

端边协同的6G内生AI网络



6G AI-Native Network with Edge-Device Collaboration

王志勤/WANG Zhiqin, 周吉喆/ZHOU Jizhe,
韩凯峰/HAN Kaifeng

(中国信息通信研究院, 中国 北京 100191)
(China Academy of Information and Communications Technology, Beijing
100191, China)

DOI: 10.12142/ZTETJ.202504005

网络出版地址: <http://kns.cnki.net/kcms/detail/34.1228.TN.20250716.0845.004.html>

网络出版日期: 2025-07-16

收稿日期: 2025-05-28

摘要: 面向6G智能终端AI业务原生、融合感知、智能协同业务需求, 提出了一种端边协同的6G内生智能网络架构。该架构通过分层设计(基础设施层、模型管理层、资源管控层、业务编排层), 实现了端边数据管控、模型动态协同及异构资源融合调度, 具备“通信+计算+数据+模型”一体化服务能力。在内生智能网络架构基础上, 提出了端边智能协同系统评估模型, 并围绕数据管控、模型协同、资源调度3个维度, 提出了高质量数据集构建、数据管控框架、参考模型库、端边模型协同、异构资源融合管控、灵活组网等关键技术, 形成端边智能协同技术体系。

关键词: 通信与人工智能融合; 端边协同; 资源管理

Abstract: To address the requirements of artificial intelligence (AI)-native service, integrated sensing, and intelligent collaboration for 6G AI terminals, an edge-device collaborative 6G AI-native network architecture is proposed. The architecture adopts a hierarchical design (infrastructure layer, model management layer, resource control layer, and service orchestration layer) to achieve core capabilities such as edge-device data management and control, dynamic model collaboration, and integrated heterogeneous resource scheduling, providing integrated "communication + computing + data + model" service capabilities. Based on this architecture, an edge-device AI collaborative system evaluation model is established. Focusing on three key dimensions—data management, model collaboration, and resource scheduling—key technologies are proposed, including high-quality dataset construction, data management framework, reference model library, edge-device model collaboration, heterogeneous resource integration and control, and flexible networking, forming a comprehensive edge-device intelligent collaboration technology system.

Keywords: integration of AI and communication; edge-device collaboration; resource management

引用格式: 王志勤, 周吉喆, 韩凯峰. 端边协同的6G内生AI网络 [J]. 中兴通讯技术, 2025, 31(4): 29–33. DOI: 10.12142/ZTETJ.202504005

Citation: WANG Z Q, ZHOU J Z, HAN K F. 6G AI-native network with edge-device collaboration [J]. ZTE technology journal, 2025, 31(4): 29–33. DOI: 10.12142/ZTETJ.202504005

1 6G端边智能协同研究背景

随着移动通信与人工智能(AI)技术的持续创新发展, 6G技术将与感知、智能、计算、安全等新兴技术深度融合, 这进一步拓宽了网络服务边界, 提升了服务效率, 在5G“万物互联”基础上实现“万物智联”^[1-2]。6G服务对象正逐步向具备融合感知、智能决策、计算能力的新一代智能终端演进, 终端形态也从传统手机扩展到AI手机、智能可穿戴、智能机器人、智能网联车等多元形态^[3], 这将促进6G智能业务多元化、泛在化发展。为满足未来泛在智能业务低时

延、高精度等多维指标与数据高效处理需求, 业界提出将模型部署下沉到边缘侧或终端侧的实施策略: 端侧负责执行本地轻量化智能任务处理, 边侧则专注于处理终端卸载的高复杂度计算任务, 通过端边协同计算显著提升智能业务指标。

当前, 全球标准化组织积极探索面向智能终端的6G端边智能协同关键技术。美国NextG联盟提出要在6G系统设计中考虑终端AI计算能力, 着力解决智能任务在终端与网络算力节点间的协同计算问题, 以应对物联网设备多元数据处理与资源受限的挑战^[4]。欧洲6G-IA协会将协作机器人等智能协作场景纳入6G重点应用场景, 强调6G需要支持AI驱动的空口智能和边缘智能, 研究包括终端在内的边缘设备分布式模型训练和部署技术^[5]。中国IMT-2030(6G)推进组提出6G移动通信网络将提供泛在的AI即服务能力(AIaaS),

基金项目: 北京市自然科学基金资助项目(L253004); 中国科协青年人才托举工程项目(2023QNRC001, 2022QNRC001)

系统开展边缘智能、分布式训练推理等关键技术研究^[6],并针对无线空口物理层领域,深入研究移动通信与AI融合的数据格式与模型设计方案,为业界研发提供模型参考^[7]。2025年,第3代合作伙伴计划(3GPP)Release 20正式启动6G网络架构研究项目,将网络AI框架设计列为核心研究内容,同时探索高效可扩展的数据管控框架^[8]。

学术界积极开展面向数据、模型、资源优化设计的端边智能协同方法研究。文献[9]提出了一种端边协同模型拆分与资源优化方法,降低了系统能耗。文献[10]基于终端移动特性,联合优化端边协同模型拆分、数据压缩和业务接入策略,降低了业务时延。文献[11]研究了云边端计算网络下的带宽、边缘算力与任务分配优化方案,提升网络整体效用。

2 6G端边智能业务特征与需求

端边智能业务以强感知、强计算、强协同为核心,通过数据融合处理、任务协作与实时决策,满足低时延、高精度、高可靠的业务多维指标要求。总体来说,6G端边智能业务发展呈现AI业务原生、融合感知、智能协同三大特征:

1) AI业务原生:大模型、智能体等AI技术与终端深度融合,将催生6G智能终端AI原生业务全新应用范式^[12]。这一趋势对网络服务供给能力提出了更高要求,促使服务需求从传统移动接入向“连接+AI”智能化服务模式转变。因此,网络需精准识别终端AI业务的异构资源需求(如通信、计算),通过一体化决策机制实现智能终端的按需融合服务。

2) 融合感知:随着智能机器人、智能网联车等终端设备自主性与智能化水平的提升,6G智能终端将深度感知周围环境^[13],具备执行智能避障、路径优化等复杂任务的能力。为支持上述业务场景,需要网络提供物理世界数据感知、采集和建模能力,并基于统一数据管控框架为智能终端交互提供高质量数据服务。

3) 智能协同:针对低空经济、智慧工厂等场景,6G智能终端行为模式从“单体智能”向“多智能协作”方式演进,通过多智能终端协同完成无人物流配送、机器人协作搬运等协作任务要求^[14]。为支持上述业务场景,需要网络提供多终端按需组网和动态任务调度能力,支持智能终端任务快速响应和协作融合决策,提升复杂协同任务效率。

由此可见,6G新型智能终端业务对网络资源供给、数据管理、协同机制提出更高要求,亟需优化支持端边协同的6G内生AI网络架构,开展6G端边智能协同关键技术研究。首先,构建内生支持数据模型管控的内生智能网络架构,支撑端边智能协同关键能力,为6G智能终端与边侧算力高效智能协同提供网络承载。其次,面向6G智能终端业务数据、

模型、资源需求,围绕数据管理、模型协同、异构资源融合管控技术维度,提出端边智能协同关键技术。

3 6G内生智能网络架构与系统评价模型设计

6G智能终端具备AI原生、融合感知与智能协同等业务特征,这对端边协同的6G网络架构提出了内生智能的设计要求。为满足融合感知业务需求,6G网络需构建统一的数据处理框架,支持数据采集、传输、存储与共享,从而为终端、网络决策及端边智能协同提供按需数据服务。6G网络需具备端边分布式多级网络全局模型管控能力,通过多节点协同提升智能业务性能指标。为满足多样化AI原生业务需求,6G网络需实现通信-计算-数据-模型的一体化智能服务调度与体验保障,支持动态按需的端边异构资源分配。

3.1 内生智能网络架构

现有移动接入与模型管理体系的分离架构难以满足端边融合感知、智能高效协同及多维服务按需调度的需求。本文提出一种“内生”数据模型管控网络架构,通过统一分布式数据模型管理与异构资源融合调度机制,有效支撑终端复杂智能业务对超低时延、超高精度等性能指标的严苛要求。

如图1所示,从逻辑分层角度,内生智能网络架构包括基础设施层、模型管理层、资源控制层、业务编排层,并通过统一接口对外支撑多元智能应用。

基础设施层由智能终端、边缘计算节点及无线接入网等分布式网络构成,通过泛在网络接入实现动态跨节点协同,为智能业务接入与端边协同提供基础资源支撑。

模型管理层实现分布式端边网络的统一数据与模型管理,核心功能包括:1)构建统一数据管控框架,实现异构设备数据的按需采集、传输、存储、处理;2)实施边侧模型全生命周期管理,包括模型训练、部署、选择、监控、更新等;3)基于终端模型能力感知,建立全局模型部署视图。

资源管控层实现通信、数据、模型、计算多维资源的统一建模与协同管控,通过原子化能力抽象与融合控制机制,支撑端到端异构资源整合及一体化调度。

业务编排层对外接收多样化应用请求,并利用内置AI模型生成“需求解析-资源映射-任务调度”的动态端边协同编排策略,实现业务体验按需保障,保持系统高效运行。

3.2 系统评价模型

当前端边协同推理策略主要聚焦推理时延、推理准确度、能耗等单一指标优化,缺少针对端边多级分布式网络的系统级评估指标,难以实现系统长期效用最优的模型管控与

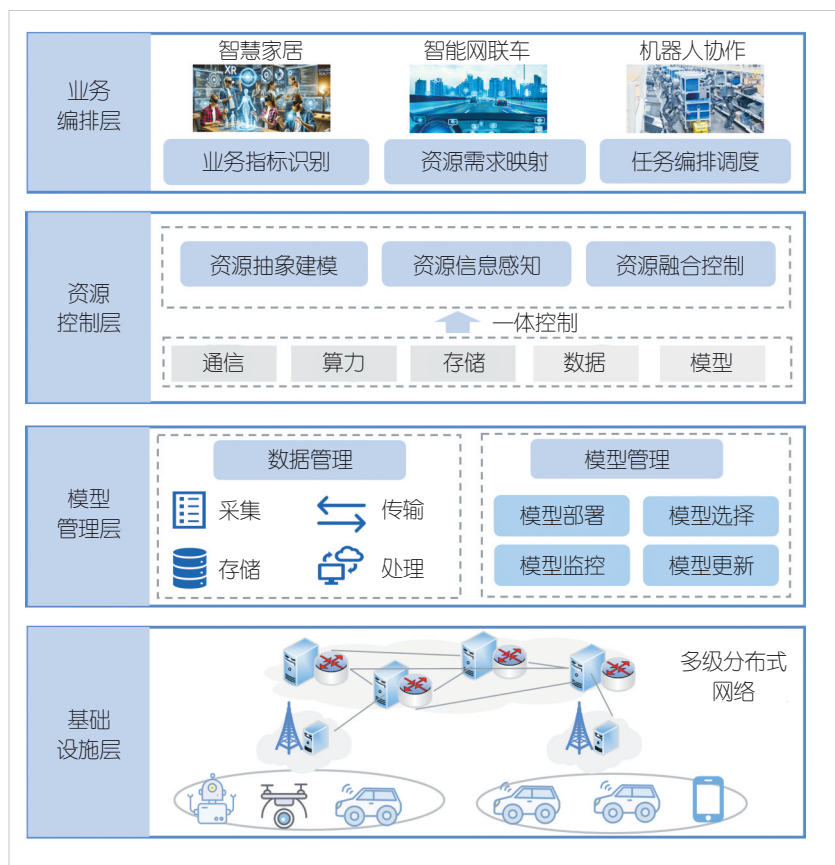


图1 6G内生智能网络架构

多维资源管理机制。针对上述问题，本节面向智能协同与业务多维指标保障需求，设计系统性能评价方法，突破传统单一维度评价的局限性，为系统整体效用优化提供模型基准。

端边协同系统评价模型涵盖基础资源开销、模型协同效率、能效比等多维指标。基础资源开销包括协同推理业务全周期的通信带宽、算力、存储等资源开销。模型协同效率需要考虑推理时延、准确度、模型复杂度等。能效比则聚焦单位任务的能耗，通过构建全过程能耗模型，覆盖通信传输、计算处理、模型更新等环节的能量消耗。为保证不同类型评价指标横向可比较，可利用动态加权、归一化等方法，将不同量纲的指标转换为可比较的数值，建立多维指标的系统综合评价模型。同时，系统评价模型具备灵活性，可根据不同业务场景特征（实时场景可能更关注时延和可靠性，离线场景可能更注重能效比），动态调整权重系数，满足不同场景下系统效用优化目标。

4 端边智能协同的6G数据管控

4.1 数据集构建

大规模、高质量数据是模型研发的基础性资源。一方

面，端边协同模型部署需要适配无线环境，而无线环境具备高时变特性和复杂传播机制，传统方法在信道参数估计精度和实时性方面难以满足端边智能协同业务低时延要求，面向真实无线环境的高质量信道数据采集存在挑战。另一方面，智能网联车、智能无人机、智能机器人等终端决策涉及动态环境中的复杂交互，其不同环境中的适应和泛化能力取决于环境数据多样性，多样化、高质量、真实环境数据采集存在挑战。由此可见，面向真实信道环境和物理世界的数据集构建，是使能6G智能终端交互的关键技术。

面向智能终端的信道数据集构建具体为：首先，基于大规模天线阵列和高性能信号处理平台，建立数据采集硬件基础；其次，开展多样化场景信道数据采集工作，并对原始数据进行降噪、格式统一等预处理；最后，利用智能终端天线方向图，结合实采无线信道参数与宽频信道响应的耦合计算方法，构建高还原度、可扩展的真实场景信道模型。

面向智能终端的环境数据集构建具体为：首先，建立包括感知基站、激光雷达等多模态数据感知平台，整合不同设备感知数据；其次，综合通信信号感知大范围环境轮廓、激光雷达点云测量姿态数据等多源数据优势，基于数据同步与配准算法，提升室内外环境地图构建效率与分辨率。

4.2 数据管控框架

端边智能协同业务需要实现跨设备数据融合与跨域数据共享，但现有数据采集框架主要面向核心网功能，缺乏端边分布式多级网络的跨域数据管控机制。为此，需构建支持数据采集、存储与共享在内的全生命周期管理框架，以保障端边多节点智能协同任务的高效可信交互。

在数据采集方面，建立统一的数据格式标准与规范化采集流程，支撑异构设备的标准化数据采集，并基于数据特征与流量需求自适应选择传输协议，优化采集传输效率。在数据存储方面，依据数据类型与实时性要求设计存储架构与索引机制，支持网络功能模块与跨域节点的高效数据调用。在数据共享方面，定义标准化的跨模块与跨节点数据共享接口，采用服务化调用机制触发共享流程，结合自适应传输协议确保共享效率。

5 端边智能协同的6G模型管理与协同

5.1 参考模型库构建

基于数据采集与建模，亟需面向不同端边智能协同场景构建AI模型。当前，不同用例、不同厂商在技术方案和模型设计方面存在较大差异，制约端边多节点模型智能协同标准化和产业落地。因此，亟需构建6G+AI参考模型库，提供典型场景基准和对比模型，并通过参考模型测试验证，指导工业界实现模型性能与复杂度的最优权衡。首先，构建面向6G典型端边智能协同用例的参考模型，给出面向用例的参考模型设计方案路线以及模型的设计架构和参数，确保模型的可复现性。其次，搭建模型验证评估仿真平台，集成开源数据集与评估工具链，实现全流程测试，同时建立模型评估标准，提供统一评估体系。

5.2 端边模型协同

随着智能网联车、机器人等新一代智能终端发展，智能应用对低时延、低开销、高精度推理等多维指标提出更高要求。然而，终端单体能力受限，无法支撑大模型高精度模型推理；同时，边侧单一静态模型推理受限于动态无线环境，无法自适应动态信道和多样化业务需求。因此，亟需构建端边模型协同机制，综合分布式、差异化端边节点能力，优化模型部署与动态模型协同推理策略。

端边模型部署方面，考虑到边缘节点、终端算力受限且分布式，需要综合考虑边缘节点分布、终端时空分布、业务请求时空分布等特征，设计时空驱动的分布式端边模型部署机制。针对具体业务推理场景，可监控模型协同指标完成情况，在完成指标下降时，基于数据管理框架按需向端边节点下发数据采集任务，并利用增强学习对融合数据和原始部署模型进行微调更新，提升模型对多样化业务适配泛化能力。

端边模型协同推理方面，侧重设计网络辅助的端边模型协同机制，利用分布式多级计算技术使能端边智能模型协同高效推理，支持终端对高算力智能业务需求。具体而言，设计适配动态信道的端侧模型拆分和推理任务分配方法，并结合模型部署状态，优化端边模型选择策略，满足端边推理业务精度、时延、能耗等多维指标要求。

6 端边智能协同的6G异构资源融合管控

6.1 异构资源融合控制调度

6G智能业务对通信、计算、数据、模型

需求多元，且终端分布和业务请求存在动态、时变特性，需要网络侧提供按需资源与服务调度，准确响应终端业务请求。当前，边缘计算业务分流主要根据业务部署和终端位置信息，模型调度主要依赖应用服务商集中管控，难以实现通信、计算、数据、模型等异构资源融合控制调度，制约资源高效使用。因此，亟需构建端边异构资源融合控制与调度机制，设计适配动态业务需求的端边异构资源融合管控机制。首先，对端侧与边侧分布式算力节点资源状态进行实时感知，生成全局资源状态视图。其次，当接收到业务请求时，基于业务需求、终端移动性等，生成最优的数据采集传输、边侧模型选择、通信计算资源分配的一体策略，在保障业务指标的同时提升网络整体资源利用率。

6.2 任务驱动灵活组网

针对机器人协作、智能车联等多智能终端协作任务场景，需要多个智能终端在完成自身智能业务的同时共同完成多终端协作任务。例如，文献[15]面向端边协同车联网场景，优化车车/车路协同下的模型拆分、组网与资源分配策略，以降低AI任务时延。由此可见，面向多智能终端协作场景，需精准识别协作任务在数据、模型等方面的需求，动态组建任务驱动的多终端-多边节点协作网络，并分配共享异构融合资源，实现多智能终端间、端边多节点间高效交互与协作。

7 端边智能协同赋能多智能终端协作

在应急救援、城市安防、智能车联等场景中，需多智能终端协作完成数据融合和统一智能决策。以城市安防为例（具体如图2所示），地面机器人完成地面目标捕捉和识别，无人机完成俯视环境感知和目标识别，智能眼镜完成人类视角画面捕捉和识别，3种终端均完成单一视角信息采集与智能识别业务，但单终端单一视角及受限算力、电池能力，很难支持城市安防快速搜索、全局判断需求。借助6G端边智能协同，可优化调度3种终端的能力，支持边缘节点分担端

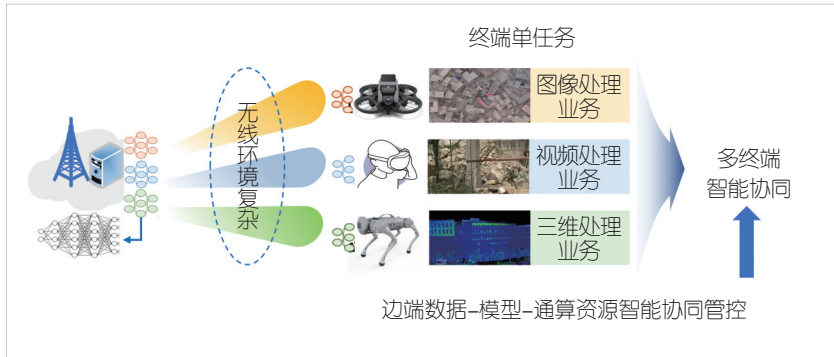


图2 端边智能协同支持城市安防业务场景示例

侧算力需求,实现多业务间协同处理,共同高效完成智能互联与融合决策任务。因此,端边协同的6G内生AI网络是解决复杂智能化业务的重要手段。

8 结束语

随着6G智能终端与智能业务演进发展,端边智能协同已成为支持6G智能终端复杂AI业务需求的有力手段。通过构建统一的端边数据模型管控框架、优化端边智能协同与异构资源融合管控机制,6G网络能够有效支撑AI手机、智能机器人、智能网联车等多形态终端的智能化、协同化业务,应对低时延、高精度等多维指标挑战。未来,随着技术持续演进和产业生态的完善,6G与AI深度融合将进一步释放智能终端的潜力,为垂直行业应用和社会数字化转型提供更强大的技术支撑,开启“万物智联”新篇章。

参考文献

- [1] IMT-2030(6G)推进组. 6G总体愿景与潜在关键技术白皮书[R]. 2021
- [2] IMT-2030(6G)推进组. 6G网络架构展望白皮书[R]. 2023
- [3] 中国信息通信研究院. 新一代智能终端蓝皮书[R]. 2024
- [4] Next G Alliance. 6G technologies for wide-area cloud evolution [R]. 2023
- [5] 6G-IA. White paper: European vision for the 6G network ecosystem [R]. 2024
- [6] IMT-2030(6G)推进组. 6G AI即服务(AlaaS)需求研究[R]. 2023
- [7] IMT-2020(5G)推进组, IMT-2030(6G)推进组. 移动通信与AI融合的数据格式和模型建议书(第一阶段:物理层领域) [R]. 2023
- [8] 3GPP. Study on architecture for 6G system: S2-2506096 [EB/OL]. (2025-05-23) [2025-06-18]. https://www.3gpp.org/ftp/tsg_sa/WG2_Arch/TSGS2_169_Fukuoka_2025-05/Docs
- [9] JIANG W, HAN H C, FENG D Q, et al. Energy-efficient and accuracy-aware DNN inference with IoT device-edge collaboration [J]. IEEE transactions on services computing, 2025, 18(2): 784-797. DOI: 10.1109/TSC.2025.3536311
- [10] LIU Z, TIAN M, DONG M, et al. MoEI: mobility-aware edge inference based on model partition and service migration [J]. IEEE transactions on mobile computing, 2024, 23(10): 9437-9450. DOI: 10.1109/TMC.2024.3366186
- [11] HE Y, FANG J, YU F R, et al. Large language models (LLMs)

inference offloading and resource allocation in cloud-edge computing: an active inference approach [J]. IEEE transactions on mobile computing, 2024, 23(12): 11253-11264 DOI: 10.1109/TMC.2024.3415661

- [12] 中国移动研究院. AI+通信业务白皮书[R]. 2025
- [13] 中国信息通信研究院. 具身智能发展报告[R]. 2024
- [14] 华为终端有限公司. AI终端白皮书[R]. 2025
- [15] LIU Z, DU H, LIN J, et al. DNN partitioning, task offloading, and resource allocation in dynamic vehicular networks: a Lyapunov-Guided diffusion-based reinforcement learning approach [J]. IEEE transactions on mobile computing, 2025, 24(3): 1945-1962. DOI: 10.1109/TMC.2024.3486728

作者简介



王志勤, 教授级高级工程师, 中国信息通信研究院副院长, 中国通信标准化协会副理事长、无线技术委员会主席, 中国通信学会无线移动委员会主任委员; 研究领域为移动通信技术、标准、产业及政策。



周吉喆, 中国信息通信研究院高级工程师、中国通信标准化协会TC5智能体通信网络子组副组长; 主要研究方向为6G网络、网络智能化技术; 主持和参与国家级/省部级项目10余项; 发表论文10余篇。



韩凯峰, 中国信息通信研究院高级工程师; 主要研究方向为面向6G的无线人工智能、通信感知一体化技术; 主持和参与国家级和省部级研究项目10余项; 发表论文60余篇。