

Mobile AI

三月二〇二六

内容

执行摘要 01

Mobile AI的价值 02

1. 经济价值：通过融合释放大规模潜力 03
2. 社会价值：负责任、包容、安全的AI进步 08

移动人工智能的内涵 09

1. 人工智能网络 11
2. 人工智能网络 13
3. 移动人工智能代理/设备 15
4. 移动人工智能应用 21

向移动人工智能

25

1. 基础设施提升 26
2. 波段保证 27
3. 技术创新 28
4. 终端进化 34
5. 标准配方 35

推荐与行动呼吁 36

致谢 39

执行摘要

数字经济已成为全球经济增长的主要引擎，这得益于移动通信技术的快速演化和人工智能（AI）的加速发展。随着5G迈向全球规模，5G-Advanced逐步走向商业化，6G研究稳步推进，移动网络不断拓展连接、效率和覆盖的边界。同时，人工智能从云端延伸至设备和网络边缘，大型语言模型（LLMs）、多模态智能和AI代理正在重塑智能的创造和消费方式。这两种力量的深度融合正在催生一个新纪元——**移动人工智能时代**。

在移动人工智能时代，移动通信网络的普遍连接性成为人工智能广泛采用和民主化发展的关键基础。基于这种无处不在的连接，人工智能正超越专业领域，实现全民可及。与此同时，与人工智能的深度融合正在重塑这些网络的核心架构。这些力量共同使得移动通信与人工智能的融合成为技术发展和产业变革的必然方向。因此，**移动AI已成为一个颠覆性的范式，是全球数字和智能发展的关键驱动力。**

移动AI建立在网络与AI之间双向赋能的原则之上，其核心价值是负责任的AI以及安全与可信。通过整合移动通信网络（5G/5G-A/6G）的普遍连接性、低延迟和高可靠性，以及AI在感知、决策和学习方面的优势，移动AI建立了一个协作的**器件边缘网络云架构**，这形成了一个提供广泛覆盖、实时响应和精确适配的智能服务系统，其总体目标是实现交付**人工智能无处不在，值得信赖且易于使用**。

Mobile AI的精髓体现在其“**三层四维**”建筑。纵向来看，基础层、执行层和应用层共同构成从基础设施到特定场景价值创造的端到端路径。横向来看，四个关键功能维度——**AI for Network, Network for AI, Mobile AI Agents /Terminals, and Mobile AI Applications** 共同努力，推动整个生态系统的持续进化。这些层次和维度动态互动，实现从技术到应用的全面整合，加速行业的智能化转型。

从全球产业视角出发，这份白皮书系统地阐述了移动人工智能的价值、定义、架构、实施路径和关键挑战。它提供了战略建议，并呼吁在整个价值链上进行协作行动，以促进全球合作、标准一致性和生态系统增长。展望未来，随着5G-A的普及和6G向商业化迈进，移动人工智能将从融合应用向原生共生进化。这一转变将释放持续的创新动力，赋能规模化的产业，并成为全球数字经济高质量发展的基石。

01

Mobile AI的价值

当前5G/5G-A网络的规模化部署和人工智能（AI）的快速扩散，正推动移动通信和AI进入深度融合的新阶段。在这一趋势下，移动AI应运而生。它不仅仅是两种技术的简单叠加，而是基于“网络与AI双向赋能”的核心逻辑。通过实现“网络赋能智能，智能优化网络”，构建了一种以无处不在的连接、实时响应和智能自适应为特征的新型服务体系。这种“双向赋能”的融合范式正在系统地重塑工业发展轨迹和社会运行。它不仅为运营商和产业链开辟了新的增长路径，还为社会治理的精细化、全行业的数字化和智能化转型以及用户日常生活的智能化提升提供了前所未有的基础支持和创新可能。

经济价值：1. 通过融合释放大规模潜力

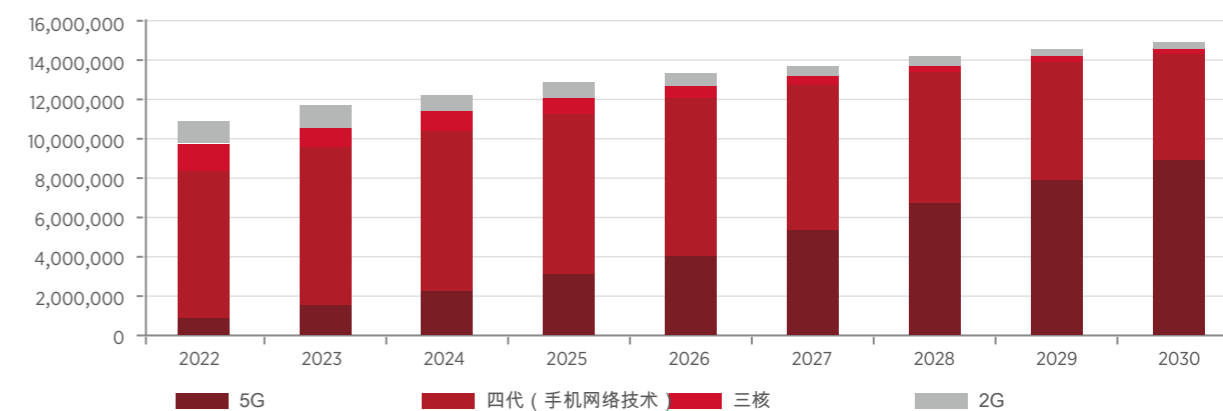
移动通信技术的迭代升级和人工智能的广泛应用趋势相互促进，成为推动全球数字经济发展的核心引擎。一方面，5G网络的广泛部署和5G-Advanced（5G-A）技术的演进为人工智能应用提供了坚实的网络基础。另一方面，人工智能技术的快速迭代和市场扩张对通信网络能力提出了更高的要求。两者的深度融合正在打破产业边界，催化新型应用场景和商业模式的出现，激发巨大的全球市场潜力。

1.1

网络基础：规模网络部署推动经济贡献

全球移动通信产业已进入5G规模发展的关键阶段。5G技术的广泛应用为其与人工智能的融合奠定了广泛的基础网络。自2019年首次商业5G部署以来，截至2025年底，全球共启动了384个商业5G网络，5G用户数量已超过30亿。从长期发展角度看，预计5G用户数量将保持快速增长。预测表明，到2028年，5G将超过

全球移动订阅预测（单位：千）



来源：Omdia

© 2025 Omdia

图1：全球移动用户预测

4G将成为全球主流的移动通信技术。到2030年，全球5G连接预计将达到88亿，占总全球移动连接的60%以上，形成一个广泛且密集互联的网络生态系统。

384

到2025年底
商业5G网络

3

十亿
5G用户数量超过

8.8

截至2030年
十亿
全球5G连接

> 60%

全球移动电话连接总数

根据Omdia的预测，世界上首个6G网络预计将于2029年开始部署和运营。到2030年，全球6G设备连接数将达到2.89亿，到2035年将增长到35亿，占全球移动设备连接总数的22.3%。伴随全球无线接入网基础设施投资的预计将从2030年的400亿美元急剧上升到2035年的2500亿美元，为移动人工智能的技术发展预留下一代网络能力。移动通信行业已经成为全球经济增长的中坚柱石。

GSMA《2025年移动经济报告》显示，到2024年，全球移动技术和服务创造了6.5万亿美元的经济增加值，占全球GDP的5.8%。这一贡献源于超过4万亿美元的生产力提升效应和1.6万亿美元的直接移动生态系统贡献，与间接经济影响相结合，形成一个完整的价值循环。展望未来，随着5G、物联网、人工智能等数字技术的广泛应用和深度融合，技术赋能的红利将持续释放，稳步提升全球生产力和运营效率。预计到2030年，移动通信行业对全球经济的贡献将增至近11万亿美元，其占全球GDP的比例将上升至8.4%，巩固其在高质量全球经济增长核心引擎的地位。

根据Omdia预测
全球6G设备连接

\$289 百万
2030

\$3.5 十亿
2035

根据Omdia预测
全球无线接入网基础设施投资

\$40 十亿
2030

\$250 十亿
2035

1.2 创新驱动：人工智能投资和市场规模加速融合

人工智能在工业领域的应用正进入一个爆发期。全球人工智能投资持续扩大，市场空间不断拓宽，其与移动通信技术的融合进一步提升了其商业价值，实现了指数级增长。

在消费者端，生成式AI (GenAI) 的采用速度显著加快。根据Omdia 2025年全球消费者调查数据显示，约75%的受访者已在日常生活、工作和社交互动等场景中使用GenAI应用程序，其中38%选择付费的高级版本，这两个数据均显示出与前一年相比的明显增长。智能手机已成为消费者接入GenAI服务的首要工具：48%的受访者使用移动GenAI应用程序进行文本搜索，14%与移动AI代理进行对话，而7%和5%的受访者分别使用移动GenAI应用程序生成图像/视频并上传到网上。这种广泛的消费者采用正在催生新的服务。

需求，全球29%的受访者表示有兴趣将GenAI应用订阅服务捆绑到他们的移动服务套餐中，为运营商指明了新的商业模式方向。



人工智能技术的进步也将推动各种新型终端设备的普及。Omdia预测，到2028年，配备生成式AI功能的智能手机将占全球智能手机出货量的一半以上 (54%)。在未来十年中，通用型具身智能机器人预计将迎来爆炸式增长，全球出货量预计到2035年将达到260万台，十年复合年增长率 (CAGR) 为85%。

根据Omdia预测
截至2028年



截至2035年

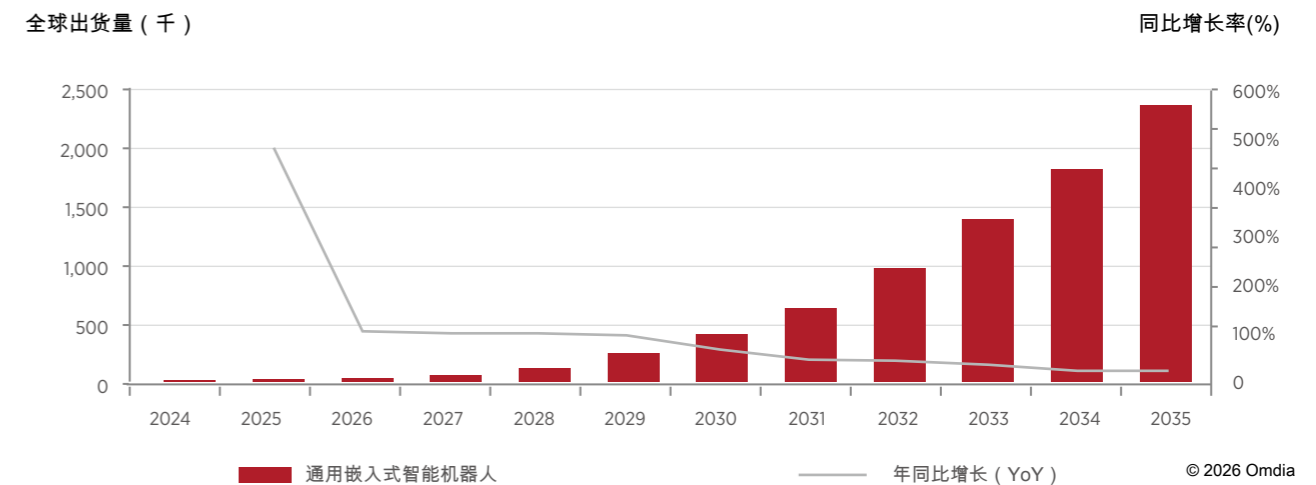


图2：全球通用型具身智能机器人出货量预测

在企业端，AI市场正呈现出爆炸式增长。Omdia预测，到2030年，全球数据中心超过70%的工作量将用于满足AI相关的计算需求。GenAI软件市场预计将从2024年的142亿美元增长到2029年的1013亿美元，到2030年将达到1228亿美元。作为AI市场内的核心增长部分，AI代理应用的发展速度甚至超过了整体GenAI市场。其市场规模预计将从2024年的1.59亿美元增长到2025年的15.1亿美元，到2030年可能达到417.7亿美元。它们在GenAI市场中的份额预计将从2025年的6%上升到2030年的31%。关键企业AI代理应用案例包括自动化代码开发/辅助、虚拟助手、智能流程自动化和智能文档处理。

根据GSMA 2025的一项专项调查，该调查涵盖了32个国家的5,320家企业，以及10个垂直行业，企业计划在2025年至2030年期间将其收入的约10%投资于数字化转型举措。在人工智能、5G连接和终端技术方面的投资排名首位，成为下一阶段产业增长的核心动力。在所有行业中，超过90%的受访企业认为生成式人工智能对其数字化转型之路具有关键的战略意义。

The GenAI software market is forecast



人工智能代理应用市场规模



1.3

运营商策略：推进移动人工智能，拓展商业价值

《GSMA 2025年移动经济报告》显示，2024年全球移动运营商收入达到10.8万亿美元。尽管收入增长受到激烈市场竞争和平均每用户收入温和增长的压力，但仍预计将呈现稳定的长久上升态势，预计到2030年，收入将增至12.5万亿美元。为支持技术迭代和业务扩展，2024年至2030年全球移动运营商的总资本支出预计为13万亿美元。这笔投资将着重于释放5G在各场景下的潜力，并推动企业数字化转型，从而为移动人工智能的融合式发展奠定基础设施基础。

人工智能应用的广泛应用正直接推动相关网络流量的爆炸性增长。Omdia预测，从2025年到2033年，全球人工智能流量的复合年增长率（CAGR）将达到73%，其中2031年将成为关键的分水岭，届时人工智能网络流量将超越传统流量。与此同时，以DeepSeek为代表的小型、高效的开源模型正推动人工智能部署向边缘设备发展。预计从2025年到2033年，流向网络边缘的新人工智能流量的CAGR将高达130%。这种流量规模的增长和流量特性的演变不仅推动

释放5G的潜力
在所有情况下

\$1.3 万亿

全球移动运营商资本总额
支出来自2024年至2030年的

推动企业数字化转型

全网流量，人工智能与非人工智能，2023-33

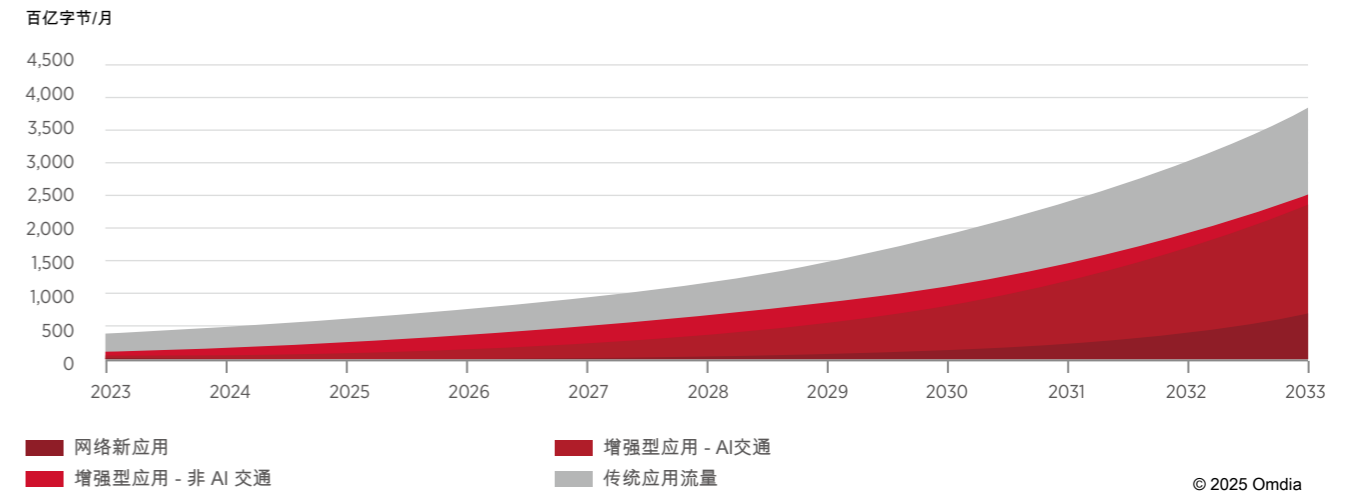


图3：网络流量预测

电信运营商需要不断提升网络传输能力，同时也触发对网络设计原则的长期优化。

面对人工智能技术发展的浪潮，全球电信运营商正积极推动战略转型。他们主动应对传统行业的挑战，即“流量增加但收入增长不成比例”，通过探索移动AI融合的三大核心维度，挖掘其价值潜力：

构建专用人工智能基础设施： 升级宽带网络以满足人工智能服务需求，建设人工智能数据中心提供GPU即服务（GPUaaS），并探索将计算能力与城域网和接入网集成布局，以促进边缘人工智能计算的规模化发展。

提升自身运营能力：利用人工智能 在诸如网络管理、客户服务和个人助理等领域部署人工智能应用，以实现整个运营过程的效率提升。

直接进入人工智能软件服务领域： 推出面向消费者、企业和垂直行业基于生成式AI的产品和服务。

Omdia的2025全球运营商调查报告显示，40%的受访全球运营商已在商业或试验网络中部署基于AI代理的自主网络解决方案，另外34%计划在未来两年内启动相关工作。运营商的这种主动参与正在加速移动通信和人工智能的融合工业化进程。

移动通信与AI的融合已从技术探索阶段过渡到规模化实施阶段。在5G和5G-A网络的支持下，并由AI市场繁荣和持续计算能力扩展的协同效应放大，它为整个产业的数字化和智能化转型提供了多元化的支持。这推动产业发展从“局部优化”跃升至“系统重构”，实现了从成本降低和效率提升到价值创造的核心进步。两者的深度融合不断拓宽产业价值边界，全面促进所有产业，从智能制造到智慧城市的数字化和智能化升级。数万亿美元的市场潜力正在加速释放。在未来五年内，移动通信与AI的融合预计将不断深化。到2030年左右全球6G商业化部署时，行业将进入一个全新的发展阶段，其特点为移动服务和AI能力的全面原生融合。

经济价值：2. 通过融合释放大规模潜力

利用其无处不在的连接性和智能协作的技术特点，Mobile AI 深度融合社会运作的核心层面。它的社会价值主要体现在三个关键维度：首先，赋能社会治理向更高精度和效率转变；其次，推动打破壁垒，实现人工智能技术的全民普及（人工智能民主化）；第三，通过实时连接和远程可控性，促进人工智能合规和安全治理。这三个维度共同协同工作，构建一个公平、包容、可靠和可持续的社会发展生态系统，为数字经济时代的社会进步注入持久动力。

推动社会治理智能化升级



依托“通信+AI”的融合能力，移动AI构建了一个具有全面感知、快速响应和精准干预的治理体系。这推动了治理模式从“经验驱动”向“数据驱动”的转变。在城市治理领域，它通过全面感知和智能分析汇聚多维度数据（如市政、交通和环境）。这为交通拥堵或公共设施泄露等痛点提供动态预警和针对性措施，从而提高运营效率，降低治理成本，推进精细化和智能化管理的演变。在公共服务中，分析供需数据有助于准确识别资源缺口，引导优质资源流向服务不足的地区，缩小城乡和区域差距。智能预约排程和流程优化等功能的运用简化了程序，提高了服务可及性和公众满意度。在应急响应方面，利用低延迟通信和智能分析创建了一个能够早期识别突发风险、快速预警和高效事件管理的综合系统。跨部门协调缩短了响应时间，减轻了灾害损失。

推广人工智能技术的普适性接入和民主化



移动AI利用广泛采用的智能移动终端，打破与AI技术应用相关的硬件壁垒、技术复杂性和场景限制。它将专业级的AI能力带给各行各业和普通家庭。这促进了AI从“专属技术高地”向“普惠资源”的转变。在诸如农村个性化教育辅导、基层医疗辅助诊断、普惠金融智能服务、老年友好型智能适配以及偏远地区数字赋能等场景中，它发挥着关键作用。它显著降低了AI技术的应用成本和使用门槛，使不同地区、不同人群和不同行业的人们都能平等地享受到AI带来的便利和价值。这实现了AI能力的全覆盖和共享，确保智能技术的开发红利惠及每一个人。

授权AI合规性、安全性和可控性



基于实时可靠移动通信网络，移动AI为具有智能体智能系统和自动驾驶汽车等先进智能设备提供远程监控和安全接管能力。它建立了一个覆盖整个AI生命周期的合规性和安全保证系统。在设备操作层面，低延迟连接启用对自动驾驶汽车和机器人等智能终端异常行为的实时状态监控和干预。在突发风险情况下，可以激活远程安全模式或接管控制，极大提高复杂环境下的系统容错性和个人安全保障。在生态系统协作层面，通过连接多个监管机构、企业技术平台和终端用户，建立了一个动态、协作响应式的AI治理网络。这促进了“研发-部署-运营-监督”闭环的形成，确保智能技术的发展始终在一个安全、可靠、可控的框架内进行。

02

移动人工智能的内涵

移动AI根本植根于“网络与AI之间双向赋能”的核心逻辑，遵循“负责任AI”和“安全与可信”的价值理念。它利用了先进移动通信网络（如高带宽、低延迟、高可靠性和无处不在的连接）的特点，并将这些特点与AI在智能感知、自主决策和协同推理方面的能力相结合。通过“设备-边缘-网络-云”的全面协同，实现了智能能力的分层部署和动态调度，构建起系统的“三层四维”架构。依靠其双向赋能、全面协同、场景特定闭环和循环演化的核心特性，实现了从技术基础到产业价值创造的全程整合和持续迭代。



图4：移动人工智能的分层四维架构

移动AI的核心架构体现在以下三个层次和四个关键维度。“三层”定义了从基础设施到价值创造的垂直实施路径，每一层通过能力供需反馈紧密相连：

基础层： 作为技术基石，它深度融合“AI赋能网络”和“智能承载网络”的双向能力，为全面连接、异构计算能力和协作智能提供核心支撑。

执行层： 作为执行和实施能力的实体，它负责将基础层的智能和连接能力封装成可部署的、交互式的功能执行单元，并为应用层提供标准化的服务接口。

应用层： 作为价值实现的层，它基于执行层提供的服务，在特定商业场景中构建智能解决方案，直接产生经济和社会效益。

“四维空间”贯穿并驱动整个系统的运作，代表了移动人工智能实现目标核心理念维度：

人工智能网络： 深度应用于通信网络的全生命周期，使网络规划、运营、维护、优化和管理实现自动化和智能化，从而提高网络效率和自主水平。

人工智能网络： 利用移动通信的高性能、高可靠性和低延迟特性，满足智能应用在数据传输、计算能力协作和无处不在的接入方面的差异化需求，提供无处不在的

智能的连接支持

移动AI代理与设备： 作为大规模托管和智能能力交互的实体，它们涵盖了机器人、AI手机、可穿戴设备、各种物联网终端和代理等多种形式。通过设备-边缘-云协同计算，它们克服了单设备的资源限制，实现了智能服务的民主化和个性化交付。

移动人工智能应用： 将融合的科技能力转变为适用于所有行业和消费者领域的特定情景解决方案。涵盖智能制造、智能交通、精准医疗和城市治理等关键领域，它们推动产业升级并提高社会运行效率。

这四个维度构成了“三层”架构中的动态循环：对基础层的持续优化推动执行层的创新和能力升级，进而催化应用层的特定场景创新。反之，应用层的实际需求驱动基础技术的迭代和执行能力的提升，形成逆向循环。这构成了“技术-执行-应用”的闭环进化系统，最终支持移动人工智能实现“智能无处不在、可信、用户友好”的愿景。

人工智能网络

AI for Network指的是人工智能技术在通信网络整个生命周期中的深度应用，涵盖规划、维护、优化和运营。其目标是提升网络性能、用户体验和运营效率，最终实现网络自主性。移动AI将通过将AI技术深度集成到通信网络架构中，变革当前“附加”和“碎片化”的网络智能解决方案的格局。这构建了自主网络感知、智能决策和闭环优化的能力，使AI成为移动通信网络的固有、原生和普遍的核心功能，从而推动网络向更高层次的智能发展。

1.1

基于移动人工智能的网络规划

数字经济中数据流量的激增给网络规划带来了挑战，如拥堵、容量管理和体验保证。虽然5G时代的AI解决方案可以利用历史数据预测容量需求并优化投资，但它们对历史数据的过度依赖导致了动态适应能力薄弱。面对新服务或施工障碍带来的突发流量激增，常常需要重复调查和调整，导致周期延长和效率低下。

移动AI推动网络规划从“静态设计”向“动态增长”转型，实现自主性、精准性和高效性。一方面，它在遵守隐私法规的前提下整合多领域数据，生成网络全景实时视图，设计合规方案，并模拟极端场景下的风险。另一方面，它通过标准化接口自动化流程，部署智能节点完成端到端操作如设备接入，并使用AI眼镜等设备实时动态收集数据，确保施工时间表按时完成。通过实现数据协同、动态执行和实时闭环，移动AI解决传统规划和建设的痛点，为数字社会的基础网络提供核心支持。

1.2

基于移动AI的网络安全运维

随着网络复杂性不断提高以及运维需求的不断攀升，传统模式已变得不可持续。随着网络规模的扩大，设备数量增加，极端天气事件频发，手动操作耗时、难度大，且易影响服务连续性。在5G时代，虽然AI解决方案能够预测单个设备故障，但它们依赖于历史数据，缺乏诊断未知或复杂故障的能力，并需要

人工干预进行维修。这导致了高响应延迟和潜在错误。AI能力在不同领域的独立部署也使得跨领域故障的协作处理变得具有挑战性。

移动AI通过持续学习和闭环协调推进运维向高效自主进化。一方面，它实现了持续学习和知识共享，使处理一个区域新故障的经验能够在整个网络中同步。同时，它动态构建因果推理图，主动预测风险并采取预防性措施。另一方面，它创建了一个“感知-决策-执行-反馈”的闭环，自动完成指标收集、异常识别、策略生成和设备编排等任务。它还促进了跨域协作，以协同处理跨域故障，并在整个网络内动态调度资源。通过赋予网络自主思考和行动能力，移动AI推动网络向自我优化和自我修复发展，推进全域自主的目标，为稳定网络运行和高效迭代注入核心动力。

1.3 基于移动AI的网络优化

网络的不断扩展和服务类型的多元化突显了静态规则与动态流量之间的矛盾。目前的人工智能优化方案大多是依赖离线训练输出固定策略的“附加工具”。模型更新落后于网络变化，不同领域之间的执行缺乏协调，难以实现全局优化或准确满足用户多样化的需求。

Mobile AI重塑网络优化系统，将基于点的离线算法升级为系统化的在线能力。通过跨域协调机制，智能代理在多个领域动态协商资源和策略，实现端到端的多目标优化。具备意图理解能力，它将业务需求转化为适应不同场景核心需求的确切优化策略。它集成了一个在线增量学习模块，实时收集数据并微调模型，实现动态、特定情境的适应（“一时一策”）。结合云计算、边缘计算、设备协同推理，这降低了响应延迟。移动AI推动网络优化向“自主、精确、高效”转型。它不仅自动维持最佳性能，还能主动预测和调整，为用户提供高品质、精细的网络保证，巩固数字服务体验的基础。

1.4 移动AI基础的网络安全

目前，人工智能技术已深度融合到网络运营的各个方面，从内容推荐和服务激活到用户服务和安全保护，形成了一个相对完整的辅助系统。然而，现有的基于人工智能的运营系统大多是为特定场景独立设计的，如客服聊天机器人、推荐引擎和安全监控系统。这些系统通常基于孤立的数据库运行，缺乏系统间的互操作性和协作。这导致运营响应延迟，不仅降低了运营效率，还可能因为决策数据的一侧性而导致误判。

移动AI构建了一个协作智能运营系统，打破系统间的壁垒，实现跨模块互操作和协作。通过收集和分析用户端、网络端和业务端的多领域数据，它构建动态用户档案，准确推送个性化服务计划和增值服务，从而提升用户体验和价值。同时，它实时监控网络和服务状态，主动识别潜在体验问题，并提前调整运营策略，以减少用户投诉并保护用户隐私和资产安全。移动AI将建立一个层级的运营架构，有机统一实时业务处理、跨域任务协调和战略规划。这促进了人脑智慧和机器智能的深度融合，为网络运营创新和升级开辟了新路径。

1.5

摘要

移动人工智能技术可以有效解决当前网络规划、建设、维护、优化和运营中的痛点，为运营商带来巨大的商业价值。



精确的产能规划和资源调度可以防止网络资源的过度配置，直接节约能源和硬件投资成本。预测性维护还可以避免因重大故障造成巨大损失，实现成本降低和效率提升。



基于用户、服务、时间和地点等维度设计的网络侧精细化服务方案和运营，提供了卓越的用户体验，提升了用户留存率、付费意愿和品牌声誉。这有助于推动新服务发展，促进商业营收增长。



网络运营产生大量数据。在合法、合规和匿名的前提下，人工智能技术可以从这些数据中提取价值。它为产品设计、营销和网络投资提供数据洞察，还可以为垂直行业提供高价值的数据产品，创造新的收入渠道并实现数据货币化。

2. 人工智能网络

人工智能与通信网络的深度融合催生了移动人工智能，推动着智能机器人、AI手机和AI眼镜等新终端的快速涌现。这些设备不再是简单的工具，而是具备交互和决策能力的智能载体。它们的大规模发展对网络连接服务提出了更高的要求。实现“感知-决策-行动”闭环的实体智能机器人对实时互动服务（如语音问答、环境分析）和远程遥控服务的网络速率和延迟有严格的要求。远程遥控还要求具备移动性和高可靠性。在智能终端层面，AI手机通过三种方式适应不同的场景：设备端（低延迟、高隐私）、云端（强大的计算能力）和混合（平衡性能和隐私）。AI眼镜依赖高清视频和图像等多模态数据的传输，对网络的延迟、速率和稳定性提出了高要求。这些新终端和服务的发展要求移动通信网络如5G-A等专门提升其连接服务能力，以匹配多样化的和严格的网络需求。



2.1 准确识别多模态服务特征



移动AI服务并非单一，它们涵盖了多样化的场景，如数据密集型、延迟敏感型和控制精度型。连接保障的核心在于移动AI服务的独特多模态特性。这意味着网络必须同时承载并保证具有不同特性的服务的传输。每种服务都有其独特的“模态指纹”和对网络资源的不同需求。这打破了传统网络以“尽力而为”为主的单一服务模式，要求网络具备模态级别的服务意识和复杂的差异化服务能力。

2.2 同时保障多维网络性能



移动AI服务要求网络超越传统优化单一指标的模式。它们必须在同一物理基础设施上实现协调的多维度保障，在有限资源约束下实现速度、低延迟和容量多目标优化。这需要灵活的调度能力，以动态分配上行/下行资源并控制延迟抖动。延迟必须根据服务优先级进行保证：核心控制和实时决策场景需要超低延迟以确保快速命令响应；实时交互场景需要低延迟以获得流畅体验；非实时数据同步场景应将延迟控制在合理范围内。带宽必须弹性适应多模态传输和多设备协作等需求，满足轻量级交互和重负载场景下的高带宽需求，并解决密集协作设备高并发传输的挑战。速率必须平衡稳定性和效率，以确保核心业务运营正常进行，适应各种场景的差异化需求。

2.3 实现跨层数据协作与共享



目前，网络和移动人工智能服务之间存在着一个由“语言不同”和“不同步”造成的协作障碍。作为底层管道，网络通常以标准化、通用方式提供“尽力而为”的服务，缺乏对服务动态需求的深入了解。相反，服务应用无法获取实时网络状态（例如，拥塞、延迟），因此难以调整其数据传输策略。这要求网络摆脱传统的“管道”传输模式，与服务终端实时协作，实现实时跨层、跨域状态感知、资源调度和协调控制。

移动通信网络如5G-A，专注于“增强连接、细粒度保证和跨层协作”。通过这三个能力的全面升级，它们为新的移动人工智能驱动的智能服务、终端和应用提供全链路支持，将人工智能技术和应用从“理论上可行”推向“大规模部署”。

2.1 多维连接增强

增强连接性

多维连接增强，提升服务体验。超越传统网络中单一指标优化的局限性，5G-A等移动通信网络全面协调速率、低延迟、容量和可靠性。这支持在同一物理基础设施上同时运行多种类型的移动AI服务，确保高质量、无缝的连接体验。上行链路增强通过上行载波聚合（CA）和补充上行链路（SUL）等技术实现。前者将多频段载波捆绑以动态分配数据流，而后者在低频段部署专用上行链路载波。两者均可提供超宽带宽和低延迟，支持智能设备实时传输高清视频、传感器数据和控制命令。智能预调度学习服务传输模式，通过评估信道质量等因素，智能分配上行链路资源。这消除了调度请求和授权的等待时间，降低了数据传输延迟。

2.2 跨层协作

跨层协作

跨网络、服务和设备的多层协同传输。类似5G-A这样的移动通信网络，利用模式级服务意识和细粒度差异化的服务能力，可以实现移动AI智能服务的“按需分配和精确适应”网络资源。这解决了传统网络“一刀切”方法导致的服务适应不足问题。无论是数据密集型的AI大模型训练、AI代理意图推理，还是对延迟敏感的自主……

2.3 颗粒级保证

颗粒级保证

为具身机器人提供实时决策和即时数据传输的驱动，或控制精度协作操作和操控具身机器人的，均可获得专用网络资源配置。另一方面，智能服务对网络上行链路有较高要求。在弱覆盖或强干扰的场景中，上行链路速率会受到严重限制。在这种情况下，需要网络设备云协调以调整处理机制。当网络确认如上行链路速率等性能指标无法满足需求时，终端和云可以协同适应，从设备云联合AI处理模式转向以设备端AI处理为主导的模式。

多模态交互差异化特性保证。通过识别服务中的多个并行模态数据流，5G-A等移动通信网络确保关键模态数据的优先级和完整传输。这使系统从“粗粒度管道”过渡到细粒度、模态级别的保证，解决复杂服务特性的挑战。新智能服务流中的模态类型包括音频、视频、控制和其它数据流，每个都有不同的服务质量要求和不同的模式特征（例如，数据量、重要性、延迟与带宽、可靠性）。网络为不同重要性和模态的数据流配置不同的服务质量要求，并分配相应的传输资源。例如，具身智能的远程操作数据流可以配置为高可靠性QoS，以保证高优先级数据传输。其他数据流接收标准QoS，允许通过空中接口进行容错传输，同时保持用户体验，从而提高传输性能，最终提升服务质量和无线的有效性。

移动AI服务场景对网络的“连接深度、协作精度和保障粒度”提出了远超传统消费服务的需求。网络必须解决服务需求与网络供应之间的适应挑战，实现“服务需求-网络能力-服务质量”的闭环。这为移动AI的大规模部署提供了坚实的数字基础。未来，设备端AI将依赖手机、可穿戴设备、车辆等设备的本地计算能力，提供低延迟、高隐私、离线功能的处理、交互和物联网协调能力，确保实时性能和个性化。网络作为能力支持中心，将从被动管道演变为智能核心，根据网络能源效率、延迟、速率和数据优先级等因素，为不同客户匹配差异化的最优保障方案。云端AI将利用强大的计算能力和大数据支持复杂任务，如大规模模型训练和全球推理。终端、网络和云将共同构成智能时代的协同核心基础。

3. 移动人工智能代理/设备

移动人工智能催生了多种智能终端形式，如具身智能设备、AI代理和可穿戴设备。这些移动人工智能代理和设备正在重塑人类与设备、设备与它们的环境之间的交互模式。智能不再是一个孤立的、依赖于单个终端计算能力的应用，而是演变成为一种持续的、通过跨设备、网络和计算领域的协作运作的的能力。在这种情况下，移动人工智能代理和设备面临关键的实现挑战：具身智能机器人终端将AI深度嵌入物理世界，需要极高的实时性能、可靠性和安全性；移动人工智能智能终端受限于计算能力、功耗和成本，需要在多重需求之间寻求平衡；终端AI代理，作为系统级网关的发展方向，必须在资源受限的情况下提供全场景服务。本章重点关注两个主要方向：移动人工智能的具身智能，以及智能手机、可穿戴设备和代理。通过结合工业级和消费级实用案例研究，系统地分析了它们实现所需的关键架构和核心能力，揭示了它们在不同应用中的共同需求和差异化价值。

3.1

移动AI具身智能

1 家庭机器人具象智能



家用智能机器人正进入多场景智能协作的新阶段。它们已经从单一功能的执行进化到智能协作，突破了结构化环境的限制。它们在社区服务和家庭陪伴等领域大规模应用潜力日益凸显。当前趋势主要体现在三个方面：形态多样化、多机器人协作以及模块化/标准化。然而，核心挑战依然存在：适应复杂环境、跨不同场景的任务泛化，以及控制和降低部署和运营成本。

移动人工智能通过“终端-边缘-云”协同混合架构解决这些核心挑战，构建起多层次能力分布体系。它采用“集中式智能和分布式能力”设计。基础的大型模型处理复杂任务感知，而终端侧的轻量级模型实现实时环境感知和基本决策，确保低延迟响应。边缘层承担区域协作计算和多机器人协作调度。云端提供大规模训练、跨场景知识库更新和全球资源优化。移动通信网络支持多模式交互和远程控制，帮助机器人跨越场景边界。这一解决方案实现了适应性、协作性和成本效益的统一价值主张。它显著增强了机器人适应复杂环境的能力，支持跨场景任务泛化和多角色协作。模块化硬件与功能软件的并行开发大幅降低了部署和运营成本。这促进了家庭机器人在家庭服务、老年护理/康复等场景的广泛应用，加速人工智能技术进入日常生活场景，有助于构建包容性智能生活生态系统。

终端

边缘

云

2 终端工厂具象智能协作



智能工厂多机器人协作必须平衡实时性能和可靠性。如大型组件的协同运输等场景，对实时机器人控制、精确协调和安全保障提出了严格的要求，依赖于物理-信息系统调度。传统的集中式云处理难以满足毫秒级响应、低带宽消耗和高并发控制的需求。加之复杂的工厂环境和终端计算能力、能耗的限制，部署AI模型和进行实时推理变得越来越困难，成为智能工厂升级的关键瓶颈。

设备

边缘

移动AI通过设备边缘协作解决工业协同控制的痛点

建筑。基于5G工业芯片平台，它构建了一个工业AI推理与控制系统，其中多个工业终端在固定周期内与基站和边缘计算平台进行无线协调。在终端侧部署轻量级AI模型以处理实时任务，如图像识别和异常检测。复杂计算如多目标跟踪和语义分割可按需卸载到MEC平台。系统支持远程模型加载和差异更新。终端可根据网络状态和任务复杂性自适应切换推理模式。5G无线通信取代了传统的有线连接，满足了工业控制对超可靠性、超低延迟和确定性通信的要求。此方案在工业协作控制效率和灵活性方面实现了双重提升。任务卸载减轻了终端的计算负担和功耗，促进了工业终端的轻量级设计，使其适用于资源受限的环境。它赋予终端自主感知和决策能力，为多机器人协作提供智能支持。它优化了带宽利用率，确保在高并发场景中关键数据的传输。它提高了运营灵活性，使AI能力能够随着业务需求的发展而发展，无需频繁更换硬件，从而有助于智能工厂的灵活生产升级。

机械导盲犬



引导视障人士需要一个既安全、可靠又包容的解决方案。传统的导盲犬需要长时间的训练和高昂的成本，这使得其难以扩大规模以满足巨大的需求。机械导盲犬必须同时满足对障碍物避让和紧急停止（对个人安全至关重要）的严格延迟要求，以及长时间、便携式使用的计算能力、能耗和成本限制。依赖于高性能芯片的传统设备端智能解决方案无法平衡可靠性和性价比。

设备边缘网络云

移动AI采用“设备-边缘-网络-云”协同架构来解决这一核心挑战，通过智能能力的分层部署实现需求平衡。终端侧仅保留本地闭环控制功能，如紧急制动和基本障碍物避让，确保在关键风险场景中快速响应。高计算任务，如复杂环境理解、多模态感知融合和多轮语音交互，通过移动网络卸载到边缘和云端。这避免了终端长时间高负荷运行，优化了能源效率，同时保持了安全底线。这实现了涵盖安全、包容性和可扩展性的多重价值闭环。终端不需要高性能GPU，显著降低了设备制造成本和功耗，明显延长了电池寿命，提高了实用性。设备端的本地闭环控制和网络端的计算能力保证为关键决策，如障碍物避让和紧急停车，提供了可预测的延迟，完全满足了安全敏感场景的需求。利用移动网络的广泛覆盖，实现了城市区域一致的智能体验。这提升了导航服务从单一设备能力到可操作、可扩展的网络服务的水平，促进了包容性AI在服务于特殊需求群体的场景中的应用。

3.2

移动AI智能手机、可穿戴设备和代理

为了更清晰直观地理解智能终端，促进对AI终端能力演进的统一行业共识，并引导终端系统的智能进化，根据终端在其功能范围内执行任务的复杂程度及其自动化程度，将终端的智能化水平分为五个等级——从一级（L1）到五级（L5）。等级越高，表示终端的智能化水平越高。

L5 完全自主级

具有超越人类专家的能力，实现全面自主。

能够提供与人类专家水平相媲美的定制服务。

L4 专家级别

终端能够全面理解用户指令和意图，具备主动明确意图的能力，拥有强大的推理能力，能够实现自动任务分解和编排；具备动态选择和自动调用工具的能力，至少在一种模式（例如文本、音频或图像）中拥有内容生成能力；并且具有短期和长期记忆能力。

L3 辅助等级

端理解用户指令和简单意图，具备基本推理能力；能够调用预设工具完成单步或明确定义的多步执行任务；在至少一种模式下拥有内容生成能力（例如文本、音频或图像）；并在单次会话中具有短期记忆。

L2 工具辅助关卡

终端理解单个直接命令，根据命令调用预定的工具，并完成单步任务。

L1. 反应层

目前，能够自主完成任务的L3人工智能代理正加速其发展。预计到2030年，终端系统有望达到L4能力。

终端代理 | 智能手机

设备边缘网络云



端智能代理已成为人机交互的核心入口，其应用场景持续拓展。它们呈现出系统级智能代理与领域特定智能代理协作的趋势，推动应用向“意图集中化”转型以及OS框架的AI驱动重构。这标志着从“命令响应”到“意图理解”的飞跃。然而，它们面临着如交互逻辑转变、协议标准化以及跨实体协作等常见的行业挑战。

移动AI，以“设备-边缘-网络-云”全面合作为核心架构，为终端代理构建了高效的运营和生态系统协同系统。系统级代理作为“智能核心”，深度融合终端硬件和操作系统能力。采用纯设备或设备-云混合AI模型，它准确解析用户从语音、文本、手势等多模态意图。结合

利用用户资料和场景特征，系统将复杂任务分解，并基于任务优先级动态分配给纵向领域代理或应用程序代理执行。利用模型量化、剪枝等轻量级技术，系统将大型模型适应终端有限的计算能力，实现高频基础任务的本地实时响应。复杂任务通过移动网络卸载到边缘或云端处理。在网络环境较弱时，系统自动切换至设备端推理以保障服务连续性。同时，系统推动以AI为中心的OS框架重构，将大型模型和代理置于系统中心，封装原子服务能力，并通过统一接口支持多个代理的高效调用。它推动了代理交互协议朝着统一和标准化演进，利用A2A和MCP等协议建立设备互连、数据共享和模型协作的渠道，打破厂商和设备间的协作障碍。系统代理还内置安全可靠的机制，采用数据加密和分层权限控制确保意图执行和数据交互过程中的隐私和安全。该解决方案推动人类-计算机交互范例的根本性转变，实现了从“命令中心”向“意图中心”的转变，提供一体化无缝智能服务体验。它推动终端代理向更高的自主性、个性化和协作性发展，重塑“硬件+智能服务”的业务逻辑，促进开放协作的代理生态系统的繁荣，并加快移动AI在消费者端的广泛应用。

包容智能 | 云终端代理

本地终端代理框架

+

特定场景的生态系统服务



对包容性人工智能的需求正在推动中低端终端的智能化升级。中等和低收入群体迫切需要性价比高、易于使用且可靠的智能技术。让AI民主化有助于弥合数字鸿沟。然而，在不显著增加硬件成本的情况下，让大量中低端终端拥有稳定且可持续演进的智能代理能力，已成为行业面临的核心挑战。

移动AI通过构建云代理架构来应对这一挑战。该解决方案在操作系统层深度集成代理能力，形成一个“本地终端代理框架+场景特定生态系统服务”的架构。轻量级模型部署在终端侧，用于处理常见的意图识别、简单推理和执行任务，确保对高频需求进行本地响应。云端，通过移动网络协调，提供复杂规划、大模型推理和知识增强的能力。利用设备-边缘-云协作和统一界面编排，它实现了跨应用、跨服务代理合作。同时，它集成了多模态交互技术，以提升场景理解和决策效率。此解决方案有效地推动了包容性AI的广泛应用，赋予中低端终端智能服务能力而无需额外硬件成本。在设备端完成高频任务降低了对外部云的依赖，即使在弱网或无网络环境下也能稳定运行，从而提升响应速度和体验一致性。跨应用协作打破了体验碎片化，促使代理从被动响应转变为主动和情境化服务。这有助于缩小数字鸿沟，丰富扩展终端生态系统中智能服务的供给。

终端设备 3 轻量移动人工智能代理eSIM连接

轻设备

重云



全球大规模部署轻量级智能终端面临着成本效益和服务稳定性双重要求。随着大型语言模型和多模态人工智能技术的发展，传统的功能性硬件正在向能够理解和推理的轻量级智能终端演进。然而，一个核心行业挑战是确保在全球化部署过程中稳定且连续地交付智能功能，同时不显著增加硬件成本。

移动AI通过结合“轻设备+重云”架构和全球eSIM连接来解决这一问题。核心概念是在云端集中部署高计算能力的任务（如多模态交互、大模型推理和知识库检索），按需、安全、低延迟地通过移动网络将任务交付给终端。在终端侧，符合GSMA标准的eSIM功能集成到蜂窝通信模块中，实现全球自动网络连接和服务访问。这消除了对高性能GPU或NPU的需求，通过计算卸载控制硬件成本。利用eSIM的智能多运营商切换功能，终端可以自动匹配最佳网络路径和最近的推理节点。该解决方案实现了低成本与全球稳定服务的统一，显著降低了终端硬件和功耗成本，使即使是性价比较高的消费电子产品也能提供高端智能体验。eSIM的全球连接能力确保了跨区域和多种网络环境下设备的连续服务，满足了AI驱动实时交互的低延迟和高稳定性要求。这推动了移动AI代理终端的全球化民主化。

合作与 4 多设备功能扩展

核心终端

边缘外设

云支持



轻量级终端面临着平衡资源限制与智能功能需求挑战。生成式AI的兴起正在推动终端向智能代理进化，儿童智能手表等可穿戴设备成为便携式交互中心。未来的场景将涵盖多种形态，如智能眼镜和耳塞等。然而，受限于尺寸、电池寿命和计算能力，单一设备难以处理复杂的AI推理任务。在这些限制条件下，实现设备与云的协作以及共享计算资源是必要的。

Mobile AI is based on a "core terminal + edge peripherals + cloud support" distributed agent system architecture. Using a smartwatch as the computing and connectivity hub on the device, lightweight multimodal models are deployed. These models collect visual, audio, and biometric data from various peripherals through short-range communication. On the cloud side, AI capabilities are leveraged to provide complex services such as in-depth companionship and learning assistance. Dynamic scheduling optimizes the energy efficiency of the "hub and peripherals," while device-cloud collaboration distributes the computational workload. Additionally, the system also supports the local storage of

代理能力，确保在多种场景下服务的连续性。此解决方案有效克服了轻量级终端的物理限制。分布式架构延长了整个生态系统的电池寿命，多模态数据协作实现了自然和精确的交互体验。在网络连接差或无网络连接的场景中，核心终端仍可驱动基本服务，保持隐私和安全的闭环。这促进了可穿戴设备从单功能工具向协同智能系统的演变，为全天候主动服务奠定基础，并促进了移动AI在可穿戴场景中的应用实施。

5 摘要

移动AI从整体架构视角来看，围绕“设备-边缘-网络-云”协作展开。通过终端和代理能力的分层部署和动态调度，它使不同形态和性能水平的终端在其各自的限制内获得稳定且可持续的智能体验。终端侧承担核心能力，如安全的闭环操作和即时响应。网络侧通过广覆盖连接和QoS保障确保稳定的服务交付。边缘和云侧集中提供复杂的推理和知识增强功能。这三层共同构成了支持代理和终端运行的底层系统。

考察典型的应用实践，尽管移动AI在不同类型的智能体上的焦点不同，但共同需求高度一致。对安全敏感的情景主要确保关键决策的可靠性。工业协作控制情景则侧重于多终端并发和确定性延迟。资源受限的终端依赖于设备-云协作来平衡体验和成本。大规模部署情景通过系统集成实现智能能力的快速覆盖和一致交付。

4 移动人工智能应用

移动人工智能利用其核心特性，即无处不在的智能、无处不在的连接以及分布式智能协作，突破了设备和空间上的限制。它推动了智能应用在消费者（ToC）和企业（ToB）领域的深度渗透。在消费者领域，通过扩展人机交互的维度，革新沉浸式体验模式，它将诸如增强现实/虚拟现实、实时语音翻译以及个性化生活服务等智能功能融入到日常场景中。这能够精准匹配用户需求，显著提升数字消费体验和生活质量。在企业领域，它利用其高可靠性和低延迟的技术优势，赋能智能制造、交通与能源以及医疗保健等垂直行业。通过优化生产流程、实现精确调度以及促进协同创新，它显著提高生产效率和工业韧性。本章将重点关注消费者和企业领域中的关键行业，探讨移动人工智能如何通过其移动、无处不在的智能能力，解决传统行业中的痛点，并推动工业的数字化和智能化转型。

4.1 用户智能消费体验



当前基于设备的智能体验面临着明显的痛点：高端智能服务受限于设备能力和空间范围，导致普遍访问困难。此外，人机交互的维度仍然较窄，场景边界也较为僵硬。依赖无处不在的连接技术和深度分布式智能协作的移动AI，有效克服了这些瓶颈。它不仅扩展了交互维度和场景边界，还使得高端智能体验能够超越设备和空间限制，实现普遍访问和无处不在。这最终引领进入一个以更沉浸式感知、更精确的服务和更高效的生活为特征的新智能阶段。

数字化消费升级与沉浸式体验创新。基于实时渲染、多模态融合技术，以及一个技术架构，该架构强调云端大型模型与终端之间的高效协作，移动人工智能也优化了智能服务的交互逻辑。它将增强现实/虚拟现实等沉浸式应用真正融入日常生活，构建一个虚拟与现实融合的社交娱乐和数字购物新生态。它实现了实时语音翻译、上下文图像识别和解说等智能功能的实际应用，打破了语言沟通和认知障碍。这使得复杂的智能服务能够以更自然、更便捷的方式触达用户，从而重塑数字消费体验的范式。

提升社会效率和提升个人生活质量。移动AI利用其广泛的连接性，结合上下文感知和意图理解技术，连接不同领域的物联网设备并构建解决方案。一方面，它精确地辨别用户需求并提供定制化服务，如在移动办公场景中的会议纪要和多语言转录，以及智能移动中的动态路线优化。另一方面，它自动化处理跨场景任务，涵盖从智能家居调整到工作流程优化的方方面面。这最终帮助用户摆脱重复性任务，专注于创造性活动，有效提升社会运营效率和个人的生活质量。

多样化的终端开发与主动智能服务的发展。移动AI运用计算卸载技术突破“大屏为中心”的限制。依托多模态传感器和自主决策系统，它填补了将数字智能应用于物理场景的差距。同时，利用设备-边缘-云的闭环和跨终端协作技术，它承担起无缝连接个人AI设备（如手机、智能眼镜）与物理AI载体（如智能汽车、类人机器人）的核心任务。这成功构建了以用户为中心的数字化服务层，形成围绕用户的协调设备生态系统。它推动智能代理从被动响应进化到主动感知和需求预测，实现全面适应全场景智能需求的前瞻性智能服务体验。

4.2

智能制造



智能制造对移动智能应用有迫切和严格的要求。生产现场对智能应用的需求至关重要，如设备预测性维护、增强现实辅助装配、移动检查和灵活排程。然而，传统的基于云的AI模型面临带宽和传输成本的压力。毫秒级的延迟可能带来安全风险。加之工厂的电磁干扰和数据隐私保护要求，单纯依赖公共云服务难以满足这些场景需求。

移动AI采用“设备端推理+边缘协作”核心架构。终端利用内置NPU计算能力，在毫秒内完成缺陷识别和异常警报，无需依赖云端。敏感数据在本地或边缘MEC进行清洗和推理，仅上报结果数据，稳固建立隐私和安全保障。对于网络覆盖薄弱的复杂车间环境，终端支持基本识别和记录功能，网络恢复后同步数据。该架构也适应工业电磁环境，增强技术稳定性和抗干扰能力，以精确匹配生产场景需求。该解决方案有效满足智能制造的实时性和安全性要求，降低带宽消耗和数据泄露风险，提高网络连接不良场景下的可用性。它推动生产过程从“辅助操作”向“无人操作”进化，促进自主移动机器人和智能代理终端之间的协作，为“无人工厂”和适应性灵活制造奠定基础，并加速制造业的数字化和智能化转型过程。

4.3

城市治理



城市治理正在向“智能化”系统转型，移动解决方案的应用场景十分广泛。从以网格为基础的移动执法和动态环境监测到应急响应，各种移动终端已成为治理的关键支持。然而，大规模非结构化视频数据的全回程传输容易导致网络拥堵和存储压力。此外，公共场所的数据收集涉及到公民隐私，使得在收集点进行实时脱敏成为核心挑战。

移动人工智能通过“计算能力前置”解决这些治理痛点，将智能处理能力延伸至治理操作的边缘。配备人工智能识别功能的终端仅在检测到异常事件时捕捉和上传证据，大幅减少无效数据传输。在数据收集源头，设备端的人工智能能够自动模糊处理个人隐私信息，实现预防性隐私风险控制。检查终端转变为能够自动生成和分发工作单的智能代理，简化了手工录入流程。同时，利用边缘计算支持确保快速本地数据处理和响应。此方案有效降低了网络和存储成本，加强了隐私和安全防御，显著提升了城市治理事件处理的效率。它推动城市治理从“感知与发现”升级到“认知预测”，从被动事件后响应延伸至主动事件前预防，为构建精细高效的现代城市治理体系提供核心技术支持。

智能交通必须平衡安全与效率，应用范围涵盖车辆、道路、人员和事件。对智能联网车辆、路边感知协作和物流跟踪的需求日益迫切。然而，高速车辆移动会影响信号稳定性；隧道和山区等弱网络环境存在安全隐患；将车辆产生的数以千计的数据完全上传到云的成本高昂且不切实际，成为行业升级的主要瓶颈。

4.4

智能交通



移动人工智能通过车路云一体化的架构解决交通运输场景中的痛点，构建了全面的安全和效率保障系统。车辆依靠车载边缘智能，利用强大的本地计算处理驾驶中的感知和决策，确保即使在无网络或弱网络条件下，也能执行紧急制动和障碍物避让等核心安全功能。用户智能手机作为移动人工智能的入口，能够无缝地将导航和娱乐信息传输到车辆信息娱乐系统，还可以作为数字车钥匙实现无钥匙进入。利用C-V2X（蜂窝车联网）技术，车辆和路边单元进行短距离通信和边缘计算，实现超视距感知，有效弥补了独立车辆智能的局限性。这一方案显著提升了智能交通的安全性和交通效率，为自动驾驶的发展奠定了基础。在网络环境较弱的情况下，安全保障能力得到显著增强，全场景互联体验不断拓展，个体车辆感知盲点得到有效缓解。它推动了自动驾驶从L2+向L3/L4级的演进，帮助车辆成为独立的智能体，通过集群协作优化交通流量和缓解拥堵，催生“无人驾驶出租车”等新型移动即服务（MaaS）商业模式。

4.5

智能医疗



智能医疗关注院外和初级护理场景，解决资源分配不均和应急响应效率低下的问题。核心需求包括5G+AI赋能的紧急服务、移动辅助诊断和个性化健康管理。挑战包括医疗影像文件体积庞大、应急场景中网络波动影响远程会诊、初级护理医生缺乏专家级诊断能力以及医疗数据传输的严格安全要求。

Mobile AI通过终端赋能、数据优化和主动干预构建了一个全面的医疗智能系统。AI辅助诊断终端集成了算法。

为了帮助基层医生实时捕捉最佳检查视角并自动测量关键指标，提高诊断准确性。在院前急救中，边缘AI设备预处理和压缩生命体征数据，优先传输关键警报信息，而车载AI模型提供即时急救建议。可穿戴设备依靠设备上的模型分析数据趋势，提供早期健康风险预警，同时通过安全传输技术确保医疗数据隐私。该解决方案有效提升了医疗可及性和应急响应效率，推动了健康管理模式的转型。它显著缓解了医疗资源分布不均的问题，降低了基层误诊率，打造了无缝的“从救护车到医院入院”的紧急生命通道，将重点从“治病”转向“防病”。它为解决老龄化社会医疗资源短缺问题提供技术支持，未来有可能培养出AI赋能的家庭医生，进一步扩大智能医疗的覆盖范围和服务深度。

4.6

能源与电力



能源行业面临着复杂且富有挑战性的检查和维护场景。核心要求包括通过无人机/机器人进行自动化检查，对电缆和设备进行精细的质量检查，以及智能电表读数与能源效率管理。然而，现场设施的网络覆盖薄弱，使得高清视频的实时传输困难；传统的手工检查效率低下且易疲劳；高压电磁环境对通信抗干扰能力提出了严格的挑战。

移动AI以设备智能和场景适应为核心，为能源行业打造专用解决方案。便携式AI驱动的质量检测一体化机器，结合5G-A和视觉算法，能在弱网或离线环境下进行360度电缆成像和设备推理，准确识别细微缺陷。配备边缘计算模块的检查无人机实时分析图像，仅在检测到异常时保存证据并触发警报。防爆机器人替代人在高风险区域工作，利用设备AI读取仪表盘并检测热成像异常，同时具有加强抗干扰设计以适应高电压环境。该解决方案大大提高了能源行业检查和维护的效率与安全性。与传统的手工方法相比，检查效率成倍提升，显著降低了遗漏缺陷和人员安全风险，同时适应现场弱网和高电压电磁条件。它推动了能源系统向“自愈电网”的演进。未来部署微型移动AI传感器将实现自主故障定位和隔离，持续提升能源系统的弹性和智能化水平。

4.7

摘要

从整体建筑学角度来看，移动人工智能依赖于“设备-边缘-网络-云”全方位协作的核心支持。利用其“三层四维”框架的协同效应，它将基础设施层的网络和计算能力以及执行层的终端执行能力转化为应用层特定场景的价值。设备侧专注于实时感知和本地响应；边缘侧处理区域协作和低延迟处理；云侧承担复杂计算和全球优化；网络侧提供无处不在的连接和精确调度。共同构成了应用实施的完整技术支持系统。

审视典型应用实践，不同行业的移动AI需求侧重点各不相同，但核心需求高度统一。智能制造关注低延迟和数据安全。城市治理强调高效数据处理和隐私保护。智能交通依赖于车路云协同和安全保障。智慧医疗强调基层资源分配和应急响应。能源电力强调适应恶劣环境和操作安全。共同之处在于，它们都利用技术解决特定场景的痛点，实现效率提升、成本优化和服务升级，从而推动各行业从传统模式向数字化和智能化范式转变。

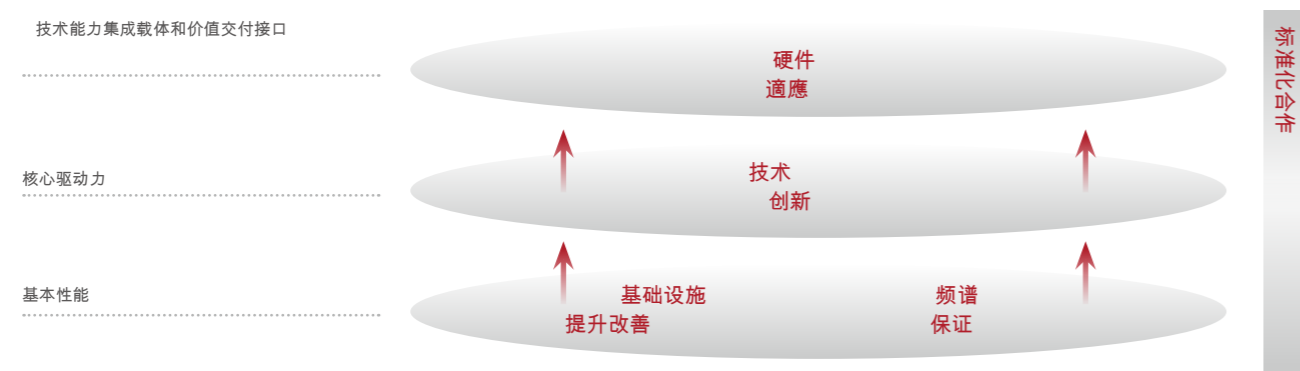
03

向移动人工智能



为了实现移动人工智能的大规模部署和价值实现，建立一套系统化、多层次的支持体系至关重要。基于产业发展需求和科技发展趋势，本章重点探讨了五个关键路径：基础设施演进、核心技术突破、终端生态协同、频谱资源保障和协同标准化发展。它系统地概述了将移动人工智能从技术概念过渡到工业应用的全面框架和潜在解决方案。具体而言，这些路径包括：基础设施提升、频谱保障、技术创新、硬件演进和标准化制定。这五个相互关联且协同进化的维度共同构成一个完整的系统，支持将移动人工智能技术能力转化为工业实践。

基础设施提升提升了网络和计算能力的根本性能，为具有高可靠性、低延迟和无处不在的连接特性的移动人工智能提供了物理层。频谱保障通过频谱资源的科学规划和动态管理，为多样化的AI服务提供了必要的无线传输资源。共同构成了移动人工智能运营的资源与能力基础。技术创新作为核心驱动力，将基础资源转化为系统能力。以AI和网络的双向赋能为中心，通过设计AI原生网络架构并结合端到端的安全可信控制，成为连接基础设施和上层应用的关键使能层。硬件适配作为技术能力的综合载体和价值交付的接口。通过硬件升级、产业协作和生态共创，将网络能力、计算能力和智能转化为用户可感知的服务，直接影响移动人工智能的体验质量和应用规模。标准化协作贯穿于上述所有方面。通过建立统一的技术规范、接口协议和评估系统，促进产业链各层级的互操作性和高效协作，降低集成复杂性，加速整体解决方案的成熟和采用。



1. 基础设施提升

基础设施提升是大规模部署移动人工智能的关键基石。它旨在通过提升基础通信网络能力及优化计算资源分配，构建一个以“自适应连接、充足计算能力和高效协作”为特征的底层保障系统。这正好满足了移动人工智能多样化场景的差异化需求，克服了传统基础设施在性能、容量和灵活性方面的限制。

提升基本通信网络能力，以动态适应多样化的场景需求。通信网络必须从“通用连接”向“场景定制连接”发展，围绕四个核心维度进行升级：高带宽、低延迟、海量连接和高可靠性，同时建立动态适应机制。通过多频段协调和载波聚合增加传输容量，以支持人工智能大模型数据传输和AR/VR等场景的需求。优化网络协议栈，引入边缘节点，以减少智能网联汽车和远程医疗等场景的端到端延迟。创新技术提高连接密度，以满足数百万终端的并发接入需求。通过冗余传输和链路自愈确保关键行业场景的可靠性。构建业务感知和资源调度联动系统，以动态匹配不同服务的实时需求和优先级。

计算能力提升以适应全场景计算需求。解决移动AI在设备、边缘和云端的差异化计算能力需求，构建了一个分层、协作的“设备-边缘-云”计算能力供给系统。云端扩展高性能计算集群以支持大规模训练。边缘侧增加节点以增强低延迟响应能力。设备侧通过专用AI芯片加强本地轻量级模型的执行。将异构计算设施集成，形成统一资源池，并预留弹性扩展空间以适应计算能力增长。建立智能能耗管理、冗余备份和计算能力监控机制以确保稳定供应。构建多样化的“通用+专用”计算能力系统以适应不同场景需求。智能调度算法实现大型模型的动态分配，提高计算资源利用率和服务响应速度。

2. 波段保证

频谱是承载移动人工智能服务数据的基础资源，也是确保无处不在的智能体验的关键前提。为了满足人工智能大模型交互和多模态数据传输等场景对网络容量和确定性的极端需求，频谱策略必须采取双轨并行演进的路径，同时追求“新频段规划”和“现有频谱优化”。

前瞻性地规划新中低频和高频乐队资源

推广中频段，包括6GHz，使之成为全球统一的新主要频段，提供连续的大带宽，主要用于保证分布式人工智能推理和高清机器视觉等服务的上行需求。同时，积极探索毫米波等高频段在热点地区的部署，利用其超大型容量和超低延迟特性，助力沉浸式XR和全息通信等前沿应用。

持续挖掘现有潜能频谱资源

通过引入人工智能驱动动态频谱管理、智能频谱聚合和其他解决方案，实现跨频、跨无线接入技术（RAT）资源的灵活调度和高效共享。这解决了如体育馆和智能工厂等场景中人工智能终端大规模并发接入的挑战。结合大规模MIMO等先进技术，可以显著提高频谱重用效率，从而在频谱资源有限的情况下，最大化网络整体容量和用户体验。

3. 技术创新

技术创新是实现移动AI的核心驱动力。它将基础资源转化为系统能力，为移动AI应用的实施提供高效支持。移动AI技术革新体系建立在原生AI网络架构之上。它发展AI向网络和网络向AI的双向技术体系，通过终端异构智能计算和多元代理协作等技术开发高效终端承载，并通过包括多模态交互和基于场景的解决方案在内的应用创新，扩大实施场景，由端到端的安全技术框架加以保障，实现网络与AI的原生共生和双向赋能。



图5：移动人工智能创新技术图谱

由于它涵盖了广泛的科技领域，本章将重点讨论与AI原生网络架构、AI用于网络、网络用于AI以及安全相关的核心内容。

3.1 人工智能原生网络架构

移动AI的原生网络架构是网络与AI实现双向赋能的基础。其核心目标是支持原生AI功能，提升网络性能和效率，并为商务应用提供无处不在、高效的通信连接以及新型服务（如计算、数据和模型）。

在网络AI方面，AI作为固有基因从最初的设计阶段就融入其中，深度融合到网络功能设计和界面交互定义。其核心特性主要包括三个主要特点：

分层集中控制 企业级集中控制和意图理解在核心网络中实施，而区域级实时控制在无线接入网络中执行，从而实现精确而灵活的多级网络管理和控制。

功能分类与聚合： 网络功能根据数据和处理等维度进行划分。它们分别承担类似所有元素数据处理等任务。

生产与消费，以及计算网络融合的集成控制与执行，协同提供高质量数据与计算能力支持人工智能模型。

弹性按需部署： 分布式执行功能根据业务场景需求动态部署，实现高效动态的资源分配，提高整体资源利用率。

在网络人工智能领域，AI原生网络架构将作为一个统一的平台，整合通信、感知、计算和智能，为外部提供智能、综合服务。基本网络能力封装成标准化的服务，并通过统一的接口公开。AI应用可以通过集中控制平面灵活调用所需服务。网络根据业务性能需求，灵活地将服务实例部署到最优节点。由此形成的“服务暴露——能力调用——邻近执行”模式可以有效支持人工智能代理和具身智能等新兴应用对严苛的低延迟和高计算能力的需求。

3.2 人工智能网络关键技术

接入网络人工智能

AI for Network aims to deeply integrate AI into the end-to-end network system, making it a key production factor for networks, and to build an endogenously intelligent network capable of closed-loop "Perception-Cognition-Decision-Execution" capability across the entire chain. Its technological innovations span all domains of the end-to-end system, including the air interface, core network, security, and cross-domain collaboration.

AI for Network旨在将人工智能深度融入端到端网络系统，使其成为网络的关键生产要素，并构建具有端到端“感知-认知-决策-执行”闭环能力的内生智能网络。其技术创新贯穿端到端系统的各个领域，包括空中接口、核

空中接口AI：

通过引入深度神经网络，网络能够对复杂信道环境进行自适应学习和优化，显著降低空中接口开销并提高频谱效率。其核心技术能力包括：

信道状态信息 (CSI) 反馈优化

人工智能算法可以智能预测和压缩信道状态信息，大大减少反馈开销，提高频谱利用率。

智能光束管理

人工智能算法实现智能光束选择和管理，快速在复杂多变的信道环境中找到最佳光束配置，从而提升连接质量和覆盖范围。

链路适配优化

人工智能算法可以根据实时信道状况和服务需求动态调整调制、编码方案和功率控制等参数，优化链路性能。

增强上行链路新技术：

移动AI中的“设备端感知、云/边缘端推理”端到端闭环对上行带宽和延迟提出了严格的要求。增强型上行新技术构成了移动AI实现设备-云协作的物理基础。它们的核心技术能力包括：

频谱重构

通过灵活的频谱分配策略，更多频谱资源被分配给上行链路，从而提高了上行链路容量。

多频段协调

上行/下行时隙比例根据业务需求动态调整，在移动AI场景中增加上行时隙的比例，以满足上行数据密集型传输的需求。

多频段频谱资源通过多频段载波聚合技术聚合，进一步增加了上行带宽。

通过增强型上行链路新技术，网络可以打破上行链路瓶颈，确保终端侧的高价值数据能够及时上传到云计算中心，满足移动人工智能端到端闭环的延迟需求。

无线网络代理：

核心目标是实现无线域内的自主闭环。依靠本地域数据和全球政策输入，它进行近乎实时的分析、推理和决策，支持无线网络的自主运行、维护和优化。其核心特性体现在两个方面：

单域自主，通过管理控制一体化的智能引擎，实现了局部网络场景识别、风险预测与预防、故障根本原因分析以及端到端自动化运维优化。

自主闭环代理，针对特定场景如运营维护、节能和体验保障，构建“感知-分析-决策-执行”全流程自主能力。

这项技术的关键驱动因素是通信大型语言模型和无线接入网络数字孪生系统（RDTS）。该通信大型语言模型经过大规模电信语料库和领域知识的预训练和微调，具备意图翻译、多目标协同优化和主动趋势分析等能力。RDTS通过构建所有网络元素的多元维度数字映射，实现对物理网络的实时数字表示。它提供高精度的建模、策略模拟和验证，以及全链可追溯能力，为无线领域的智能决策提供准确的支持，并显著提高网络自主性水平和决策的可信度。

核心网络人工智能

核心网络充当服务策略控制和网络拓扑的中心。为适应个性化的服务需求和动态资源优化，核心网络迫切需要智能升级。它需要支持网络元素和网络智能化，同时利用数据和孪生技术巩固这一变革的基础。

AI原生网络元素

传统的核心网络政策决策逻辑基于静态规则，无法完全满足用户个性化的多样化需求，也无法实现网络资源的最佳实时动态配置。有必要引入由AI模型驱动的推理和决策。通过将AI原生嵌入到网络元素的函数逻辑和界面交互中，可以在通用场景中实现网络的自适应能力。网络元素中的AI原生不仅体现在AI能够部署到每个网络元素的感知、推理/决策（例如，服务特征分析或政策调整）层面，还体现在可扩展到网络级资源调度和配置（例如，动态切片或网络组成）层面。为确保网络操作的高可靠性，网络元素中的AI原生还体现在网络内AI模型的自主性。尽管离线预训练模型获得了基本的泛化能力，但训练数据样本（例如，类别不平衡）中的不可避免的质量偏差和在部署环境（例如，用户或网络资源分布的差异）中的特征变化，需要AI模型能够使用实际网络部署环境中的数据样本进行周期性或基于实时反馈的在线学习更新。这持续优化模型性能和场景适应性，确保网络中运行的模型推理效率和准确性。

网络代理化

通信网络中的中枢神经体系，承担服务订阅管理、策略执行和网络元素协调等核心功能，核网可基于基础能力演变为网络代理商。这使得它能够理解用户/服务需求、进行复杂任务计划、工具调用及自动闭环执行。与通用人工智能代理商相比，核心网络代理商具有具体的能力要求；环境管理引入独立的数据平面和先进的检索增强生成（RAG）技术提取关键信息；策划与推理需要具备实时、可靠的远程任务策划、调度和闭环迭代的能力；使用工具涉及网络元件函数的模块化重组织，通过有限解码提高调用精准度；互操作协议涉及制定网互协议（NetMCP）以实现网络能力的统一抽象和标准化调用。此外，需要引进多智能体合作技术以分解复杂任务。同时，电信行业正在推行A2A-T（电信智能代理商间）协议的标准，以构建跨领域、跨 vendors 合作框架。这个框架将自然语言的商务意图解析为网络全范围的协作任务，推动网络迈向以意图驱动、生态协作的智能转变。

核心网络运营大型模型

通过实时分析和处理网络运营数据（如流量、信号等），大型模型可以全面感知关键信息，包括服务质量、网络负载状况和用户行为。利用其强大的特征提取和相关性分析能力，它们可以突破传统的基于规则的判判决策方法，实现高效的网络运营和提供高质量、个性化的网络服务。与小型网络AI模型相比，大型模型在泛化和准确性方面具有显著优势。一方面，通过“基础模型+局部微调”的方法，它们实现了跨区域应用的快速迁移和适应。统一的模型框架大大降低了重复模型开发的成本。另一方面，利用模型的巨大参数数量和高知识含量，它们可以整合多维网络信息，实现全面优化决策，使网络运营更加精细和准确。

网络操作数据平面：

数据平面是支撑核心网络智能协作的核心数据枢纽，承担着管理整个数据生命周期并释放其价值的关键责任。为了更好地提升网络运营效率，最大化网络数据价值，有必要通过实时数据收集和智能分析技术构建综合数据平面架构，将网络数据深度转化为推动业务闭环运营的核心资产。网络运营数据平面需要解决四大技术挑战：

高效数据收集：

促进多源数据的按需订阅、去重和协调，集成来自网络元件、终端、服务平台及其他全场景来源的数据。

实时数据处理：

通过智能清洗、实时聚合、动态脱敏和其他技术提高数据质量。

高效存储与检索：

通过分层/分类存储和多维度关联技术，支持高效访问大量结构化和非结构化数据。

安全、可信、开放共享：

通过权限控制和标准化接口实现安全合规的数据共享。

数据平面实现网络数据的有效供应，减少冗余数据收集，解决数据孤岛问题，推动网络从“规则驱动”向“数据与AI自适应”转变。这有助于运营商转型为全面的“网络+计算+智能”服务提供商，巩固移动AI从融合应用到原生互惠共生演变的数据基础，并加速数字经济的高质量发展。

网络数字孪生与服务孪生：

核心网络人工智能将依赖于核心业务数据，如服务体验数据、配置数据和网络状态数据，来构建网络数字孪生模型。这些模型准确映射物理网络的拓扑结构、资源状态和服务流量路径，实现了数字空间中对物理网络的精确表征。根据不同的建模数据域和虚拟-物理连接目标，它们可以分为网络数字孪生和服务数字孪生。

网络数字孪生：通过构建动态网络的AI仿真系统，它深度融合了终端行为、设备操作和故障演化的多维度子模型。结合任务并行调度和统计向量优化等技术，提高了仿真效率和场景逼真度。基于网络可用路径求解算法，它能快速定位故障网络元素和路径，推动运维从“被动感知服务退化”升级到“主动隔离故障网络元素”，为网络高稳定性和运营效率提升提供核心支持。

服务孪生：以大数据统计挖掘和大型模型模拟预测作为双核基础，专注于服务体验分析、高速铁路场景分析和媒体中继等多种商业场景。它深入关键能力提升路径，如识别服务体验瓶颈、服务配置优化和资费计划推出评估，探索构建“精准网络规划/建设和高效服务实施”的迭代增强生态系统。其核心包括空间时间多维业务分析基础的技术优化、AI智能孪生基础核心算法的研究、大规模部署实施路径、多维核心业务视图和综合分析系统建设方案，以及解决平台触发评估、数据订阅和报告、智能预测展示的全流程闭环管理技术挑战。它为业务创新迭代和稳定收入增长提供坚实的理论支撑和成熟的技术解决方案。

人工智能协作 终端-边缘-云

终端-边缘-云协作的核心目标是实现业务需求与网络和计算资源实时状态的优化平衡，从而在终端、边缘和云之间动态选择最合适的执行位置。基于5G-A/6G网络，终端可以与边缘或云AI协作，实现复杂任务的分布式协同计算，将终端转变为集环境感知、网络协同编排、本地处理和结果展示于一体的关键整体智能中心。终端-边缘-云协作的核心技术包括协同推理、协同微调和网络辅助卸载决策。

协作推理：协作推理包括推理卸载和分割推理。推理卸载：终端设备通常首先进行预处理，然后上传压缩特征到边缘或云端进行主要推理任务，降低链路负载并提高整体效率；通过在模型结构中选择合适的分割点，分割推理实现了终端-边缘协同执行，达到了延迟和准确度的动态平衡。

• 协作微调：协作微调是指涉及复杂训练和微调的场景。通过在终端侧执行轻量级更新，

完成边缘或云端的复杂度高训练，不仅能降低数据泄露风险，还能提高模型迭代效率。

• 网络辅助卸载决策：高效的终端-云-边缘协作需要综合考虑排队延迟、链路质量、终端电池和热状态、隐私域约束以及业务质量体验等多维因素。这些因素需要在任务执行过程中持续更新，以确保推理卸载策略保持最优。在移动场景中，系统可以根据终端位置和网络条件将推理会话迁移到更近的节点，或在必要时回退到本地推理，以最大化服务连续性。

3.3 人工智能关键技术研究网络

优化 1 代理连接

作为通信网络中的新型终端和应用，智能代理的大规模普及正在推动通信协议的升级和网络架构的变革。网络迫切需要引入新的技术系统以满足代理通信对高并发、高性能、低延迟和高安全性的需求。代理通信技术系统涵盖了连接优化、计算赋能和可信管理等多维技术能力。

智能代理连接优化技术的目的是满足新型交通模式、新型连接模式和新型网络方法中代理交互的需求。这包括设计新的服务质量（QoS）机制和协议设计、流量识别和特征分析、以及动态多模式流量传输技术，以满足新信息元素（例如，AI模型、令牌）的传输需求、新的交通方向（例如，代理之间的东西向交通）以及新的交通特性（例如，多模式突发）。设计了注册/发现和按需广域连接等新机制，以支持以任务为导向的点对点网络连接。构建了一种灵活的任务驱动多代理网络机制，以实现异构代理之间的群体智能协作。

赋能 2 计算

计算赋能技术旨在通过使网络能够提供新的数据和计算服务来赋能智能。这包括设计一个端到端网络协作计算服务框架，以支持计算资源的管理、选择和卸载。它将代理计算任务选择并卸载到终端、网络、边缘或云端的适当计算资源，以满足代理通信的高效和大规模计算需求。通用计算资源协作调度技术支持在复杂的多维约束（如延迟、吞吐量、资源利用率和计算负载）下，对计算资源进行高效和协作的调度。构建了一个端到端统一的数据服务框架，以支持全生命周期数据管理和多模态异构数据传输，赋能代理AI模型训练和数据共享等需求。

管理 2 可信的代理

代理可信管理技术采用统一身份识别系统，将属于同一用户的多种形式的代理进行关联和管理。这包括适用于具身智能/代理应用的统一身份机制、代理-用户关联映射机制，以及基于任务粒度的代理权限动态授权机制（而非静态订阅），以实现可信且安全的代理通信。

3.4 网络人工智能安全关键技术

安全技术 1 多代理系统

移动AI的安全包括两个方面：首先是如何确保移动AI系统的安全性，其次是如何利用移动AI服务于安全操作。

多智能体系统的安全保护与传统的通信系统安全保护有根本的不同。随着这些系统被部署在各个领域，它们的复杂性和攻击面持续扩大。需要防范诸如直接和间接的即时注入、模型中毒以及智能代理AI工具调用的漏洞等风险，这些风险是移动人工智能面临的。

已成为行业的核心关注。当移动AI拥有智能体特征时，多智能体协作增强了系统灵活性和能力，但它也允许攻击者利用智能体之间的交互关系发起复杂攻击。攻击面从单个智能体扩展到系统协调级别。多智能体安全解决方案覆盖了智能体的整个生命周期，从创建到退役，其核心涵盖设计、开发、部署和操作四个阶段。根据“左移安全”和“持续保证”的治理原则，安全措施被提前到设计和开发阶段，并在整个生命周期中持续监控和评估其安全合规性。基于OWASP多智能体系统威胁建模指南，MAESTRO（多智能体环境、安全、威胁风险和结果）框架可以用于系统地分析多智能体协作通信环境中的安全威胁、风险和潜在后果。这有助于有针对性的对策，有效地增强移动AI系统的安全性和可靠性。

For Security Assistance 移动人工智能技术运营

生成式AI在网络安全领域得到广泛应用，以提升防御能力，并逐渐融入各种安全产品中。在专业大型模型领域，已出现针对安全防护的专用模型和用于攻击测试的模型。同时，生成式AI正从通用辅助工具演变成成为专门的安保代理和解决方案，推动网络安全智能化。生成式AI显著提高了安全运营中心（SOCs）的效率，在那里它已用于自动或半自动完成警报分类、威胁情报分析、事件响应和规则生成等任务。由AI驱动的网络推理系统可以自动发现和修复源代码中的漏洞。随着通信领域模型能力的提升，它们可以降低安全测试的成本和门槛。未来，将可实现更频繁、主动和全面的网络安全测试和补丁管理。对于防御者来说，这将对安全运维模型带来重大变革，推动网络安全进入以智能化和高频为特点的新阶段。

4. 终端进化

终端是移动人工智能服务连接用户和产业的最终载体。其能力的演变和生态系统的成熟度直接决定了无处不在的人工智能体验的边界和商业化的规模。终端支持的核心目标是通过升级硬件能力的升级来适应多样化的场景需求，通过行业协作构建标准化生态系统，推动移动人工智能终端从试点探索到大规模商业部署，并在终端层面为整个产业链的实施提供坚实支撑。

硬件升级以适应所有移动人工智能场景下的差异化需求

移动AI应用的多元化对终端硬件提出了超越传统通信需求的全方面要求，促使向多模态感知、全频段可靠连接和高效本地智能处理进化。终端需要集成和高效处理多模态数据，如语音和视觉，平衡传感器精度、功耗和集成密度，为各种AI应用提供统一的感知基础。现有和未来的移动通信频段需要全面优化，加强5G-A支持，同时确保与新的6G频段兼容。通过5G与Wi-Fi和卫星通信的智能融合构建广泛覆盖。应深化AI模型轻量化和专用AI处理单元的集成研究，推动推理和决策能力迁移到终端，形成一个支持本地实时智能和减少对云的依赖的智能分层系统。

创新端型，以释放移动AI在不同场景中的服务潜能。

为满足具身AI和包容性AI的发展需求，终端形态正朝着多样化、协作化和普及化方向发展，打造全场景自适应终端系统。我们优先考虑具身智能终端的创新，以用户为中心的终端研发

全产业链合作，助力移动人工智能终端大规模商业化

聚类和轻量级泛在智能终端，丰富终端产品矩阵。同时，我们以多形态终端为载体，培育全新服务模式，通过终端形式创新推动业务创新，激活移动人工智能在各领域的应用价值。

终端制造商、芯片公司和运营商需要建立深度合作机制，以解决如技术标准碎片化等瓶颈，构建“统一标准、共创功能、共享生态繁荣”的产业格局。这三方应共同制定移动AI终端技术规范，明确如功耗、延迟和连接可靠性等核心指标，并统一接口协议和适配标准，以降低产业链上的协作成本。应建立跨行业创新平台，共享研发成果和场景需求，形成“芯片-设备-网络-应用”的协同创新链。运营商发挥其网络和场景优势，而终端制造商和芯片公司则专注于硬件研发，推动软硬件和网络能力的深度融合，加速终端技术的成熟和商业推广。

5. 标准配方

标准化统一是打破移动人工智能产业合作壁垒、促进规模化发展的制度保障。通过标准化技术接口、统一应用规范和明确评估维度，它降低了产业链合作成本、加快场景实施、确保用户体验一致性，并为产业的健康有序发展提供坚实支撑。

推广全球统一的行业标准，以促进行业共识

聚焦于移动人工智能跨层、跨设备和跨场景的协作需求，应努力推动全球统一的标准化技术体系建立。关键领域包括标准化终端、边缘、网络和云之间的互联互通接口协议，定义多模态交互框架和数据传输规范，以及为AI代理、具身智能和智能网联汽车等复杂应用提供统一的技术语言。通过国际标准化组织和行业联盟的共同努力，可以形成全球技术共识，消除由技术碎片化导致的协作障碍，确保不同厂商产品之间的兼容性和适应性，降低产业链上的研发和适配成本，并提升全球行业协作效率和经验一致性。

完善行业标准，加速标准化场景实施

针对能源、建筑、制造、医疗健康和交通等关键行业的差异化需求，应制定行业特定应用标准。行业标准引导移动人工智能应用向规范化、标准化发展，减少冗余的研发和试错成本，搭建技术与产业需求之间的适应通道，推动解决方案从试点验证到大规模复制，释放产业数字化和智能化转型的价值。

建立可量化经验评估标准，以与全链价值相一致

产业链资源应整合，构建一个涵盖场景建模、指标定义和能力适应的量化、可实施的移动人工智能体验评价系统。应建立统一的、客观的多模态交互体验评价标准，解决主观评价效率低和定量基础不足等问题。应定义AI应用关键质量指标（KQI）与关键性能指标（KPI）如网络和终端性能之间的明确映射规则，以实现体验需求与技术能力的精确匹配。基于场景的测试案例应标准化，将网络传输、终端性能和算法效率等全链因素纳入评价范围，以确保评价结果与真实用户感知相一致。在现有标准化组织项目基础上，AI-MOS（AI平均意见评分）评价系统应持续改进，为行业优化升级提供客观依据。

04

建议与行动呼吁



随着人工智能与移动通信技术的深度融合，移动AI已成为推动全球数字经济发展和加速数字智能转型的基石发动机。两大趋势正在塑造其演变。

首先，AI原生网络架构的兴起为网络为AI和AI为自我演进智能系统确立了双通路。

其次，以人工智能驱动的网络迅速扩散，形成了早期的终端设备——从AI手机和AI眼镜到AI车辆、机器人，甚至AI玩具——正将大量流量推向网络边缘，使边缘智能成为不可逆转的行业趋势。

与此同时，智能代理网络开始成形，移动AI赋能应用在各垂直行业加速发展。

从商业角度来看，运营商正从以流量为核心的运营转向以体验驱动的盈利模式。由Open Gateway推动的“网络+计算+数据”综合服务架构正成为主流。新的商业模式，例如分级计费、能力编排和生态系统扩张正在涌现，为可扩展、可重复的商业价值循环铺平道路。

Mobile AI now stands at a **关键时刻——从技术可行性到商业可行性的跨越**。然而，关键挑战依然存在。

标准	基础设施	商业化
缺乏统一评估 人工智能产品体系不一致的数据格式，不可靠的代理协议及不成熟的API规范导致碎片化并且上升的整合成本。	上行链路瓶颈，激增东西向交通，以及严格的延迟和弹性分布式人工智能的需求施加重大压力当前网络架构。	价值观共享机制保持不发达。许多解决方案并不充分解决行业痛点并且不明确的申请情景阻碍了释放AI的完整经济潜力。

为了应对这些挑战并释放移动人工智能的变革价值，我们提出以下六项行业范围内的建议：

培养战略共识

移动互联网AI的发展既是技术的必然趋势，也是政策协调和产业努力的产物。决策者应采用开放包容的方法，优化频谱规划，激活数据资源，加强数字基础设施建设，并完善治理体系，以确保环境稳定和可预测。运营商应加速向科技巨头演进，在计算网络编排、开源平台和价值分配等领域打造核心竞争力。终端制造商应深化设备-边缘网络-云的协同，并培养强大的代理生态系统。AI企业必须专注于现实场景，并提供轻量、可访问的解决方案。产业用户应提供高质量数据，明确场景需求，参与标准化，并共同推动整个行业的发展。

建立一个全球统一标准化体系

充分利用6G标准化窗口以避免发展碎片化。国际组织如GSMA、GTI、ITU和3GPP应共同推动包括统一标准在内的统一。

建立人工智能终端的智力评级标准。

- 定义AI生成内容的标签规则。
- 制定人工智能车辆和机器人的远程安全控制要求。

Mobile AI

建议与行动呼吁

- 标准化人工智能模型调用方法、代理协议、工具调用API和多模态数据格式
- 制定云-网络-边缘-终端计算能力协作的统一规范
动态适应网络质量的模型流式传输。
- 推广标准化的API，以暴露网络、计算和数据能力，从而实现新的商业模式和挖掘新的商业价值。

预留和优化 频谱资源

开展前瞻性频谱规划，重点加强上行链路扩展。加快部署6GHz和毫米波频段，并协调潜在的6G频谱以确保长期能力。考虑减少费用或豁免以鼓励生态系统发展。

加强基础设施建设：

构建针对移动人工智能需求的下一代连接性 - 极低确定性的延迟，大规模上行能力，始终在线连接，以及超高可靠性。加速5G-Advanced和6G关键技术的实施，例如AI原生和代理沟通。降低持续升级和迭代的监管障碍。提高边缘节点部署和支持分布式推理以及联邦学习。构建集成网络、计算能力和数据方面的编排能力，由一个开放平台支撑生态系统

激活数据为核心 生产要素：

推动安全跨领域和跨境数据流动。开放高质量、场景特定的数据集，同时建立明确的数据所有权和利益共享框架。鼓励隐私保护机制，如数据可用但不可见，以建立信任并提高数据共享意愿。发展行业特定数据集和数据资产市场，推动新的“网络+计算+数据”业务模式。

推动一个良性循环 行业创新：

鼓励开放API的应用，以实现跨APP意图理解和任务编排，发挥智能代理技术的全部潜力。利用有针对性的补贴和资金刺激对移动人工智能终端的需求。在关键垂直领域启动移动人工智能试点项目，以催化生态增长。推广包容性、云辅助的智能终端，以扩展轻量级、低成本的AI体验。引导精确的行业投资，建立多利益相关者价值共享机制，加速技术迭代，并加强“行业+通信+AI”领域的人才培养。

只有通过政策、技术和商业的联合力量，在“道德AI与可靠连接”的指导原则下，才能将移动AI的颠覆性潜力转化为共享的数字繁荣。让我们共同努力，构建一个开放、创新和负责任的移动AI生态系统——并在移动通信与人工智能综合演化的新篇章中开启新的篇章。

这份白皮书是在和的联合倡议和协调下开发和发布的。



贡献机构

我们对以下组织为这份白皮书的研究、起草和审阅做出的重大贡献表示衷心的感谢：

中国移动 高通 | 中国电信 中国联通 | |
| 诺基亚 | Omdia | 中国宽带 华为 中兴
| 荣誉 | E& 阿联酋 | | Droi | Showmac | 威振咨询

支持组织

我们也对以下组织表示深深的感谢，感谢他们在本文件的编制过程中提供了宝贵的支持，包括提供技术见解和行业专业知识（按字母顺序列出）：

AGrandTech 通信微 | 爱立信 | HKT Keysight 快商云 NXCLOUD | OPPO |
| 联合社 |
| vivo |
YTL 小米 罗德史瓦茨
| Z.AI | Zain KSA | Zain 科威特



1天使巷伦敦EC4R 3AB
英国

电话 : +44 (0)20 7356 06
00 传真 : +44 (0)20 7356
0601