



白皮书

Mobile AI: AI 与移动通信融合的新生态

目 录

1	迈入 Mobile AI 时代	1
1.1	移动网络的发展	1
1.2	Mobile AI 带来网络流量深刻变化	1
1.3	Mobile AI 网络新需求	2
1.4	Net4AI 与 AI4Net 双向奔赴	3
2	Net4AI：应用与技术的双向奔赴	5
2.1	应用需求	5
2.1.1	AI 应用的产业趋势和政策	5
2.1.1.1	AI 手机：耳聪目明的 AI Agent	6
2.1.2	智能驾驶	6
2.1.3	智能机器人	7
2.1.4	AIoT	7
2.2	网络新型技术	8
2.2.1	大上行技术	8
2.2.2	全频段 M-MIMO	9
2.2.3	连接增强	9
2.2.4	通算融合	10
3	AI4Net：AI 与移动网络的双向赋能	11
3.1	迈向 Mobile AI	11
3.1.1	网络智能化发展趋势	11
3.1.1.1	智能化应用案例	11
3.1.2	技术需求小结	13
3.2	构建自进化的通信大脑	13
3.2.1	通信大模型	14
3.2.2	数字孪生技术	15
3.2.3	技术需求小结	15
4	构建 Mobile AI 商业新范式	17
4.1	连接+算力双向协同	17
4.2	体验变现	17
5	未来展望	19
5.1	流量流向边缘，网络重心由中心向边迁移	19
5.1	从管道到平台，运营商向 TechCo 转型需掌控生态主导权	19
	术语表	20
	参考文献	21

5G-A 作为 5G 演进的重要阶段，不仅带来了通信能力的进一步跃升，也为人工智能与移动通信的融合提供了关键契机。运营商为应对 AI 业务快速增长，亟需推动“以网兴智（Network for AI）”与“以智赋网（AI for Network）”双向融合：一方面构建原生支撑 AI 的网络能力，另一方面通过通信大模型、数字孪生等关键技术提升网络的自学习、自优化能力。同时，围绕“连接+算力+服务”融合供给，也需要探索通信资源（Byte）与计算资源（Computing）双轮驱动的新型商业范式，实现运营商向科技型公司的转型，推动移动通信迈入“Mobile AI”阶段。

本白皮书由中国电信与华为技术有限公司联合撰写，围绕 Mobile AI 时代的核心特征展开分析，指出网络流量结构正在由“人-人”通信向“人-机”与“机-机”智能交互转变，AI 业务带来的大上行、多模态、低时延、高算力的需求，对网络架构提出系统性挑战。在此背景下，我们将在持续迭代 5G-A 向 6G 网络演进的同时，构建具备通智算融合能力的无线网络，支撑未来 AI 应用的敏捷部署与高质量体验，携手业界合作伙伴共同探索 AI 与网络的深度融合。

1 迈入 Mobile AI 时代

1.1 移动网络的发展

自 20 世纪 80 年代 1G 开启“移动语音”时代以来，移动通信技术经历了从语音通信（1G）、数字传输与基础数据服务（2G）、移动互联（3G），再到多媒体融合与深度应用（4G）的不断跃升。现如今，5G 技术的商用不仅带来了高速率、低时延、广连接的飞跃，更推动了“万物互联”格局的高效落地。伴随人工智能（AI）的技术突破与广泛部署，移动网络正加速迈向“Mobile AI”时代。Mobile AI 不仅指移动设备的智能化，更是在网络各类节点深度协同引入 AI 技术基础上，驱动通信技术体系从“万物互联”迈向“万物智联”的结构升级。同时，在 AI 赋能下，移动网络不仅实现了人与人、人与物、物与物之间更高阶的智能交互，而且也为智能助手、智慧交互、沉浸式体验等全新应用场景提供了基础性支撑。这一趋势的发展背后，更是广泛且便捷的智能计算资源作为关键底座的持续演进的过程。展望未来，6G 将进一步打破“通信+计算”的边界，使终端演进为具备感知、认知、决策能力的智能体。移动网络也将实现从连接通道向智能载体的跃迁，支撑通信与算力一体化、大模型、数字孪生等颠覆性技术的发展，构筑人机融合时代的智慧底座。

1.2 Mobile AI 带来网络流量深刻变化

智能设备的快速普及与 AI 模型的大规模部署正推动移动网络的流量结构和业务范式发生深刻变革。网络连接模式正从“人-人协作”向“人-机协同”以及“机-机智能体交互”演进，AI 驱动的网络业务日益增加，推动通信网络加速向智能化、服务化、平台化方向升级。首先，南北向（垂直）流量结构发生显著变化。随着终端侧智能能力增强，设备与网络之间的数据交互频率和内容复杂度迅速提升，特别是视频、图像、文本等多模态数据的上行传输激增，上行流量需求渐进化提升。据统计，单个 AI 应用所需的上行速率超过 20 Mbps，传统网络架构面临性能瓶颈。其次，东西向（水平）流量快速跃升。分布式 AI 系统要求多个智能体之间频繁交换模型参数、推理结果和数据片段，东西向流量的内容由传

统多媒体扩展为“模型-数据-算力-token”等智能化载荷，这一变化对网络横向传输能力提出了更高要求。最后，AI 业务的实时性特征使网络响应面临严峻的挑战。实时语音交互、具身机器人操控等场景要求网络确保端到端往返时延(RTT)小于 20 ms,以支撑高质量的人机交互体验,这一变化推动网络架构向超低时延、高确定性方向演进。以上一系列流量演进,依托的是一套可动态流转、贴近场景、持续演进的算力保障体系。AI 应用对推理响应速度和带宽负载的双重要求,意味着网络架构必须具备更加灵活高效的算力供给能力与调度机制。

1.3 Mobile AI 网络新需求

移动网络正从“连接承载”向“智能底座”转型。Mobile AI 业务的兴起引起了网络架构、资源协同、服务模式的全面演进,网络需要只有在算力体系、数据调度、服务供给等方面实现新型能力提升,才能更好地满足 AI 业务所要求的普遍性、实时性和可扩展性特征。具体来看,新需求主要体现在以下三方面:

1) 算力调度成为网络和业务的双重需求

在 Mobile AI 时代,算力成为贯穿网络各层的基础能力。就网络自身而言,需要在不同层级的节点部署 AI 能力,以提升调度、感知和控制的智能水平;就业务应用而言,需要根据模型规模、响应时延和隐私要求灵活选择执行路径。AI 业务的分布式推理、弹性迁移、持续更新等特性决定了算力必须“随业务而动”,且同时需具备快速、敏捷、高效的调度能力。基于此,移动网络需具备对多源异构算力的统一抽象和高效调度能力,实现网络与算力的深度融合,才能更好地支撑 AI 服务的敏捷部署。

2) 数据与模型的网络级协同处理能力

在 Mobile AI 服务场景中,数据的生成与消费分布更加动态化、碎片化,模型的执行位置也从集中处理演变为分布协同。以语音助手、增强现实、具身智能等场景为例,系统需要实时采集环境数据,并在最合适的节点快速完成模型推理以返回响应。面对数据和模型分布的不确定性,网络需要实现数据路径与算力路径的协同规划,支持数据在本地处理、区域聚合与跨域调度之间灵活切换。同时,AI 模型也应具备跨层灵活部署的能力,根据业务需求和网络资源实现高效协同,以此推动数据、模型、算力在网络中的融合演进。

3) 新型商业模式的探索

在 Mobile AI 发展背景下,传统以流量计费为核心的商业模式正面临高原期,运营商亟需探索连接、算力、服务融合的新价值体系。面向 6G,通信不再是简单的被动的传输手段,而是将演进为支撑智能服务体验的主动平台,连接和算力能力将共同成为核心经营要素。运营商可基于具体业务场景,灵活组合网络能力、算力资源和 AI 服务,形成多元化的产品供给和新型商业模式。以上技术升级变革对网络能力提出了更高的要求:需要支持从“面向连接的资源调度”向“面向服务的能力编排”转型,构建统一的资源抽象、实时感知与灵活计费机制。在此基础上,网络应具备原生支撑智能服务的架构与接口,推动通信即服务、算力即服务的协同供给,加速运营商从“连接提供者”向“智能服务平台”的角色转型。

1.4 Net4AI 与 AI4Net 双向奔赴

在 Mobile AI 成为引领未来信息社会的重要力量的背景下,产业界需要协同推进网络、算力、平台、应用等要素的融合创新,把握新一轮技术变革和价值重构的机遇,实现产业的迭代升级。5G-A 作为 Mobile AI 时代的关键支柱,其演进进程决定了 AI 应用的可达性和产业落地深度,因此应加快产业结构升级,释放活力,实现商业资源整合新价值。

1) Network for AI: 以网兴智,打造 Mobile AI 的坚实基础

当前,Mobile AI 业务呈现出多维度和深层次的变革趋势,对网络的带宽、时延和上行能力等领域提出了新的挑战。就沉浸式 3D 交互技术而言,需要支持双眼 8K、120 fps 以及六自由度的动作,这带来了至少每人 1 Gbps 的下行带宽需求;同时在多模态 AI 助手的人机交互方面,单轮生成的 token 数量可超过 2000 个,大型模型处理时延超过 320 ms,这就要求网络空口时延必须控制在 20 ms 以内,以确保整体交互时延低于 400 ms。

在这些挑战下,网络不再仅是数据管道,更是 AI 体验的基础保障和服务平台。在面向消费者(ToC)领域,运营商需从“流量经营”升级为“体验经营”,构建可量化、可保障的网络体验。例如,国际某些运营商针对直播用户推出了上行带宽保障套餐;国内部分运营商融合流量、算力和 OTT 应用权益,推出“云卡”套餐等新模式,以此来拓展算力服务维度。在面向企业(ToB)领域,AI 推

动行业服务从“连接”延展到“连接+服务”。运营商可以依托自身网络基础设施的资源优势，叠加云计算、感知、AI 分析等能力，提供端到端的集成解决方案的落地。例如，面向智能制造场景提供“网络+边云协同+模型推理”的一体化方案；在智慧交通、低空经济等场景中，拓展出通信与感知融合、空域数据服务等新型能力，实现网络价值的深度延伸^[1]。

2) AI for Network: 以智赋网，全面提升网络智能化水平

随着网络规模和业务复杂性不断上升，AI 正成为推动无线网络智能演进的核心驱动力。通过在网元层、网络层和业务层全面引入 AI，网络技术将实现从性能自优、运维自简、运营自智，逐步实现自治网络 L4 级的高阶智能阶段的可持续发展。

以通信大模型为代表的智能体系统正在构建新一代智能网络架构。例如，面向网络运维人员的 Copilot 类 AI 助手具备知识问答和任务辅助能力，可有效提升运维质量和效率；面向复杂网络场景的 Agent 类智能体通过意图识别和任务自动编排，实现对网络问题的自主感知与闭环处理的应用落地。在业务开通部署中，AI 可以模拟拓扑、流量与环境分布，构建精准的体验视图，支持网络的快速部署优化；在网络调优中，AI 可实时感知覆盖、速率、能效等指标，通过多目标优化实现用户体验与能耗的兼顾提升；在故障处理过程中，Agent 智能体能够利用数据理解和自动化任务编排，大幅减少告警冗余并解决大量软参数配置问题，而 Copilot 则辅助工程师高效完成复杂硬件故障的闭环处理，实现技术突破，网络结构的治理完善。

2 Net4AI：应用与技术的双向奔赴

2.1 应用需求

2.1.1 AI 应用的产业趋势和政策

如图 2-1 所示，随着人工智能技术的蓬勃发展，生成式 AI、多模态问答 AI 以及智能体（Agent）等技术正引起社会深刻变革。



图 2-1 Net4AI 场景

1) 生成式 AI

在内容创作领域，生成式 AI 技术展现出巨大的潜力。从文本生成到图像、视频的自动化创作，生成式 AI 正在革新广告制作、社交媒体等多个行业。以 Mid-Journey、Stable Diffusion 等工具为代表的生成模型，大幅降低了内容创作成本，提高新质生产力。这些技术的广泛应用也在智能客服、企业管理等领域激发创造力，嵌入服务场景。

2) 多模态问答 AI

多模态 AI 通过融合图像、文字、语音等多种数据形式，推动了人机交互方式的技术多元化。它突破了传统单一输入输出模式的限制，使 AI 系统能够理解并响应更为复杂的需求。例如，在金融、医疗等行业的智能服务中，引入多模态问答 AI 显著提升了服务的效率和质量。

3) 智能体技术

智能体作为能够自主决策并执行任务的 AI 技术,正在广泛应用于自动驾驶、智能家居等场景^[2]。区别于传统机器学习模型,智能体可以在动态环境中自我适应,高效完成复杂任务。智能体技术的普及不仅提高了 AI 应用的灵活性,也推动了更多场景的智能化升级。随着 AI 技术的不断成熟,未来的 AI 系统将能更好满足不同群体的需求,适应复杂多变的环境,提供更精确的预测与决策,推动社会实现深度数字化转型。与此同时,支撑这些 AI 技术的核心基础设施——移动网络,也需要不断加强自身的系统韧性,以便将来更好地服务 AI 应用。

2.1.2 AI 手机：耳聪目明的 AI Agent

Mobile AI 的应用逐渐从高端消费市场走向大众市场,智能手机成为 Mobile AI 的核心载体。这些设备不仅能够提供语音助手、实时翻译、视频编辑等基础功能,而且通过 AI 技术可实现智能推荐、个性化服务,深度嵌入用户的日常生活。目前,苹果、三星、华为等主流厂商纷纷推出具有 AI 能力的高性能手机。例如,华为 Mate 70 系列和三星的 Galaxy 系列机型集成了先进的计算能力和深度学习模型,使智能手机不仅具备强大的基础算力,还提供了丰富的 AI 应用功能。Mobile AI 赋能下的智能设备不仅能够实时处理图像、视频等多模态数据,而且可以优化用户体验。Mobile AI 的应用场景日益广泛,从智能客服、智能家居到健康监控,AI 手机的功能不断拓展。

2.1.3 智能驾驶

智能驾驶技术正在引领交通领域的革新,自动驾驶技术正从试验走向现实生活应用。L3 级和 L4 级自动驾驶系统依赖高精度传感器、强大的计算能力以及低时延的网络连接,通过实时感知和高效决策,自动驾驶车辆能够在复杂道路环境中安全行驶,一定程度上提升交通安全性和出行效率。目前,车辆上的计算平台需要具备强大的处理能力,以应对来自道路环境、其他车辆和基础设施的实时数据。实现自动驾驶面临的核心难题在于如何处理海量传感器数据并实时针对数据做出可行性决策。此外,低时延的网络连接对于车与车、车与路侧基础设施之间的协同至关重要,必须确保信息在毫秒级内传输,减少整体反应时间。随着技术进步,智能驾驶的应用场景将不断拓展,从高端私人车辆延伸到共享出行,自动驾驶将逐步普及,推动出行方式的革命性变化。

2.1.4 智能机器人

智能机器人正成为现代社会的重要组成部分，在服务业、工业制造和家庭生活中得到广泛应用。具身智能技术使机器人不仅能够感知环境，还可以与人类进行更加自然的交互，极大拓宽了机器人的应用边界^[3]。在商业服务领域，具身智能机器人已广泛应用于酒店、商场、机场等场所，提供导引、信息查询和客户服务等服务功能。在工业制造领域，机器人借助视觉和触觉感知执行复杂的生产任务，极大提升了生产效率。在家庭生活中，智能家居机器人正逐渐成为集清洁、安防、娱乐于一体的多功能核心设备。但是需要注意的是，智能机器人的运行不仅依赖强大的本地计算能力来处理复杂环境数据，还需要通过高速网络与云平台协同工作，来提升工作效率和响应能力。因此，移动网络的低时延和高可靠性成为支撑机器人广泛应用的关键因素。

2.1.5 AIoT

物联网（IoT）技术正在加速推进智慧物流、城市治理、农业生产等领域的智能化进程。在物流领域，IoT 结合射频识别（RFID）、传感器和定位系统，实现了对货物的精确追踪和环境调控，将库存周转效率提升了约 35%。智能决策系统使得自动引导车辆（AGV）与自动分拣设备高效协同，减少人工干预，将处理效率提高了数倍。通过数据驱动的供应链优化，物流行业有效降低了库存峰值和运输成本，朝着更智能、更精细的方向迈进。在城市治理领域，IoT 的全域感知能力显著提升了资源管理的精准度，推动垃圾处理、空气质量监控、交通管理向智能化转型。在农业领域，IoT 的实时监控优化了作物种植密度，提高了农作物产量，同时节约了水资源。随着 IoT 与 AI 技术的深度融合，AIoT 不断拓展应用场景，增强感知能力与智能决策，在零售、健康监测等行业推动数字化升级。此外，随着大模型技术和具身智能机器人的不断发展，智能世界的构建日趋完善，人们的工作效率和生活质量得以提升。AI 不仅能够优化无源物联网等应用，还推动了零售、健康监测等领域的新兴创新。未来，AI 将进一步助力无源物联网的广泛应用，释放海量数据的潜能，为各行业带来更高效的解决方案。

2.2 网络新型技术

随着 5G 及后续 6G 的发展，智能终端与网络之间的连接需求逐渐增加，5G-A 后续演进路线需要满足大规模、高效能、低时延等业务需求。其中，充分利用全频段频谱资源至关重要，频率资源的有效利用并结合网络架构的升级能满足多元化的用户需求。借助载波聚合（CA）、多小区频谱合作（MBSC）等技术在 5G-A 网络中的尝试，充分验证 5G-A 网络根据需求智能聚合频谱资源的能力，提供更大的带宽和更高的吞吐量，实现在不同场景下灵活调度频谱。上下行解耦、补充上行（SUL）和频谱聚合（FSA）等创新技术的研究深入，将进一步提升网络的稳定性和智能化水平，使 5G-A 网络不仅是一张通信网络，更成为支撑智能应用的多功能平台。未来，5G-A 将迈向真正的“万物智联”，打破传统通信与计算的边界，集成通信、计算与智能能力。借助通信与算力一体化技术，物理世界的实时数字化将成为可能，支撑无人驾驶、车路协同、低空经济等前沿应用。融合算力的 5G/6G 网络，为实时智能交互提供低时延、高效能的推理能力，并为业务提供更加灵活、智能的支持。

2.2.1 大上行技术

为保障高质量的 AI 交互体验，网络上下行能力均衡发挥重要的作用，尤其是在 AI 交互对实时性和数据传输双向高要求的驱动下，提升网络上行能力成为提升体验的关键途径。因此，大上行技术的应用在后续网络演进中具有重要的作用，下面介绍两个相关技术——上行链路增强技术和上行载波聚合：

上行链路增强技术（SUL）通过高频与低频协同的机制，优化 5G 网络中的上行频谱资源配置，有效缓解当前上行带宽受限的瓶颈问题。该技术融合智能频谱调度策略，能够在远距覆盖不足等复杂环境下显著提升上行速率与覆盖能力，特别适用于智能制造、远程医疗等对上行通信质量要求较高的场景，有助于实现更高质量、更强鲁棒性的 AI 交互体验。

上行载波聚合（UL CA）则通过聚合多个上行载波来增强上行链路的总吞吐能力，显著提升无线资源利用效率。该技术支持 TDD 与 FDD 频段的灵活联合调度，使频谱资源配置更加高效灵活，能够满足多样化业务场景对上行通信的需求。随着 UL CA 技术的持续演进与成熟，预计其将在更多垂直行业中实现规模化部

署，进一步提升网络的上行承载能力，为 AI 业务提供强有力的通信支撑。

2.2.2 全频段 M-MIMO

随着无线网络向更高频段演进，利用大带宽频谱资源技术成为提升用户体验的关键。然而，高频段仍存在覆盖能力弱、部署密度高、能效要求严苛等问题，增加网络建设难度。超大规模天线阵列（ELAA）作为 Massive MIMO 的演进技术也是 5G 超大规模空分技术的核心，旨在通过扩大阵列规模和集成密度，提升高频段的覆盖能力与网络性能^[4]。ELAA 技术通过多项轻量化、小型化的创新，突破了超大规模阵列部署的工程瓶颈。例如，基于超轻一体化天线单元和信号注入直馈技术（SDIF），ELAA 在提升覆盖增益的同时缩小了天线阵列的体积，实现了高集成度与低重量的平衡；在算法层面，引入自适应高分辨率波束赋形方法，实现“对得准、跟得快、配对好”的动态波束管控，可提升边缘用户体验和多用户调度效率，进一步扩展小区容量。

2.2.3 连接增强

蜂窝无源物联网技术正加速应用于制造、仓储、物流、资产管理等行业，逐步成为支撑低成本、高效率物联接入的重要手段。相比传统 RFID 等方案，该技术不仅具备低功耗、轻量化、易部署等优势，还可复用现有蜂窝网络资源，因而在覆盖能力和系统兼容性方面表现更为出色。但与此同时，其在覆盖范围、穿透能力和定位精度等方面仍面临诸多挑战，尤其是在大空间、多层级的复杂场景下，对系统性能提出了更高要求。为推动蜂窝无源物联网从“基础连接”迈向“精准感知”，需要在组网、覆盖和定位等方面进行增强。在组网方面，结合实时全双工干扰抵消、多基站联合接收、拓扑自适应搜索以及 SuperCell 小区级站间协同机制，可提升对平面仓库、立体库房等复杂场景无缝覆盖能力。在覆盖性能方面，融合如授权频谱、高性能信道编码以及多基站协同收发等技术，并通过提升标签端的能量采集灵敏度，使系统整体的覆盖增益较传统方案提升超过 20 dB，视距条件下的通信距离也实现了约 10 倍的拓展。在定位能力方面，设计三边信号强度测量与指纹识别算法，能实现 5~7 米级的室内定位精度，为物资查找、状态监测等精细化管理场景提供保障。。

2.2.4 通算融合

在 AI 驱动的新业务需求下，传统的中心化计算架构正在向更加智能、网络原生的体系转变，网络的角色已不再局限于连接，而是扩展为支持实时计算和推理的智能平台。新一代智能业务通常分布广泛、交互频繁且时延敏感，这使得网络的计算能力面临更高的要求。为满足这些业务需求，网络节点需要具备超越数据转发的能力，能够支持 AI 推理、模型部署、任务调度和状态感知等功能。此外，算力资源需要具备弹性调度和按需分发的能力，根据业务负载、网络状态和用户需求进行动态调整，确保灵活供给和就近处理。

网络节点的角色也在发生转变，逐步从“被动通道”向“主动智能”发展，能够保障关键任务的连续性和服务质量。网络与算力的协同作用正在重塑资源的组织方式。不同层级的网络节点通过数据、模型和控制的协同作用，形成一个高效的感知、推理与决策闭环，推动“网络即智能”理念的落地。通算融合不仅扩展了 AI 服务的覆盖范围和响应能力，还为下一代智能应用落地提供了更加敏捷、高效的支持。

3 AI4Net: AI 与移动网络的双向赋能

3.1 迈向 Mobile AI

3.1.1 网络智能化发展趋势

移动网络的业务形态日益多样化，从传统的语音、数据、视频服务，到裸眼 3D、XR 等新兴业务对低时延和高带宽的要求，以及行业用户对大上行、精准定位等新需求，都对运营商在保障用户服务的同时提出更高的要求。另外，新型无线站点、频谱和技术的引入带来了结构性挑战：频谱资源日益丰富（如 Sub-6 GHz、C 频段、毫米波、U6G 等），站点类型更加多样（宏基站、杆站、微站等），进一步增加了网络规划和运维的复杂性。随着大模型、数字孪生等 AI 技术的快速发展，5G 与 AI 的融合正在提速，网络智能化正从外挂叠加向内生一体转变。越来越多的运营商将无线网络智能化纳入战略并持续投入，全球领先的电信运营商也展示了无线网络智能化的成功实践。借助 AI 能力，运营商重构了网络运维和运营支撑系统，提升了网络数据分析的精细化和实时性，并通过开放 API 简化了业务应用与网络功能的集成协作。在全球标准化方面，3GPP、GTI、TM Forum、ETSI 等组织已经启动了相关工作。3GPP 在最新版本规范中研究了 AI 在电信网络中的应用，关注运维智能化和网元智能化，并推动 AI 与 5G-A 网络的结合。TM Forum 的自治网络（Autonomous Networks）项目发布了相关行业指南，致力于推进智能技术在资源、服务和业务运营层面的应用，提升网络效率和可持续发展能力。ETSI 的 ZSM 发布了端到端网络与服务管理架构，并展开对意图驱动闭环控制、数字孪生等技术的研究。

3.1.2 智能化应用案例

2024 年，根据 TM Forum 对多家运营商的调研显示，网络智能化在无线网络的降本增效方面发挥着关键作用。在增收方面，智能化广泛应用于提升用户体验、加速业务上线和保障业务连续性等场景；在降本方面，智能化主要体现于降低运维成本、优化网络规划和提高节能效率等方面。

1) 多意图多维度协同节能

智能节能的核心目标，是依托智能化技术实现用户体验与能耗代价的全局最优平衡。通过云端、网络、网元三层智能的协同，网络实时感知业务变化并动态调整站点、频谱、载波等网络资源，在保障用户感知的前提下实现能耗最优。

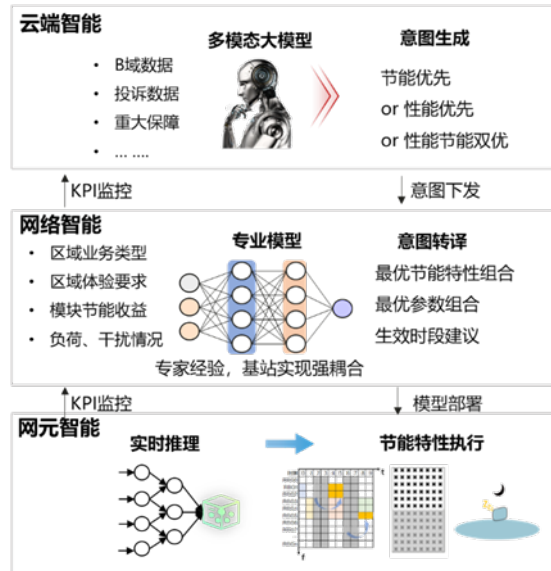


图 3-1 三层智能节能架构

云端智能（运营商网络中枢）完成场景到意图的精准感知和决策。依托智能节能平台的多意图引擎，结合场景感知能力，针对不同场景与时段的业务特征差异，制定差异化节能意图。网络智能（厂商网管）完成意图到动作的智能转译与编排，动态计算适配场景的业务体验目标（如用户速率阈值、连接时延指标），基于目标精细化编排载波调整、通道管理等原子节能功能。网元智能（基站网元）完成动作到执行的精细落地和迭代。借助网元内生智能技术，在基站侧实时进行业务感知和多维度节能功能集建模。通过多频协同、多站协同等创新技术，实现网络资源“按需关断-精准唤醒-灵活调整”的精细操作。基于上述三层智能节能架构，在保障用户体验的前提下，在陕西、江苏、广东、安徽实现节能提升 10%以上，同时业务体验维持稳定，达成真正的体验能效双优。

2) 差异化体验保障

网络内生智能能力能够实时感知生成的分层级业务体验目标，进一步落实服务水平协议（SLA）的确定性保障。无线网络智能分析模块实时感知业务特征（如 AR/VR 的低时延需求、大视频的带宽敏感性）和业务质量指标（卡顿率、首帧时延），结合核心网给定的用户服务质量等级标识（QCI）等级标签，针对用户+业

务的分级策略输出分级分类的体验质量（QoE）保障目标（例如，超高优先用户的云化业务设定 40 ms 时延目标，普通用户的流媒体业务设定 200 ms 缓冲阈值）。智能无线网络在载波资源层和时频资源层动态生成差异化的资源调度策略。通过多层次（载波-小区-时频）、多维度（预测-优化-控制）的智能资源管理，最终达成不同层级用户与业务的确定性 SLA 要求。

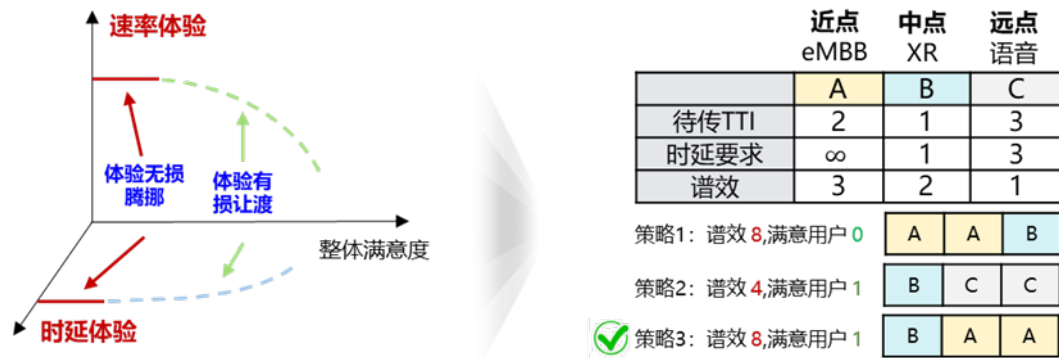


图 3-2 差异化体验保障

3.1.3 技术需求小结

AI for Network 技术正推动网络由传统的自动化管理向智能化自优化演进。借助 AI 技术，网络能够更高效地分析处理海量数据，自动识别问题并进行调整，对增强网络性能、提升用户体验和保障系统韧性起到了至关重要的作用。随着 AI 深度融入网络，网络调控不再依赖静态规则，而是能够根据实时数据自动优化流量、配置资源、预测需求，显著提升网络管理效率和用户体验，实现动态化管理。在 AI 应用的远景中，AI for Network 将通过深度学习和强化学习等技术，实现更加智能的自优化功能。需要注意的是，AI for Network 的演进不仅推动了网络智能化水平的提高，也对算力提出了更高要求——网络需要强大的计算资源来支持 AI 模型的实时推理和数据分析，尤其在边缘计算和云计算平台中，需要加速计算以保证网络的高效运行。因此，运营商需在网络架构中合理配置算力资源，以促进 AI 技术的可行性发展。

3.2 构建自进化的通信大脑

在 Mobile AI 时代，AI 逐步演化为通信系统内部不可或缺的核心决策引擎。AI for Network 实现人工智能深度嵌入网络运行全生命周期，包括通算资源调度、

干扰管理、故障预测、节能控制等环节，无线网络逐步从“模型驱动”取代“规则驱动”。相较于依赖固定阈值和静态规则的传统网络优化方法，AI 可从已有网络数据中持续学习并优化分析，使网络具备自学习、自调整、自愈合的智能化特征。AI 优化方法依赖实时数据反馈，具有周期性、低时延和差异化的特点，这对支撑 AI 优化的算力体系提出了新的要求——算力不再局限于集中部署，而是需要分布在网络的不同层级，并贴近场景进行决策。为了满足这一需求，通信系统正在逐步构建任务分级的算力分布架构，对于时延要求严格、场景耦合度高的任务（如小区级负载调节、功率控制优化），需要在基站侧部署轻量级模型实现快速推理；而对于跨域分析和全局策略制定等任务，则仍依赖于核心平台进行模型训练和策略输出。

通过 AI 模型算法与网络架构的深度融合，AI for Network 使网络具备了“运行中感知、优化中学习”的能力，推动通信系统从被动响应向主动优化转变，加速实现自驱动、自适应、自闭环的目标。这不仅提升网络的运行效率和用户体验，也为后续网络的演进提供智能决策的参考。

3.2.1 通信大模型

大模型是指基于超大规模预训练语料，且参数规模达到数亿甚至数十亿级的深度学习模型，如 ChatGPT 和 DeepSeek 等。大模型在性能和泛化能力的提升，使其在许多任务上的表现超越了小模型。例如，大模型在自然语言处理领域的成功，极大推动了 AI 在语言理解与生成方面的应用。

在电信领域，随着大模型的引入，网络智能化正在加速向 L4 级意图自适应网络演进。大模型具有强大的自然语言理解能力，能够将传统规则驱动的命令交互方式转化为意图驱动的交互模式，从而实现网络配置和操作的智能化，降低人工操作的门槛。同时，大模型能够学习大量的运维案例与规则，替代基于经验的故障分析，提供智能辅助运维功能，提高故障诊断的准确性与效率。大模型的推理能力支持自主决策和闭环处理，使网络管理能够在复杂环境中自动响应与优化，推动网络管理向更加智能和自主的方向演进。

通信大模型的体系可分为无线基础大模型和场景化大模型两大类。无线基础大模型涵盖无线网络相关的语言理解、性能预测、策略决策等通用能力，构建了面向无线领域的智能基座。而场景化大模型则针对特定应用场景（如无线感知、

网络仿真、故障诊断、网络优化等）进行专门设计和训练，具备高度的任务适配性。这些大模型的应用可以帮助电信运营商在无线网络的各个环节实现更精细的调度和优化，提升整体网络性能。

为了更好地适应无线网络的特点，通信大模型需要具备对全量专业知识的覆盖能力。通过基于亿级规模语料和思维链训练，通信大模型能够深刻理解无线网络的各类业务流程，并在网络部署、优化和运维中提供全局视角的认知与推理辅助。此外，通信大模型还需支持接口适配和操作指令生成。

3.2.2 数字孪生技术

无线数字孪生系统（RDTS）是一项应用于无线网络的数字化建模技术^[5]，旨在通过构建与物理网络高度映射的虚拟模型，实现对网络状态、性能、行为及操作过程的实时仿真与预测。该系统能够在数字空间中近似还原真实网络的运行机制，辅助运营商对网络进行更及时、全面的感知和监测，并提前识别潜在风险，为网络管理提供了更加主动的辅助手段。

数字孪生技术的核心优势在于其为网络优化提供了可量化、可模拟的决策基础。通过数字模型的实时推算能力，运营商可在虚拟环境中开展多种优化策略的预评估与验证，有效降低实际部署过程中的调整风险与资源消耗。同时，数字孪生系统与 AI 技术的深度融合，使得网络在应对复杂多变的场景时具备更强的判断与调整能力，从而提升网络运行的效率与稳定性。通过将物理网络中的关键运行数据同步至虚拟模型中，可以较为准确地反映网络的流量状态、性能指标及资源使用情况，从而为流量调度、通算资源配置及故障预防等操作提供辅助支撑。

伴随 AI 与大数据技术的升级，数字孪生系统的能力也在持续迭代。通过机器学习与深度学习算法，数字模型能够不断从历史数据中学习和调整，增强其对网络行为的理解和未来趋势的预测。未来的无线数字孪生将成为网络系统中的关键组成部分，为实现网络的自适应调整、故障预警与动态优化提供持续支持，助力运营商迈向更高水平的智能化管理。

3.2.3 技术需求小结

AI for Network 致力于将人工智能深度嵌入网络运行的各个环节，使网络在自学习、自调整、自愈合方面具备智能化特征，从而提高网络的自动化管理能力。

在 AI for Network 的驱动下，通信网络也需要实现对 AI 应用的优化支撑。尤其是大模型与数字孪生技术的应用，需要网络具备高吞吐、低时延和高可靠性，以满足 AI 任务对数据传输与算力调度的严苛需求。因此，通信网络正在逐步构建面向 AI 的智能底座。网络的算力分布架构应当根据不同任务的时延需求、数据处理能力和资源消耗进行合理部署。

4 构建 Mobile AI 商业新范式

4.1 连接+算力双向协同

在 5G 时代，电信运营商的主营模式以流量经营为主，而云计算等算力业务多由互联网厂商主导，通信资源与计算资源长期处于割裂状态。面向 6G，AI 终端和智能应用的兴起，传统单一的流量经营模式已难以满足日益增长的智能业务需求，必须实现连接与算力的深度协同融合。带宽、可靠性、推理时延与准确率等指标将直接决定用户体验。运营商需要基于不同场景的业务特性灵活编排、提供定制化的网络能力服务，从而挖掘连接流量更深层次的价值。算力和网络的融合促使新的商业模式的发展，利用通信资源（Byte）与计算资源（Computing）的统筹管理与协同调度，构建“B（Byte）+C（Computing）双轮驱动”的新范式。新型商业模式需要流量、算力、服务三者深度融合，实现统一感知、动态编排、按需分配。

在智能眼镜、具身机器人等新型 AI 业务场景中，AI 推理和实时环境感知已成为核心需求。运营商通过提供低时延、弹性可调度的算力，并保障大带宽、广覆盖的基础连接，能够以统一的“B+C”服务形态，赋能新型 AI 业务，实现“通信即服务”与“算力即服务”的双向协同。此外，即便是传统业务（如直播、电商等），也可以在原有流量服务的基础上叠加本地 AI 推理能力（如实时画质增强、人像美化、内容分析等），进一步拓展运营商的服务边界，提升用户的整体体验。通过通算一体化的资源经营，运营商不再依赖单一的流量计费模式，而是实现以连接、算力和服务三位一体为核心的新型商业模式。

4.2 体验变现

当前，流量规模本身已不再是唯一的增长点，用户对智能体验、响应速度和 AI 服务质量的敏感度显著提升。运营商需要从“以量变现”转向“以体验变现”，通过对智能服务质量分级定价、差异化提供 AI 能力，开辟新的价值增长曲线。基于 B+C 协同机制，可探索以下体验变现路径：1）面向终端的智能增强服务：针对智能眼镜、具身机器人等新兴终端，可按照推理响应时延、识别精度、任务

完成度等维度进行分层计费，形成“AI 加速服务套餐”，实现按需付费与高品质体验溢价。2）基于 AI 的实时内容处理：在直播、电商等传统高流量业务中，引入 AI 能力（如内容增强、实时翻译、人像分割等），按调用次数或增强时长计费，形成新增收入来源。3）事件驱动型智能计费：面向机器人导航、远程控制等业务，按照导航规划次数、动态路径更新频率、任务完成可靠性等进行智能计费，将智能能力作为体验价值的一部分，拓展传统连接套餐的边界。通过体验变现，不仅提升了用户的感知价值，也突破了传统按流量或带宽简单收费的模式，实现了从资源经营向体验经营、从单向连接向双向智能服务的根本转型。未来，运营商可以通过体验分级与服务定制，差异化满足多元化用户需求，增强整体业务黏性与品牌竞争力。

5 未来展望

5.1 流量流向边缘，网络重心由中心向边迁移

面向未来，业务生成与响应的重心正加速向网络边缘迁移。智能终端产生的数据更趋本地化，处理需求更加即时化，这推动流量在无线接入网侧聚集，使边缘成为网络价值的新承载点。流量重心的转移不仅是物理传输路径的变化，更代表着网络服务逻辑的重构。边缘不再只是数据的“中转站”，而要成为用户体验的“生成点”。展望 6G 时代，运营商需重塑流量感知与控制策略，推动网络流量组织从“云为中心”向“边云协同”演进，实现边缘功能的跃升。随着内容消费、AI 推理、多终端交互等业务向边缘侧聚拢，传统核心网的负载压力有望得到缓解。边缘不再仅是网络连接的边界，也是智能服务的锚点，成为承载新型业务体验的战略前沿。

5.1 从管道到平台，运营商向 TechCo 转型需掌控生态主导权

运营商要从传统的“连接提供者”向“平台赋能者”转型，不能仅依赖网络基础设施来维持商业价值的持续增长。未来的运营商将不再局限于提供通信连接和传输能力，而是要向科技公司（TechCo）演进，主动构建和主导以智能为核心的业务生态。与传统以资费和容量为主导的运营逻辑相比，TechCo 模式更注重技术开放、服务接口标准化、能力商品化和生态协同机制的建设。这意味着运营商需从纵向一体化的运营模式，转变为横向聚合的生态协同模式。通过开放 AI 模型平台、提供算力调度接口、共建边云服务平台等措施，运营商可以吸引终端厂商、AI 开发者、垂直行业用户等多方角色加入生态圈，构建“网络+算力+AI 能力”的多边平台型价值体系。运营商需要具备算网一体的资源管理、业务编排和价值分配协调等核心能力，真正实现从网络中心向生态中心的转变，成为驱动 AI 业务落地的关键使能者。

术语表

缩写	全称	说明
3GPP	3rd Generation Partnership Project	第三代合作伙伴计划
5G	the Fifth Generation	第五代移动通信技术
AGV	Automated Guided Vehicle	自动引导车辆
AI	Artificial Intelligence	人工智能
AIoT	Artificial Intelligence Internet of Things	人工智能物联网
API	Application Programming Interface	应用程序编程接口
CA	Carrier Aggregation	载波聚合
ELAA	Extremely Large Antenna Array	超大规模天线阵列
ETSI	European Telecommunications Standards Institute	欧洲电信标准协会
FDD	Frequency Division Duplex	频分双工
FSA	Frequency Spectrum Aggregation	频谱聚合
GTI	Global TD-LTE Initiative	TD-LTE 全球发展倡议
MBSC	Multi-Cell Bandwidth Sharing and Cooperation	多小区频谱合作
MoE	Mixture of Experts	混合专家模型
QoE	Quality of Experience	体验质量
RAN	Radio Access Network	无线接入网络
RDTs	RAN Digital Twin System	无线数字孪生系统
RFID	Radio Frequency Identification	射频识别
SDIF	Signal Direct Injection Feeding	信号注入直馈技术
SLA	Service Level Agreement	服务水平协议
SUL	Supplementary Uplink	补充上行
TDD	Time Division Duplex	时分双工
TechCo	Technology Company	科技公司
TM Forum	TeleManagement Forum	电信管理论坛
XR	Extended Reality	扩展现实

参考文献

- [1] Xu M, Du H, Niyato D, et al. Unleashing the power of edge-cloud generative AI in mobile networks: A survey of AIGC services[J]. IEEE Communications Surveys & Tutorials, 2024, 26(2): 1127-1170.
- [2] Xi Z, Chen W, Guo X, et al. The rise and potential of large language model based agents: A survey[J]. Science China Information Sciences, 2025, 68(2): 121101.
- [3] Duan J, Yu S, Tan H L, et al. A survey of embodied ai: From simulators to research tasks[J]. IEEE Transactions on Emerging Topics in Computational Intelligence, 2022, 6(2): 230-244.
- [4] Zhu Y, Guo H, Lau V K N. Bayesian channel estimation in multi-user massive MIMO with extremely large antenna array[J]. IEEE Transactions on Signal Processing, 2021, 69: 5463-5478.
- [5] Khan L U, Han Z, Saad W, et al. Digital twin of wireless systems: Overview, taxonomy, challenges, and opportunities[J]. IEEE Communications Surveys & Tutorials, 2022, 24(4): 2230-2254.



如有咨询或合作意向，欢迎联系：
contact_wistar@chinatelecom.cn