 DOI: https://doi.org/10.26549/bdai.v4i4.14315

Discussion on Computing Power Service Model of Telecom Operators in the AIGC Era

Kaikai Yin Jun Zhai Zhaohui Sun

China Telecom Company Limited Beijing Branch, Beijing, 100032, China

Abstract

The rapid development of AIGC applications represented by ChatGPT brings new opportunities and challenges to the information and communication industry. As operators of new digital information infrastructure such as cloud, network and computing power, telecom operators also usher in new opportunities for the development of computing services. This paper introduces the concept, development history, technical architecture and application scenarios, development trend and challenge of artificial intelligence Generated Content (AIGC), and combs out the concept and main characteristics of computing service. On this basis, it is proposed that the development of enterprise computing service model of telecom operators will go through at least three stages, and eventually evolve from computing resource leasing service to computing platform and large model as a service, and finally put forward relevant development suggestions.

Keywords

artificial intelligence-generated content (AIGC); computing power; computing service; large model as a service (MaaS)

AIGC 时代电信运营商算力服务模式探讨

殷凯凯 翟骏 孙昭晖

中国电信股份有限公司北京分公司，中国·北京 100032

摘 要

以ChatGPT为代表的AIGC应用快速发展，为信息通信行业带来新的机遇和挑战。电信运营商作为新型云、网、算力等数字信息基础设施运营者，也迎来算力服务发展的新机遇。论文介绍了人工智能生成内容（AIGC）的概念、发展历程、技术架构与应用场景、发展趋势与挑战，梳理了算力服务概念及主要特征，并在此基础上提出了电信运营商企业算力服务模式发展至少经过三个阶段，最终会由算力资源租赁服务向算力平台和大模型即服务演进，最后提出相关发展建议。

关键词

人工智能生成内容（AIGC）；算力；算力服务；大模型即服务（MaaS）

1 引言

2022 年 11 月 30 日，全新对话式 AI 模型 ChatGPT 一经发布便在全球范围内引发大模型和人工智能热潮，AIGC（AI Generated Content，人工智能生成内容）类应用迅速爆火，截至 2023 年 8 月，据专业媒体统计，中国已发布的大模型超过 100 个。2023 年 8 月 31 日包括百度（文心一言）、智谱 AI（GLM 大模型）等中国共 11 家大模型产品通过《生成式人工智能服务管理暂行办法》备案，获得 AIGC 牌照，可向公众开放服务。随着 AIGC、大模型等算力新应用、新业态不断涌现，全社会对算力的需求快速增长。电信运营商作为数字信息基础设施建设的主力军，也迎来算力服务发展的新机遇。论文分析了 AIGC 和算力服务的相关概念、现状和发展趋势，提出了电信运营商提供算力服务的三种模式，最后给出发展建议。

【作者简介】殷凯凯（1986–），中国山东东营人，硕士，从事数据中心、智算中心等算力基础设施规划、建设、运营管理，算力服务管理等研究。

2 人工智能生成内容（AIGC）研究背景

2.1 AIGC 概念与内涵

目前对于 AIGC（AI Generated Content，人工智能生成内容）的普遍认识是继专业生成内容（Proessional Generated Content，PGC）和用户生成内容（User Generated Content，UGC）之后，利用人工智能技术生成内容的新型生产方式 [1]。论文认为 AIGC 是指利用人工智能技术，如生成对抗网络（GAN）、扩散模型（Dffusion）、Transformer 预训练大模型等，来生成内容的新型内容生产方式，又可称生成式 AI。

2.2 AIGC 发展历程梳理

AIGC 早期概念可追溯到 20 世纪 50 年代。随着人工智能技术发展，AIGC 发展可大致分为 4 个阶段，分别是早期

萌芽阶段、沉淀积累阶段、快速演进阶段和应用爆发阶段。每个阶段发展特点及典型应用如表 1 所示。

2.3 AIGC 技术架构与应用场景

AIGC 类应用借助大模型的跨模态综合技术能力，可以实现提升内容生产效率，降低内容生产成本，激发创作灵感，提升内容多样性，实现数据优化等作用。目前在包括办公、文本、图像、音频、视频、游戏、代码、生物技术等方面海内外已涌现出诸多 AICG 应用产品。支撑 AIGC 拥有广泛应用场景的是其 4 层技术架构，分别是基础设施层、框架模型层、AIGC 服务层和 AIGC 应用层，如图 1 所示。

2.4 AIGC 发展趋势与挑战

AIGC 应用发展迅速，产业化方向众多，未来前景广阔，将呈现三方面发展趋势：一是由于数据提优，硬件升级和算法改进，AIGC 的性能将越来越强大，大模型参数体量显著提升，复杂度将提高；二是随着 AIGC 技术发展，其内容创作的基础能力将显著增强，可以更逼真地复刻虚拟世界；三是产品类型和应用领域将更加丰富，将涉及教育、医疗、政务、金融、传媒等更多行业和领域。

AIGC 发展受算力、数据、算法、网络、能源、隐私及安全、知识产权等方面影响。算力、数据、算法是人工智能和 AIGC 发展的核心三要素，其中算力资源决定 AIGC 发展高度。以 ChatGPT 为例，根据网络公开资料 ChatGPT 训练阶段总算力消耗约为 3640PF-days（即 1PetaFLOP/s 效率训练 3640 天），总训练成本约为 1200 万美元，在推理阶段则也会有较大算力消耗。由此可见以 ChatGPT 为代表的 AIGC 应用发展需要强大的算力支撑，且成本高昂。

3 算力服务概念及主要特征

3.1 算力服务概念及内涵

算力，通俗理解即计算能力。算力包含通用算力、智能算力、超算算力及前沿算力（如量子计算、光子计算）等 [2]。通用算力以 CPU 芯片输出的计算能力为主，超算算力主要是以超级计算机输出的计算能力为主。智能算力则以 GPU、FPGA 和 AI 芯片等输出的人工智能计算能力为主，具备渲染、推理和模拟能力，可面向智能驾驶、人脸识别、大模型等人工智能应用提供智算服务的一种算力服务形态 [3]。

算力服务由云计算服务演进而来，可理解为以多样性算力为基础，以算力网络为连接，以供给有效算力为目标的算力产业新领域，通过全新计算技术实现异构算力统一输出，并与云、大数据、AI（人工智能）等技术交叉融合，最终将算力、存储、网络等资源统一封装，以服务形式完成算力交付 [4]。

3.2 算力服务主要特征

算力服务作为数字技术能力的主要输出方式之一，以多样性算力资源为基础，以算力网络为连接，日益成为支撑数字经济发展的关键。根据业务对算力服务的需求，更好支撑典型应用场景，算力服务需要具备绿色、共享、智能、可信等特征 [5]，同时算力服务呈现“普惠化”“泛在化”“标准化”特征，推动算力成为社会基础公共资源 [4]。

表 1 AIGC 发展历程梳理

阶段	早期萌芽	沉淀积累	快速演进	应用爆发
时间	1950—1990 年	1990—2010 年	2010—2021 年	2022 年至今
发展特点	受限于技术水平，AIGC 仅限于小范围实验	AIGC 从实验性向实用性转变，受限于算法瓶颈，无法直接进行内容生成	深度学习算法不断迭代，人工智能生成内容百花齐放	国内外 AIGC 类应用迎来集中爆发，大模型持续火爆
典型案例	1.1950 年，图灵提出著名的“图灵测试”，给出判断机器是否具有“智能”的方法； 2.1966 年，世界第一款可人机对话机器人“Eliza”问世	1.2007 年，世界上第一部完全由人工智能创作的小说 1 The Road 问世； 2.2012 年，微软展示语全自动同声传译系统，可将英文语音自动翻译成中文语音	1.2017 年，微软人工智能“小冰”推出了世界首部 100% 由人工智能创作的诗集《阳光失了玻璃窗》； 2.2019 年，DeepMind 发布了 DVD-GAN 模型用以生成连续视频	1.2022 年 11 月 30 日，Open-AI 发布 ChatGPT； 2. 截至 2023 年 8 月，中国已发布的大模型超过 100 个

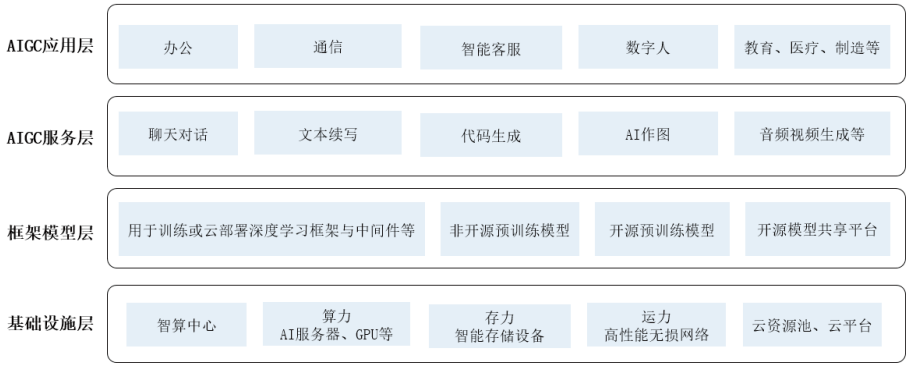

图 1 AIGC 整体技术架构示意图

4 电信运营商算力服务模式探讨

4.1 中国数字信息基础设施发展阶段

电信运营商作为新型云、网、算力等信息基础设施服务运营者，自身拥有优质网络、算力和云服务能力和产业链优势[6]，是中国数字信息基础设施建设的主力军。中国数字信息基础设施从 1970 年代到现在，短时间内走过了通信基础设施、互联网基础设施、移动互联网基础设施、云基础设施四个阶段，目前还处于云基础设施大发展阶段，整体呈现从连接向算力升级，现在正处于以智算中心为代表的算力基础设施的起步阶段。

4.2 中国运营商算力建设运营现状

算力基础设施方面，中国电信布局智算中心资源池，在东数西算枢纽节点建设集中式大规模训练资源池，支持多机多卡、并行训练框架，基于天翼云打造算力调度管理平台，为人工智能发展提供高品质的云、网、算力等服务。中国移动面向基于大模型的智能服务，构建新型智算中心算力底座，集约化建设 E 级超大规模新型算力基础设施。大模型应用方面，三大运营商已全部发布大模型应用。中国电信发布中大语言模型 TeleChat 和星河通用视觉大模型 2.0。中国移动发布九天・海算政务大模型和九天・客服大模型。中国联通发布“鸿湖图文大模型 1.0”。算力套餐方面，2023 年 5 月 17 日中国电信在业内率先发布算力套餐，包含“基础算力＋算力连接＋算法模型＋算力安全”的一体化服务产品。

4.3 电信运营商算力服务模式探讨

电信运营商作为新型云、网、算力等信息基础设施服务运营者，自身拥有优质网络、算力和云服务能力，应逐步探索由 IDC、云计算服务向算力服务转型，积极推进算力服务及产品布局。论文认为算力服务至少分为三个阶段，分别是算力资源租赁服务、算力平台服务和大模型即服务（Maas）三个阶段，如图 2 所示。

4.3.1 资源运营阶段：提供算力资源租赁服务

算力资源租赁是算力服务的初级模式，以算力基础资源（智算中心、算力网络、AI 服务器、GPU 卡等）租赁为主。此阶段主要目标是通过自建资源或生态合作模式，积累算力基础资源，提前卡位算力市场，抢占市场份额。服务模式主要有两方面特征：一是市场需求方面，算力基础设施资源市场供给较为紧俏，客户尤其是大模型训练客户对资源的需求旺盛；二是产品形态方面，此阶段主要以裸金属服务器、GPU 卡的对外租赁为主，需求集中，但缺乏长期、稳定性需求。此阶段服务对于运营商和客户技术能力有一定要求。

4.3.2 平台运营阶段：提供算力平台服务

平台运营是算力服务的进阶模式，以云计算管理服务平台和算力调度管理平台运营为主，需要电信运营商具备一定技术能力储备。此阶段主要目标是积累算力运营经验，搭建多梯度 / 异构算力平台，形成统一算力资源分配、调度、管理的能力。服务模式主要有两方面特征：一是云计算管理服务平台方面，主要提供 AI 服务器训练环境的预置能力、算力资源运维能力、资源管理能力等 IaaS 和 PaaS 服务；二是算力调度管理平台方面，借助运营商云公司能力（如中国电信天翼云“息壤”算力调度分发平台）提供自建 / 合作算力资源的统一管理、调度以及使用，实现一点开通、一点管理。此阶段要求运营商具备提供 IaaS（异构算力、融合存储、无损网络、统一资源管理调度等）和 PaaS（训练框架、大模型库、自动配置、任务托管、数据处理等）技术能力。对客户技术能力要求较低，能够调用平台接口即可

4.3.3 生态运营：提供大模型即服务（MaaS）

大模型应用快速发展正在重构现有的商业模式，目前 AI 公司、云商公司已经推出大模型即服务。OpenAl 在 2023 年 2 月 1 日推出 ChatGPT Plus 试点订阅计划以及 API 付费调取服务，这是典型的 MaaS 订阅制收费服务。腾讯云基于

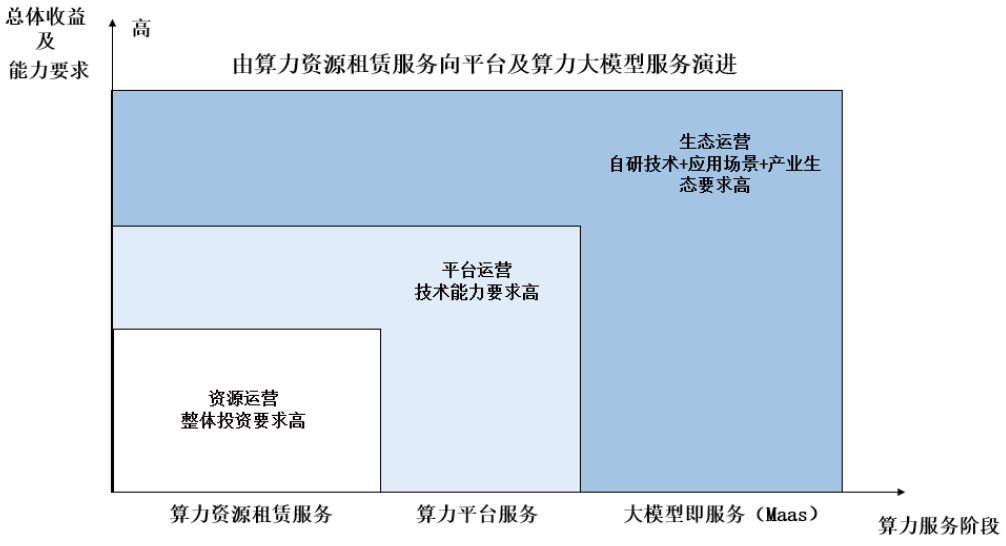

图 2 算力服务演进路径示意图

大模型高性能计算集群和服务能力，为客户提供 MaaS 一站式服务和行业模型解决方案，助力客户构建专属大模型及智能应用 [7]。

对于电信运营商而言，MaaS 是算力服务的高阶模式，以应用场景和生态运营为主，需要电信运营商具备较强的自研技术和生态合作能力。此阶段服务特征主要有两个方面：一是 MasS 服务方面，主要目标客户是行业大模型，提供基于电信的基础算力资源及市场上通用大模型和行业大模型能力；二是数据服务方面，可提供特定行业模型的数据清洗训练，以行业大模型的能力服务行业客户。

运营商提供 MaaS 需具备各类行业 / 场景大模型开发能力，供 AI 应用提供商进行能力调用，同时最佳的可落地路径是基于行业大模型优先切入垂直行业落地试点。总体来说运营商提供 MaaS 服务挑战较大，目前应着力开展能力建设，提高技术水平，培养专家队伍，联合产业生态伙伴共同探索场景应用和商业模式落地。大模型即服务总结如表 2 所示。

表 2 大模型即服务总结

服务目标	参与大模型和算力网络建设，打造算力生态合作体系，满足客户大模型业务全生命周期、全要素业务需求
典型产品	通用大模型服务、行业大模型服务、行业数据标记、整理、分析
运营商服务能力要求	为政府、央国企等客户提供模型训练服务。对于数据敏感客户，可灵活提供属地化部署服务。可结合技术能力，提供系列化 AI 工具集及技术咨询服务
客户技术能力要求	基于客户对数据安全和模型能力需求，客户技术能力要求不同
目标客户	政府、企业、科研等客户

5 结语

以 ChatGPT 为代表的 AIGC 应用快速发展，开启了人工智能新一轮增长，并带来新型业务模式，未来将重塑所有行业，也给信息通信业带来新的机遇和挑战。对于电信运营商而言，需要积极把握行业发展趋势，聚焦行业需求，基于自身优质的“云—网—算力”资源，探索由算力资源租赁向算力平台和大模型即服务演进路径和模式。同时提高企业全栈自研技术水平，加强产业联盟生态合作，共同建设普惠可用、安全可控、智能便捷的 MaaS 一体化平台和服务体系，打造运营商第二增长曲线，推动社会数字化转型，助力产业升级。

参考文献

[1] 中国信息通信研究院,京东探索研究院.人工智能生成内容(AIGC)白皮书(2022年)[R].2022.

[2] 李正茂,王桂荣.论算力时代的三定律[J].电信科学,2022,38(6):13-17.

[3] 郭亮.数据中心发展综述[J].信息通信技术与政策,2023,49(5):2-8.

[4] 中国信息通信研究院云计算与大数据研究所.中国算力服务研究报告(2023年)[R].2023.

[5] 赵倩颖,邢文娟,雷波,等.算力时代下的算力服务需求与特征思考[J].信息通信技术,2022,12(2):14-18+26.

[6] 刘亮,张琛,杨学燕.生成式人工智能技术对通信行业的影响研究[J].邮电设计技术,2023(7):1-7.

[7] 腾讯研究院,同济大学.大模型时代的AI十大趋势观察[R].2023.