 91-94.
- [15] 陈俊明, 张洁, 左罗. 运营商AI能力建设及演进探讨[J]. 邮电设计技术, 2021(12): 83-88.

作者简介:

孟利超, 博士在读, 主要研究方向为计算机视觉、领域自适应、人工智能; 李晶晶, 教授, 博士生导师, 博士, 主要研究方向为机器学习、迁移学习、小样本学习、多媒体、计算机视觉、领域自适应、人工智能等。

