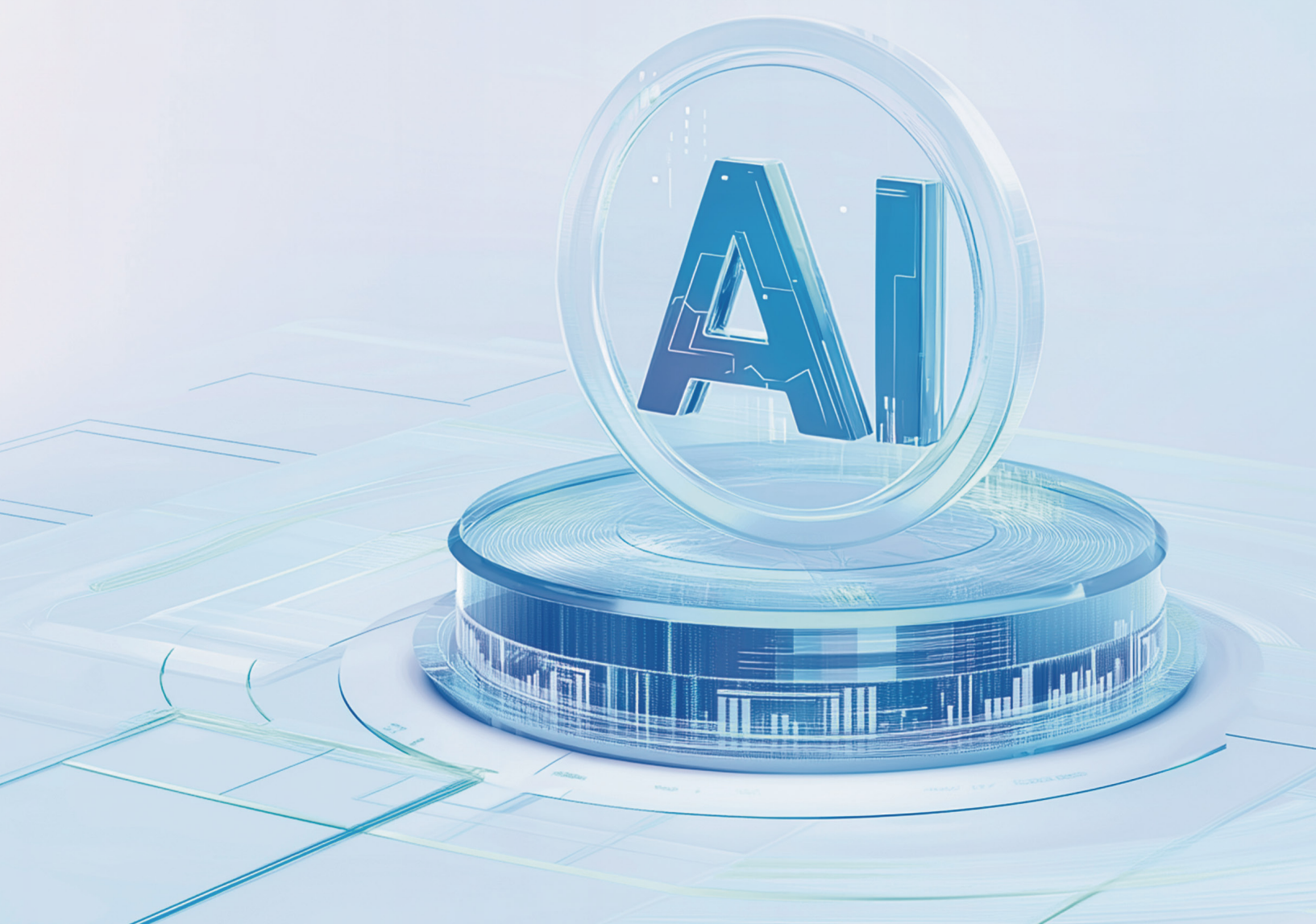
ZTE中兴

中兴通讯

AIR RAN 白皮书

AI RAN先锋助力全场景业务拓展与创新



目录

1 AI关键技术蓬勃发展，商业落地加速 02

2 AI与移动网络加速融合 03
2.1 无线网络加速拥抱AI 03
2.2 AI新业务带来对移动网络的变革性需求 03
2.3 电信运营商承载边缘AI具有天然优势 06

3 AI与RAN融合，在挑战中孕育新机 07
3.1 AI RAN带来的全新挑战 07
3.2 运营商加速布局，积极抢滩AI for RAN红利 08
3.3 AI RAN商业价值体系发展趋势 09

4 AIR RAN基于通算智融合的前瞻性布局，引领行业发展 10
4.1 ToC网络应用 10
4.1.1 优体验，提升连接价值 10
4.1.2 优能效，降碳排促发展 11
4.1.3 优维效，网优提质降本 13
4.2 ToB行业应用 14
4.2.1 智能工厂升级 14
4.2.2 多流视频传输优化 14
4.3 站点全域数智化 15
4.3.1 天线数智化 15
4.3.2 传输数智化 15
4.3.3 电源数智化 16

5 AI与RAN深度融合，持续拓展AIR RAN生态 17
5.1 AI for RAN：分层注智，全面提升无线网络性能 18
5.1.1 底层以高性能驱动 18
5.1.2 高层以多任务驱动 19
5.1.3 顶层以强智能驱动 19
5.2 RAN for AI：连接、感知与计算的全要素供给 20
5.2.1 连接服务助力具身智能云边协同 20
5.2.2 低空、海域、NTN全场景感知赋能 21
5.2.3 连接感知计算协同赋能自动驾驶 21
5.3 群智协同：泛在Agent服务 22

6 未来展望：引领无线网络的长期价值 23
6.1 AIR RAN愿景 23
6.2 AIR RAN路径 24
6.3 AIR RAN生态 24

7 缩略语 26

参考文献 27

序言

在信息科技日新月异的当下，人工智能（AI）作为第四次工业革命的核心驱动力，正以前所未有的深度和广度重塑全球产业格局。从医疗领域的智能诊断与精准治疗，到金融行业的智能风控和高效交易，AI的创新与应用已成为各行业转型升级的关键力量，深刻改变着社会经济的运行脉络。

在这一变革浪潮中，人工智能的广泛应用不断推动无线接入网络（RAN）与AI的深度融合。一方面AI成为移动网络迈向高阶智能的核心引擎（AI for RAN），另一方面网络带宽、速率、连接数等连接能力的大幅提升以及通信、感知、计算的融合为泛在AI应用提供了更广阔的舞台（RAN for AI）。

中兴通讯推出了AIR RAN（AI Reshaped RAN）解决方案，持续深耕AI for RAN和RANforAI领域，不断探索智能化网络的无限可能。具体来看，在体验增强方面，差异化业务保障为流量经营向体验经营的转型奠定了基础；在运维提效方面，利用智能算法实现故障的精准预测与快速修复，大幅提升网络的可靠性和稳定性；在能效提升方面，通过实时分析网络流量和用户行为，动态调整基站功率，有效降低能耗，推动通信行业绿色可持续发展；在拓展AI新应用方面，更是催生出如智能体、具身智能、车联网等方向，满足用户日益增长的多样化、个性化需求，使移动网络从单纯的数据传输管道转变为智能服务的赋能者。

该白皮书紧扣AI与无线接入网（RAN）融合发展脉络，深度剖析其中技术挑战与机遇，并从融合趋势、关键技术及商业探索等方面展开阐述。旨在为从业人员提供专业洞察，把握先机，共同推动行业迈向智能新未来。



01 AI关键技术蓬勃发展 商业落地加速

人工智能近年来取得了巨大的进展，尤其是在机器学习、自然语言处理和计算机视觉等关键技术领域。这些创新正在迅速推动人工智能从研究实验室走向实际应用，其商业化进程在各个行业中都在加速。特别是大语言模型如GPT-o1，处于这一变革的前沿，彻底改变了AI与语言、知识和决策的互动方式。大模型不仅为人工智能技术的普及奠定了基础，还在多个行业释放了前所未有的价值，包括医疗、金融、制造、娱乐等领域。

大语言模型技术的发展为这一变革打下了根本的基础，模型的迅速进步极大地提升了机器理解和生成类似人类语言的能力，彻底改变了从客户服务到创意产业等各个领域。最新的大模型有着数千亿乃至万亿参数，这些模型展现了惊人的语境理解力、语气感知和意图识别能力，使得基于AI的应用更加高效、直观和友好。AI正在多个领域被视为价值创造的驱动力。在医疗领域，人工智能已经取得了显著进展，帮助提供更准确的诊断、预测分析和个性化治疗。据麦肯锡预测，到2026年，人工智能将在医疗保健领域每年带来超过1500亿美元的价值[1]。在金融领域，AI驱动的工具正在简化风险管理、欺诈检测和算法交易，提高效率和盈利能力。人工智能在制造、零售和物流等行业的影响力也不可忽视，AI驱动的自动化帮助降低了运营成本、提升了客户体验，并增加了生产力。

通信网络也同样受益于人工智能，系统厂商和运营商正在共同推动利用AI增收和降低成本。通过在预测性维护、网络优化和客户服务等领域应用AI，电信公司能够显著提高效率。AI帮助移动运营商预测网络拥堵、优化带宽分配，甚至通过预测和解决技术问题来减少故障。根据数据显示，2021年全球电信行业的AI市场规模为12亿美元，预计到2026年将增长至超过63亿美元，复合年均增长率达到38%[2]。AI驱动的聊天机器人和虚拟助手也提升了客户服务质量，不仅提高了用户满意度，还降低了人力成本。此外，基于AI的数据分析使移动运营商能够提供个性化服务，挖掘新的收入来源，并增强客户忠诚度。

另一方面，AI应用的进一步发展、更大价值的创造和普惠，必将对于通信网网络特别是移动网络提出更多更高的需求，使得人们可以在各种场景下都能够访问AI、利用AI，重塑用户习惯，转变人们的生活和生产方式，而AI的进一步普及和多种多样的AI应用发展，也将对于网络的带宽、时延和可靠性等产生更多需求，催生云端、边缘和终端的多样化协同，促进网络技术的创新和发展。



02 AI与移动网络加速融合

2.1 无线网络加速拥抱AI

GSMA的《2024年移动经济》报告[3]显示，到2030年5G连接比例预计将达到56%，移动流量年复合增长率（CAGR）达到23%，意味着移动运营商仍然需要保持持续的5G建设投资以应对网络流量的变化。大规模部署5G MIMO多天线技术将带来巨额投资成本，预计到2030年，总资本投资将达到1.5万亿美元。然而，运营商的投资回报率预计将从2022年的19%下降至2030年的14%。为了降低成本并提升效率，运营商亟需加速网络及服务的自动化。一方面实现网络运维优化AI技术下沉，利用人工智能技术优化网络运维，提升网络质量，提供精准的优化洞察，智能识别和预防故障隐患。另一方面，通过智能化结合业务识别等手段，实现精细化节能，以及关键业务的流量激发和5G-A新套餐等差异化体验保障。

从技术发展角度来看，通信、云计算、人工智能、数字孪生等技术的融合，已经成为下一代网络演进的主要趋势。随着5G商用进入深水区，基于基站能力拓展算网一体化服务也是运营商关注的重点场景。一是面向消费者（2C）业务，通过基站算力池化共享组网可满足极低时延，以及大带宽应用的本地计算需求，例如XR本地业务渲染、具身机器人以及沉浸式交互应用等。二是面向企业（2B）业务，通过本地化通信计算一体化服务，实现业务灵活敏捷部署，同时有效降低行业应用的综合成本。典型应用场景包括工业视觉AI检测和车联网中的智能碰撞预测等。

2.2 AI新业务带来对移动网络的变革性需求

AI新业务的快速发展正在推动移动网络的全面升级。从交互对象的演变来看，人与人的互动逐渐扩展为人与AI代理、AI代理之间，以及AI代理与其他设备的交互；从交互内容来看，多模态融合以及与现实世界的深度交互催生了更加丰富和高质量的GenAI内容形式。这些变化对网络性能和架构提出了全新的要求，推动移动网络从传统的连接服务向高性能网算结合的智能服务转型。

变革1：GenAI拉动内容数量，视频画质，用户参与度三方面爆发式上升，激发网络流量增长

AI在内容生成领域的应用正快速扩大，AI生成视频的市场规模2024年到2030年，年复合增长率达到19.9%[4]，并且随着AI算法的进步，生成高分辨率视频所需的计算效率显著提高。例如，GANs（生成对抗网络）等技术已经可以高效生成高分辨率的视频内容。当前，AI生成的大多数短视频已实现2K分辨率，而StyleGAN3等先进模型在静态图像生成上已突破8K分辨率，视频生成的分辨率也正向这一目标迈进。

另外，GenAI重塑沉浸式扩展现实（XR）和3D内容生成，打破内容生成效率低，价格昂贵的生态瓶颈，迎来实质性增长。沉浸式内容需要超高速的数据传输支持，例如，XR应用单用户8K分辨率需要至少下行200Mbps的带宽，而未来6G时代对单用户带宽的需求可能超过20Gbps。

AI驱动的超级个性化和更加沉浸的体验，极大提升了用户参与度。Spotify 使用AI为用户提供个性化推荐，高级订阅注册量增加20%；Snapchat推出集成GenAI的全新AR，实时生成AR体验，使用AR滤镜的每日活跃用户增加40%。

根据GSMA的预测，移动网络流量到2030年将增长4倍多，而AI带来的颠覆性变革，极有可能打破既有的流量增长路径，带来更多的变化。

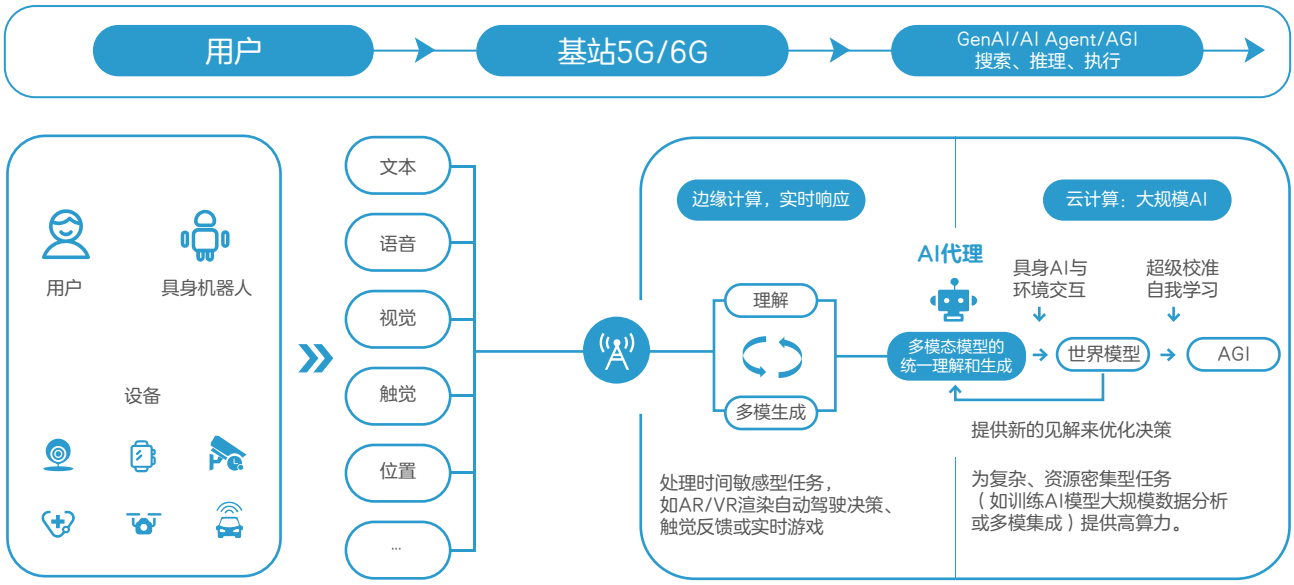
变革2：AI Agent将成为普遍性应用，需要同时满足高带宽和低时延要求，并带来多向数据流模式变化

AI代理从个人助手到企业智能客服，再到工业自动化中的决策引擎，这些场景中的实时交互对网络性能提出了严苛要求。网络数据流模式从传统的不区分业务类型的用户数据到业务服务器单一路径传输，升级为用户到边缘服务器，AI Agent，云服务器，以及各节点之间的多次交互的多路径传输。常见的AI Agent实时交互需要同时支持高速的上下行数据传输：

智能体类型	吞吐量	延迟	应用场景
个人助手	10-20Mbps（上行）	<50ms	智能音响，移动助手
企业智能客服	20-50Mbps（上行）	<20ms	客户服务、业务支持
工业自动化	<100Mbps（上下行）	<10ms	无人车间、机器协作、远程维护
复杂多模态智能体	20Mbps-100Mbps (上行) 100Mbps-500Mbps (下行)	多模态感知：<10ms 音视频流媒体：<200ms	全息通信，虚拟人，具身机器人

变革3：云边端算力融合的协作模式是AI业务发展的探索方向

AI业务的多样化和复杂性使得传统单一的计算架构难以为继。从实时性到算力规模，再到数据隐私和协作效率，AI业务对计算资源的需求呈现出多维度的增长。



云主导型协作：适用于计算量大但对延迟要求相对较低的场景，云端负责完成大部分计算任务，边缘设备承担任务分发和数据缓存，终端则主要用于轻量化的计算。例如AI模型的复杂训练，云端提供超大规模算力，而边缘负责处理用户请求的本地化响应。

边缘主导型协作：适合对延迟要求极高的业务，并可节省大量的有线传输资源。例如自动驾驶和具身机器人。在这种模式中，实时任务的处理主要由边缘设备完成，而云端则负责长期的数据存储和模型优化。例如，通过边缘设备与智能汽车间数据，算力，模型的深度实时协同，完成路况识别，多车联合监测危险事件，完成边缘判决等，而云端则对全局路径进行优化。

端智能协作：终端设备承担简单的计算任务，边缘和云端则协同完成更为复杂的计算。例如，在AR/VR场景中，AR眼镜可以完成简单的本地渲染，而复杂的环境建模则交由边缘计算完成，云端则生成全局环境模型。

2.3 电信运营商承载边缘AI具有天然优势

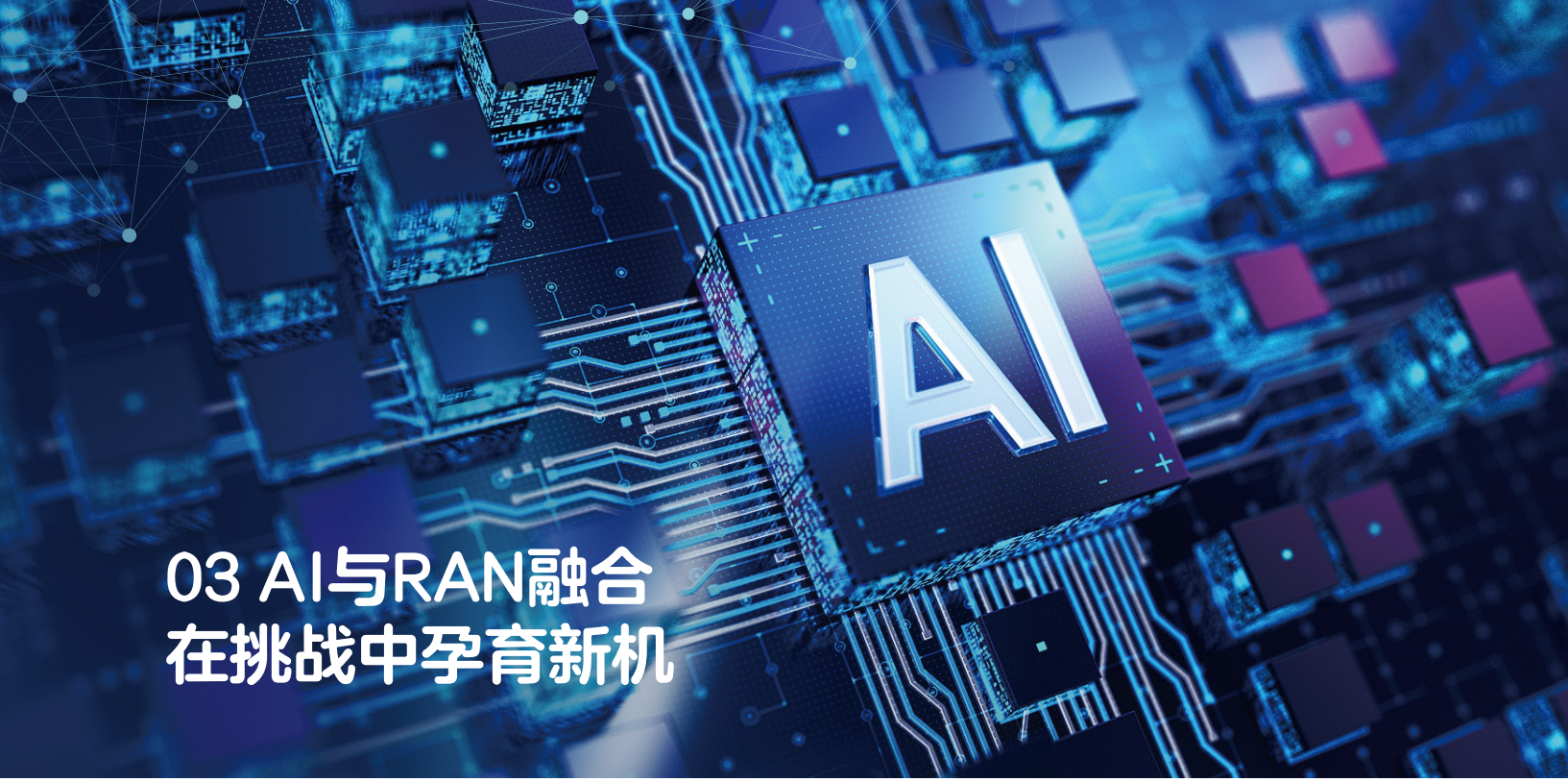
电信运营商的移动网络天生是一个广域的分布式网络，这与移动业务、AI业务的分布有着天然的匹配性。借助网络边缘的优势，运营商也早早开始边缘算力的探索，MEC就是一个典型的代表。运营商通过把MEC下沉化、边缘化，可以实现业务的本地化部署，从而有效地降低业务时延、带宽开销和终端成本，提升业务体验和数据安全。中兴通讯在边缘算力的探索上，一直走在业界的前列，2020年推出了NodeEngine基站级算力引擎，通过在基站上新增一块单板即可提供边缘算力，实现本地分流和边缘应用的灵活部署；2023年，中兴通讯推出业界最简的UniEngine算网一体机，在提供算力的同时实现了5G专网网元的全集成。2024年中兴推出AIREngine智算引擎，通过在基站中增加智算单板实现基站内生AI。

相比于云端远端算力，运营商的边缘算力在如下几类场景中更具优势：

局域场景：典型的如各行业专网。企业对数据的敏感性和安全性有较高的要求，数据不出园是通用的需求，运营商在构建专网的同时，同时提供算力，算网一体化的供给，便于企业本地应用包括AI应用的园区级部署。再深入到工控领域，传统的ISA-95正在向云边端架构演进，随着机器视觉质检、设备预测性维护等场景的引入，工业现场侧存在大量的实时性的控制指令和非实时数据需要处理，边缘控制的概念应运而生，边缘控制是集工业设备边缘、运营商网络边缘和计算边缘于一体的新型控制，运营商可以基于通算融合基础设施为工业现场提供IT/CT/OT合一的设备。例如，中兴通讯的NodeEngine算力基站/UniEngine算网一体机在物流自动分拣场景中，通过内置算力大幅降低了现场工控机数量，实现了PLC控制的集中化和云化。

广域场景：实时性要求比较高的业务场景。通常服务于人的业务，相对来说对时延不敏感，业务云端处理的收益更大。而服务于控制类的物联网，对时延往往十分敏感。典型的如车路云一体化场景，需要对路侧的感知（包括摄像头、雷达、红绿灯信息）等进行感知分析，再通过V2X云控下发信息给车辆提供驾驶辅助信息，该场景对时延敏感，将算力下沉，在基站侧完成路侧感知分析和决策，将有利于保障驾驶安全。类似的，随着大模型在沉浸式设备、具身智能等场景的应用和普及，将渲染和控制大脑边缘化部署，降低时延可带来更好的实时体验。





03 AI与RAN融合在挑战中孕育新机

AI与RAN的融合在为通信行业带来前所未有的机遇的同时，也对RAN的数据采集、算法设计及算力利用带来相应的挑战。

3.1 AI RAN带来的全新挑战

在数据精细化方面，AI RAN需要处理多样化的数据种类，如车联网、XR等，这些数据的采集周期必须实时化，以满足AI业务对数据实时性的要求，而这与传统网络中数据种类固定、采集周期较长的需求形成鲜明对比。以车联网为例，鬼探头等避让决策需要在车辆高速移动和时延分布不均的环境下进行实时判断，对数据采集的实时性提出了更高的要求。

在算法匹配度方面，传统模型设计强调可解释性和易于推理实现，但参数配置复杂且人力消耗大。AI模型设计虽然在可解释性上较差，但其参数配置简单，能够获得统计特性下的全局最优解，其不足则是AI模型在应用场景变化时可能需要重新训练，这会导致性能下降和额外的开销。例如，在自动驾驶领域，AI模型需要根据不同的道路条件和交通状况进行动态调整，以确保安全性和可靠性。

算力使用效率则是AI RAN面临的另一个重要挑战。AI终端形态多样，从XR眼镜到汽车再到具身智能设备，其算力需求从Nx10TOPS到Nx100TOPS不等；话务的潮汐分布特性对网络的动态资源分配提出了更高要求，如住宅区和CBD在白天与晚上的流量差异需要通过动态资源分配才能带来更高收益。这些变化都对基站设备的算力演进和算力分配提出了新的需求，基站算力将会从专用设备向异构算力方向演进来满足差异化的算力需求和场景应用，例如，在专用硬件上扩展通用算力板，通过智能调度算法优化资源分配，实现高峰时段和正常时段资源的高效利用。

AI与RAN的深度融合正引领通信行业迈向智能化与高效化。通过持续优化数据处理、算法设计和算力利用，电信运营商正从AI赋能无线网络向AI与RAN双线协同的方向演进。在这一变革过程中，AI RAN成为电信运营商提升用户体验的重要支点，

带来更便捷高效的服务。未来，随着技术的进一步成熟与应用场景的不断拓展，AI RAN有望成为驱动通信行业变革的核心力量，并催生全新的商业机遇。

3.2 运营商加速布局，积极抢滩AI for RAN红利

随着AI技术的不断发展，移动运营商已通过AI赋能无线网络运营，在用户体验提升、能效和运维效率方面都带来了显著的收益，并将助力运营商拓展新场景。

- 直播、游戏等新业务兴起，用户网络体验需求多样，流量需求涨但ARPU增长慢，运营商面临获客与增收难题。基于AI的端到端套餐经营，精准感知业务、分配资源，提升用户体验和网络资源利用率，助力运营商增收与行业智能化升级。

- 通信行业能耗大，5G普及加剧无线网络能耗问题。AI技术引入节能管理，精细化调控，在保障服务质量的同时提高能源利用率，推动行业绿色发展，如商业区基站的智能节能调控。

- 智能体革新运维领域，改进运维流程、提升系统能力，实现多智能体协同，精准预测故障，大幅提高运维效率，推动运维向高阶智能化转变，业界也在探索RAN 与AI融合的新机遇。

在当前的技术发展趋势下，业界主要围绕着RAN连接和AI应用这两条主线场景展开积极探索，致力于研究RAN和AI的新融合方式以及新架构的构建，对于未来的AI与RAN深度融合的场景，主要分为以RAN连接为主的场景，和AI应用为主的场景。

- 在以RAN连接为主的场景中，业界对AI在RAN提效的研究方向正逐渐发生聚焦和转变。研究方向从无线应用层L3、无线链路层L2聚焦到物理层L1。IMT-2030（6G）推进组[5]针对AI技术在6G无线物理层、无线链路层、应用层的应用已经展开了分析和研究，尤其是深度学习在物理层的优化能力和价值挖潜。与此同时，关于在L1层是采用模块化多个AI小模型，还是端到端单一模型的技术路线选择，将会对RAN产品架构产生趋势性的影响。

- 在AI应用场景中，云端协同已成必然趋势，业界也在积极探讨边侧AI的应用价值与云边端协同的必要性。无线运营商凭借广泛分布的无线网络优势，在边缘侧高效部署算力，与端侧AI业务实现数据、模型和计算的深度协同，能满足数据隐私要求、低时延高效响应、实时获取数据且低成本快速部署。不过，对于必须云边端协同的AI应用，如具身机器人、沉浸式多模态交互应用、自动驾驶等领域，业界仍在细分场景挖掘确定性价值。而在以AI应用为主、兼顾RAN连接部署的场景中，是否采用软银主导的通用架构[6]，以及无线运营商改造广泛部署的无线网络基础设施能否形成确定性投资回报，都有待深入研究。

3.3 AI RAN商业价值体系发展趋势

人工智能正在引领一场深刻的技术变革，移动网络在这场变革中不仅是基础设施的提供者，更是通过开放自身的算力和AI能力，成为创新价值的积极赋能者。随着5G和6G技术的发展，移动网络将其强大的算力资源和AI处理能力开放给各类应用，催生全新的商业模式和服务场景，推动数字经济的蓬勃发展。

在多个领域，移动网络不仅提供高速连接，更可以将自身的边缘计算和AI推理能力开放给行业应用，帮助企业降低部署门槛，加速数字化转型。通过网络侧的算力支持，这些行业可以更经济高效地实现智能化升级。

对于机器人应用，移动网络可以开放自身的AI视觉分析、实时决策等能力，协助机器人完成复杂的识别和规划任务。网络的边缘节点可以为机器人提供近场算力支持，既确保极低时延，又降低终端成本。泛在的网络覆盖则让机器人可以随时接入这些能力，在任何场景下实现智能化操作。

在XR领域，移动网络不仅提供高带宽传输，还可以开放云渲染、空间计算等能力，帮助XR设备实现更逼真的虚拟体验。通过网络侧的算力支持，可以大幅降低XR终端的硬件要求，让更轻便的XR设备也能呈现高质量内容。

对于自动驾驶，移动网络可以将自身的AI感知、决策能力开放给车辆及交通管理部门，协助完成复杂环境下的实时分析。网络的边缘节点可以为车辆提供补充算力，增强自动驾驶的安全性和可靠性。泛在网络则确保车辆随时都能获取这些能力支持。

移动网络通过开放自身的算力和AI能力，正在从单纯的连接提供者升级为创新使能者。这种转变不仅满足了新兴应用对算力、时延和泛在性的需求，更重要的是通过能力开放激发了更多创新，创造了全新的商业价值。随着技术的持续演进，移动网络将与AI深度融合，携手推动各行各业迈向更智能、更高效的未来。





04 AIR RAN基于通算智融合的前瞻性布局引领行业发展

人工智能（AI）与无线接入网（RAN）的深度融合，不仅推动了技术的创新与变革，还催生了全新的商业模式和市场机遇，然而，这一过程中也不可避免地面临诸多挑战。为此，中兴通讯率先在AI for RAN和RAN for AI两大领域进行技术探索，推出AIR RAN（AI Reshaped RAN）解决方案，构建面向多场景的通算智融合产品体系，打造全域数智化底座平台，在RAN领域的智能化价值提升和ToB产业通算一体数智化升级等方向，持续推动技术创新与实践，引领行业发展。

4.1 ToC网络应用

在移动互联网流量增速放缓、ARPU值下降以及套餐同质化加剧的背景下，运营商面临着提升竞争力的挑战。中兴通讯AIREngine产品通过在5G-A中引入AI技术，为行业提供了全新的破局之道，不仅优化用户体验，还推动绿色低碳发展和运维效率提升。通过智能调度和能耗优化，AI有效降低无线接入网设备能耗，助力实现碳中和目标；在复杂环境中的运维中，AI提升了高铁和海域等场景的路测效率，减少了时间和人工成本；同时，AI赋能网络故障溯源和安全分析，使数据处理更高效，响应速度更快，从而保障网络稳定性与可靠性。AI技术的深度应用为通信行业注入创新动能，开辟了绿色、高效与差异化发展的新路径。

4.1.1 优体验，提升连接价值

提供卓越体验是增强用户粘性和市场竞争力的关键。在激烈的通信市场竞争中，用户已不再满足于基础服务，而是更关注附加价值和优质体验。通过AI技术，运营商可智能识别用户场景并动态优化策略，确保热门权益（如视频或游戏平台会员）享有流畅体验，从而提升用户满意度和忠诚度。与此同时，运营商需从低价竞争转向价值定价，通过AI赋能专属权益和智能服务，明确费用与价值的匹配，既稳定客户群体，又提升市场份额，成为未来通信服务竞争的核心方向。

业务差异化：网络可根据不同业务的特性和诉求进行差异化的保障，以获得与业务匹配的服务质量。针对视频业务，可优先保障带宽、优化协议和调度策略，减少卡顿、提升观看体验；游戏类业务对时延敏感，侧重降低延迟、增强连接稳定性；通话类业务则需动态调整传输路径确保流畅互动；而对于直播类业务，上行速率需求高，需重点保障上行大带宽。网络通过智能化识别典型业务、精准评估业务的关键质量指标（KQI），在匹配网络能力的前提下，平衡多重业务体验，并确保显著提升业务质量。与此同时，网络侧可从多个维度评估业务体验，助力网络指标从关键性能指标（KPI）到业务关键质量指标（KQI）的跃升。

用户个性化：运营商根据不同用户对网络差异化的诉求定义了不同的用户分类，也就要求网络能够根据用户的具体分类提供差异化服务。同时，由于网络中同时存在多种类型的用户，其用户分布也是随机的。在网络资源受到限制时，需要智能化地调配网络资源分配，以实现综合保障效果最大化。

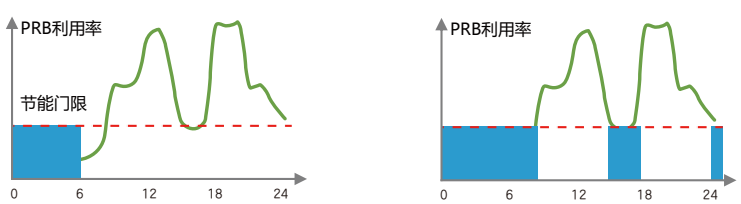
体验连续性：引入“用户为中心”的理念；以用户为中心的小区间深度智能协同，提供稳态用户体验。智能协同方案通过用户移动轨迹动态成簇，洞察用户、业务和小区覆盖，预测用户移动轨迹。根据移动轨迹和业务需求，进行动态聚类成簇，簇内联合调度收发，簇间使用多点协同降低/规避干扰。

服务确定性：针对特定政企用户、新兴业务或特定行业应用场景，网络侧的连接质量（QoS）提出了更高、更精准的要求。从速率、时延、可靠性维度给出明确的保障标准，特别是在部分行业场景下，对确定性的QoS要求尤为突出。网络侧可学习业务特征、启用与业务特性匹配的保障策略，通过保障智能体对实时传输的特性进行智能分析，制定和优化决策调控策略。通过计算预测业务量变化、优化资源调度、平衡负载，联动满足优先级用户需求，确保提供高品质服务。

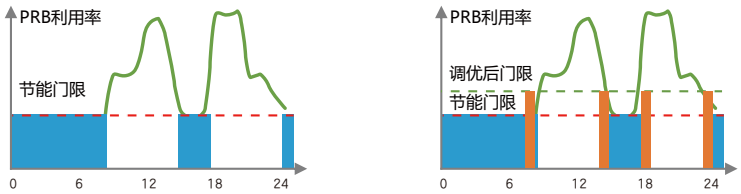
4.1.2 优能效，降碳排促发展

在基站节能过程中，面临的主要问题包括如何在保证网络性能的前提下，最大化降低能耗。传统方法往往依赖于固定的节能策略，难以适应复杂多变的网络环境。AI技术通过深度学习和大数据分析，能够实现场景识别和动态调整节能策略，从而有效解决这些问题。具体而言，AI技术可以实现以下几点：

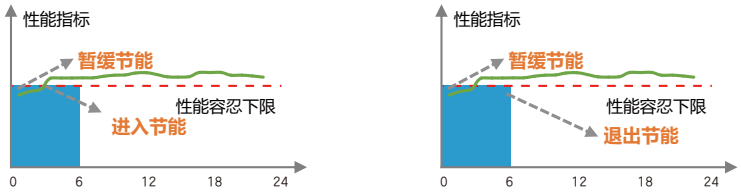
负荷精准预测：算法准确度>95%



参数迭代寻优：寻找节能与性能拐点



感知快评保障：近实时保障用户感知



场景识别与负荷预测：不同应用场景的小区话务分布呈现出明显的规律性。例如，高校和写字楼的话务趋势呈现周内高周末低的特点，而商场和公园则在周末和晚间出现高峰。通过K-Means算法对时间序列波形进行聚类分析，可以准确识别不同场景的负荷模型。这不仅有助于预测未来的负荷变化，还能为制定个性化的节能策略提供依据。

动态节能策略生成：基于小区业务变化趋势的预测结果，AI技术可以生成最优的节能策略。例如，在低负荷时段自动关闭部分基站或降低发射功率，而在高负荷时段则保持正常运行。此外，AI还可以区分容量层与覆盖层，筛选共覆盖的小区作为补偿小区，用于分摊业务负荷。这种动态调整不仅减少了不必要的能耗，还确保了网络性能不受影响。

实时评估与自优化：AI技术通过实时监测网络KPI（关键性能指标），可以及时评估节能效果，并根据实际反馈进行秒级自优化。例如，如果某个节能策略导致网络性能下降，AI系统会自动调整策略，以达到能耗与性能的最佳平衡点。这种闭环优化机制使得节能策略更加灵活和高效。

多维度能效评估：传统的能效评估主要关注单一的流量指标，但随着新业务和新场景的不断涌现，单一指标已无法全面反映网络能耗和效率。AI技术通过构建多维能效模型，从场景化、归一化、多维度等角度持续扩展能效评估体系。这不仅有助于更全面地获取网络能效信息，还能指导后续的节能策略选择。

通过在5G-A中引入AI技术，运营商可以在保证网络性能的前提下，实现极致节能的目标，预计全天候开启网络降耗下降15~25%，这不仅有助于降低运营成本，还能为实现绿色低碳目标做出重要贡献。

4.1.3 优维效，网优提质降本

在网络维效过程中，主要面临故障分析数据不足、位置定位不准、现场运维技术要求高等问题，导致传统维效的准确率和效率均较低。AI技术借助深度学习和大数据分析，可实现场景识别和动态调整节能策略，有效应对这些难题：

故障助手：为了最大限度降低一线现场运维人员的学习成本和技能要求，提升故障排障的处理效率，其以自然语言为交互方式，使用大小模型结合完成故障处理相关的能力，提供知识问答，动态数据查询和引导式排障，故障看板等特性。

支持从标准协议，运营商规范，直至设备厂家产品型号&特性，故障现象的识别和处理建议，以及相关案例知识的回答；

提供以自然语言提问的方式对在线设备硬件&链路&时钟等模块进行诊断的能力，从而获取最新的设备状态和运行数据来辅助维护人员进行综合判断和故障定位；

提供了排障流程的系统级指引，运维人员通过和故障助手以自然语言方式的多轮对话，抽丝剥茧得到故障根因；

网优专家：以多目标联合寻优和数字孪生为关键技术的网优专家大模型，自动关联告警隐患、操作日志等信息，对小区级、栅格级性能感知质差小区进行综合分析，智能输出覆盖、干扰、容量等问题准确根源和参数自闭环优化建议，并通过孪生仿真，预测未来网络需求，提供精准扩容和网络提升方案。

看网专家：看网专家提供人机交互式多维综合看网及讲网能力，通过转换自然语言至查询语句，按照业务类型召集多维智能体专家，进行看网和洞察总结，给出规建优维专业处理建议。看网和讲网能力依赖多维网络洞察任务的执行结果，根据问题现象，结合知识库给出问题解决建议。例如：通过业务感知与空口质量的单域拼接，实现视频时延等指标的精准识别，并支持通过AI的感知模型完成业务质量指标感知；支持时间维度的用户轨迹回放，结合GIS栅格打点呈现，关联主服务小区并区分室内外环境，为运维提供可视化支持。

知识助手：通过大模型加外挂知识库（包含运维帮助文档、网元业务文档）的方式给用户提供了内容丰富、即时更新、准确可靠的知识问答服务。可以通过自然语言问答的方式，查询运维系统中的告警、性能数据以及网元信息数据，并对获取到的数据进行总结，以提供更加人性化的结果展示形式。

保障专家：是面向无线网络保障场景的大模型运维解决方案，应用于演唱会等重大活动、突发应急、业务潮汐等保障场景。通过业务识别技术，能够深度剖析网络流量，精准识别短视频、云游戏、长视频、网页浏览、扫码、即时通信、移动游戏等业务类型，并触发大模型开启用户体验保障，并结合小型模型协作智能生成保障方

案，保障人效投入降低30%。

通过5G和AI的融合，实际测试提升质差分析效率9倍，支持虚拟路测，数据&能力开放的新模式。

4.2 ToB行业应用

随着工业数字化转型的深入，智能工厂面临大规模数据处理、实时监控与控制、人工智能算法支持以及边缘与安全计算等多重挑战。中兴通讯的 NodeEngine 算力基站与 UniEngine 算网一体机，通过通信、算力与 AI 的深度协同，为 ToB 行业带来智能化升级，全面提升生产效率与竞争力。

通过在5GA中引入AI技术，运营商可以在保证网络性能的前提下，实现极致节能的目标，预计全天候开启网络降耗下降15~25%，这不仅有助于降低运营成本，还能为实现绿色低碳目标做出重要贡献。

4.2.1 智能工厂升级

在钢铁行业，无人天车系统通过无线通信和边缘计算实现物料搬运自动化，解决设备老化、人力不足及安全风险等问题。该系统依托实时数据采集、机器视觉技术和AI算法，完成任务调度、路径优化和故障诊断。无人天车系统可以利用人工智能和机器学习的技术，综合各环节各设备的数据，以及物料的重量、形状、位置等参数，计算出合适的吊钩高度、速度和角度，以防止物料在空中摇晃或碰撞；根据生产计划、物料库存、设备状态等信息，动态调整运输任务的优先级和顺序，以避免出现物料堆积或缺乏的情况。

3C制造领域高度依赖自动化与精确的质量管理。机器视觉质检系统通过图像处理与AI算法高效检测产品表面缺陷、装配不良等问题。系统需应对大规模图像数据的高效处理与低延迟响应，通过并行计算与优化算法满足实时性需求。质检数据不仅用于即时反馈，还为生产优化和质量追溯提供依据，帮助企业实现数据驱动的持续改进。

5G通信为智能工厂提供高带宽、低延迟和高可靠性的数据传输能力，而边缘计算通过就近处理任务减少延迟，确保实时性。AI技术进一步提升了系统智能化水平，从数据洞察到设备控制贯穿工业流程。无论是钢铁行业的无人天车，还是3C制造的视觉质检，通信、算力与AI的深度结合为工业场景注入创新动能。

4.2.2 多流视频传输优化

在密集IP摄像头场景下，多流视频并发时可能出现多I帧同时传输的现象（称为I帧碰撞），导致传输带宽需求骤增，远超视频流的典型码率。这种情况会引发网络传输链路拥塞，导致时延突增和视频卡顿，严重影响传输质量和业务稳定性。

为保障视频传输的稳定性和流畅性，可在算网一体机上部署I帧检测AI算法服务。通过预训练模型识别视频流中的I帧特征，实时检测摄像头的I帧碰撞概率，并对碰撞概率高的摄像头进行调整。调整指令通过摄像头管理协议下发至摄像头，从而有效避开I帧碰撞，降低网络带宽需求，确保视频传输质量和链路稳定性。

4.3 站点全域数智化

随着AI和RAN的持续深度融合，站点作为通信网络关键的硬件基础，也面临向数智化演进的重大变革，来推动网络向更高阶自智网络演进。

高阶自智网络需根据用户及业务需求实时调整站点资源，来提升业务感知和运营效率，降低能耗，其硬件基础是站点具备实感知和多维调整能力。从整个站点资源看，核心硬件包含了BBU+RRU+天线三部分，同时还包含了电源、传输、机柜等配套。其中BBU和RRU已经具备数智化基础能力，天线、电源、传输和机柜等仍需要向数智化演进，构筑整个网络智能化底座。

4.3.1 天线数智化

当前广泛使用的传统天线作为无源设备是“哑设备”，感知和调整能力极弱，仅具备垂直倾角调整。相关工参获取和天线调整主要依靠人工方式进行，不仅效率低下，也无法适应业务环境的变化。作为基站和终端用户之间的唯一链接桥梁，天线的数智化演进是必然趋势，关键特征包含实时工参获取、多维波束调整和网络智能感知，来实现大幅降低运维和网优时间，提升用户感知。

实时工参获取：实时精准获取方位角、机械倾角、经纬度等站点工参信息。

多维波束调整：包括水平方向角调整、垂直方向角调整、水平波宽调整、垂直波宽调整，能调整的范围越大，可独立调整的频段越多，越能发挥网络优化效果。

网络智能感知：自动获取天线阵列和RRU通道以及小区的映射关系。

4.3.2 传输数智化

随着CRAN部署比例的不断提高，BBU和RRU之间的长距离拉远变得更加普遍，前传光路也更加复杂。前传光路是无源链路，传统的光模块只能感知链路的光功率大小变化，无法感知整个前传光路的健康状况。当前传光路故障时，只能依靠人工上站逐点排查，费时费力，而且效率低下。前传链路数智化是解决传统问题的必然路径。通过智能光模块，配合智能软件算法，实现前传光路故障自动诊断、根因智能分析，一次上站即可精准解决问题，大大提升运维效率，提高网络可用性。

4.3.3 电源数智化

采用gSDU，电源和主设备数智联动，实现站点绿色低碳。

配电端口名称可以软件定义，自动识别负载设备小区关系；根据负载设备规格型号自动调整断路器过流保护阈值，一种断路器可以适配各种型号的基站设备，实现无差别配电；每个端口都可以远程上下电。

智能感知业务：实现AAU\RRU智能启停，实现基站设备节能感知双优；同时智能预测站点设备功耗，动态调整电源整流器休眠数量，保证整流器转换效率最高。

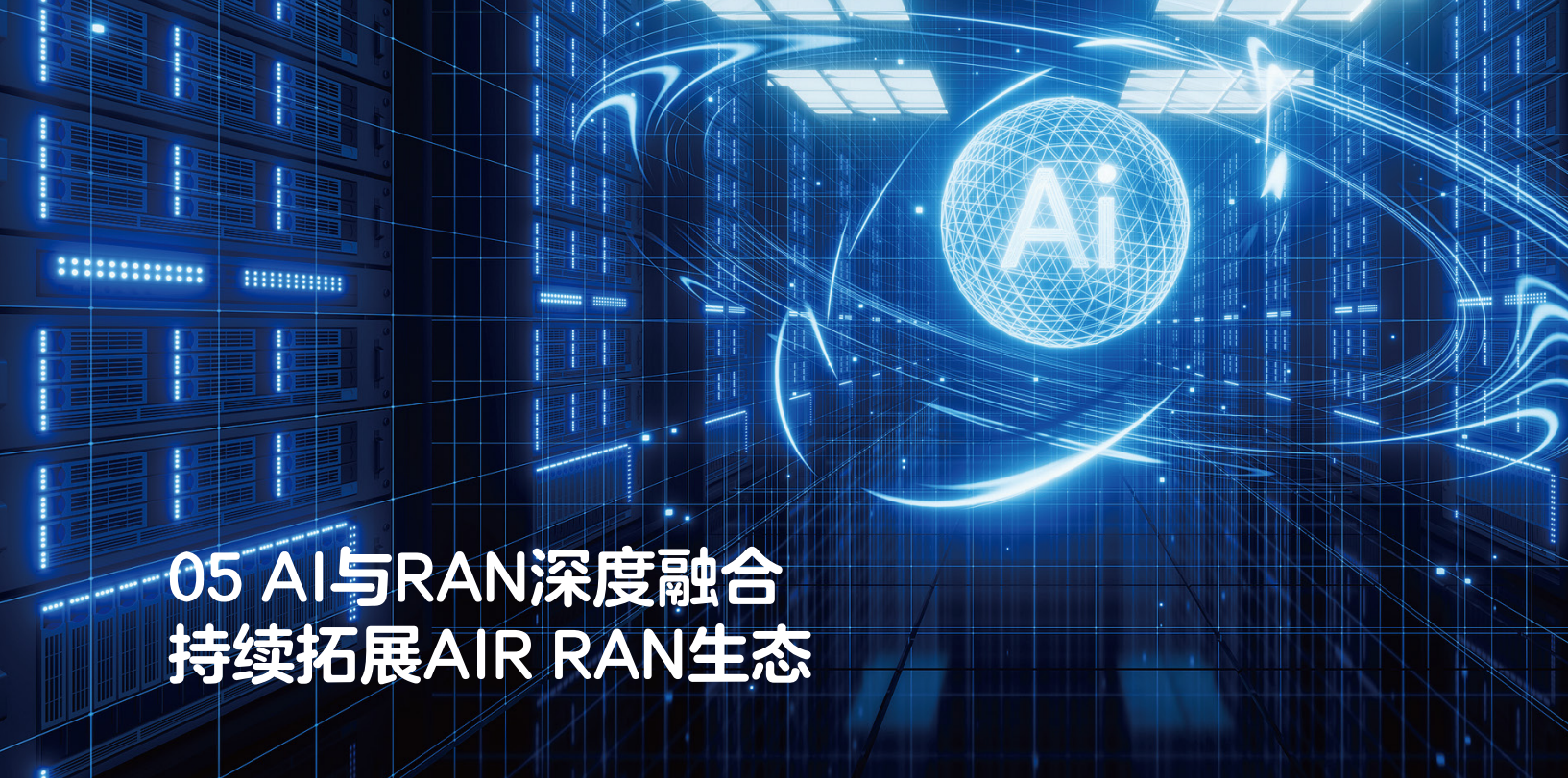
分级备电：备电最小颗粒度为设备级，设备增加，电池容量可以不增加；管理第三方电源通讯，获取电源电压、电流、电池容量和电池剩余容量，智能预测备电时长，智能动态调整下电电压，实现智能备电；

智能绿电：根据天气预报预测太阳能发电量，根据预测数据，智能动态调整基站设备输出功率和锂电池充电电流，尽可能延长站点服务时间。

智能错峰：错峰用电功能支持在电价低谷期为锂电池充电，电价高峰期改由锂电池供电，从而降低基站的整体电费；据历史功耗预测站点功耗需求，计算备电所需容量，从而计算峰值电价时段可以用来放电的电池容量，锂电池放电时长智能动态调整，确保基站安全运行。

采用风液混合近端制冷技术，BBU与机柜温控系统联动，动态跟踪分析柜内能耗变化和气候变化、进行智能节能寻优；机柜在无线网管可视、可感、可控。





05 AI与RAN深度融合 持续拓展AIR RAN生态

如前文分析，AIR RAN架构演进的核心目标是支撑网络的泛在智能，推动网络中智能体之间的智能互联，最终将网络打造为连接物理世界和数字世界的智能中枢。

在产品架构层面，当前业界主要采用两种形式：一种是从RAN专用加速器发展而来的通专异构XPU架构，另一种是基于通用CPU+GPU的架构。两种形式各具特点，异构架构定制化设计，因地制宜地在不适合AI化的RAN模块和部分适合AI加速的模块中灵活选择最为适合的计算架构，在计算资源调度与性能优化方面能够提供更高的运算效率，同时在成本与功耗上展现出明显优势；通用架构的优势在于能够支持第三方AI应用的快速落地，但在实现智能化RAN时，算子需要在不同硬件单元中来回切换才能完成经典RAN模块和部分AI模块的配合，这将造成资源的严重浪费。整体来看，通用架构在当前阶段具备一定的探索意义，但异构架构凭借其在整体方案中成熟度、兼容性、定制化及功耗等方面的突出优势，将成为AIR RAN长期演进更具潜力的方向。

在技术架构层面，不论上述哪种硬件产品形态，都应从以下几个维度考量：

AI for RAN：在RAN协议栈中引入AI为L1/L2/L3精准注智。AI在无线网络性能提升中的应用，包括底层、高层和顶层的优化策略。底层以高性能驱动，利用传统算法和AI算法的优势组合；高层以多任务驱动，通过局部AI任务优化实现多任务协同；顶层以强智能驱动，利用大模型和语言理解能力提升网络操控智能化。

RAN for AI：通过通信与AI技术深度融合，RAN不仅承担传统的连接功能，还通过感知与计算能力扩展，赋能具身智能、车路协同和低空无人机等场景。基站利用通感技术感知环境与动态目标，结合AI实现实时优化，提升资源利用效率与服务质量，为全场景智能化升级提供支持。

群智协同：网络智能体从单一智能体向群智协同演进，通过智能体间的协作分解复杂问题、优化全局决策，展现群智协同优势，显著提升系统的适应性、鲁棒性与创新能力。AI for RAN提供性能支撑，RAN for AI贡献数据与交互，二者互促，推动全局智能升级。

5.1 AI for RAN：分层注智，全面提升无线网络性能

整体上说，RAN的各层设计在理论基础、优化目标和计算方式上有所差别，因此在AI注入智能的时候，需要根据底层、高层和顶层的不同特点进行更精准的优化演进。

5.1.1 底层以高性能驱动

对于无法准确统一数学表达、也无法统一为典型数学分布的局部模块，例如多场景非导频位信道的时频域估计、信道压缩，不同硬件非理想性拟合等，传统算法难以获得理论最优解或最优解复杂度过高，根据特征在时、频域的呈现，发挥AI的“见多识广”的特点，提高准确率。然而，考虑到AI计算复杂度和实时性要求，在硬件、功耗对模型规模的约束下，泛化性是否可以达到性能要求，暂时看有非常大的挑战。

	泛化方式	优势	劣势
传统算法	从数字信号处理、矩阵理论、概率论等经典数学原理出发，以典型数学分布作为假设前提设计系统，即基于数学原理与分布特点做先验泛化性推演。在不同的时间、地点和站型下，由于数学规律或分布特点不变，则新场景做出的预测有效。	数学模型、分布保障了一致性，可解释性、稳定性好；一步直接映射到目标空间，不依赖或较少依赖数据，代价较少，推理复杂度低。	对于场景间很难统一数学表达、或无法较精确统一数学分布的假设前提，例如不同环境、干扰下的信道时频域的变化、不同硬件非理想性等只能分类建模，颗粒度大、准确性低。
AI算法	以数据的相关性作为假设前提，将各种类型的数据，提取特征得到的相关性规律记录储存为参数。在不同的时间、地点和站型下，AI通过计算新场景数据的相关性，匹配相似的既有数据，量化出先验预测的概率。	通过训练数据学习各个场景数据的相关性关系，对于个性化的数据分布可以更广泛和精细的匹配到场景的相似性。对于既有场景的规律总结更完善。	没有不变性假设前提，暂时无法定量的确认新场景中涵盖的不同数据组合是否已经被既有数据体现，泛化的确定性较低；由于需要小步长，长期迭代搜索，需要的大量的数据和高维计算，代价也较高，推理复杂度高。

举例来说，无线信号分别在时频域上有不同的特征呈现，时域的多径、频域的氧衰等，经典底层框架和算法设计基于此，通过傅里叶变换灵活在时频域切换，找到最佳特征呈现域进行信号处理。而如果简单粗暴地直接进行端到端AI推理，将空口时域数据为输入，那么频域明显特征将弥散隐藏在时域里，特征挖掘的代价不言而喻。另一方面，传统算法通过公式推演可一步映射到目标空间，AI算法则需从数据中广泛积累规律，这就导致目前广泛研究的AI模型的推理复杂度，相较传统算法高出数倍甚至2~3个数量级，在时间约束下硬件规模和功耗的急剧上升。

因此，对底层而言，更具潜力的策略是将传统算法和AI算法做更细化的优势融合，利用传统算法的理论泛化优势和AI的场景学习优势，以高性能驱动和复杂度约束进行算法的优势组合，实现不同场景下的性能最优。

5.1.2 高层以多任务驱动

经典高层算法主要是根据不同的信道、业务特点、场景目标来做系统控制，形成控制决策和动作。如果以AI的方式实现，需要考虑两个主要的方面：学习空间，即确认输入特征空间和输出动作空间交织出来的待学习空间；评价体系，即确定最终的输出动作以什么方式评价，是否在工程代价上和反馈速度上可支持遍历性学习。

- **学习空间**：高层控制的学习空间非常大，涉及到众多输入特征和输出动作维度的组合，工程上几乎无法遍历。

  输入特征空间：即使已经将原始的数据流根据经验压缩成相关特征，也有非常大的输入特征空间。以信道为例，每个终端都有接收场强、干扰、时间变化、多径情况、终端实现等多种因素，所有终端排列组合，相关输入维度大小呈几何倍数增长。

  决策动作空间：即使通过定义帧结构、标准流程缩小动作空间，对于每一个最小的时、频、空域资源单位，都可以分配给不同的终端做不同调制的数据传输或不同目的的信道测量，此时的动作空间将是百万量级以上。

- **评价体系**：一方面，由于电磁波的相位特性敏感性高，目前仍不具备在输入特征维度和输出动作维度上和实际外场对齐的孪生环境。另一方面，在实际传输环境中反馈，探索未知动作空间可能造成现网问题。

综上，高层全盘或大模块AI替代，会面临学习空间过大，评价体系难以建立的问题。通过减少模块任务范围来缩小学习空间，以专家经验为基础，拆分高层系统为多任务协作架构，减少单个任务的学习空间，并找到可准确孪生或在线低代价快速评价的部分，进行AI智能化升级，最终实现局部AI任务优化，传统、AI多任务协同的整体高层框架。

5.1.3 顶层以强智能驱动

以统计特性预测和信令语言理解为基础，结合人的意图和既有运维经验，完成更智能的网络操控。一方面，顶层的数据建立在统计意义上，往往体现出较强的规律性；另一方面，顶层确定性流程、跟人交互更多，所以已经将很多的信息压缩到信令空间或码流空间等，学习空间更小，并且已经有一些成熟的专家经验以文字形式记录，这些特征都使得顶层适合AI化。值得注意的是，主要以模型推理为主的RAN顶层应用计算模式，在量化和模型压缩方面有更多空间，需要以应用性能为基础，选择更经济的计算方式来支撑强智能。

5.2 RAN for AI：连接、感知与计算的全要素供给

随着AI技术与通信技术的深入融合，通信网络正全面迈向智能化与协同化。基站不仅是连接终端与网络的枢纽，还在感知与计算领域扮演核心角色，为AI驱动的业务提供高效支持。通过无线边缘计算和多模态感知技术的引入，基站实现了从数据传输到智能决策的跃升，显著提升了网络性能、隐私保护和资源利用效率。在此基础上，基站赋能人形机器人、车路云、NTN、低空经济和海域等前沿场景，推动无线网络的创新应用加速落地。

连接服务增强：在具身智能，交互式AI等应用场景中，AI计算能力向终端进一步下沉，分布式AI的计算协作促使连接传输带宽增加。同时具身智能场景还存在频繁交互的诉求，对网络连接的时延和可靠性也有新的要求。

感知全场景赋能：通过基站通感技术与AI结合，实现NTN、海域及低空场景中对环境与目标的精准感知和动态优化，显著提升通信覆盖、资源利用及用户体验，推动智能化网络服务的全面升级。

计算服务拓展：在连接和感知的基础之上，边缘算力可以为行业机器人做大脑赋能，同时自动驾驶场景下，部署在边缘的应用可以统筹管理多车和路侧感知信息，避免各个端侧AI重复推理消耗算力，保护个人驾驶信息的隐私。为端侧提供计算服务、可靠链接服务。

5.2.1 连接服务助力具身智能云边协同

为推动机器人发展，需将部分算力卸载至边缘或者云端，以提升续航、降低计算成本并增强多机器人协作能力。通过部署在 AIR RAN 上的多模态大模型及智能体服务，在提供无线连接的同时，赋予机器人更出色的自然语言理解与交流、融合感知、分析决策及任务规划能力，助力其从本地智能迈向具身智能与群体智能。具身智能结合物理身体与智能算法，强调通过身体与环境互动增强感知决策能力，让机器人借助多模态感知适应复杂环境并完成任务。在 RAN 边缘部署具身智能大脑能力，与机器人本地小脑边端协同，既能实现低时延分析决策，又能节省功耗、保护数据隐私。

以某人形机器人为例，全身拥有28个自由度，其中旋转关节14个，直线关节8个，头部6个。其灵巧手自由度22个。全身配备28个执行器，包括高扭矩电机和无刷有齿槽电机。另有深度视觉相机和激光雷达。核算下来与边缘算力传输的数据量大约为上行>30Mbps，下行以控制指令和语音为主，流量较小。这些流量通过RAN网络进行连接和传送，RAN需要对其数据进行时延和抖动的保障。以语音交互端到端<2s的要求来考虑，网络时延需小于20ms。

用途	内容/指令	上行速率(Mbps)	下行速率(Mbps)
控制	身体电机	1M(State)	0.1M(Cmd)
	手部	1M(State)	0.05M(Cmd)
	里程计信息	0.5M	--
语音	语音报文	1M	1M
视频	视频报文	10M	--
激光雷达	激光雷达报文	20M	--
总计		**33.5M**	**1.15M**

5.2.2 低空、海域、NTN全场景感知赋能

随着通信技术的快速发展，基站的功能已从传统通信服务扩展到通信与感知融合的全新领域。基站天线通过反射信号感知外部环境，实现对物体位置、运动轨迹及环境特征的实时监测。结合AI技术，通感能力在NTN、海域通信和低空无人机场景中展现出巨大的潜力。

在NTN中，基站利用信号反射感知地面用户设备的运动状态和环境特征，并通过AI动态优化波束方向和频谱分配，不仅提升网络覆盖，还支持动态导航和紧急通信。在海域场景中，基站通过感知海洋环境、船舶位置及气象条件，结合AI预测通信链路稳定性和负载需求，动态调整资源分配，有效支持远洋通信、航线规划和船舶避碰。低空无人机场景下，无人机飞行对地理信息和障碍物更新的实时性与精确性要求极高，静态地图航路规划及机载雷达 / 摄像头难满足需求，基站天线感知无人机高度、速度和路径，AI则对感知数据进行实时处理，优化频谱分配与路径规划，确保无人机群的高效协同与安全通信。

以低空场景为例，通感技术结合AI在多领域发挥重要作用。在物流配送领域，助力无人机精准定位、自主避障，安全快速送货，降低物流成本、提高配送效率；在智慧交通方面，实时监测空中交通状况，避免拥堵与安全事故，为飞行器提供精确导航，优化飞行路线，提升便捷性与安全性；在低空安防领域，可识别非法无人机并实时监控打击，还能对合法网联无人机进行预定航路监控及敏感区域保护。

通过AI驱动的实时决策，基站通感能力使RAN从传统通信平台向智能化感知平台转型，在复杂环境中实现动态优化和资源高效利用，全面赋能全场景的智能化升级。

5.2.3 连接感知计算协同赋能自动驾驶

基于5G的车路云一体化方案中，路侧感知是方案的重要组成部分，为了发挥5G网络的优势，需要把5G网络的性能和路侧感知AI分析、边缘计算等技术进行深度融合。

多模态感知融合：实时性是路侧智能系统的关键要求。路侧需快速整合毫米波雷达、摄像头等多传感器数据，边缘实时AI通过数据融合，精准提供车辆的3D位置、速度和方向角等信息，支持交叉路口等复杂场景下的快速决策。相比传统云端处理，由于传输距离远易导致网络拥塞和时延增加，将处理部署在5G基站能够大幅缩短传输路径，显著降低时延，使融合数据更快应用于车辆控制。例如，高速行驶中，车辆可及时根据路侧信息调整状态，确保行驶安全与效率。

目标动态跟踪：路侧感知系统需具备实时性和低时延，以应对道路上车辆和行人状态的快速变化。边缘实时AI结合RNN和LSTM技术，可迅速检测和识别目标物体，预测其轨迹并更新状态，帮助车辆提前规划路径，例如在行人横穿马路时及时制动或避让。相比云端处理可能因传输和处理延迟导致目标信息滞后，边缘部署能够本地快速处理数据，将时延控制在极低水平，确保车辆在复杂交通流中实时掌握目标动态，有效保障行车安全。

目前，中兴通讯与济南移动联合打造的车路云样板点已落地，覆盖5公里路线、14个路口（8个重点路口），基于5G、路侧感知AI分析和边缘计算技术的“5G + 车联网系统”完成部署。车端可实时接收路口红绿灯、超视距事故预警及行人闯入预警信息。经实测，该系统降低道路事故发生率约37%，提升交通态势感知效率约20%，超视距事故预警探知距离超3公里。云端实现交通实时监管，为智能汽车及其用户、管理机构提供车辆运行、基础设施和交通管理等动态基础数据。

5.3 群智协同：泛在Agent服务

通信AI小模型主要解决单个通信目标问题，构建、训练等环节依赖人类专家，行为相对被动、静态且局部。而未来无线系统面临通感算智信等多目标复杂任务。

智能体是高阶智能实体，能主动感知、理解、规划等，具备自主性等特征，基于特定大模型完成任务并自我完善。“网络智能体”（NW Embedded AI Agent，NEA）是内嵌部署在网络侧的智能体，既可集成在RAN网元节点内，也能以独立节点存在。它不仅服务所在网元节点，还与其他NEA协同，完成复杂任务。

NEA面向泛化无限域问题，主动采集数据，深度感知决策，不依赖外部专家指导，行为主动、动态且全局。结合自身存储的知识经验与工具，其判定决策比通信 AI 小模型更具可解释性、鲁棒性和可信度。

6G无线系统追求高阶网络运维自治，从“外挂式自治”走向“内生自智”。NEA在规划组网方面，致力于环境态势自发现、按需灵活组网、可信泛在接入；在网络维保方面，聚焦智能策略生成与验证、任务编排协作、会话管控、目标自寻优等。在NEA助力下，未来网络运维与运行将融合，无线系统能更精准适配环境与诉求，朝优化方向演进。

未来，NEA将从单域单一智能体向跨域群智协同发展。通过群智协同分解复杂问题，

优化决策，避免单一智能体局限。不同智能体相互补充，提升系统多目标决策能力与鲁棒性，还能涌现群体智慧，产生创新解决方案应对复杂网络环境。

在未来6G时代，移动网络不仅是泛在连接的载体，更是实现人，机，物，智间无缝交互的智能化平台。6G的核心技术愿景是实现“通感算智超融合”，即实现对通信、感知、计算和AI智能在系统架构层面的深度融合和应用，以推动全社会数智化、可持续化发展。AIR RAN作为无线接入网的智能化深度进化，是支撑实现上述愿景的重要部分。



06 未来展望：引领无线网络的长期价值

6.1 AIR RAN愿景

旨在实现6G“万物智联、数字孪生、绿色高效”等愿景目标，6G RAN需要支持超高带宽、超低时延、超大连接数，以及复杂的动态环境，RAN智能化成为不可或缺的技术基石。AIR RAN赋予无线接入网深度智能感知、自主决策和动态优化等多维度能力，诸如：

通信智能化：AIR RAN通过AI驱动RAN优化和无线资源管理，实现更高效的通信能力。

感知实时化：AIR RAN将具备环境智能感知能力，为实现数字孪生提供数据基础。

计算内生化：在AIR RAN中，计算是RAN内生功能，通过边缘计算和分布式AI降低计算延迟提升效率。

智能进化化：AIR RAN核心新特性是“自学习、自优化、自管理”，它将使RAN从被动响应转向主动感知预测演进。

计算能力延伸：通过深度边缘计算和分布式计算架构，在接入网层面为AI推理和训练提供算力支持。

数据处理枢纽：AIR RAN将成为AI应用的数据处理节点，为实时性要求极高的应用（如自动驾驶、工业控制、远程医疗等）提供低时延的数据处理服务。

智能协同中心：AIR RAN将通过多维资源调度、任务分发和模型协同，为分布式/边缘式AI应用提供高效的运行环境。

智能生态融合：AIR RAN将成为跨行业、跨场景的智能中枢，赋能垂直行业数字化转型。

6.2 AIR RAN路径

AIR RAN发展和大规模应用并非一蹴而就，它需要经历上述“短中长期”的发展历程，具体为：短期着眼于“以智强网，逐层推进，异构协作，体验优化”方面，中期着眼于“以智筑网，整体内生，智算融合，价值拓进”，长期着眼于“以智生网，全面一体，多维超融，万物智联”。在上述历程中，客户收益和产业价值潜能将逐步地被释放。

短期（2025-2027）：广泛引入AI实现RAN深度优化，解决传统网络痛点，如频谱利用效率低、能耗高、算力资源利用率不足等问题。构建基础的RAN AI算法框架，赋予RAN初步的自学习优化能力。AIR RAN不仅能够满足用户对高质量连接的需求，还能提升RAN的灵活性适应性。另一方面，包括车路云、具身智能以及边缘运算也将在这一时期完成验证测试。

中期（2028-2030）：实现RAN智能协同与动态管理。AIR RAN将成为通信、算力、模型和数据交互的核心枢纽，支持多要素网络协同，具备深度感知和实时优化能力的智能平台。同时也会作为AI边缘服务器，协助端侧和云侧共同构建AI生态。

长期（2030以后）：全面迈向“通感算智超融合”的网络生态，实现全场景、全域的智能化连接，RAN能够自主感知、实时决策，支持复杂的多维交互。AIR RAN将真正成为构建数字孪生世界的核心基础设施，推动人类社会迈向全连接的智能化未来。

6.3 AIR RAN生态

AIR RAN的发展，既依赖技术突破，更离不开全行业的紧密协作。AIR RAN生态演进是一个从技术到试验到产业、从愿景到实践到收益的全方位进化过程。

AIR RAN愿景与6G系统“通感算智超融合”的目标高度契合，是6G网络与生态的关键。通过AIR RAN，未来6G网络将实现从“连接驱动”到“智能驱动”的转变。面对AIR RAN未来的机遇挑战，呼吁产业各方积极参与，共同加速推进产业化进程：

强化技术研发：加强RAN侧AI算法、边缘计算和网络架构的创新，推动AIR RAN技术从实验室走向商用，支持具备AI训推能力的智能基站。

增进跨界协作：通信设备商、运营商、云服务商、AI企业等需要形成深度协作，共同推动AIR RAN生态发展，打造基于AIR RAN的边缘智算服务。

开放联合创新：建立开放的AIR RAN测试床和实验平台，鼓励更多中小企业和研究机构一起参与到AIR RAN的技术开发和实践中。

标准化合作：推动AIR RAN相关标准的制定，形成统一的技术框架和接口规范；推动AI模型在RAN部署应用的标准化，确保异厂家间互操作性。

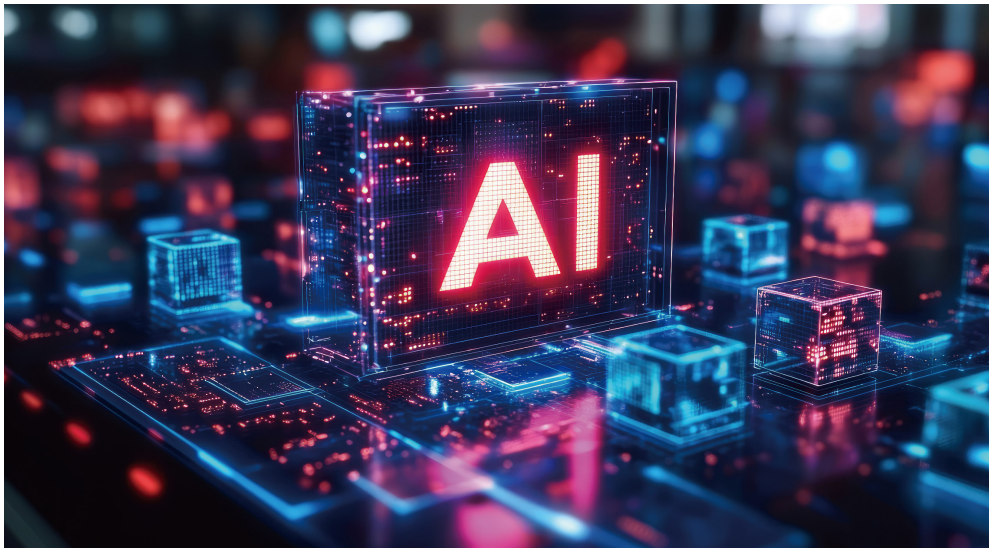
商业模式探索：探索AIR RAN的商业化路径，例如通过“网络即服务”（NaaS）模式实现价值变现；推动和垂直行业深度合作，共同构建AI新商业生态。

绿色生态建设：共同构建开放共赢绿色的产业生态，推动AIR RAN在更广范围内的应用。

伴随着AIR RAN的演进，它将逐步释放客户收益和产业价值，推动网络从“以网带智”到“以智带网”的转型，为全面智能化社会奠定坚实的基础。

未来，随着技术的持续创新与突破，在数据处理方面，将实现跨领域、多源异构数据的深度融合与实时分析，确保AI应用在不同场景下的精准响应与高效运行。算法设计上，有望研发出兼具高效性、可解释性和强适应性的新一代AI模型，极大提升网络资源管理和业务调度的智能化水平。算力架构将进一步朝着异构、分布式协同的方向演进，充分挖掘边缘计算与云计算的互补优势，形成无处不在且弹性可扩展的算力网络，满足 AI 业务爆发式增长的算力需求。

全球范围内，中兴通讯将携手产业合作伙伴形成紧密的技术创新及价值创造共同体，共同制定统一的行业标准与规范，促进技术的跨区域、跨平台共享与协作，加速AIR RAN技术在全球的普及与应用，缩小数字鸿沟，让智能网络服务惠及世界各个角落，开启一个万物智联、高效协同、创新无限的未来世界。



07 缩略语

缩略语	全称
AI	人工智能（Artificial Intelligence）
RAN	无线接入网（Radio Access Network）
5G	第五代移动通信技术（5th Generation Mobile Communications Technology）
6G	第六代移动通信技术（6th Generation Mobile Communications Technology）
MIMO	多输入多输出（Multiple Input Multiple Output）
XR	扩展现实（Extended Reality）
AR	增强现实（Augmented Reality）
VR	虚拟现实（Virtual Reality）
MEC	多接入边缘计算（Multi-access Edge Computing）
KPI	关键性能指标（Key Performance Indicator）
KQI	关键质量指标（Key Quality Indicator）
QoS	服务质量（Quality of Service）
L1	第一层（Layer 1，物理层）
L2	第二层（Layer 2，链路层）
L3	第三层（Layer 3，网络层）
RNN	循环神经网络（Recurrent Neural Network）
LSTM	长短期记忆网络（Long Short-Term Memory）
5G-A	5G增强技术（5G-Advanced）
NTN	非地面网络（Non-Terrestrial Network）
NEA	网络智能体（Network Embedded AI Agent）
NaaS	网络即服务（Network as a Service）
OPEX	运营支出（Operating Expense）
CAPEX	资本支出（Capital Expense）
GSMA	全球移动通信系统协会（GSM Association）
IMT-2030	国际移动通信2030（International Mobile Telecommunications-2030）
GPT	生成式预训练变换器（Generative Pre-trained Transformer）
V2X	车辆到一切（Vehicle-to-Everything）
BBU	基带处理单元（Baseband Unit）
RRU	射频拉远单元（Remote Radio Unit）
K-Means	K均值聚类算法（K-Means Clustering Algorithm）
GAN	生成对抗网络（Generative Adversarial Network）
3C	计算机、通信和消费电子产品（Computer, Communication, and Consumer Electronics）

参考文献

[1] Accenture, https://www.accenture.com/au-en/insights/health/artificial-intelligence-healthcare

[2] Research and Markets, https://www.prnewswire.com/news-releases/the-world-wide-ai-in-telecommunication-industry-is-expected-to-reach-6-3-billion-by-2026--301401190.html

[3] GSMA，The Mobile Economy 2024 https://www.gsma.com/solutions-and-impact/connectivity-for-good/mobile-economy/wp-content/uploads/2024/02/260224-The-Mobile-Economy-2024.pdf

[4] Grand View Research，“AI Video Generator Market Size, Share & Trends Analysis Report By Component (Solution, Services), By Application (Marketing, Education), By Organization Size, By Source, By Region, And Segment Forecasts, 2024 - 2030” https://www.grandviewresearch.com/industry-analysis/ai-video-generator-market-report

[5] 无线人工智能（AI）技术研究报告. IMT-2030（6G）推进组

[6] SoftBank，AI-RAN：Telecom Infrastructure for the Age of AI https://www.softbank.jp/corp/set/data/technology/research/story-event/Whitepaper_Download_Location/pdf/SoftBank_AI_RAN_Whitepaper_December2024.pdf

ZTE中兴 中兴通讯股份有限公司
ZTE CORPORATION

地址：深圳市高新科技产业园科技南路中兴通讯大厦
电话：+86-755-26770000　　邮政编码：518057
传真：+86-755-26771999　　网址：www.zte.com.cn