



IMT-2030 (6G)推进组
IMT-2030 (6G)Promotion Group

6G AI即服务 (AaaS) 需求研究



2023 年 4 月

目 录

CONTENTS

图目录 /5

表目录 /5

引言 /6

第一章 现状和发展趋势 /7

1.1 AI 的概念和应用场景	7
1.1.1 AI 的概念及解读	7
1.1.2 当前 AI 的应用场景	7
1.2 AI 应用的现状	8
1.2.1 MEC AI	8
1.2.2 Cloud AI	8
1.2.3 5G 网络中的 AI	9
1.2.4 当前 AI 应用的不足	9
1.3 AI 技术发展趋势及对应用 / 业务的影响	10
1.3.1 AI 技术发展趋势	10
1.3.2 AI 技术发展对业务 / 应用的影响	11
1.3.3 AI 场景、用例和业务发展趋势	12

第二章 6G AI 即服务 (AlaaS) 概念与典型应用场景 /13

2.1 6G AlaaS 概念	13
2.2 6G AlaaS 典型场景	13
2.2.1 智慧城市	13
2.2.2 智慧家庭	15
2.2.3 智赋工业	15
2.2.4 智赋农业	17
2.2.5 精准医疗	17
2.2.6 普智教育	18
2.2.7 智慧能源	19
2.2.8 超能交通	19
2.2.9 网络金融	20
2.3 6G AlaaS 典型应用	21
2.3.1 无人出租汽车出行	21
2.3.2 工业机器人	22
2.3.3 食物循环网络	23
2.3.4 居家健康	25
2.3.5 无人放牧	26
2.3.6 疾病诊断	27
2.3.7 虚拟课堂	28
2.3.8 电网智能巡检	29

2.4 6G AlaaS 典型服务	30
2.4.1 模型推理服务	30
2.4.2 模型训练服务	30
2.4.3 模型生成服务	31
2.4.4 模型优化服务	31
2.4.5 模型共享服务	31
2.4.6 模型部署服务	32
2.4.7 模型管理服务	32
2.4.8 模型压缩服务	33
2.4.9 数据管理服务	33
2.4.10 计算服务	33
第三章 6G AI 即服务 (AlaaS) 性能指标 /34	
3.1 6G AlaaS QoAIS	34
3.1.1 QoAIS 定义	34
3.1.2 QoAIS 指标体系	34
3.1.3 QoAIS 映射模型	35
3.1.4 QoAIS 具体场景指标分析	36
3.2 6G AlaaS 关键性能指标	38
3.2.1 性能指标定义	38
3.2.2 性能指标模型	40
3.2.3 性能指标评估与计算方法	41
3.2.4 性能指标评估示例	46
第四章 6G AI 即服务 (AlaaS) 的技术优势、产业价值与潜在挑战 /48	
4.1 技术优势	48
4.1.1 时延优势	48
4.1.2 算力调度优势	48
4.1.3 成本优势	48
4.2 产业价值	48
4.2.1 对网络自身的价值	48
4.2.2 2B 产业价值	49
4.2.3 2C 产业价值	49
4.3 潜在挑战	50
缩略语简表 /52	
参考文献 /53	
贡献单位 /54	

目 录

CONTENTS

图目录

图 2-1 智慧城市	14
图 2-2 智慧家庭	15
图 2-3 智赋工业	16
图 2-4 智赋农业	17
图 2-5 精准医疗	17
图 2-6 普智教育	19
图 2-7 智慧能源	19
图 2-8 超能交通	20
图 2-9 网络金融	20
图 2-10 食物大循环	23
图 2-11 AI 模型推理业务	30
图 2-12 分布式 AI 模型训练业务	30
图 3-1 QoAIS 和 AlaaS 关系图示例	34
图 3-2 QoAIS 指标分解到各资源维度上的 QoS 指标	35
图 3-3 三种不同编排方案的 AI 任务分布图	36
图 3-4 AI 模型推理服务关键性能指标体系	40
图 3-5 分布式 AI 模型训练服务关键性能指标体系	41
图 3-6 AI 模型推理服务处理示意	41
图 3-7 左：分布式模型训练 右：集中式模型训练	43
图 3-8 AI 数据服务流程	45
图 3-9 VGG-16 模型分割示意图	46

表目录

表 3-1 AI 训练服务的 QoAIS 指标体系	35
表 3-2 AI 训练服务性能 QoAIS 到各资源维度的映射	36
表 3-3 用户分布预测准确度	37
表 3-4 性能指标界对比	37
表 3-5 开销对比	37
表 3-6 AI 服务的关键性能指标	38
表 3-7 通信相关的关键性能指标	39
表 3-8 AI 模型和计算相关的关键性能指标	39
表 3-9 AI 数据相关的关键性能指标	39
表 3-10 通信参数	46
表 3-11 节点计算能力参数	46
表 3-12 VGG-16 模型参数	46
表 3-13 不同模型分割情况下的性能指标	47
表 3-14 VGG-16 模型分布式训练参数	47
表 3-15 VGG-16 模型分布式训练的关键性能指标	47

引言

半导体芯片生产成本逐年下降，芯片的算力体积比不断提升，在更小体积的芯片中得以运行更多的任务。人工智能（AI）得以逐渐摆脱算力的桎梏，衍生出更多更具实用性的算法、模型和架构。而随着信息化进程推进，以及消费电子类产品和各类型传感器的更新和广泛应用，各类系统趋向自动化，为 AI 渗透到人类生活的各个方面提供了土壤。

深入分析 AI 近十年的发展，会逐渐看到一个重要的现象：人工智能在感知方面取得了重要进展，已经逐渐接近人类的水平，在语音、图像、视频处理等多个方面，人工智能处理的效率和效果都已经超越了人类。随着近两年大模型的快速发展，人工智能在语言理解与处理方面也取得了重大进展。比如，美国 OpenAI 研发的聊天机器人程序 ChatGPT，作为人工智能技术驱动的自然语言处理工具，它通过海量语言材料的学习，能够实现理解并基于人类的语言来进行对话，还能根据聊天的上下文进行互动，真正像人类一样来聊天交流，甚至能完成撰写邮件、视频脚本、文案、翻译、代码、写论文等任务。

当前，人工智能在通信生态系统中的架构融合与功能应用在 3GPP、ETSI 中已有标准化定义，通信人工智能的相关网元如 NWDAF 尚处于测试阶段。通信人工智能面向网络管理、业务管理以及应用层的应用较多，也取得了阶段性成效。面向 2030 年，6G 将与先进计算、大数据、人工智能、区块链等信息技术交叉融合，实现通信与感知、计算、控制的深度耦合，实现 eMBB、uRLLC、mMTC 的不断演进，并在此基础上新增人工智能服务和感知服务两大超越通信的应用场景。未来 6G 网络的各个节点将具备提供分布式的高效、节能、安全的 AI 服务的能力。

本报告旨在推进 AlaaS 的应用场景与技术指标等的研究。AI 即服务需求研究的目的在于梳理 AlaaS 的潜在应用场景、典型应用、典型服务等，并在此基础上完善 AlaaS 的技术指标分析；从技术、商业、生态等的角度分析对比，给出 6G AlaaS 区别于现有 AI 服务的优势场景与应用；从网络 AI 服务能力的角度给出对应业务的需求和指标要求；针对未来 AI 技术的发展（大模型等），给出终端和网络侧能力演进的趋势和评估等。本报告在当前 AI 现状和发展趋势进行调研分析的基础上，结合部分成员单位在 AlaaS 领域的研究工作，对 AlaaS 的场景、应用、用例等做了详细地分析和梳理，针对 AlaaS 的 QoAIS 和性能指标等做了分析讨论，给出了相关的示例，最后提出了 AlaaS 的技术优势、产业价值与潜在挑战。

1

第一章 现状和发展趋势

1.1 AI 的概念和应用场景

1.1.1 AI 的概念及解读

AI 是研究、开发能够模拟、延伸和拓展人的智能的理论、方法及应用系统的一门新的技术科学。它涵盖了训练计算机使其能够完成自主学习、判断、决策等人类行为的范畴。人工智能的目标是让机器能够像人一样理解、思考和学习，即用计算机或其他智能体去模拟人的智慧，甚至具备与人类相当的智慧。除计算机科学外，人工智能还涉及信息论、控制论、自动化、仿生学、生物学、心理学、数理逻辑、语言学、医学和哲学等多门学科。人工智能的研究范围主要包括机器学习、语音识别、计算机视觉、自然语言处理、数据挖掘等。人工智能是更广泛的概念，即让机器能够以人们认为“智能”的方式执行任务，而机器学习则是实现人工智能的一种方法。作为一种数据驱动的决策方法，机器学习使计算机能够自动解析数据、从中学习，在没有人工干预的情况下从中检索出事件发展的内在规律，从而对真实世界中的事件做出决策和预测。

本报告聚焦于以机器学习（ML）为核心的 AI，它包含 AI 模型 / 算法和数据等基本要素，以算力资源作为支撑，依托于 AI 模型的生成、训练、推理等而实现智能化目标。AI 模型 / 算法从历史数据中学习规则，并基于学习到的规则对新数据做出预测或决策。AI 模型 / 算法指以机器学习模型为代表的各种算法，它通常具备可学习的参数，因此通信中传统的确定规则的物理层或 RRM 算法就不是 AI 算法；AI 模型 / 算法对推理结果的精准性和安全性有特定的要求。AI 数据指和 AI 模型 / 算法强相关的“训练样本”、“参数数据”和“输入输出数据”等，因此一般的网络或用户数据不一定是 AI 数据；AI 数据对数据格式和质量保障有特定的要求。AI 算力指算力中用于处理分析 AI 数据，执行 AI 模型 / 算法的那部分算力，因此用于传统基带算法的算力就不是 AI 算力；AI 算力对计算硬件资源类型属性有特定的要求。

1.1.2 当前 AI 的应用场景

在农业场景，主要包括有作物管理、害虫和杂草处理、疾病管理、土壤管理、产量预测和管理等。作物管理，主要提供作物选择，施肥建议，使得作物免受恶劣天气影响等；害虫和杂草处理，即识别害虫和杂草，提供处理害虫和杂草的相关建议，推测害虫行进路线和繁殖规模 and 速度，推测杂草的生长状态和发展等；疾病管理，即预测、识别分类作物病害；土壤和作物管理，包括评估作物表面土壤湿度，预测天气，结合天气预测结果进行灌溉等；产量预测和管理，根据气候，季节等因素提供最佳播种时间建议，并预测最佳收成时间和最终产量等。其主要运用的 AI 技术最开始是基于规则的专家系统，发展到后来的模糊推理系统和人工神经网络的结合。主要涉及模式识别，图像识别等用例 [1][2]。

在自动驾驶场景，分为感知和决策两类主要场景，感知类包括汽车定位、静态障碍物映射、移动障碍物检测跟踪、道路映射、交通信号检测和识别；决策类包括路线规划、路径规划、行为选择、运动规划和控制等 [3]。其运用到的 AI 技术主要是基于规则的推理模型，启发式算法，近似推理，类人推理等。主要涉及的用例是图像和语音识别，分析推理，分类，模型迭代等 [4]。

在医疗卫生场景，主要有疾病诊断预测、临床和患者护理。疾病诊断预测包括前期疾病诊断预测，例如糖尿病、癌症、结核病、精神疾病的预测诊断，和就诊中的医学成像和图像诊断，例如肿瘤学影像、光学相干断层扫描等。临床包括临床试验参与、机器人辅助手术系统和计算机辅助手术，减少用药错误。患者护理包括心力衰竭监测、术后健康监测、肿瘤患者的健康监测，检测例如血氧浓度，体温，心率等参数并分析。除此之外近年来 AI 技术也在应用到新药研制场景中得到应用。主要使用到的 AI 技术为神经网络、专家系统等。主要涉及的用例是图像识别、分析推理、分类等 [5][6]。

在网络金融场景，主要包括信用预测和评估，根据客户的历史交易数据、资金流动情况智能运营聊天机器人应答客户部分疑问，智能手续柜台满足客户部分需求。交易风险预警，大数据分析历史诈骗交易行为，提前为

客户预警，进行报警等。涉及的 AI 技术较为广泛，包括支持向量机、决策树、神经网络、K 邻近遗传算法、随机森林、Boosting 算法、集成学习、聚类等。主要涉及的用例是分析推理、模式识别、分类等 [7][8]。

在智慧城市场景，主要有公共设施控制，例如通过控制交通信号灯等设施帮助监控交通流量和增加城市连通性并减少交通事故和交通违法，通过人脸识别门禁和安检识别控制重要场所通行；灾害预测，例如预测建筑物火灾风险；人员识别和寻踪，通过收集公共摄像头的视频图像，分析其历史行动轨迹等，锁定犯罪分子或失踪人员。主要涉及的 AI 技术有神经网络、演化算法、专家系统、分布式人工智能、机器视觉、决策网络等。主要涉及的用例是图像识别、推理分析等 [9]。

在电子商务场景，主要用例是推荐系统、欺诈识别、营销活动、产品退货预测。推荐系统又可以细分为关系发现和购买预测，关系发现包括商品之间的销售量关系发现和客户之间关系发现，购买预测包括购买预测和回购预测。推荐系统将根据内容或商品之间的关联度、用户特征和商品的关联度、甚至用户之间的关联度，向用户发起推荐。欺诈识别是通过分类或分析交易双方交易行为识别非法交易或异常交易，识别交易欺诈 [10][11]。

此外，AI 在网络安全、物流运输、教育等方面均迎来蓬勃发展。

1.2 AI 应用的现状

1.2.1 MEC AI

MEC 与 AI 的结合，一般来说，中心侧与边缘侧对 AI 要求各不相同，越靠近边缘侧，AI 推理程度增加，越靠近中心侧，AI 训练程度增加；两者互为补充，一般来说，MEC 对于 AI 来说，MEC 部署在靠近数据源侧，便于适应边缘原生的 AI 类应用对时延、安全、隐私性的要求；AI 对于 MEC 来说，在网络资源复杂不确定性较强的情况下，AI 可在准确的时间和地点辅助边缘节点资源分配，提升边缘节点自治能力。

AI 类应用场景，涉及行业众多，常见场景例如智能制造、智能安防、智慧医疗等：

(1) 智能制造，在某园区部署 MEC 平台，基于该平台可实现预测性维护和 AR 远程协助 / 指导。预测性维护系统通过传感器采集、监控生产设备运行状态，实现设备运行数据可视化，根据历史数据和采集的实时数据，基于 AI 算法对设备潜在故障进行预测，针对故障定位及部件损伤程度安排备品备件，制定检修计划，可提前预知重大故障隐患，避免突发停机风险。

(2) 智能安防，针对现代安防的特点结合 5G MEC 技术，MEC 平台可提供开放能力调用接口，例如 AI 视频分析、调度算法等以 API 的方式直接调用，大幅降低合作伙伴的开发难度和缩短研发周期。实际实施过程中通过在园区内布置监控摄像头以及巡检机器人，实时巡检监控，结合 AI 人工智能视频分析，遇到异常情况及时告警，同时联动巡检机器人到达现场排查，实现智能快速处理、实时响应；在小区入口处设置门禁系统，通过识别来访人的特征来判断是否有允许进入园区的权限等。

(3) 智慧医疗，一般来说 5G+MEC 智慧医疗可以为医院提供确定的网络环境、算力保证和安全性保证，来确保医院业务的实用性、稳定性和安全性。针对院内医疗应用，可将远程会诊、PACS 影像系统、AI 分析、电子病历、HIS 系统等逐步迁移上 MEC 平台上，可通过低时延的 AI 辅诊发现和提示疑似病患情况提升阅片的速率。

1.2.2 Cloud AI

目前在 Cloud AI 技术架构中，边缘节点侧重多维感知数据采集和前端智能处理；边缘域侧重感知数据汇聚、存储、处理和智能应用；云中心侧重业务数据融合及大数据多维分析应用。数据从边缘节点到边缘域，实现“聚边到域”；从边缘域到云中心，实现“数据入云”。域和云中心可多级多类，根据不同应用，边缘域汇聚的数据和传到云端的数据在模型和内容上也会不同；Cloud AI 将提供 AI 资源的可调度、数据的按需汇聚、应用的场景化响应、运维的一体化建设，随着云原生技术的快速更迭，业务需求得到进一步的挖掘，技术边界将从云上延展到边缘，服务形态多元化，云原生将整合人工智能技术形成服务能力的调度中枢，促进算力服务和运营。

未来将从传统的 Cloud AI 向 Network AI 转变，智能从应用和云端走向网络，实现网络的自运维、自检测、自

修复，智能在云-边-端-网间协同，提升网络总体效能，促进网络内生智能。

1.2.3 5G 网络中的 AI

当前，在利用 AI 赋能和增强网络各方面，在 5G 系统中已有着较多的应用和价值探索（即 AI for Network），它涉及到多个层面和不同的业务领域：

首先，在非标准化方面，AI 可以通过各个厂家私有实现的方式应用在无线接入网域、承载网域、核心网域和网管域；无线接入网域，gNB 基站可基于 AI 实现基站的动态节能、MIMO 多天线权值寻优和基站故障快速定位和预警等；承载网域，可基于 AI 实现云化超强管控和意图承载快速部署；核心网域，可基于 AI 实现智能切片和 SLA 智能拆解等；网管域，可基于 AI 实现网络智能监控、错误根因分析、系统行为趋势预测、全域全量数据关联的数字化运维和端到端感知保障等。上述诸多应用已经能够彰显 AI 手段在网络智能化方面的价值意义。

其次，在标准化方面，3GPP SA2、SA5 已针对 5G 系统化智能和管理架构方面，已有着手并拓展，标准化定义了 NWDAF 和 MDAF 功能和相关机制流程。3GPP RAN3 和 RAN1 针对各自识别的关键用例，尝试定义和引入可标准化的 AI 操作范式。

总体来说，5G 技术标准将 AI 技术引入网络，可实现 5G 网络自运营和运维；AI 智能将数据转化为信息，从实战中学习积累知识和经验，提供分析决策，为网络使用者提供零等待、零接触、零故障的网络服务体验，也为网络运营者打造自配置、自修复、自优化的运维能力，实现通信网络高度自治，实现无处不在的智慧内核。此外 5G 网络通过切片来实现能力按需部署，也可针对不同的业务自动化智能化调动网络资源和算力，针对不同业务进行差异化配置资源，满足各行各业的差异化需求。总之，网络设计走向智能泛在，通过有效连接局部数据，实现不同智能体之间的语音、视觉等交互需要的泛在化，为用户提供更加极致的网络服务体验。

1.2.4 当前 AI 应用的不足

虽然当前国内 AI 技术逐渐成熟提升，服务种类不断丰富，但是在 MEC 与人工智能结合进行开发和应用时仍

然面临着技术不达预期的问题。究其原因，主要归纳为以下几方面：

(1) MEC+AI 类行业应用，存在应用部署障碍，究其原因主要是生产环境对 AI 基础设施、算法及数据质量要求较高，但是企业 IT 基础设施在海量数据参与运算及分布式架构条件下，存在算力不足以及不兼容等问题。

(2) MEC+AI 类行业应用，存在投入产出比不足，企业应用人工智能的回报不及预期；在成本支出方面，企业所需的 AI 开发人才及 IT 资源相对稀缺，价值回报方面，虽然部分场景效果得到验证，但是在不同企业和具体工作环境下，效果并不稳定。

(3) Cloud AI 技术架构中，边缘域与云中心侧对 AI 计算能力要求不同，云侧在满足运算效能的前提下，希望针对不同网络有更优化的性能表现；边侧，在能耗为首要要求情况下，更加注重推断运算的性能；此外，数据入云过程中，由于数据流通不畅、数据质量良莠不齐和关键数据缺失问题，会影响有效的智能分析结果。

(4) 在 AI 赋能 5G 网络方面，目前由于大数据和 AI 的智能化水平并不完美，算法模型研发整体规划不足，各种 AI/ML 算法模型基本都是针对已识别的特定通信问题弊端缺陷而开发，缺乏系统性、全局性、可解释性方面的考虑，限制了系统 AI 能力服务的可拓展性、迭代增强性和 AI 模型泛化应用能力；在实际场景中，设备功耗、计算能力、应用场景等多种因素限制了 AI 算法在 5G 系统中真正部署限；当前 5G 系统中的各种 AI 资源能力（涵盖 AI 算力，AI 算法和 AI 数据方面）适配性待完善，尚不具备开放可交易性和服务化的特征，大部分仅仅限于 5G 系统内的应用。

综上，AI 技术发展已进入行业深水区，亟待进一步向前发展。

1.3

AI 技术发展趋势及对应用 / 业务的影响

1.3.1 AI 技术发展趋势

基于深度学习的人工智能方法在许多场景取得了重要突破,但仍然存在模型可解释性差、对抗样本鲁棒性差、数据与算力需求大、理论基础薄弱等问题。未来,人工智能需要突破的主要领域有:可解释人工智能和鲁棒人工智能理论与方法、通用人工智能理论与方法、安全、可靠、可信及可扩展的人工智能技术、类脑智能等。

从感知智能走向认知智能是人工智能技术发展的趋势。认知智能是人工智能技术的高级阶段,旨在赋予机器数据理解、知识表达、逻辑推理、自主学习等能力,使机器成为人类改造世界、提升能力的得力助手。在从感知智能到认知智能的发展过程中,语义及知识的表达和逻辑推理是进行认知的重要手段,而多模态学习则是获得信息融合和协同的重要手段。通过构建多模态的大规模基础模型,可以学习多种信息的融合表征,建立模态转换和协同关联,从而提高 AI 系统对复杂环境的认知和理解能力,进而获得多场景多任务的 AI 应用能力。

► 通用基础大模型:

从弱人工智能到强人工智能发展的路线上,提高机器解决问题的“泛化”能力,是重要的手段。在场景泛化、模态泛化、任务泛化等方面,通过大规模基础模型的领域通用方法,赋予 AI 系统解决多问题的能力。随着数据规模逐渐扩大,数据模态进一步丰富,预训练大模型充分学习不同模态数据的表征,通过预训练 - 微调的模式,完成多种类型的任务。预训练模型与领域数据结合,基于多场景数据进行学习,将成为一种完成下游任务的基础模型,应对更为复杂的智能决策场景。传统的 AI 是企业或组织根据自身已部署数据收集系统来收集数据,根据执行系统设计执行动作,从根本上导致的同一问题的模型多样化,然后这样会导致重复的模型训练工作,并且一个模型只能用在系统中,导致资源浪费。为提高资源利用率,同时得益于多点协作等技术的应用带来的通用基础设施和通用数据集,AI 模型亦将趋向通用化。

► 边缘人工智能:

面向手机、可穿戴设备、传感器等海量边缘设备,通过模型压缩与优化加速技术,将智能赋予边缘设备。基于预训练大模型,根据边缘设备的计算、存储、功耗等限制,设计特定的低精度和稀疏化等模式,并通过知识蒸馏等方法,实现小尺寸、低复杂度、低功耗等目标。进一步,边缘设备可基于本地数据训练的模型优化大模型的性能。

► 自动自治 AI:

目前,深度学习的开发及应用并未突破主流监督学习的模式,数据清洗、数据标注,模型的设计、开发、训练和部署等都需要大量人力投入。迁移学习、小样本、零样本、自监督、弱监督、半监督、无监督及主动学习等新方法将推动人工智最终实现“自治”,解决模型训练、迭代、设计对人工的依赖。未来 AI 自治使得模型更加归一,数据规模进一步扩大,不再需要人工干预,模型可以在线学习吸收新的数据知识,实现自身能力的迭代提升。数据规模扩大及在线学习将使模型的生产更加集约化,各行业的业务模型会汇聚成几个甚至一个超大模型。

运用 AI 技术的本质就是要减少人工参与,甚至消除人工参与,当前 AI 在应用过程中往往与多种传感器系统和联网的策略和执行功能组件配合,避免过多人工收集处理数据作为 AI 进行自我更新的输入,也减少人工参与 AI 输出结果的分析和执行。人类社会是实时变化的,固化的算法模型无法满足即便是单一的业务,因此,为保障整个系统的有效性或准确性,AI 系统需要实时接收反馈,不断更新自身。以便系统向精准化迈进。

► 隐私保护学习:

数据匿名化、联邦学习、差分隐私等一系列隐私保护学习方法,可以解决包括用户数据在授权范围以外被使用、模型训练后存在数据记忆等隐私泄露问题。联邦学习将数据保存在本地,通过模型融合,实现多个参与方在保护数据隐私要求的前提下进行机器学习,合作进行模型训练,实现用户间的数据联合建模。更加灵活高效的保护用户数据隐私的 AI 学习方法将帮助 AI 算法在各种场景中持续地进化,让数据被有效利用,充分发挥其商业、科学和社会价值,为用户提供高效优质的服务。

► 类脑人工智能:

类脑人工智能期望借鉴和模仿生物神经元的工作模式，通过构建功能更加丰富的神经元，建立具有事件触发、脉冲编码、时间和空间信息协同处理的能力；利用神经动力学原理，可实现短时可塑性和长期记忆，在开放环境中具备自适应调整和学习能力；借鉴生物脑的稀疏连接和递归特性可大大减少能耗。如果能够突破相关技术，未来五到十年类脑计算可能会在众多计算任务中展现出性能和功耗优势，并在智能终端、穿戴式设备、自动驾驶等领域得到应用。

► 大小模型端云协同进化：

云上大模型全局知识向端上小模型有效降维迁移和大规模端上小模型学习所得新知识向大模型升维融合。异构终端轻量化模型弹性构建、端侧模型个性化增强学习、端云协同模型训练框架设计和分布式多引擎计算，构建完善的大小模型端云协同进化技术体系，形成包含端云一致高性能模型运行容器、低时延端云模型迁移通道和端云协同进化部署的研发工具链。

► 多智能体博弈智能：

多智能体多回合复杂合作和冲突博弈可以实现自我学习和自我演化，以强化学习为基础的学习推理模型发展出面向进化的多维评估管理方法，结合作和非合作博弈智能模型实现对复杂问题的人机协同求解。

► 3D 视觉：

3D 视觉技术是视觉 AI 领域多年热点研究方向之一，核心任务是对三维空间、物体及环境进行真实还原与重建。随着相关算法与硬件计算能力的不断升级，3D 视觉算法效果得到大幅提升，三维几何重建更加精细，表面纹理重建更加清晰，带来更加逼真的视觉观感。基于 3D 虚拟形象的舞台演出、直播带货、教育互动等应用层出不穷，成为 AI 内容产业全新发展方向。

► 公平性 AI 算法：

由于数据偏差、算法本身缺陷、甚至是人为偏见的存在，现有 AI 算法普遍存在对于某些特定人群效果不公平的“歧视性现象”。随着 AI 算法在社会各行业的广泛落地应用，作为辅助人们决策的重要工具，算法的公平性问题正受到越来越多的关注。过去的几年业界已在逐步探索一些针对性的解决方案，包括构建更公正的数据集、

算法训练中引入公平性约束损失、提高机器学习算法的可解释性等。

► 安全 AI：

近年来算法后门攻击、对抗样本攻击、模型窃取攻击等针对人工智能算法的攻击技术持续发展，通过篡改构造特殊数据诱骗人工智能应用产生不可信的错误结果，带来了更大的算法安全风险，因此保障人工智能应用安全可靠的需求日渐迫切。未来人工智能技术将向着安全智能方向持续演化，一方面从算法的可解释性入手提升模型的鲁棒性，另一方面化被动为主动，通过主动安全检测机制对各类攻击进行侦测与拦截，最终实现人工智能可用性与可信性双轨并重的现实需求，推动人工智能技术在更广泛领域的安全落地。

► 可解释性：

随着 AI 深入到生活的各个领域，特别是涉及多个系统，多个责任对象的领域，安全可信和责任相关问题凸显。要求 AI 不可再作为黑盒子存在，而应该呈现 AI 在发挥作用的每一个阶段的具体作用，是否符合法规等。例如自动驾驶领域，近期的特斯拉交通事故。可解释性也为工程师提供优化模型的依据。因此，学术界和业界正在 AI 的可解释性方面进行了较大的投入。

► 多点协作：

在 AI 技术发展过程中，算力、数据和模型作为人工智能三要素被呈现，然而实际上位于大型数据中心或者说云端的大算力节点不具备访问位于边缘的具备海量数据的节点的权限；位于边缘的具备海量数据的节点不具备将数据转化为模型的算力，导致单一节点通常因为未能全部满足三要素导致难以完成任务。此外，5G 网络部署扩展以及未来 6G 网络部署，为多节点的互通提供更多的通信资源。这些因素促使 AI 技术向多点协作的方向发展。

1.3.2 AI 技术发展对业务 / 应用的影响

无处不在的智能应用：通过 AI 技术重构智能边缘，在移动通信网络 2030 架构中，6G 网络将综合 AI 面向业务的感知能力构建原生智能。原生智能要支持基于 AI 的业务感知能力：面向消费者将针对全感全息类通信业务

提供高效编解码、传输优化、体验保障、协同调度的能力；面向行业可基于确定性操作系统，强化系统调度框架，为千行百业提供业务保障。如 6G AI 推理服务，以机器视觉处理为例，在边缘侧采用 AI 图像特征识别的处理方式，可以降低传输带宽要求，并提高业务实时性。

分布式学习与推理：如果把所有原始数据全部上传到中心云，不仅会对网络带宽造成很大压力，也会给中心云的算力带来巨大挑战。分布式地组织 AI 任务，并结合协同 AI 和机器学习，可以减轻这些压力。随着智能手机和物联网设备的计算能力越来越强，用户对数据隐私的保护越来越重视，移动设备会进行更多的 AI 计算和训练。分布式学习将使 AI 离用户和数据更近，能更迅速地进行迭代升级。

深度实时边缘计算：网络节点和机器人之间的低时延通信至关重要，为了打消数据隐私方面的顾虑，同时也为了保证低时延，可以使用边缘节点，而不是把所有信息都发送到云端。边缘节点在物理位置上更靠近 AI 应用，一定程度上可以提供一些特殊的即时服务。

网络高带宽全连接需求：分布式学习将汇聚大规模节点的集体智能，并使智能普惠到所有人。分布式训练对各智能节点的 AI 模型进行融合进化。模型的持续增大促进智能持续增强，需要网络提供大规模的高带宽连接，支持节点之间的实时模型传输。分布式推理通过网络与终端的协作，将智能带给所有终端，特别是计算能力低的设备，需要网络通过全连接为人、机、物、环境等提供实时智能服务。

1.3.3 AI 场景、用例和业务发展趋势

综上，AI 的场景、用例和业务的发展趋势呈现以下几种特性：

- 场景扩张到各方面，用例与日俱增，业务百花齐放。AI 技术被应用到生产生活的各个方面，囊括农业，自动驾驶，医疗卫生，网络金融，智慧城市，消费者业务等多个场景。

- 加强了业务 / 用例细分。多种模型协作。在 AI 发

展过程中，人们逐渐意识到单一的 AI 算法和模型几乎不可能完全解决或应对某一大概念的场景，往往将大的场景划分为更细分的业务，例如在农业场景，某一种算法模型可能只识别一种害虫。所以种植某一种植物的 AI 系统通常通过多种算法模型协作，也包含多个方面的细分业务。

第二章 6G AI 即服务 (AlaaS) 概念与典型应用场景

在业界广泛开展的关于 6G 愿景的讨论中，“6G 将为人和物提供更好的连接，推动人联、物联向智联转变，开启智能社会”目前已成为一个基本共识 [12]。在大模型机器学习和大数据分析支持下，普惠智能将是未来商业和经济模式的重要基础。6G 将是使能该服务的基础平台，在 5G 三大应用场景基础上，新增人工智能作为应用场景。

普惠智能要求 6G 网络支持原生智能架构、对用户数据的隐私保护、可信的网络和支持消费者和垂直市场的多元化生态系统。6G 网络将通信、信息和数据技术以及工业智能深度集成到无线网络，并且具备大规模分布式训练、实时边缘推理和本地数据脱敏的能力。在 6G 网络中，网络边缘运行的分布式 AI 在提供高性能的同时，也能解决个人和企业都十分关心的数据所有权问题。普惠智能与深度融合的 ICT 系统相结合，在网络边缘提供多样化的连接、计算和存储资源，将成为 6G 的固有特征。

2.1 6G AlaaS 概念

6G AI 即服务 (AlaaS)：6G 移动新网络提供泛在的智能服务。6G AI 即服务是通过网络（包括 6G 核心网、6G 无线接入网和 6G 终端）内的连接、计算、数据、模型等资源和功能，构建分布式的高效、节能、安全的 AI 服务（包括 AI 模型训练、推理、部署等）和开放生态。

6G 将为 AI 相关业务和应用提供高效的端到端支持，将分布式智能体智能地连接起来，以便在各行各业大规模部署 AI。AI 业务和应用既包括为 6G 网络自身性能优化提供的智能能力，即 AI4NET (AI for Network)，如利用端到端 AI 实现空口和网络的定制优化和自动化运维，提供满足多样化需求的最佳解决方案；也包括向第三方业务提供的智能能力，即 NET4AI (Network for AI)，如通过 6G 网元将具有的原生集成通信、计算和感知能力，加速云上集中智能向深度边缘泛在智能演进，为 AI 提供分布式学习基础设施。

6G AI 服务主要面对高实时性能、高安全隐私或低综合能耗等需求，在网络内进行 AI 训练或推理，提供适应不同应用场景的智能能力。6G 网络作为原生智能架构，通过网络内的通信、计算、数据集、基础模型等资源和功能，具备大规模分布式 AI 的高效训练或推理能力，为用户提供无所不在的高性能 AI 服务，同时降低由于超大规模 AI 模型相关的训练和部署导致的通信和计算开销，提供一个低碳节能的开放生态。

2.2 6G AlaaS 典型场景

AI 场景是指 AI 技术或 AI 资源要素发生作用的场合和情境。它可有不同的维度划分方式，例如：地理区域、环境特征、领域行业等。6G AlaaS 典型场景是 6G 网络 AI 发生作用的典型场合和情境，包括但不限于工业、农业、生活、教育、医疗等不同行业和领域。

2.2.1 智慧城市

城市是一个囊括住房、交通、卫生、公用事业、土地使用、商品生产等系统的庞大而复杂的实体 [13]。智慧城市将从城市生命、城市健康、城市发展、城市幸福等领域出发，促进城市的各个系统数字化，智能化。政府部门，企业，相关组织甚至个人得以更便捷更高效地参与城市的生产生活。

5G 网络的部署为城市中个体和组织的互联提供较大的带宽，较强接入能力，物联网终端或传感器以此散落到城市中的部分角落。一些独立的智能系统开始在城市中萌芽，例如智能应急调度系统，智能市政照明，智能燃气抄表等，然而这些系统和部分数据等基本不互通，其 AI 模型亦没有共享的意愿。这些 AI 系统虽然对城市有一定的优化作用，但随着城市的继续发展，这种优化将到达瓶颈。

城市中的各个部分或系统是相互交叉，相互渗透的。6G 网络将为城市的各个部分或系统提供相互交互的通道，

为城市时刻生产的数据提供存储之处，也为这些数据提供处理它们的算力。基于 6G 网络的 AlaaS 也以此对城市的各个领域发挥作用。



图 2-1 智慧城市

2.2.1.1 城市生命

城市存在的前提是人存在，人们在城市中生存，能源、水、食物必不可少。依托 6G AlaaS 平台的管理系统将更实时地监控分析城市各处电表、水表、燃气表，燃气站出入账，食品仓库和食品交易市场出入账本，每个人或家庭的能源、食物、水的账本，实时获得城市各处的能源、水、食物输入和消耗状况，及时发现缺口或者异常，从而优化能源、食物的供给路线和数量，优化水管道压力和分配，及时修补能源或水管泄露等。例如与 6G 网络中的交通系统交互，提前为运输能源的管控交通，规划路线。例如与仓储系统交互，根据各区的消耗情况及时在各个仓库之间调度食物，甚至提前控制仓库内部摆放顺序，优化进出仓库的能源和时间的消耗。例如与金融系统交互，通过分析输入城市的资源和人们消耗的资源账本，发现资源浪费甚至相关违法犯罪行为。同时 AlaaS 平台依据实时收集的数据不断优化算法模型，以期进一步优化相关管理措施。

2.2.1.2 城市健康

城市健康必须兼顾人文环境和自然环境两个方面。基于 6G 网络的 AI，将与智能警务系统交互，协同分析监控摄像头数据，重点场所进出数据，银行金融交易数据等，同时从 AlaaS 平台中获取优秀的相关分析模型，以识别不文明行为违法犯罪，维护城市安全和文明。基于 6G 网

络的 AI 也将复用交通系统中的传感器、摄像头，协同车联网 AI 和车载摄像头等，分析推算路面建筑健康度，提醒相关部门及时执行维护工作。6G 网络 AI 也将通过与环境监测系统交互，分析空气环境变化，生成适用于城市整体的环境治理模型，促进环境治理自动化智能化，降低环境治理的成本等。

2.2.1.3 城市发展

城市发展又有城市基建扩展，城市经济发展，城市文化发展三个要素。基于 6G 网络的城建系统将通过 AlaaS 平台分析当前城市的气候变化、地质情况、城市已有的基础设施数据，以及政府和人民的需求，发现城市基建缺陷以及为新区域规划提供相关建议等，以此保证城市基建扩展合理，也符合需求；基于 6G 网络的政务系统通过 AlaaS 平台将收集更细化的税务，消费相关数据，也将以强大算力分析相关数据，发掘当前城市产业发展的薄弱之处或者未来发展的经济风险，为相关组织或系统生成改进建议等，为城市经济发展保驾护航；AlaaS 系统也将会更高效实时地收集分析城市人口的实际构成数据，人口的流动数据，为相关部门调整教育政策和人才政策提供参考，同时也通过相关城市广告系统和政务发布系统，向不同区域的人员推送相关公益材料，以促进城市文化发展。AlaaS 将承载城市各系统之间的数据交互，相互借鉴相关数据和模型，减少重复工作，进一步推理城市发展状态，优化城市发展投入。

2.2.1.4 城市幸福

城市生命，健康，发展是城市幸福的前提。医疗住房是当今城市幸福的两个突出问题，依托 6G 的 AlaaS 系统将以分布式 AI，在保护每位市民隐私的前提下分析市民的病例或相关汇总数据，发现区域性疾病，合理分配医生以及器材，及时将高需求药品和器材纳入医保系统等。AlaaS 系统将实时分析人们的收入变动、物价变化、货币通胀率等经济因素，生成合理的房产指导价，或通过分析城市交通数据，为市民生成合理的通勤路线或购房区域建议。AlaaS 也将通过分析政务评价系统收集相关反馈，以进一步优化模型。此外，依托 6G 网络，AlaaS 将能收集、处理海量数据，生成巨量模型，满足人们各个方面的需求，提升城市幸福度。

2.2.2 智慧家庭

家庭住宅环境是人类生活的主要活动场所，每个人平均下来可能有近一半以上的时间在居家环境中度过。智慧家庭是以居家住宅类型的环境为平台，将通信、检测等技术结合起来，形成一个高效的居家事务的管理与更加舒适和智能的生活体验。目前，越来越多的智能网络设备已经开始渗透入居家生活当中，而家庭环境的智能性现在主要体现在被动的控制与检测，如通过人的语音或用手机操作对家用电器进行控制，或者通过视频监控设备进行监控。随着社会与技术的进步，人们对居家生活的智能性也提出更高的要求，在需要更广泛的交互式、主动式的服务同时，也需要及时、准确和智能化的服务，从而我们的居家生活也不断地向更加便利、安全，具有好的隐私性以及舒适性等发展，极大地改进生活体验。



图 2- 2 智慧家庭

AI 技术无疑是居家智能化的重要技术和巨大推力，这使得在 6G 时代网络 AI 会在居家生活中具有非常广阔的应用场景。网络 AI 的技术将会被用于家居控制，安防监控和行为监控，居家健康服务和家庭娱乐等各个方面。而在 6G 的智慧家庭的智能性，主要体现在利用人工智能的技术，可以通过对人的生活模式、喜好等进行智能化的分析与了解，并作用于每个人的生活中，就像一位智能化的管家一样。这要求人工智能技术有长期不断学习、准确理解人类行为模式的能力。而在其他不同的应用方面，对于 AI 的服务与网络计算的要求也有所不同。如对于家用电器等的一般控制，与家庭健康的日常检测，对于通信的

及时性与 AI 的计算力要求相对不是很高。然而在与安全与突发健康时间有关的，如陌生人闯入、老人突然跌倒等，一方面要求 AI 技术可以准确捕捉到相关时间，另一方面也要求及时通信告知与响应。而对于家庭娱乐的一些应用，特别是虚拟现实等技术的家庭化，对 AI 的算力以及通信能力都会有较高的要求。与此同时，在智慧家庭场景中，家庭既有的网络设备应该被充分利用来实现居家的智能化，这样也可以降低家庭购买专用设备的开销，从而更加经济地让智能生活走进每个人的家庭。

2.2.3 智赋工业

工业生产的全生命周期业务流程主要涵盖产品的前期设计、生产制造、运输、运维、仓储以及用户交付等。6G AlaaS 具有广泛覆盖、密集连接、极速传输等一系列特点，在未来场景下，可以智赋工业，为工业生产提供 AI 数据、AI 训练、AI 推理、AI 验证等服务，支持其朝向数字化、网络化、智能化的方向加速转型。

2.2.3.1 智赋工业机器人

在未来工业制造、运输等场景下，工业机器人将被广泛使用。例如建筑机器人能在远程控制下完成工地建设任务。而特别是在高温、高空等极端工作场景下，工业机器人也将被考虑使用以代替工人完成高危作业，人们只需利用手中的终端设备，实现真正的“无人工厂”。在完成智能机器人的 AI 训练之后，需要在 AI 模型部署之前对其进行验证，最终实现智能机器人安全高效地生产任务。

6G AlaaS 为机器人提供 AI 训练服务，例如通过机器人收集数据，应用模型训练进行多智能体学习，并将训练后的模型分发给机器人。而 6G 的极致传输能力将支持机器人与网络侧之间进行快速模型训练与参数交换。6G AlaaS 还提供多系统联动能力，各类终端设备都能够与机器人进行灵活组网，人们利用电脑、手机、VR 设备即可远程遥控。6G 移动通信系统的覆盖对象还从传统的用户终端扩展到了环境感知设备，机器人能够通过 6G AlaaS 的感知技术实现与环境的连接与交互，甚至还能向不同部门的机器人跨环境、跨系统学习，共享经验。同时，6G 网络可利用数字孪生网络提供 AI 验证服务，通过性能预验证迭代训练鲁棒性更强，性能更优的 AI 模型。6G

AlaaS 为训练好的模型提供 AI 验证服务，要求数字孪生网络产生比物理环境更多场景下的样本数据，降低对物理网络数据采集开销和性能影响，进一步提高 AI 模型的稳定性和性能。

智赋工业机器人对算法维度性能指标界、训练耗时和算力维度计算精度、时长、效率的服务质量存在需求。随着智能终端类型近年来不断丰富，5G 可能将逐渐无法承担更多设备的 AI 训练需求，这要求 6G 提供超大连接、低时延、稳定的 AI 训练服务；AI 验证服务要求 6G 网络为数字孪生网络的虚拟环境提供海量数据采集、精准定位和超低时延传输等能力。此外，智赋工业机器人场景对自治维度的服务质量指标存在需求。工业机器人的安全性，可以避免很多因人为疏忽或疲劳工作引起的隐患，这要求 6G AlaaS 提供完全自治、部分人工可控和全部人工可控三方面的高质量服务。

2.2.3.2 智赋生产过程

在将训练好的 AI 模型部署到各类智能生产设备（机械操作臂、机器人等）后，工厂开启智能运行模式，智能生产设备进行 AI 推理获得高效的生产决策。当智能生产设备需完成一项模型推理任务，而自身算力又不够时，6G 网络侧提供算力资源和 AI 推理的协助，将推理模型进行分割，部署在端侧和网络侧，共同完成推理任务，满足推理的性能需求。

为及时将经过训练好后的 AI 模型部署到生产线，6G AlaaS 提供强大的推理服务能力，通过云、边、端分布式部署的协同算力进行网络内生的 AI 推理，实时按需动态调整各类生产设备的配置参数，实现各种工业生产场景的灵活适配。

智赋生产过程场景对算法性能指标界、计算开销、传输开销、能耗等服务质量存在需求。6G AlaaS 使能工业生产中的 AI 快速推理过程，而时延和准确度是 AI 推理服务的两项重要性能指标。模型分割点的决策将影响终端侧和基站侧分别负责的计算任务量，以及空口需要传输的数据量，这要求 6G AlaaS 在算法性能指标界、计算开销与传输开销方面提供高质量服务。影响这两个指标的还包括终端侧计算资源分配、空口无线资源分配和基站侧计算资源分配等环节，需要进行通信与计算资源的协同调度，这与能耗服务质量指标则有着密切的联系。尤其是当终端和基站侧均存在多种异构计算资源时，计

算单元类型和数量的分配会对推理时延产生直接影响。而 6G 以网络极简为升级目标，将通过简化架构、功能和协议来降低开销、节约能耗，实现轻量级的无线网络。

2.2.3.3 智赋用户交互

当工业产品经过设计、制造、运输等步骤，最终出厂交付给用户后，智赋工业的全生命周期流程并未完全结束。随产品携带的传感器将收集产品在用户使用过程中的运行状态，此类用户信息将反馈传输回企业，用于进一步优化产品，改善用户体验，以适应个性化、定制化的工业制造趋势。

6G 网络提供 AI 数据服务，为上述过程提供智能的数据采集、处理、流转策略，按需采集数据，防止出现数据冗余的情况。在采集海量用户数据后，6G AlaaS 进一步提供数据特征提取服务，通过 AI 模型挖掘数据价值，精准定位用户使用工业产品过程中的问题，这要求 6G 网络提供海量数据的 AI 处理能力。

智赋用户交互场景对性能指标界、计算开销和时长等服务质量指标存在需求，6G 网络超高速的数据传输速率与超低的时延抖动将保障交互过程的实时、高效和智能化，对于用户而言，通过产品与企业进行良好交互，减少甚至消除故障，产品使用满意度将得到大幅度提升。另外，智赋用户交互涉及海量隐私数据的特点，还要求 6G AlaaS 能够保障安全与隐私维度存储、计算、传输安全和数据隐私等级四方面的服务质量，6G 网络无限接近有线传输的可靠性和安全性将确保收集和处理后的数据能够精准、可信、不冗余地通过网络传输到企业，使企业更好地洞察用户需求，提前优化研发设计。



图 2-3 智赋工业



图 2-4 智赋农业

2.2.4 智赋农业

农业包括种植业、林业、畜牧业、渔业、副业五种产业形式，当今世界农业的发展需求是提高土地产出率、农业劳动生产率、农产品商品率和国际市场竞争力 [14]。农民和农业组织为此不断地追求科学化、自动化、精准化，学界和产业界因此逐渐尝试将 AI 技术应用到农业生产活动中。

传统基于单体 AI 的农业智能化系统构建成本极高，体现在搭建数据采集网络，购买租用算力基础设施，聘用 AI 工程师训练 AI 模型。农民 / 农业组织是成本极度敏感的个体。除却巨型农业组织或企业，大部分农民或农业组织不具有相应的经济实力，造成模型通用性差，农业 AI 的模型种类也不多。针对单一的农业问题，企业也因其系统的独立性，无法与其他企业共享，不得不大量重复投入。

基于 6G 网络的 AlaaS 平台允许农民或农业组织以较低成本租用运营商已有基础设施，例如算力节点和 6G 通感一体网络中的数据收集节点；还能通过 AlaaS 平台与遥远的农业生产实体共享 AI 模型；农民或农业组织也可通过 AlaaS 与高校等研究机构合作，农民提供试验环境，研究机构提供 AI 人才，双方共享研究成果。AlaaS 将从数据收集到模型生成，极大地降低了农业 AI 实施成本。降低的 AI 实施成本终将更多追求精准化自动化的农民和组织统一到 6G 网络中，6G 网络的其他单元也能藉此更深度地参与农业生产活动。更多的数据，更多的场景也会带来更丰富的模型，AlaaS 平台本身也将有相对应的增长。

2.2.5 精准医疗

未来，智慧医疗在时间维度上，涉及到各类疾病的预防、预判、诊治、推理、监控、临床手术、病患护理和疫苗药物研制等全生命周期的诸多环节。智慧医疗在地域维度上，涉及到集中式各级医院、分布式各级社区和家庭医疗场景。智慧医疗在专业维度上，涉及到跨不同的医疗学科之间的知识经验信息的融合利用。智慧医疗总体愿景是：使得个体能够以更低成本、更便捷高效方式获得对自身健康平安的更好呵护、治愈和保障。6G 新系统不仅能够更好地支撑智慧医疗相关的海量信息传输和同步，更能直接为医疗信息的处理和决策赋能。

医疗关乎个体的生命健康安全，因此预判、诊疗动作结果的准确性、可靠性显得很重要。传统基于单体 AI 的预判诊疗执行方式，由于受到 AI 模型算法成熟度和病例数据样本的限制，预判诊疗动作结果的准确性、可靠性、实时性等指标还不是很理想。通过基于 6G 网络 AI，各个地理分散的医疗机构个体医者，能够更广泛地互联，汇聚利用更多的 AI 模型算法和病例数据样本，高速传输同步医生和患者的相关信息，加深联邦学习和群体学习，从而不断迭代提升预判诊疗动作结果的准确性、可靠性、实时性，增强个体对智慧医疗前景的信赖。智慧医疗还能够基于 6G 网络 AI，实现对各种医疗资源的更合理规划和优化分配使用，极大减轻医患者的体力心理压力，避免“看病难”“排队难”“过度医疗”等弊端问题。此外，对某些疾病预判和诊疗结果的隐私保护，也是非常重要的；通过 6G 网络 AI 分布式技术，可以更好地保护个体医疗信息隐私。



图 2-5 精准医疗

2.2.6 普智教育

未来的教育，将依托 6G 网络和 AlaaS 智能内生能力，在消除教育鸿沟，提高教育手段方面实现质的飞跃。在 2030 年代实现普智教育。

面向基础教育方面，城市与乡村之间，以及不同地区之间的教育差距历来是一个难以逾越鸿沟，其中最重要的是师资与实验设备的差距。在 2030 年代 6G 网络广域覆盖的基础上，依托 AR/VR 等智能交互技术，AlaaS 将让教学名师和教学仪器设备“真实”的走进每一个课堂中。即使在偏远的乡村，学生可以与教学名师实时互动，操作实验设备，达到身临其境的学习。6G AlaaS 还可以在课堂上实时的分析每个学生的学习状态，一方面监督促进学生学习；另一方面，针对不同学生的状态与兴趣，智能的安排个性化的学习课程与学习进度，依托智能分析实现个性化教育。

面向行业智能教育方面，基于 6G AlaaS 内生智能能力，利用 XR、数字孪生、智能交互等技术，构建面向各行各业的教学与实践环境。将贵重、复杂、庞大的行业装备，如航空航天、芯片制造、医疗器械、交通运输等行业的装备，展现在行业教育的虚拟实践环境中。学生的实际操作动作和设计作品，通过 6G 网络 and 智能内生的能力，智能分析反馈操作行为结果，针对设计作品分析其优点与缺点，不断提高学生的设计能力，切实提升行业尖端技术的教育水平。

教师可以使用 AI 系统浏览学生的论文，并检查语法、句子结构和抄袭，查看学生是否准确理解了他们的作业，以及他们是否正确地写下了给他们的提示。教师可以让 AI 在进行自己的复习之前先通过每篇论文，而不是花费无数小时仔细浏览一页又一页的论文，让他们有更多的空闲时间与学生一对一地工作或完成其他必要的任务。

2.2.6.1 实时笔记

通过 AlaaS 创建实时笔记 AI 服务，录制讲座和使用语音转文本技术来生产实时笔记。此外，如果学生因为忙于做笔记而在课堂上难以集中注意力，人工智能可用于创建与课堂录音同步的文本，以便学生可以回去听或阅读老师在特定时刻所说的话。这也为教师腾出了时间，因为他们不必为学生提供自己的课程回顾和笔记。通过 6G 的泛在全连接，学生可以随时随地进行笔记学习和回顾。

2.2.6.2 学生和班级理解

借助人工智能驱动的见解，教师可以根据作业和测试的个人和集体分数快速了解学生或班级对主题的理解程度。然后，教师可以使用这些信息来确定他们是否需要在课堂上再次讨论这些主题，或者根据个人知识差距以某种方式个性化学生的学习。同样，这些信息可以帮助管理人员确定这些主题是否在教室和其他学校中得到普遍理解，从而为课程变更提供信息或确定需要改进的地方。6G 网络需要通过高宽带的实时通信能力，AlaaS 提供秒级别的反馈延迟。

2.2.6.3 个性化学习

每一个人，无论是孩子还是成年人，都有独特的思维特点和学习方式。个性化学习就是通过收集和分析学生的学习数据，用人工智能勾勒出每个学生的学习方式和特点，然后自动调整教学内容、方式和节奏，使每个孩子都能得到最适合自己的教育。随着时间的推移，数据积累逐渐增多，人工智能也就越“聪明”，对学生学习的适应也就越精准，这样形成良性循环，效果越来越好。除了通过提供分析能力和存储计算能力，6G AlaaS 还会通过安全措施保障个人隐私。

2.2.6.4 虚拟导师

虚拟导师其实也是一种自适应学习系统，只不过焦点在学生的课后自学和答疑，而不是课堂教学上。这个领域目前还在早期发展阶段，进一步的推动则依赖于一些关键技术的发展，如语音语义的识别，以及数据的进一步采集和分析。虚拟导师也许几年内还不能代替真人辅导，但一定会在课后辅导行业中逐渐占有一席之地。虚拟导师把教学从一对多模式变成了一对一模式，这需要 6G 网络提供高带宽低延迟的网络保障交互过程的用户体验，同时，对于虚拟导师形象进行结合本地特征的渲染保障交互过程的真实性。

2.2.6.5 教育机器人

用机器人充当教育者的角色直接和孩子交流，已经不只属于科幻小说了。目前，一些新兴的创新公司正在开发可以成为孩子的老师和朋友的机器人。机器人在听到孩子的问题之后，可以自动连接网络寻找答案，并且通过和孩子的交流逐渐学习和了解孩子的情绪和个性。目前，机

器人能够流畅交流的话题还是相对有限，需要持续开发研究。6G 网络互联了海量的知识仓库，供机器人搜索和自动学习答案生成，构建强人工智能如 Chat GPT。

2.2.6.6 场景式教育

将虚拟现实（VR）和增强现实（AR）运用在教育中，想象空间是不可估量的，益处也是显而易见的。课堂不再局限于小小的教室、白板和 PPT，而是整个宇宙。6G 网络为提供高带宽低延迟的传输能力，同时 AlaaS 提供基于边缘和基于云的计算，AI 建模分析能力。



图 2-6 普智教育

2.2.7 智慧能源

未来的电力系统，将包括大量分布式的基于可再生能源的发电装置，并通过大规模输电网络将电能交付给消费者。由于大多数可再生能源本质上是动态性的，而需求也呈现高度的动态性，为了实现供需之间的平衡，必须让电力系统更加灵活。智慧能源将通过智能优化和调度分布式能源系统，实现高效的清洁能源利用，降低甚至取代化石燃料的电力 [15]。智能既体现在通过人工智能算法优化能源的生产，如通过 AI 控制光伏面板角度提升发电量、通过智能摄像头进行设备状态检测等；也体现在能源的智能分配和使用，如通过算法实现对分布式能源的生产预测、基于历史数据的负荷预测、动态需求响应的实时分析判断等。通过人工智能技术的加持，目标是实现能量在生产和消费者之间按需自由流动，最终实现整个能源系统的绿色低碳、安全稳定。6G 系统作为未来信息通信技术的新基础设施，将为智慧能源相关数据的实时传输和智能决策提供支持，实现全网智能化。



图 2-7 智慧能源

AI 模型的性能依赖于可用于模型训练的大量数据、强大的模型和推理时数据的实时可靠传输，而智慧能源的不同业务对 AI 模型的准确性、时延等可能具有不同的要求。6G AlaaS 一方面通过融合大量分布式能源生产设备或能源消费节点基于本地数据训练的模型，在保护各节点数据和隐私的前提下，提供高准确度的 AI 模型，为各个分散的能源生产节点提供更高效的配置算法，为分散的能源消费节点提供更准确的预测算法；另一方面，通过实时高速汇聚能源生产节点的状态与消费者的需求，并基于网络内的 AI 模型推理能力，为智慧能源的生产和消费提供实时的优化，实现极致高效的能源利用。

2.2.8 超能交通

未来，超能交通在时间维度上，涉及到高清地图下载、车辆环境感知、环境预测和路径规划等环节。超能交通在地域维度上，涉及到高速公路、城市道路、封闭区域等场景。超能交通的总体愿景是：使得智能车辆以更低的成本为用户提供安全、高效、舒适的乘车体验。6G 系统作为未来信息通信技术的新基础设施，具有超低时延高可靠传输、超强算力、精准感知等一系列特点，将为超能交通相关数据的实时传输、处理和智能决策提供支持，实现全网智能化。

在超能交通场景中，以自动驾驶为例，6G 系统可以提供基于 AI 的数据服务、AI 算力卸载服务和基于 AI 的环境预测或路径规划。其中，基于 AI 的数据服务是指 6G 系统为自动驾驶车辆提供基于 AI 的环境感知结果，例如，当自动驾驶车辆存在感知盲区时，6G 系统通过传感器或无线信号采集感知数据，再通过 AI 模型推理，得到盲区



图 2-8 超能交通

环境的感知结果，反馈自动驾驶车辆，从而提升驾驶安全性。基于 AI 的算力卸载是指自动驾驶车辆将一部分 AI 模型推理或训练的算力卸载到 6G 系统。例如，当车辆行驶到路况复杂路段时，自动驾驶车辆对 AI 模型推理所需算力急剧增高，造成时延无法满足需求时，车辆可以将一部分 AI 计算任务卸载到 6G 系统，从而降低了对自动驾驶车辆处理芯片的需求。

面向海陆地空立体交通方式，通过 6G 网络广域覆盖及其智能能力，实现车辆的自动驾驶和智能交通管理、船舶和航空领域的智能调度，对水上、陆地、地下及空中交通工具实现统一的控制管理调度。在 6G 网络具备的深度学习、多传感器融合等方法下，形成车辆、路侧、云端的全场景一体化感知决策架构，进一步推动出行智能化、服务泛在化、管控全局化的新时代智能交通愿景的实现。在交通出行方面，实现个人最优出行方案以及整个交通的最高效使用的有机统一。

2.2.9 网络金融

网络金融，即通过网络技术支持全球范围内的金融服务活动，包括网络银行、网络证券、网络保险、网络信托、股票交易等。网络金融安全和网络金融监管等底层保障也重度依赖通信技术自身的安全防护能力。

网络内的连接、高性能计算、大数据传输、和泛在分布的 AI 能力赋能高频金融交易和保障交易数据的安全可靠传输，降低金融恶性交易风险。通过 6G AlaaS 提供智能数据路由和转发自适应，降低交易数据传输时延，保证金融交易的实时性，助力金融交易发展。

2.2.9.1 高频金融交易

以股票交易为例的高密度、高交易额的金融交易，在交易日内的任意时刻股票价格都是动态变化的。6GAlaaS 可以在高频金融交易中发挥高速交易数据传递的关键作用。利用 6GAlaaS 提供的服务选择适当的数据转发路由算法进行金融交易业务数据转发模型的训练、优化及部署。6G 智能化的数据传输能力，能够让用户第一时间获得近实时的交易现状数据，在确定交易时机后，通过 6G 网络快速回传交易请求。基于 6GAlaaS 的智能路由能力，避免将交易信息流转到拥堵的网络链路中，优化全体用户的交易速率和频率。

2.2.9.2 电信金融诈骗

电信诈骗一直是网络金融领域一直难以杜绝的难题，4G/5G 阶段实现双向互认后改善了电话诈骗的风险。科技的进步带来的便利，也被不法分子利用到诈骗中。近年来，电信诈骗正在从以电话诈骗逐步向网络诈骗转变。网络诈骗具有多样化的特性，诈骗分子可以通过窃取用户社交账户、向网络用户发送诈骗链接、APP 夹带诈骗链接等实施诈骗，手段层出不穷。用户登录的网站多样，网址派生快，APP 夹带的广告网址具有随机性。用户社交软件被窃取利用的难易程度、用户自身的社交圈、用户接入网络的目的、习惯、用户的支付手段、用户的反诈骗意识等不同对于辅助用户反诈的难以程度也有所不同。网络诈骗具备持续周期短、往往用户还没反省过来，财产已被分散无法追回了。

AI 在广连接、泛算力算力的支持下提供更快的数据处理分析能力，以应对区域内全用户、全异常网站的监测



图 2-9 网络金融

数据分析。6GAlaaS 可以提供更高速度的信息采集传输能力,通过对用户恶意消费 AI 模型的不断训练优化,提供更加精准的恶意消费判别,更低的模型反应周期,应对诈骗分子登录周期短,无法定位犯罪分子位置信息的问题。具体的 6GAlaaS 基于其自身高安全级别和异常网址、异常交易分析的敏捷性,可以辅助监管机构降低网络诈骗发生的概率。6GAlaaS 配合金融机构对用户的异常登录、异常消费、短时间的跨地域支付、高频次的支付等进行特征分析,即时将异常用户推送给金融监管机构,配合监管机构阻断资金外流。6GAlaaS 可配合国家有关网络安全监管机构对网址进行优劣甄别,对恶劣网站判定后具备网页拒绝访问。

2.3

6G AlaaS 典型应用

AI 应用是指向用户提供相对可管可控的 AI 服务,通常是基于多个 AI 服务的集合。6G AlaaS 典型应用是指在 6G 网络内面向终端用户、网络运维人员、第三方客户等对象提供的 6G AlaaS 服务。

2.3.1 无人出租汽车出行

2.3.1.1 背景

个人打车出行先后经历了出租汽车、网约车的时代,两个时代中都面临着打车需求与载客服务不匹配的问题,即经常面临客户打不到车、司机空跑无客的情况。随着无人驾驶 / 自动驾驶技术的成熟,个人打车必将进入无人驾驶出租汽车的时代。由于个人出行需求、个人定位 / 轨迹等敏感的个人信息,与无人出租汽车的位置信息、以及巡游路线是两座无法直接相通的信息孤岛,基于网约车提交订单、接单匹配的打车方式,个人快速打车的需求,以及出租汽车高效使用的需求不能高效结合。

6G 智能内生网络作为智能普惠时代的纽带和基石,将通过 AlaaS 服务,面向个人和无人出租汽车,分别提供数据管理、模型训练(可选)、模型优化等 AI 服务,个人与无人出租汽车数据脱敏后的联合训练,以及 AlaaS 的开放服务,实现个人个性化出行需求的快速服务,以及无人出租汽车的高效管理。

此外,在无人车辆驾驶过程中,6G 智能内生网络 V2X 场景功能,可以为无人出租汽车提供更安全的自动驾驶辅助服务。

2.3.1.2 角色和前提假设

(1) 角色:无人出租车、打车乘客、无人出租车传感器、通感融合基站、道路传感器(V2X)、提供智能服务的网络运营商。

(2) 前提假设:

- a. 网络为本场景的运行提供执行特定任务的 AI 能力。
- b. 无人出租车、用户终端、基站、传感器、网络运营商网络之间能够按需传递相关信息。在提供智能服务的网络运营商网络能够根据乘车人的请求、当前位置,选派一辆满足乘车人需求的无人出租车;
- c. 无人出租车传感器、通感融合基站、道路传感器,都能够及时完成感知,并将感知信息及时传递到网络;
- d. 无人出租车(含传感器)、通感融合基站、道路传感器、用户终端之间,可以直接通信。

2.3.1.3 应用需求

(1) 业务需求

- a. 根据道路信息、历史区域需求、脱敏用户位置和需求信息、目前无人出租车的运行情况等,智能编排空载车辆的巡游路线或待客区域。
- b. 接收到打车乘客需求信息后,根据路况信息和车辆运行情况,智能安排汽车前往上车点。
- c. 在无人汽车前往途中,以及接到乘客后前往目的地途中,根据路况信息,智能规划路线,以及调度更换无人汽车。
- d. 行驶过程中,根据不同汽车的传感器生成更及时的道路信息,如突发状况,包括检修、事故、坑洼等,以及训练无人汽车面对不同突发状况的反应模型。

(2) AlaaS 需求

- a. 数据服务:网络提供 AlaaS 的数据服务包括:
个人的与区域的用户数据的管理与脱敏,包括标识信息、定位/轨迹信息、习惯信息等;
道路状态信息的收集和更新,包括通感融合基站感知道路信息、突发情况等。
- b. 计算服务:网络提供 AlaaS 的计算服务包括:
无人出租车的卸载数据的计算,包括感知数据计算,突发情况规避动作计算等。
- c. 模型训练:网络提供 AlaaS 的模型训练服务包括:
根据用户数据,训练用户出行模型;
根据区域用户数据,训练区域出租车需求模型;

根据通感融合基站和道路传感器数据，训练高效导航路线，包括巡行、前往上车点、前往目的地等。

d. 推理决策：网络提供 AlaaS 的推理决策服务包括：推理区域动态出租车出行需求。

e. 模型优化：网络提供 AlaaS 的训练优化服务包括：根据道路交通状况，无人汽车巡行路线 / 待客区域的优化；

根据区域出行需求，调度区域出租车数量及出租车分布；

根据基站 / 道路感知数据，优化出租汽车自动驾驶和突发情况驾驶动作。

f. 开放服务：网络提供 AlaaS 的开放服务包括：

数据开放：基站感知数据及道路传感器（V2X）数据的开放；

算力开放：网络算力的开放。

模型开放：用户出行模型推理数据脱敏后的开放，包括个人需求推理、目的地推理、区域需求推理。

2.3.1.4 潜在技术需求

(1) 隐私数据管理与脱敏：涉及乘车人的身份信息、目的地信息、行程信息；

(2) 算力服务：网络为无人出租车提供算力；

(3) 模型训练服务：用户需求轨迹预测、出租车运营负荷 / 路线预测等；

(4) 模型优化服务：针对上述模型，进行优化，提升准确性等；

(5) 开放服务：算力、数据、轨迹模型等方面的开放给需要相关方；

(6) 超低时延高可靠传输：或者归为时延确定性传输，针对保证行车安全、感知等信息，需要确定的时延抖动等方面的保证。

2.3.2 工业机器人

2.3.2.1 背景

随着工业生产的数字化和自动化的转型，具备高度灵活性的机器人开始被广泛应用于工业，如搬运机器人、焊接机器人、装配机器人、质检机器人等。传统的工业机器人需要进行生产设备操作的预设，一个工业品的生

产流程需要多种类机器人的配合，特定的生产环节需要设计专用的机器人，这会增加生产的成本，并且无法处理未知情况以及生产过程中的随机因素。因此，承担不同生产操作的通用型机器人对降低成本、提高鲁棒性十分重要。通用性工业机器人可以使用多种生产工具，来进行不同的操作。同时，一个生产环节可能也需要多个同种类机器人的配合，因此机器人之间、机器人和环境之间需要通信与交互，并且期望机器人可以根据交互过程中的信息改进自身的决策，在复杂的环境中提高生产效率。利用人工智能，特别强化学习，机器人可以具备智能决策能力，学习如何操作、规划移动路线，从而可以执行许多任务并适应新环境。

考虑到智能机器人的感知环境范围受限和移动性等特征，6G 网络可以为工业机器人的协同生产提供 AI 服务。工业生产拥有较高的时延敏感性，需要通过无线网络为移动中的机器人提供确定性时延的传输条件。为了实现协同生产，需要广泛收集不同机器人的数据进行训练，考虑到工厂生产的隐私性，需要采用分布式学习的架构。这样的学习架构可以让 6G 网络收集不同工厂机器人的训练中间数据，来扩大样本的丰富性。机器人可以根据自身感知到的数据进行本地训练，适应环境的变化并做出快速的决策。随后机器人将模型（而不是数据）传到基站，由网络聚合多个机器人的模型进一步提高决策效果。

2.3.2.2 角色和前提假设

(1) 角色：配备感知能力（传感器、摄像头）、操作能力（操作臂、轮 / 机械腿）和微型计算机的机器人，具备计算能力的基站、提供智能服务的网络运营商。

(2) 前提假设：

a. 网络应用为机器人运行提供执行特定任务的 AI/ML 模型推理能力；

b. 机器人能采集观测数据，包括传感数据、动作数据、通信数据、信道状态信息等，并通过上行传输传输到 AI/ML 服务器；

c. AI/ML 服务器可以选择一组机器人进行协作，并确定训练配置，如训练算法、学习率、批大小、迭代次数等；

d. AI/ML 服务器能收集多个机器人的观测数据，训练 AI/ML 模型，并通过下行传输将 AI/ML 模型分发给机

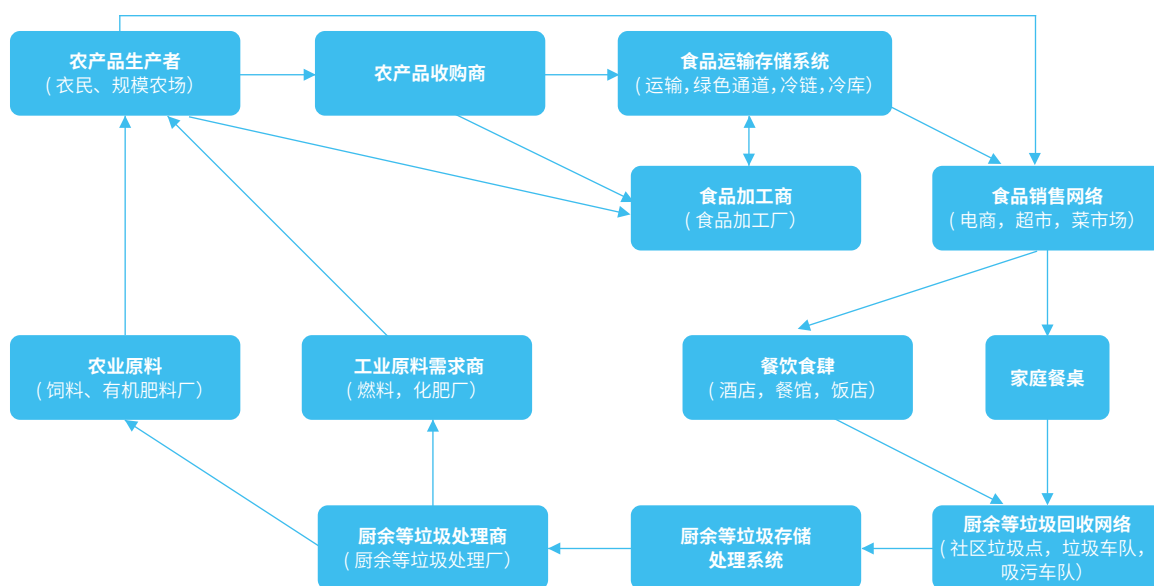


图 2-10 食物大循环

器人；

e. 机器人可以通过机器人之间的通信来交换信息。

2.3.2.3 应用需求

(1) 业务需求

a. 工厂根据产品生产的流水线的逻辑关系，进行生产环节分解，如不同环节的生产操作、运送（生产环节的衔接）、质检和包装，每个环节中存在多个机器人协同完成同一个任务；

b. 机器人在操作过程中进行数据采集与处理、本地训练、推理决策（操作臂执行、机器人行进路线规划），并将模型上传给基站；

c. 网络收到工厂生产线设置参数和机器人上传的 AI 模型，进行训练优化，发送给机器人。

(2) AlaaS 需求

a. 数据采集与处理：机器人采集观察到任务过程中的数据，包括感知环境、产品参数数据、图片数据、动作数据等，进行数据预处理和储存；

b. 机器人训练：针对每一个生产环节，同一组机器人训练相同的 AI 模型（包括图像识别、智能决策），在本地根据自身采集的数据进行训练；

c. 推理决策：机器人根据实时感知的数据，输入到自身 AI 决策模型，生成实时的操作决策；

d. 训练优化：机器人定期将训练模型报告给网络运营商，运营商接收来自多个机器人的训练模型，进行分类聚合，并将新模型分发给机器人。

2.3.2.4 潜在技术需求

(1) 时延确定性传输：确定的时延抖动要求传输时延有个明确的范围，不能高也不能低，强调的是准时达到。

(2) 机器人实时推理：机器人本地具有一定的计算能力。

(3) 网络辅助推理优化：网络支持分布式训练架构，可以进行模型的聚合和协同训练。

2.3.3 食物循环网络

2.3.3.1 背景

食物是维持人类生命，也是维持城市生命的一部分。食物的运输、加工、存储、销售，这些过程关系到城市的生命力。而后产生的一系列厨余垃圾、有机和无机废物的回收、处理、成品再销售等过程将影响城市健康和发展，最终影响城市幸福，也就是人类的幸福度。食物本身具备变质的特性，食物在运输存储过程中耗费时间越长，营养和经济价值通常也会越低，部分食物由于产出地区等原因无法被所喜爱的人食用。因此需要高效地人性化地处理整个食物循环过程，以期缩短时间的消耗，

减少食物的损耗，及时满足城市更多市民的需求。城市管理者对此已经采用一系列手段和方法或系统，但是效果仍有待提高。这是因为食物大循环中涉及海量的农民、农场、海量的运输车、工厂、销售 / 回收网络、食品加工烹饪点、家庭。这些角色或节点之间需要交换信息，并在生产活动中协作，交换信息的网络需满足巨大的带宽，海量的接入需求，而多样化而庞大的系统协作则需要高效的大脑。

基于 6G 网络的 AlaaS，将依托 6G 庞大的网络，为食物大循环中的所有系统、网络、人、生产设备、运输设备、传感设备提供足够多的接入点，足够大的带宽。6G 网络也将为从这些接入端收集的数据提供存储的空间，处理分析他们的算力，以及提供一系列针对各个环节的 AI 模型。AlaaS 平台将会提供对食物仓库数据、食品运输车 / 垃圾车的行进路线和容量数据、食客和家庭的食物账本数据、销售网络的收支数据收集管理分析等智能服务，将生成食品运输车 / 垃圾车的行进路线规划建议，家庭食物购入和营养摄入建议等。这个基于 6G 网络 AI 的食物大循环系统极大地减少食物或者说能量的浪费，也减少营养的浪费，保障人们饮食的健康和合理。提升餐余 / 厨余垃圾的回收效率和重利用率。

2.3.3.2 角色和前提假设

(1) 角色：食品运输车队定位系统，食品仓库出入口系统和监控，食品销售网络出入账系统，个人或食品食物消费账单系统，垃圾回收网点传感系统，垃圾回收车定位系统，垃圾处理厂传感器系统，个人或家庭用健康监控管理设备。

(2) 前提假设：

- a. 参与食物循环网络的所有系统都应接入 6G 网络，并通过 6G 网络进行交互。
- b. 参与食物循环网络的所有终端可不都接入 6G 网络。每个系统可以只通过一个出口节点与其他角色通信。
- c. 社区具备垃圾回收点传感器。
- d. 具备权限的节点可通过 6G 网络直接或间接访问个人或家庭健康设备数据。

2.3.3.3 应用需求

(1) 业务需求

a. 食品运输车队定位系统，食品仓库出入口系统和监控，食品销售网络出入账系统，个人或食品食物消费支付应用垃圾回收网点传感系统，垃圾回收车定位系统，垃圾处理厂传感器系统均将通过 6G 网络接入互联网，这些系统的各个节点也通过 6G 网络进行互联。6G 网络 AI 根据实际情况周期性或实时地收集食物的交易转运，餐余和厨余垃圾的收集转运，垃圾处理厂的相关数据，将数据分配到边缘或云节点 6G 网络将根据需求调配相关算力和模型，分析食物的实际转运情况，回收到的厨余垃圾的物质比例，判别食物的实际利用率和厨余垃圾再利用率，食品 / 垃圾运输和存储效率，也利用相关数据进一步优化模型；

b. 6G 网络 AI 将根据分析结果，向食品转运和垃圾回收车队发布最优行进路线建议，向食品销售网络发布食物网点分配建议，以此减少食物的转运和存储时间，也减少厨余垃圾在社区的停留时间，加快食物循环的速度。为市民提供更鲜活更健康的食物来源和更清洁的社区环境还将根据智能医疗系统提供的个人和家庭的健康状况，结合地理位置，食物购买成本，为个人 / 家庭提供营养摄入、食物购买类型和地点建议。降低食物购买成本的同时，提升每个家庭成员食物摄入的营养价值。

(2) AlaaS 需求

a. 数据服务：网络提供 AlaaS 的数据服务包括：

食物转运分配数据的收集和更新，包括通感融合基站食物转运车队的行进路线，食物仓库的出入账单数据，个人和家庭食物购入账单等数据。

商业相关数据的管理与脱敏。包括食品销售网络销售数据，食品转运行进路线，食品仓库数据等。

专业数据库服务：根据垃圾处理厂的存储区的传感器数据，访问物质分析数据库服务，以便后续分析厨余垃圾的组成成分和比例。

数据共享：非敏感数据共享。

b. 计算服务：网络提供 AlaaS 的计算服务包括：

根据城市各个区域的食物交易数据，仓库进出数据，

食品转运车队的位置和运力，生成食品转运分配方案。

根据各个社区垃圾回收点的传感数据，各个垃圾处理工厂的仓储情况规划垃圾回收车队的转运时间和路线。

根据个人或家庭的食物消耗和健康状况，乃至收入状况生成食品采购建议。

c. 模型训练: 网络提供AlaaS的模型训练服务包括:

根据城市各个食品销售和仓储节点的出入数据等，训练合理分配食品到每个节点的食物分配模型

根据食品来源地生产情况，食品存储销售网络的数据，结合城市的交通状况生成食品转运路线规划模型。

根据垃圾回收区的传感信息，结合物质分析库，训练分析垃圾成分和比率模型。

根据个人或家庭的食物消费和摄入数据与健康状况，训练食物 - 健康类医用模型。

d. 推理决策: 网络提供AlaaS的推理决策服务包括:

预测食物浪费和实际消耗情况，推理食品种植生产方案，生成食品的加工、分配、转运、存储、垃圾回收决策，实时调整食品转运或垃圾回收车队的运动姿态策略。

推理城市中每个人的健康状况变化，实际营养摄入量、推理食品消费清单建议。

e. 模型共享: 网络提供AlaaS的模型共享服务包括:

根据不同社区，不同城市之间市民健康度，出生率，平均寿命等共享全局性的食品供给模型，优化再共享。

为具备共性的市民共享食品 - 健康医疗模型。

2.3.3.4 潜在技术需求

(1) 通感一体化网络: 通感结合，减少整个食物循环网络对监控摄像头和传感器的成本投入。

(2) 智能任务分割技术: 将复杂智能服务任务分解为通用的子任务，由较为简单的基础模型分别处理，简化数据收集、模型训练的难度，降低数据收集设备、模型训练节点的要求

(3) 协同推理优化: 支持分布式多实体训练架构，可以进行模型的聚合和协同训练，也方便进行模型共享和数据收集设备共用，降低 AlaaS 建设成本。

2.3.4 居家健康

2.3.4.1 背景

健康与医疗问题是人类社会生活必需要面对的问题之一。通过通信技术和人工智能技术的结合，在居家环境中对人类健康进行持续有效的检测、智能的分析以及健康的管理，一方面可以让人类保持良好和健康的生活状态，另一方面可以对一些健康问题及时的干预，有效的进行预防，也可以减少家庭医疗的负担与缓解紧张的社会医疗资源，从而从整体层面提高全民健康。特别是随着社会进入老龄化，以及独居的普遍性，在居家健康方面的需求会越来越高。而网络AI将在居家健康方面带来颠覆性的革命。

对于智能化的监测，除了需要一些额外设备的被动式的监控，无线信道的波动信息等也可作为一种手段提供对健康指标或突发事件的监控。如利用AI技术上可以对无线信道的波动进行分析，可以对人体的呼吸以及心率等健康指标进行持续有效的监控。当发现有人，呼吸和心率异常时，能及时进行治疗和干预，网络AI能在嘈杂的无线信道中抓取有用的健康监测信息。如对于老人跌倒等突发事件，通过网络AI技术可以及时捕捉到，从而进行第一时间的救援。当然，通过AI和网络的结合，也可以对其他方面，如日常运动、睡眠质量、情绪等各个方面监测。

而智能化的分析，是通过对监测等搜集到的数据，进行准确个性化的分析，对人体的健康状态以及生活习惯等有充分的了解。如对睡眠时长以及质量的分析，来了解人作息时间以及周期性变化；对呼吸、心率数据的分析可以判断人体的健康状况，甚至发现潜在的健康问题；对个人情绪方面的分析可以得知其情绪变化，是否有情绪或精神方面的问题，来进行及时干预。

而对于健康管理，则是在分析结果的基础上进行的。可以通过与人的交互，来提供有效的管理方案与注意事项。在存在健康潜在问题的时候，可以通过医院网络服务对其进行进一步诊断，并自动通知亲属等。当然在出现突发的健康问题，如老人突然摔倒，呼吸停猝时，需要紧急自动通知医疗机构进行干预。

通过利用网络AI的相关技术，可以为居家生活的健康保障提供第一条防线，来改进人们的生活质量。

2.3.4.2 角色和前提假设

(1) 角色：个人用智能设备、家庭检测设备（血压计等）、提供智能服务的运营商

(2) 前提假设：个人用智能设备、家庭检测设备（血压计等）等连接到 6G 网络

2.3.4.3 应用需求

(1) 业务需求

a. 家庭检测设备收集用户数据储存在本地，并不断进行本地数据处理；

b. 家庭检测设备等具有较低的 AI 能力，可以进行数据预处理、初步的模型推理（故障检测）、本地训练，并将初步推理结果或模型上传到服务中心；

c. 6G 网络对初步推理结果进行模型推理，将推理结果传输给居家健康中心；对上传的模型进行融合，发送给对应的家庭检测设备。

(2) AlaaS 需求

a. 数据采集：家庭检测设备对患者数据进行数据预处理和传输；

b. 模型推理：家庭检测设备根据实时采集的数据，输入到自身 AI 模型，生成初步检测结果，传输给基站；基站基于自身 AI 模型对初步检测结果进行进一步处理，生成实时检测结果；

c. 模型训练：家庭检测设备根据检测目标，确定相同的 AI 模型，根据自身采集的数据在本地进行训练；

d. 模型融合：家庭检测设备定期将训练模型报告给网络运营商，运营商接收来自多个健康设备的训练模型，进行模型融合，并将新模型分发本地设备。

2.3.4.4 潜在技术需求

(1) 网络实时推理：网络支持与家庭智能设备的协同实时高准确度推理。

(2) 分布式训练：网络支持分布式训练架构，可以进行模型的聚合和协同训练。

(3) 高可靠低时延传输：推理数据和推理结果需要在规定时间内成功传输。

(4) 上行大带宽传输：大量的 AI 模型参数需要上传到网络进行融合。

2.3.5 无人放牧

2.3.5.1 背景

传统放牧通常牧民骑马带着牧羊犬餐风宿雪，日夜兼程，把牲畜从一个草场赶往另一个草场，这个过程中，牲畜可能患上不知名的疾病，而茫茫草原极度不利于寻医问药，还需防范牲畜走失，狼群等野兽袭击。此外，草场变换完全基于牧民经验，草场可能未被完全利用使得下一周期新长牧草过老，或者过度利用使得草原局部沙化，引起一系列连锁反应。在这种环境下，牧民保证自身收益已然不易，难有优化草原，优化养殖的精力和时间。当前，得益于卫星通信和地面移动网络的部署和发展，牧民可在局部地区接入网络，沟通各方或在电商平台出售牲畜产品，也有使用无人机驱赶羊群的场景出现，但是仍需牧民近距离操作无人机。牧民的处境并无太大改变。

一个可以在荒芜人烟的地区接入，控制集群无人机驱赶，计算牲畜群，发现和驱逐野兽，监控分析草场状态并及时轮牧，识别牧草病害和识别牲畜疾病，发现优质牧草，预测出栏量，出肉量的，优化育种的系统将被放牧业所需。

草原广阔，6G 网络可为放牧提供 AI 服务，通过周期性飞行的无人机，放牧无人机、通感一体的固定基站，固定摄像头，遥感卫星和牧民手持终端收集草原，牲畜、气候相关数据。6G 网络将基于这些数据训练用于优化轮牧的轮牧模型，用于识别病虫害的识别模型，用户发现牲畜病变的识别模型，用于优化牧草和牲畜育种的育种模型等等，牧民甚至可以通过 AlaaS 平台与其他牧民共享相关模型和数据，也无需额外花费时间精力部署和维护独立的基础设施。由于 6G 网络 AI 的共享算力，数据等特性，牧民的支出成本也将大大降低。

2.3.5.2 角色和前提假设

(1) 角色：固定摄像头、地面通感融合基站、卫星基站，气象 / 环保卫星，放牧无人机、牧民、提供智能和通信服务的运营商。

(2) 前提假设：

a. 无人放牧系统中的固定摄像头、地面通感融合基站、卫星基站，气象 / 环保卫星，放牧无人机、牧民终端均接入 6G 网络。

b. 6G 网络中的具备可提供服务的算力网络系统。

c. 牧民或当地畜牧业协会已和相关高效算法模型研

发机构达成协议。

- d. 卫星网络在荒芜的草原中可用。
- e. 通感一体基站部署完成。

2.3.5.3 应用需求

(1) 业务需求

a. 分布在草原各处的固定摄像头、地面通感融合基站，扫描草原地面的卫星基站，气象 / 环保卫星，跟随牲畜群的放牧无人机等将通过 6G 网络接入互联网，6G 网络 AI 根据实际情况周期性或实时地收集羊群和草原数据，并将数据上传到边缘或云节点。

b. 6G 网络将根据需求调配相关算力和模型，分析相关数据，识别草原和羊群情况，发现病虫害，判别草原利用效率，也利用相关数据进一步优化模型；

c. 6G 网络 AI 将根据分析结果，向牧民和相关部门或组织发出警报等相关信息，并根据实际情况和牧民决策为放牧无人机生成相关指令，以控制牲畜群行进路线，从而实现牧场轮替，避开恶劣天气和病虫害草场。

d. 6G 网络将收集海量牧民的放牧数据，训练针对全局的放牧相关模型，并且支持牧民共享局部数据和模型。

(2) AlaaS 需求

a. 数据服务：网络提供 AlaaS 的数据服务包括：

牲畜群和草原的收集和更新，包括通感融合基站感知牲畜群数量信息，运动状态，草原状态等数据；

商业相关数据的管理与脱敏，包括牲畜群健康度，牲畜群数量，行进路线等；

专业数据库服务：访问病虫害生物数据库服务；

数据共享：非敏感数据共享。

b. 计算服务：网络提供 AlaaS 的计算服务包括：

根据固定摄像头和放牧无人机的摄像头等图像数据，计算牲畜群数量，识别牲畜群的健康度和牧场的病虫害，根据卫星和遥感信息感知草原的实际利用率等，生成合理轮牧方案等；

根据卫星遥感数据计算局部天气变化。

c. 模型训练：网络提供 AlaaS 的模型训练服务包括：

根据全局特性的数据例如天气和应对病虫害的措施的数据，训练全局模型；

根据历史牲畜群轮牧情况数据训练合理时间间隔和合理区域的轮牧模型；

根据实时的牧场影像和牲畜影像，结合生物数据库

识别病虫害；

根据历史放牧和牧民放牧行为，结合牲畜群行为，训练放牧无人机群运动路线和轨迹的协作模型；

根据牲畜群进食状态训练优质牧草识别模型。

d. 推理决策：网络提供 AlaaS 的推理决策服务包括：预测牲畜群状态，局部天气状态，推理轮牧方案，生成牲畜群停留时间和移动路线和区域决策，实时调整放牧无人机群运动姿态策略；

推理牧场出栏量，出肉量。

e. 模型共享：网络提供 AlaaS 的模型共享服务包括：为牧民和高校研究机构提供模型的共享，优化再共享；为牧民之间共享优秀的针对单一问题的模型，并相互验证以促进模型优化；

为牧民提供全局模型共享。

2.3.5.4 潜在技术需求

(1) 通感一体化网络：通感结合，进一步减少牧民对监控摄像头和传感器的成本投入。

(2) 卫星网络：宏观收集草原和牧场数据，也为人烟稀少且广阔的草原提供网络接入。

(3) 协同推理优化：支持分布式多实体训练架构，可以进行模型的聚合和协同训练，也方便进行模型共享。

2.3.6 疾病诊断

2.3.6.1 背景

精准的医疗一定是以“患者 / 病者诉求”为中心，实现医疗服务提供方，能按需、客观、精准、高效、可信地提供相关的医护解决方案。举应用实例说明，某患者感染上了某种皮肤炎症，他先在家通过自己的终端设备进行拍照和采样，可本地进行脱敏处理，之后基于 6G 网络上传至 MEC 边缘节点上的医疗服务器上。边缘医疗服务器 APP 基于上传的脱敏样本，先进行皮肤病例识别 AI 模型的再训练优化，和 AI 初步推理。如果推理结果不佳或不确定，可进一步上传到更上游的医疗服务器节点上（例如在核心网中），进行更大范围的二次推理和确认。6G 网络基于 AlaaS 获得初步的皮肤炎症病例推理报告，第一时间反馈给患者进行参考。如果患者对 AI 推理结果不确信或不满意，可隔日继续重复上述的请求和服务过程，最终获得自己较为确信的“皮肤病病理结果”。之后患者拿着“病理结果”去最近的当地医院，通过远

程视频会诊方式和预约的医生，进行远程诊断和相关治疗操作。患者不再担忧“看病难”“排队难”“过度医疗”等弊端问题，能以最低的成本和最高的效率获得最佳的治疗方案结果。

2.3.6.2 角色和前提假设

(1) 角色：医疗服务器节点、患者、医疗服务人员、医疗设备、提供智能服务的运营商

(2) 前提假设：医疗服务器节点、患者终端、医疗服务人员终端、医疗设备等连接到 6G 网络

2.3.6.3 应用需求

(1) 业务需求

a. 医疗设备对采集的用户数据，脱敏后上传到医疗服务器节点进行智能化分析；

b. 医疗设备等具有较低的 AI 能力，可以进行数据预处理、初步的模型推理（故障检测）、本地训练，并将初步推理结果或模型上传到服务中心；

c. 6G 网络对初步推理结果进行模型推理，将推理结果传输给医疗中心；对上传的模型进行融合，发送新模型给对应的医疗设备。

(2) AlaaS 需求

a. 数据采集：医疗设备对患者数据进行数据预处理和传输；

b. 模型推理：医疗设备等根据实时采集的数据，输入到自身 AI 模型，生成初步诊断结果，传输给基站；基站基于自身 AI 模型对初步诊断结果进行进一步处理，生成实时诊断结果；

c. 模型训练：根据检测目标，确定相同的 AI 模型，医疗设备等根据自身采集的数据在本地进行训练；

d. 模型融合：医疗设备等定期将训练模型报告给网络运营商，运营商接收来自多个医疗设备的训练模型，进行模型融合，并将新模型分发医疗设备。

2.3.6.4 潜在技术需求

(1) 网络实时推理：网络支持与医疗设备的协同实时高准确度推理。

(2) 分布式训练：网络支持分布式训练架构，可以进行模型的聚合和协同训练。

(3) 高可靠低时延传输：推理数据和推理结果需要在规定时间内成功传输。

(4) 上行大带宽传输：大量的 AI 模型参数需要上传到网络进行融合。

2.3.7 虚拟课堂

2.3.7.1 背景

通过 6G AlaaS 构建虚拟课堂，学生可以通过 VR/AR 等方式接入课堂，该课堂通过元宇宙技术构建虚拟教师和虚拟同学形象构建虚拟学习环境，并将学习内容赋予有趣的游戏环境中，帮助学生建立感兴趣的学习环境。

通过在线测试，虚拟教师获得学生的学习能力和兴趣指数，AlaaS 智能评估系统根据学生的能力和兴趣，自动定制学习计划，确定学习线路图和内容，并提供个性化的学习资源。

学生在学习过程中，可以与虚拟教师进行交互，获得实时的学习指导和反馈。如通过采集学生的表情信息，经过 AlaaS 智能评估系统建立情绪分析模型，结合学习软件设定的反馈机制构建学习的及时反馈结果。教学端根据学生学习的反馈结果进行整合学习，AlaaS 可以根据反馈结果决策教学流程和内容。

学生在完成课程学习后，可以通过模拟实际场景的虚拟现实技术进行实践，提高实际操作能力。AI 智能评估系统会根据学习成绩和反馈，自动生成学习报告，并通过短信或邮件发送给家长和教师。

学生完成学习后，AI 智能评估系统会根据学习成绩和反馈，自动生成学习报告，并通过短信或邮件发送给家长和教师。

教育的好坏有三个核心组成部分：课程、教学、评价。基于 6G 的 AlaaS 智能实时评价来促进学习，高精度的分析结果用于预测学习者是否跟上教学节奏。AlaaS 将提升学习分析的精确度，以获得跟学习者学习完成状态、应对挑战和自身需求等方面的相关信息和数据，进而用于改进学习。为学习者提供及时干预措施，辅助其学习，既可采用技术支持、教师特别关注，也可以是两者的结合。可以分析数据集，以帮助教师了解学习者如何得出答案。帮助教师更好地理解认知过程，如遗忘曲线，以及这些对学习效果的本质影响，确定学习者是否感到困惑、无聊或沮丧，提高改善学习者的情绪准备。

2.3.7.2 角色和前提假设

(1) 角色：学生学习系统，终端设备（电脑，手机，AR、VR 设备），虚拟教师，提供智能服务的网络运营商。

(2) 前提假设：终端设备连接到 6G 网络。

2.3.7.3 应用需求

(1) 业务需求

a. 学生通过 6G 网络连接学生学习系统，进入个性化学习空间，创建自身的虚拟角色；教师接入教学系统创建虚拟自己的虚拟角色；网络运营商构建虚拟课堂。

b. 通过终端设备收集学生的学习数据，包括摄像头捕获的表情数据，学生的操作习惯，测试应答情况等，对数据进行本地化预处理、初步的模型训练，并将初步推理结果或模型传递给网络运营商；

c. 网络运营商智能服务模型根据搜集到的数据建立学生画像，构建学习档案，学习路线图等。

(2) AlaaS 需求

a. 数据管理服务：通过终端采集交互数据、测试应答数据，通过摄像头采集学生情绪数据，进行数据预处理和传输；数据管理服务提供数据的安全存储、传输、和隐私保护工作。

b. 模型推理：根据收集的数据对学生进行用户画像学习。

c. 模型训练：针对海量学生数据，进行画像调整需要进行大规模的训练；根据海量数据对准确性进行提升。

2.3.7.4 潜在技术需求

(1) 上行大带宽传输：大量的 AI 模型参数需要上传到网络进行融合。

(2) 分布式训练：网络支持分布式训练架构，可以进行模型的聚合和协同训练。

(3) 隐私保护：对数据进行脱敏处理。

(4) 高可靠低时延传输：推理数据和推理结果需要在规定时间内成功传输。

2.3.8 电网智能巡检

2.3.8.1 背景

在电力网络中，电力从发电厂经过电网传输到用电中心，然后分配到居民楼、写字楼等用电端点。由于用电者本身的分布式特性和未来风能、太阳能分布式电源生产者的增加，由高压线路和变电站等构成的输电和变电网络覆盖面积异常大。通过巡检，排查故障隐患，是确保电网安全可靠运行的前提。而长距离大面积的供电线路导致故

障检测非常困难，传统基于人力的监测效率很低、巡检不全面、安全风险大。利用 AI 的视频分析能力，通过摄像头、无人机等，可以实现更准确、更及时、更安全的智能巡检，包括杆塔 / 线路的故障、变电站人员、设备和业务识别等。

考虑到大规模分布的电力线路和杆塔等，6G 网络可以为智能电网运营者提供大规模的 AI 实时推理、分布式模型训练等服务。无人机等将采集到的视频数据经过处理后，传输到具有 AI 能力的 6G 网络，6G 网络一方面可以对视频进行实时处理，检测出故障或非法入侵，并传输给控制中心；另一方面，6G 网络可以收集多个基于本地数据训练故障检测的 AI 模型，通过模型融合提升检测的性能，同时保护各个电网运营者的数据隐私。

2.3.8.2 角色和前提假设

(1) 角色：无人机、机器人、监控摄像头、监控中心、具备计算能力的基站、提供智能服务的网络运营商

(2) 前提假设：无人机、机器人、监控摄像头、监控中心连接到 6G 网络

2.3.8.3 应用需求

(1) 业务需求

a. 监控中心定期或者根据突发状况，向无人机、机器人或监控摄像头等发起巡检指令；无人机或机器人等根据巡检路线对电力线路等进行视频数据采集；

b. 无人机和机器人等具有较低的 AI 能力，可以进行数据预处理、初步的模型推理（故障检测）、本地训练，并将初步推理结果或模型上传到基站；

c. 6G 网络对初步推理结果进行模型推理，将推理结果传输给监控中心；对上传的模型进行融合，发送新模型给无人机或机器人等终端设备。

(2) AlaaS 需求

a. 数据采集：无人机或机器人等对采集任务过程中的视频和传感数据，进行数据预处理和传输；

b. 模型推理：无人机或机器人等根据实时采集的视频数据，输入到自身 AI 模型，生成初步检测结果，传输给基站；基站基于自身 AI 模型对初步检测结果进行进一步处理，生成实时检测结果；

c. 模型训练：根据检测目标，确定相同的 AI 模型，无人机活机器人等根据自身采集的数据在本地进行训练；

d. 模型融合：无人机或机器人等定期将训练模型报告给网络运营商，运营商接收来自多个无人机或机器人的训练模型，进行模型融合，并将新模型分发给无人机或机器人。

2.3.8.4 潜在技术需求

- (1) 网络实时推理：网络支持与无人机、机器人、监控摄像头的协同实时高准确度推理。
- (2) 分布式训练：网络支持分布式训练架构，可以进行模型的聚合和协同训练。
- (3) 高可靠低时延传输：推理数据和推理结果需要在规定时间内成功传输。
- (4) 上行大带宽传输：大量的 AI 模型参数需要上传到网络进行融合。

2.4 6G AlaaS 典型服务

AI 服务是指向被服务方提供 AI 技术、AI 业务或 AI 资源要素。6G AlaaS 典型服务包括 6G 网络为用户提供大规模分布式模型训练、推理、生成、优化等服务。服务对象可以是终端用户、第三方用户、网络内运维等。

2.4.1 模型推理服务

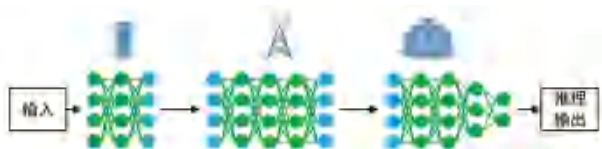


图 2-11 AI 模型推理业务

模型推理是功能执行的基础。6G 智能内生网络基于全域覆盖和泛在的基础设施，提供近实时、实时不同需求的模型推理能力。在分布式模型推理服务中，网络利用融合通信与计算的能力，通过模型协作，为用户提供实时的高精度模型推理业务，让计算能力受限的终端也可以享受高性能 AI 服务。不同于 5G 只作为数据传输管道的服务模式，6G 模型推理服务可以充分利用边缘网络节点的计算和模型资源，提供更低时延的高精度模型推理服务。一种典型的 AI 模型推理业务如图 2-11 所示。大模型切割后将计算需求高的部分部署在网络侧，利用网

络强大的算法，为终端用户提供联合的模型推理服务。

2.4.2 模型训练服务

模型训练是 AI 能力的关键，模型的好坏关系到 AI 能力的高低。6G 智能内生网络面向用户、行业和网络自身的不同特点，提供更合适的模型训练算法，以及更合理的编排模型训练算力，提升训练模型的速度与精度。在大规模分布式模型训练中，网络作为管理平台，为分布式智能体之间的数据或模型参数交互提供高速通道和高效调度机制，支持快速的模型融合与分发，同时实现高效模型训练与用户隐私保护。一种典型的联邦学习训练业务如图 2-12 所示。终端利用本地数据训练模型，网络将各终端模型进行融合分发，实现数据不出用户的联合学习。

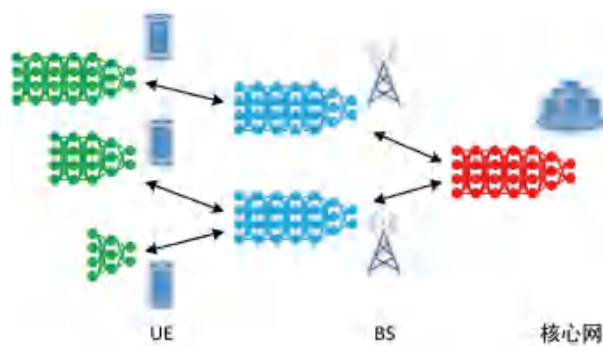


图 2-12 分布式 AI 模型训练业务

模型训练（model training），是在神经网络（neural network）模型初始化和训练数据收集、清洗之后，通过数据驱动的方式进行机器学习，也就是神经网络中参数的优化的过程，来找到输入、输出数据之间的映射关系。

从数据采集和网络训练的实时性角度，模型训练可分为离线训练（Offline training）和在线训练（Online training）。离线训练是在预先收集好的静态数据上进行模型的训练，模型训练的节点或设备可以与模型部署和推理应用的节点或设备不同。这种训练通常是非实时的，所以可能无法适应实时数据的变化。训练是指在生产环境中使用实时数据对模型进行更新，模型训练的节点或设备多数在模型部署和推理应用时的节点或设备上。这种训练具有实时性的优点，可以应对模型对数据的实时变化，然而在线训练对计算资源和开销的消耗较大，并且存在过拟合与慢收敛的风险。因此目前多数模型训练

可能会基于离线训练的方式。对于离线训练出的模型与实际部署环境的差异，也可以采用先离线学习、后在线微调更新的方式，而在线微调的过程和在线学习过程近似。训练方式也可以通过不断的训练更新，以及训练数据的清洗与迭代，进一步演化为持续性的模型训练方式。

从机器学习的方式与目标的角度，模型训练可分为监督学习（Supervised learning），无监督式学习（Unsupervised learning），强化学习（Reinforcement Learning）半监督学习（Semi-supervised learning），迁移学习（Transfer learning）和元学习（Meta learning）等。其中前三种主要从数据中进行模型的学习来改善系统性能，而后三者则是强调利用已有的知识和数据来提高模型的效率和泛化能力。监督学习训练时需要输入的数据和标签，来构建其之间的映射关系；无监督学习没有预先给定的标签，而是通过对数据自主的学习来发现数据中的隐藏结构与模式；在强化学习中，通常需要与环境的进行交互，来学习如何做最好的决策；半监督学习则需要模型训练时有标签的输入数据和无标签的输入数据，来提高模型性能；而迁移学习与元学习则更强调模型通过学习，如何提高自身的泛化能力与快速适应新任务的能力。

在通信系统中，从模型是在单一设备上（基站或 UE）进行还是双方共同（基站和 UE）进行，模型训练也可以分为独立训练和协作训练。独立训练时，模型只在一方的设备上就可以完成，如用机器学习来优化信道估计的模型训练，只需要在接收机上就可以完成。而协作训练方式则需要发射机和接收机共同完成，如信道状态信息的压缩与反馈（CSI Feedback），模型训练的过程中需要对 UE 端的编码器（Encoder）和基站端的解码器（Decoder）一起训练。

模型的训练方式可以分为集中式训练和分布式训练。集中式训练是指将所有训练数据和计算资源集中在单数据中心或节点上。在集中式训练中，数据的存储、训练是集成的，易于实现与管理，并不需要大范围的通信，但其计算资源受限，并存在单一节点故障的风险。分布式训练是指将训练任务分配给多个计算节点进行并行计算。将模型在多个去中心化的边缘节点上进行训练，每个节点用本地数据对模型进行训练，最后把模型合并为一个模型。在分布式训练中，每个节点用本地数据对模型进行训练，数据中心之间不需要交换训练数据，仅通过交换模型参数或中间结果的方式来构建全局模型。

2.4.3 模型生成服务

模型生成服务是面向业务开发者，由 6G 网络针对不同应用场景训练 AI 模型。业务开发者不需要做额外的模型开发工作，只需要选择所需的应用场景，向 6G AlaaS 输入业务配置参数和训练数据；模型生成服务利用 6G 网络内的模型和算力，自动进行算法组合和训练，输出给业务开发者所需应用场景的 AI 模型。

2.4.4 模型优化服务

模型优化是 AI 生命周期管理中的重要步骤。6G 智能内生网络除了对模型针对性的提升模型精度、泛化性能等优化能力以外，更能在大范围以及跨场景的情况下发挥优势，合理利用不同数据，对区域内或功能集合进行整体最优的模型优化。

2.4.5 模型共享服务

模型共享服务是指在相同的场景、需求或任务下 6G AlaaS 复用已有模型的服务，各个模型生产者将已经训练好的优秀模型发布，而 6G AlaaS 根据 AI 服务消费者的需求选择适合的已有模型。模型共享将是 AlaaS 的基础服务，将为模型生产者提供变现渠道，也降低 AI 服务消费者的服务使用支出，也将降低 6G 网络总体的算力使用压力。是构建 AlaaS 重要一环。

模型共享服务将判别模型生产者的模型产品的通用性程度，从而判别模型的可共享范围，AlaaS 收到新的请求后，根据收集到的数据、请求的实际需求等，判别是否可以共享已有的旧模型，而非直接接触发生成新模型的流程。如果已有的旧模型库与实际需求不匹配，先对需求进行分割，尽可能分成多个子基础任务，从而复用部分已有的优秀的基础模型，而后为当前基础模型无法应对的子任务进行少量的新模型设计训练工作。

AlaaS 的模型共享有多种实现方式。首先为了兼容当前已有的商业模式和传统模型设计训练方法，算法模型提供商针对具备价值的场景自行设计训练模型，而后上传到模型共享平台上，以供消费者付费使用。在不远的未来，算法模型生产者和消费者，还有消费者之间将会通过 6G 网络相互连接，消费者们通过 6G 网络上传自身数据甚至直接通过 6G 网络的通感一体能力收集相关数据，而后在算力节点中将数据和算法模型生产者设计的基础模型交汇。最终生成的成品将由各方共享。当然，

一些特殊的算法模型生产者例如高校研究机构或行政部门，可能不直接与算法模型消费者直接共享经济效益，而是希望获取相关行业的数据经过模型处理后的结果。

2.4.6 模型部署服务

AI 模型近年来通过诸如搜索、广告、推荐、NLP、CV 等多个领域，不断渗透进入用户的日常生活，并不断突破传统模型的性能瓶颈。AI 模型在训练完成之后需要交付给用户，供用户便捷使用，这需要对训练好的模型进行部署管理。模型部署是交付给用户 AI 产品时必不可少的一个环节，而 6G 网络将在模型部署的过程中提供重要支撑。

模型部署一般会按照部署平台分为三大类。

第一类是将 AI 模型在后端服务器上部署，也可以称作云端部署。常见的服务器模型部署的相关用例主要有在线搜索、在线翻译、在线识图等等。这类业务的请求量相对较大，因此高精度和大吞吐量更为优先，而对于时延并不敏感。以在线翻译应用为例，用户并不要求要在输入文字一秒之内得到翻译结果，两至五秒都是可以接受的，所以这类服务并不需要 AI 模型实时地返回结果。但出于精度和数据量考虑（比如翻译内容的精准度和语言的多样性），此类模型较为复杂，无法部署在一些资源有限的移动平台上，因此必须要在服务器上进行部署。在此场景下，基于内生 AI 与数据服务紧密结合的特点，6G 网络 AI 高质量的数据服务能够获取、处理和存储海量、多源、多样的数据，支持云端模型部署。

第二类部署平台是基站、边缘计算设备。以网络节能服务为例，部署于基站上的 AI 模型对基站参数、用户位置信息和业务分布状态进行推理演算，在保证通信网络业务性能的前提下，自主生成节能策略，完成通道关断、基站开关、多维资源调度、载波频段和负载功率调整等一系列操作，以实现有效降低基站能耗。随着 6G 算网融合的不断发展，“端边协作”的模式将会在未来成为主流，此类模型部署将发挥更大的作用。例如当算力有限的终端需要执行一项任务时，基站或边缘计算设备上部署的 AI 模型将通过通信网络为其提供算力协助，根据终端的要求分配计算资源以达到性能需求，最终协同终端一同完成任务。对 6G 网络而言，基站、边缘计算设备模型部署的进一步发展，意味着更多 AI 处理流程将在边缘计算

层实现，而无需交由云端。由于边缘计算层更加靠近用户，响应会更加迅速。同时，这还能够大幅度提升处理效率，减轻云端负荷。

第三类部署平台是设备终端，或者说嵌入式端。在终端部署 AI 模型时，设备的资源是有限的，能耗、体积相较于服务器端的限制大很多。此时主要通过 SDK 部署的方案，将模型打包封装到 SDK，集成到嵌入式设备，在手机以及各类 IoT 终端设备上执行 AI 模型推理功能，如门禁闸机中的人脸检测系统、手机拍照中的笑脸自动抓拍系统。近年来发展迅速的手机在线会议也是典型服务用例，手机需要对参会人进行人脸识别，并对视频进行实时的抠图、替换背景，这要求部署的模型兼顾精度与速度。6G 网络能为本地的模型部署带来实时性更高，综合能耗更低的智能化服务。同时，大规模分散部署于不同设备终端的模型，能够通过 6G 网络网元间的分布式协作，支持本地实时性要求较高的应用场景，实现单一设备终端有限算力所无法完成的 AI 任务。因此，6G 网络很好地解决了端侧算力不足的问题，更大程度地发挥设备终端模型部署功耗低、模型小、部署灵活、隐私性高等优势。

2.4.7 模型管理服务

模型管理服务，是指 AlaaS 提供的模型开发、部署、优化、迭代的全生命周期管理过程。包括如下过程：

- 基于业务目标出发构建 AI 学习问题（场景，QoAIS, KPI，监控需求）；
- 收集数据、准备数据，包括特征工程、清理等任务；
- 分析、训练、测试和评估模型，构建和自动化机器学习管道；
- 部署模型，静态部署或动态部署。
- 模型应用，如预测和推断。
- 监控、优化和维护模型。

由于 AI 技术的局限性，如数据质量变化，模型衰减，数据局部性等问题，要求模型能够持续迭代。对于不同的 AlaaS 服务，需要支持不同的 QoAIS，以及基于任务的管道和资源的编排。只有通过模型生命周期管理所有步骤中实现自动化和监控，并建立模型持续优化和改进流程，使能 AlaaS 全场景自动闭环，自诊断，自修复和自优化，将促使 6G 网络实现 L5 级“自动驾驶”网络的自治管理模式和自我演进。

2.4.8 模型压缩服务

现有研究表明，很多深度神经网络中都存在着显著的冗余性，AI 模型中存在着一种过参数化的现象，即不是所有的参数对模型的表达能力都起到关键作用，模型压缩是减小模型规模的有效途径。模型压缩方法，就是利用网络结构的冗余特性对 AI 模型进行重构、简化并且加速的技术，在不影响原任务完成度的情况下，得到参数更少、结构更经典的模型。目前主流的模型压缩方法有剪枝、量化、知识蒸馏等几类。剪枝和量化主要针对模型中的冗余参数进行且最小化精度的损失；知识蒸馏则是先训练一个大模型，再训练一个更紧凑的模型，将已训练好的大模型所包含的知识，蒸馏到小模型里以达到相同效果。

未来 6G 网络中，AI 业务愈发普及，各类 AI 模型将被广泛应用到丰富的领域与场景，AI 模型的复杂度、计算能力、学习能力不断提高，在性能表现上也会更加卓越。但是随着神经网络层数的加深、参数规模的增大，模型对存储设备和计算设备的要求也将增加。高性能服务器具有强大的算力可以承载复杂的 AI 模型，但是对于部分对推理时间或隐私要求较高的场景，需要在资源有限的端侧上

（移动端或嵌入式设备）部署 AI 模型进行 AI 推理，例如智能手机、智能穿戴设备、人脸闸机等。这些设备受到体积的限制，计算能力和存储能力相对较弱，在性能上与服务器有较大差距。如果将复杂的 AI 模型直接部署在端侧设备上，其性能无法实现正常的 AI 推理。考虑到端侧推理的性能，如时延、能耗以及内存占用等，计算量大、存储成本高以及模型复杂度高等问题将会使端侧部署难以有效的进行。同时，不仅对于端侧部署，云端部署也希望能够使用更小存储空间和更少的算力，获得相同或者更好的效果。

6G AlaaS 以算网融合的网络架构为基础，支持“端 - 边 - 网 - 云”协同发展，为各类部署场景提供 AI 模型压缩服务，运用 6G 网络中无处不在的内生 AI 算力，加快剪枝、量化、知识蒸馏的流程，最大程度降低模型复杂度与存储空间，并促进对建模的训练和推理。6G AlaaS 提供计算与通信的深度融合，有望实现模型压缩策略的实时调整。例如，当连接带宽受限，但本地算力充足时，增加本地计算量，对本地用户所需的 AI 模型进行高保真度的压缩；当连接带宽充足、质量稳定，但本地算力受限时，减少本地计算量，通过增加周边节点的协作，在云端或边

缘计算设备上完成模型压缩任务。内生 AI 也有望为模型压缩制定更智能的结构选择策略。当前的模型压缩方法，无论是剪枝、量化或是知识蒸馏，都是以现有 AI 模型的主干网络结构作为基础，再手动选择策略缩小模型规模。

未来，6G AlaaS 能够利用内生的强化学习等策略进行自动网络结构搜索，跳出现有框架，从而得到更紧凑的网络结构。相较于目前的 AutoML，6G AlaaS 能够运用对特定应用场景下数据的感知能力，为压缩模型的选择提供更加丰富的数据来源，还能够利用分布式的高性能 AI 算力，降低大规模模型压缩所产生的能耗开销。

2.4.9 数据管理服务

当前以数据驱动的 AI 范式中，高质量数据是 AI 能力要素的根本。6G 智能内生网络开放的数据管理服务包括数据收集、数据预处理、数据脱敏、数据存储、数据开放等。实现用户隐私数据的安全管理，并能够在脱敏处理下，与关联行业进行数据的联合使用。

2.4.10 计算服务

计算服务是指专门为高算力要求的 AI 应用提供专门的算力支撑服务。包括分布式算力集群的算力分配与调度，满足延迟、吞吐量指标的算力网络协同服务。大模型现在已经成为通向通用人工智能可以预见的一条道路，因此 AlaaS 需要能支持构建大模型的能力，而这些就需要强大的算力支持。6G 智能内生网络针对特定大模型构建定制化的算力部署，对外提供 API 接口等，提供基础算法库，并能根据特定应用生成新的算法模型。

3

第三章 6G AI 即服务 (AlaaS) 性能指标

现有移动网络的设计目标主要是面向连接的数据传输，其关键性能指标主要包括连接的传输速率和传输时延等。6G AlaaS 基于 6G 网络内的连接、计算、模型、数据等资源，提供适应不同应用场景的 AI 能力。不同于传统的移动网络，6G 网络需要配置除连接之外的资源，为用户提供高性能 AI 服务。相应地，需要从多个维度来构建 6G AlaaS 的性能指标和服务质量保障体系。一方面，6G AlaaS 需要新的包含连接、计算、模型等维度的 QoS，保障 AI 服务质量。另一方面，6G 网络在设计时，需要考虑如何将通信连接能力、AI 模型和计算能力等进行融合，构建面向 AI 服务的综合性能指标体系和评估方法，为系统设计和网络资源配置提供指导。下面从 6G AlaaS QoAIS 和系统关键性能指标两个方面分别阐述指标定义和评估方法。

3.1 6G AlaaS QoAIS

3.1.1 QoAIS 定义

QoAIS (Quality of AI Service) 是对 AI 服务质量进行评估和保障的一套指标体系和流程机制 [16]。未来，6G 网络将不再只是服务于传统通信业务的管道，6G 网络将构建内生于网络的 AI 能力，形成一套可服务于多种智能应用场景的能力体系，即 AlaaS。QoAIS 反映了用户层面对于 AI 服务质量的需求，并将需求量化、导入 6G 智能内生网络，使网络能够按照用户提出的指标提供相应的 AlaaS 以保障其需求。

3.1.2 QoAIS 指标体系

6G 时代将出现各种新型智能应用场景（如网络高水平自治、行业用户智能普惠、用户极致业务体验、网络内生安全等），不同场景对 QoAIS 将有着不同的需求，因此需要一套指标体系通过量化或分级的方式表达用户层面的需求，以及网络编排控制 AI 各要素（包括算法、算力、数据、连接等）的综合效果。

传统通信网络的 QoS 主要考虑通信业务的时延和吞吐率（MBR、GBR 等）等与连接相关的性能指标。6G 网络除了传统通信资源外，还将引入分布式异构算力资源、存储资源、数据资源、AI 算法等 AI 服务编排的多种资源元素，因而需要从连接、算力、算法、数据等多个维度来综合评估网络内生 AI 的服务质量。同时，随着“碳中和”和“碳达峰”政策的实施、全球智能应用行业对数据安全性和隐私性关注程度的普遍加强，以及用户对网络自治能力需求的提升，未来性能相关指标将不再是用户关注的唯一指标，安全、隐私、自治和资源开销方面的需求将逐渐深化，成为评估服务质量的新维度，而不同行业和场景在这些新维度上的具体需求也将千差万别，需要进行量化或分级评估。因此，QoAIS 指标体系从初始设计时，即需要考虑涵盖性能、开销、安全、隐私和自治等多个方面，需从内容上进行扩展。

6G 网络内生的 AI 服务可以分为 AI 数据类、AI 训练类、AI 推理类和 AI 验证类，每一类 AI 服务均需要一套 QoAIS。表 3-1 提供了一种针对 AI 训练服务的设计方式。



图 3-1 QoAIS 和 AlaaS 关系图示例

AI 服务类型	评估维度	QoAIS 指标
AI 训练	性能	性能指标界、训练耗时、泛化性、可重用性、鲁棒性、可解释性、损失函数与优化目标的一致性、公平性
	开销 *	存储开销、计算开销、传输开销、能耗
	安全 *	存储安全、计算安全、传输安全
	隐私 *	数据隐私等级、算法隐私等级
	自治	完全自治、部分人工可控、全部人工可控

表 3-1 AI 训练服务的 QoAIS 指标体系

其中，“性能指标界”是评估模型性能好坏指标的上界和下界，如模型错误率、查准率、召回率等性能指标的范围。“泛化性”指模型经过训练后，应用到新数据并做出准确预测的能力。“可重用性”是模型在应用场景变化时能够继续使用的能力。“鲁棒性”指在输入数据受到扰动、攻击或者不确定的情况下，模型仍然可以维持某些性能的特性。“可解释性”是指模型能支持对模型内部机制的理解以及对模型结果的理解的程度。“损失函数与优化目标的一致性”是指模型训练过程中，对损失函数的设计与 AI 用例的优化目标的一致程度，比如函数中考虑的变量个数是否完全覆盖智能优化场景的优化目标指标。

“自治”指对 AI 数据 / 训练 / 验证 / 推理服务的工作流中自主运行部分和人工干预部分的要求，反映了用户对 AI 服务自动化程度的要求。自治分为三个等级：完全自治（全流程自动化的 AI 服务，全程无需人工干预）、部分人工可控（AI 服务的工作流在部分环节自动化，部分环节要求人工辅助）、全部人工可控（AI 服务工作流的各环节均要求人工参与）。

3.1.3 QoAIS 映射模型

QoAIS 是网络内生 AI 编排管理系统和控制功能的重要输入，网络内生 AI 管理编排系统需要对顶层的 QoAIS 进行分解，再映射到对数据、算法、算力、连接等各方面的 QoS 要求上。

图 3-2 展示了 QoAIS 各指标维度和各资源维度上的 QoS 之间的映射关系。AI 服务的 QoAIS 整体指标拆解到各指标维度上的 QoAIS 指标，再进一步映射到各资源维度上的 QoS 指标，由管理面、各资源维度的控制面和用户面机制进行保障。各资源维度上 QoS 指标可分为适合量化评估的指标（如各类资源开销）和适合分级评估的指标（如安全等级、隐私等级和自治等级）。在前一类指标中，有部分指标的量化方案已成熟或较容易制定（如训练耗时、算法性能界、计算精度、各类资源开销等），部分指标目前尚无定量评估方法（如模型的鲁棒性、可重用性、泛化性和可解释性等），如表 3-2 所示。

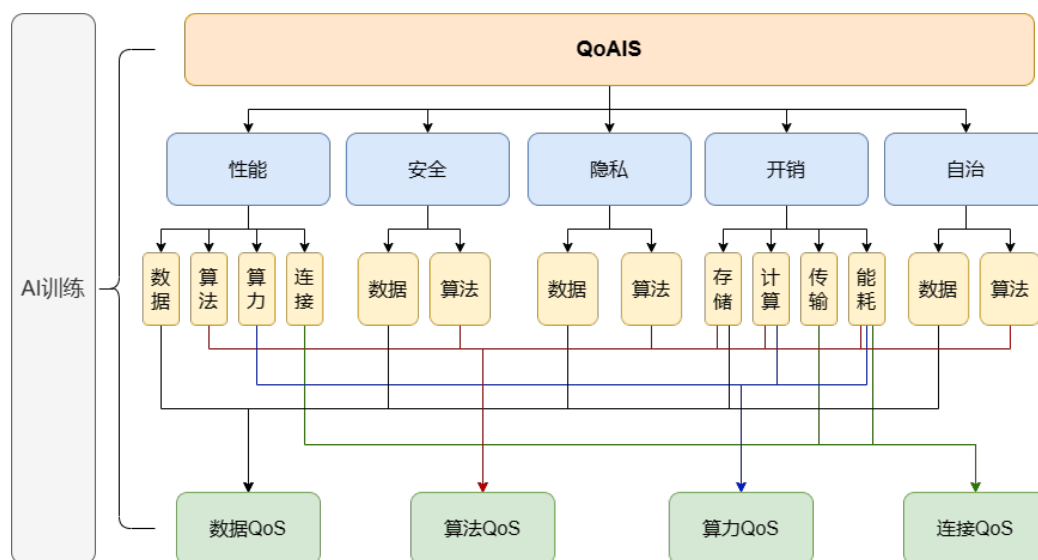


图 3-2 QoAIS 指标分解到各资源维度上的 QoS 指标

指标维度	QoAIS 指标	资源维度	可量化指标	尚无量化方案指标
性能	性能指标界、训练耗时、泛化性、可重用性、鲁棒性、可解释性、优化目标匹配度、公平性	数据	特征冗余度、完整度、数据准确度、数据准备耗时	样本空间平衡性、完整性、样本分布动态性
		算法	性能指标界、训练耗时、是否收敛、优化目标匹配度	鲁棒性、可重用性、泛化性、可解释性、公平性
		算力	计算精度、时长、效率	无
		连接	带宽及抖动、时延及抖动、误码率及抖动、可靠性等	无

表 3-2 AI 训练服务性能 QoAIS 到各资源维度的映射

3.1.4QoAIS 具体场景指标分析

根据前文所述，QoAIS 在性能、开销、安全、隐私、自治五个评价维度都有相应的 QoAIS 指标，特别是在性能层面，数据、算法、算力、连接四大资源维度下均有适合量化评估的服务质量指标映射。因此，下面将在多小区天线波束联合赋形的场景用例下，对内生 AI 网络架构中基于 QoAIS 的 AI 工作流编排方案（包括集中式、分布式和协同式）展开分析，选取多个典型的可量化 QoAIS 指标进行介绍。

图 3-3 展示了集中式、分布式、协同式三种 AI 工作流编排方案下，上述 AI 任务在集中式云脑和多个分布式边脑中的分布情况。该用例利用 6G 智慧内生 AI 架构解决在多个基站覆盖的区域中出现人群聚集场景时，基站天线的波束权值动态调整方案。因此，设置三种不同的场景，场景一、场景二和场景三分别含有 10%、50% 和 100% 的动态用户。对于每一种场景，分别使用集中式、分布式和协同式三种方案进行波束选择，并分析三种方案对 QoAIS 相关指标的满足程度。

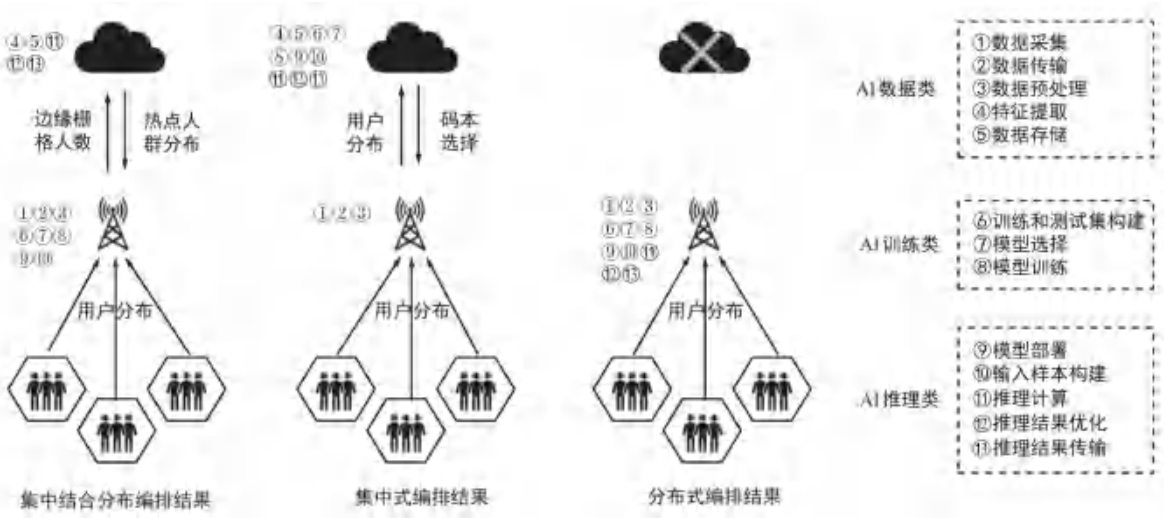


图 3-3 三种不同编排方案的 AI 任务分布图

3.1.4.1 性能指标界

在该场景用例下，内生 AI 能够在遇到人群热点时，通过对人群分布作出准确的预测，及时指导基站天线作出决策。因此，以用户分布预测准确度作为衡量性能指标界的关键因素。将区域栅格化，用户分布预测准确度相当于利用三种编排方案预测结果与真实分布结果数值相同的栅格数与真实分布中的栅格数之比。

场景	用户分布预测准确度 /%		
	集中式方案	分布式方案	协同式方案
场景 1	98.50	97.48	97.64
场景 2	98.77	97.73	98.14
场景 3	98.35	97.47	99.19

表 3-3 用户分布预测准确度

如表 3-3 所示，集中式方案在场景一和场景二下的预测准确度较高，协同式方案在场景三下的预测准确度较高，分布式方案在三种场景下的预测准确度较低。

三种编排方案的性能指标界如表 3-4 所示，协同式方案的预测准确度上界最高，分布式方案的下界最低。

性能指标界	用户分布预测准确度 /%		
	集中式方案	分布式方案	协同式方案
上界	99.75	97.50	99.77
下界	98.47	96.40	97.59

表 3-4 性能指标界对比

3.1.4.2 优化目标匹配度

一种量化优化目标匹配度的方式是计算模型训练过程中损失函数的参数变量对 AI 用例优化目标指标的覆盖程度。比如，在本用例场景下，若优化目标指标包括 RSRP 覆盖性能和 SINR 覆盖性能，而损失函数的设计仅包含 RSRP，则并非完全匹配。具体的计算公式可设计如下：

$$\theta = \alpha \times f(RSRP) + \beta \times f(SINR)$$

其中， α 代表 RSRP 的权值， β 代表 SINR 的权值， $(\alpha + \beta = 1)$ ， θ 代表优化目标匹配度。 $f(RSRP)$ 与 $f(SINR)$ 作为损失函数是否包含相应优化指标的 0 - 1 函数（包含则为 1，否则为 0），在本用例中，三种编排结果的优化目标都是相同的，故而三种编排结果的匹配度相等。

3.1.4.3 鲁棒性

鲁棒性用来衡量方案结果的抗干扰性，针对该场景用例，以用户分布预测模型准确度的方差表示方案的鲁棒性程度，预测结果的方差越小，鲁棒性越高。三种编排结果的鲁棒性分别为：集中式方案 31.58×10^{-7} ，分布式方案 25.14×10^{-7} ，协同式方案 1.082×10^{-7} 。集中式方案的鲁棒性差于分布式方案，协同式方案的鲁棒性最好。

3.1.4.4 传输、存储、算力开销

对于该场景用例中的 QoAIS 开销类指标，从数据传输、存储和算力开销三方面对比。

从表 3-5 可以看出，集中式方案在传输数据上传量和训练数据的存储量上需要较大的数据开销，同时需要较多的算力资源。分布式方案因为数据决策都在本地执行，所以没有传输数据量，数据的存储和算力消耗也比较少。协同式方案由于云边之间的反馈，所以传输数据下发量需要的开销较大，其余指标与分布式方案相同。

指标列项		开销		
		集中式方案	分布式方案	协同式方案
传输数据量	上传数据量	3330kB	0kB	463kB
	下发数据量	1.2kB	0kB	1924kB
存储数据量	训练数据	12.3GB	1.99GB	1.99kB
	推理数据	9990kB	9990kB	9990kB
	模型大小	1kB	17kB	17kB
算力消耗 (单位：MFLOPS)		1150	268.6	268.6

表 3-5 开销对比

除了以上评估维度，QoAIS 也可以包括智能应用的性能指标。以信道压缩为例 [17]，可以选择归一化均方误差（Normalized mean square error, NMSE）或是余弦相似度作为信道恢复精度的 KPI，也可以选择链路级 / 系统级指标（如误比特率或吞吐量等）作为反映信道反馈精度对系统性能影响的 KPI。此外，QoAIS 还可以包括 AI 服务的可获得性、AI 服务的响应时间（从用户发起请求到 AI 服务的首条响应消息）等与 AI 服务类型无关的通用性评价指标。

3.2 6G AlaaS 关键性能指标

6G AlaaS 利用网络内的通信、模型和计算等资源，使能高效的模型训练和实时高精度模型推理等 AI 服务。AI 服务质量同时取决于模型能力、用户和网络侧的计算能力，以及用户与网络连接的通信性能等，因此需要从端到端服务、通信、模型和计算多个维度来综合评价 6G 性能。另外，随着“碳中和”政策的实施，碳排放也成为了评估网络性能的新维度。因此，亟需在 6G 网络形成一套基于通信能力与 AI 能力的 AI 服务关键性能指标和评估体系，从而为 6G 系统的设计和通信、计算等资源的配置提供指导。

3.2.1 性能指标定义

从 6G 网络提供 AI 模型训练和推理等服务、AI 数据服务的能力的角度来对 AlaaS 的性能指标进行定义，我们认为其提供的 AI 服务的性能同时取决于系统底层所能提供的如 AI 模型相关的能力、用户和网络侧的计算能力，以及用户与网络连接的通信能力等。我们首先有如下表 3-6 中所综合定义的 AI 服务性能指标列表：

	名称	定义	单位
AI 服务性能指标	服务准确度	用户体验的 AI 服务准确度	%
	服务覆盖率	给定链路预算下，满足服务准确度的区域面积占比	%
	服务响应时延	端到端 AI 服务响应时延	ms
	系统能效	单位能量消耗所能处理的数据量	bit/J
	服务密度	单位区域内的服务速率	bps/km ²
	服务完整度	数据服务的完整度	%
	服务冗余度	数据服务的冗余度	% 或 bit

表 3-6 AI 服务的关键性能指标

1) **服务准确度**：反映了用户所能体验的 AI 服务的精度，如目标检测或物体识别的精度。对于不同的业务，该准确度的要求可能不同，如对于自动驾驶业务，物体识别的精度要求要远高于用户对花草的分类。

2) **服务覆盖率**：反映了网络中的用户能享受到 AI 服务的比例，高的服务覆盖率既要求网络提供良好的链路覆盖，也要求网络部署能够覆盖更多数据的模型和计算，以提供满足服务准确度要求的 AI 性能。

3) **服务响应时延**：反映了用户从请求到得到 AI 服务的时间，对于不同业务要求也不相同，如自动驾驶业务中物体识别的时延要求远高于普通消费者日常识别花草的要求。

4) **系统能效**：反映了网络提供 AI 服务的能量效率，包括计算和通信的能耗。AlaaS 需要追求以总体最低的能耗提供更好的 AI 服务，以实现绿色生态。

5) **服务密度**：反映了网络能够同时提供的 AI 服务的能力，即能够同时提供 AI 服务的用户数。

6) **服务完整度**：反映了网络提供 AI 数据服务的质量，高质量的数据既要求数据本身不能存在缺失、错误项，还要求数据的标签标注准确。

7) **服务冗余度**：反映了网络提供 AI 数据服务包含的有效信息占比。对于 AI 模型训练，冗余的相似数据不仅会造成额外存储、通信和计算开销，还会导致模型易于过拟合。AlaaS 需要追求更高效的数据服务。

对于 AlaaS, 不仅涉及到用户和网络的 AI 模型和计算，还涉及到用户与网络的连接，包括用户与网络之间数据或者模型的传输，用户侧和网络侧的数据采集、处理、评测和监测等。因此，在表 3-6 所定义的 AI 服务性能指标需要由通信性能和 AI 模型 / 计算性能两个方面综合决定。例如，AI 服务响应时延既包括 AI 模型推理所需要的计算时延，也包括传输数据或者推理结果所需要的通信时延。

与表 3-6 中所列 7 项主要指标对应，下面尝试将 AlaaS 涉及到的通信相关的性能指标、AI 模型 / 计算、AI 数据相关的性能指标分列在表 3-7、表 3-8 和表 3-9 中。

	名称	定义	单位
通信性能指标	可靠性	传输链路的可靠度	%
	链路预算	达到特定数据速率时基站与终端之间的最大耦合损耗	dB
	数据速率	用户体验的数据速率（特别是上行）	bps
	区域传输容量	单位区域内的数据吞吐量	bps/km ²
	通信能效	单位能量消耗所能传输的数据量	bit/J

表 3-7 通信相关的关键性能指标

	名称	定义	单位
AI 模型 / 计算性能指标	推理准确度	模型推理结果的准确度	%
	数据分布覆盖范围	达到模型推理准确度的数据分布所能覆盖的范围	km ²
	训练或推理时延	模型训练达到目标准确度（训练）或者模型推理一次前向传播的时间（推理）	ms
	训练或推理数据密度	单位区域单位时间内模型能处理的比特数	bps/km ²
	训练或推理能耗	单位能量消耗所能支持的训练或推理的浮点运算次数	FLOPs/J

表 3-8 AI 模型和计算相关的关键性能指标

	名称	定义	单位
AI 数据性能指标	数据完整度	数据集的无损数据占比	%
	数据冗余度	数据集的冗余数据占比或数据冗余程度	% 或 bit

表 3-9 AI 数据相关的关键性能指标

从通信的角度来看，考虑影响端到端 AI 性能的关键指标，包括数据速率、可靠性、区域传输容量、链路预算和通信能效，相应的定义如表 3.7 所示。对于这些性能指标，网络需要配置带宽和传输点密度等资源，以满足服务需求。而涉及到 AI 模型和计算相关的关键性能指标包括模型的推理准确度和数据分布覆盖范围、模型训练或推理的计算时延、数据密度和能耗，相应的定义如表 3-8 所示。对于不同的 AI 服务，关键性能指标可能会不同，如推理服务更关注准确度和时延。对于这些性能指标，网络需要配置 AI 模型和计算等资源，以满足服务需求。涉及到 AI 数据相关的关键性能指标包括数据完整度和数据冗余度，相应的定义如表 3.9 所示。对于这些性能指标，网络需要配置数据采集的功能，配置数据存储的存储资源和数据处理的计算资源等，以满足服务需求。



图 3-4 AI 模型推理服务关键性能指标体系

3.2.2 性能指标模型

按 3.1.1 节中定义的系列指标，我们以 AlaaS 中最为典型的模型推理和模型训练业务为例，具体讨论 AI 服务的关键性能指标模型，包括与通信和 AI 模型 / 计算的关系。总的来说，由通信的 KPI 和 AI 模型的 KPI 共同确定 AI 服务 (AlaaS) 的 KPI。

1) 对于 AI 模型推理服务，用户希望从网络获取实时的高准确度的 AI 推理结果。根据前面所描述的 AI 模型推理服务流程，用户先将本地数据或者部分 AI 模型的中间推理结果上传给网络，网络基于另一部分 AI 模型对数据或中间结果进行处理后，得到最终推理结果，反馈给终端。

如上图 3-4 所示，从以下 5 个方面具体讨论 AI 模型推理服务的关键性能指标的影响因素，包括：

- **准确度：**用户能够获得的 AI 服务准确度既取决于 AI 模型推理的准确度，还取决于通信的可靠度，包括数据 / 中间推理结果和最终推理结果反馈的可靠度。AI 模型推理的准确度通常随着模型参数的增加而升高，而模型参数的增加也意味着对计算需要的增加。

- **时延：**用户能够获得的 AI 服务响应时延取决于数据传输速率决定的传输时延和用户与网络进行 AI 模型计算的推理时延。传输时延还取决于 AI 模型的分割点确定的数据量，而推理时延同时取决于用户与网络的模型和计算能力。

- **覆盖：**网络能够提供的 AI 服务覆盖取决于链路预

算确定的连接覆盖和模型的覆盖。其中，模型的覆盖是能够满足推理准确度要求的数据覆盖区域，即所部署的模型能够覆盖该区域内的数据分布。对于当前以过参数化为特征的深度神经网络来说，推理准确度通常受限于模型大小和训练时所使用的数据分布，而模型越大，达到相同性能时所能覆盖的数据也越多。

- **能效：**服务能效取决于推理过程中的通信能效和计算能效。通信和计算的能耗也取决于 AI 模型分割点确定的传输数据量和用户与网络的计算能力。

- **密度：**服务密度即能同时提供 AI 服务的数据密度，取决于区域传输容量和推理数据密度，前者决定了能够同时传输的数据量，后者决定了能够同时进行 AI 推理的数据量。

2) 对于 AI 模型训练服务，用户希望从网络获取高准确度的 AI 模型。AI 训练服务既可以采用集中训练模式，即网络首先收集所有用户的数据，然后基于网络的计算能力进行 AI 模型的训练，然后将训练好的模型发送给用户。AI 训练服务也可以采用分布式训练模式，如前面所描述的联邦学习。在分布式学习模式下，用户基于本地数据训练本地模型，将训练得到的本地模型上传到网络；网络对多个用户上传的模型进行融合，得到全局模型，并下发给用户；用户基于新的全局模型进行下一轮训练。用户通过与网络的多轮模型传输，实现用户之间数据中知识的共享。由于不需要用户上传原始数据，分布式学习可以更好地保护用户隐私。



图 3-5 分布式 AI 模型训练服务关键性能指标体系

如上图 3-5 所示，据此可以分析 AI 模型训练服务的关键性能指标的影响因素，具体包括：

- **时延**：用户能够获得的 AI 模型训练服务响应时延取决于训练模式、用户与网络的计算能力、数据量和数据传输速率等。如对于分布式模型训练，时延包括用户在本地进行模型训练的时间和与网络交互模型的时间；而对于集中式模型训练，时延包括网络从用户收集数据的时间和进行集中训练的时间。

- **覆盖**：网络能够提供的 AI 模型训练服务的覆盖取决于链路预算确定的连接覆盖和模型的覆盖。链路预算决定了用户是否能够及时与网络进行数据或模型的交互，模型覆盖决定了是否能够满足准确度的需求。

- **能效**：服务能效取决于训练过程中的通信能效和计算能效。训练过程涉及模型参数更新所需要的多次迭代计算所消耗的能量，以及用户与网络进行数据或多轮模型传输所消耗的能量。

- **密度**：服务密度即能同时提供 AI 模型训练服务的数据密度，主要取决于区域传输容量和训练数据密度，后者由模型大小和计算能力确定。

3.2.3 性能指标评估与计算方法

针对上述的 AI 模型训练与推理服务、AI 数据服务，以及定义的相应的关键性能指标，以模型推理和训练服务、AI 数据服务为例，这里分别给出相应的性能指标评估与计算方法。

3.2.3.1 AI 模型推理服务

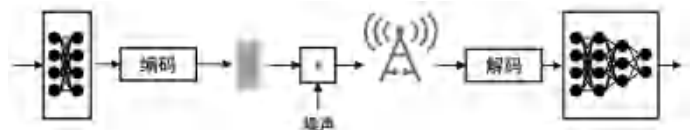


图 3-6 AI 模型推理服务处理示意

一个用户享受 AI 模型推理服务所涉及到的处理如图 3-6 所示，既涉及到用户和网络侧的模型和计算，也涉及到用户与网络间的传输。以分割推理为例，训练好的深度神经网络在某一层分割后，分别部署在用户和网络侧。1) 用户对待推理的数据使用分割的用户侧 AI 模型进行前向计算，得到中间推理结果；2) 用户将中间推理结果编码后发送给网络；3) 网络解码后，使用分割的网络侧 AI 模型对中间推理结果进行前向计算后，得到最终推理结果；4) 网络将最终推理结果反馈给用户。根据分割位置的不同，用户也可能直接把数据上传到网络，由网络进行模型推理。AI 模型推理服务的关键性能指标具体评估方法如下。

▪ 服务准确度

服务准确度由 AI 模型推理的准确度和传输的可靠度决定。记模型推理准确度为 a_m ，链路可靠度为 r_c ，则服务准确度为 $a = a_m \cdot r_c$ 。

模型推理准确度由模型和数据共同决定，对于充分

训练的模型，模型越大，即参数量越多，准确度越高。模型越大也带来更大的计算开销和时延。链路可靠度为在给定时间内成功传输数据的概率。

▪ 服务响应时延

服务响应时延包含 AI 模型前向计算的时延 T_{comp} 和通信时延 T_{comm} ，即：

$$T = T_{comp} + T_{comm} = T_{comp}^{NET} + T_{comp}^{UE} + T_{comm}$$

其中，计算时延包含网络侧模型计算时延 T_{comp}^{NET} 和用户侧模型计算时延 T_{comp}^{UE} ，分别由网络和用户侧的模型运算量和计算能力确定：

$$T_{comp}^{NET} = \frac{O_{comp}^{NET}}{F_{comp}^{NET}}, \quad T_{comp}^{UE} = \frac{O_{comp}^{UE}}{F_{comp}^{UE}}$$

其中， O_{comp}^{NET} 和 O_{comp}^{UE} 分别是网络和用户侧 AI 模型前向计算的浮点运算次数（FLOPs）； F_{comp}^{NET} 和 F_{comp}^{UE} 分别是网络和用户侧计算芯片的浮点运算速度（FLOPs）。通信时延可以计算如下：

$$T_{comm} = \frac{D}{R}$$

其中， D 和 R 分别是传输的数据大小（bit）和链路的数据传输速率（bit/s）。

模型分割的位置会影响用户和网络侧的模型大小，以及需要传输的数据大小，进而影响用户和网络侧的计算时延和通信时延。模型分割需要综合考虑用户和网络侧的计算能力以及链路的通信速率，选择最佳的模型分割点，以降低服务响应时延。如果 AI 模型只部署在用户侧，则不需要与网络通信，即 $D=0$ ；如果 AI 模型只部署在网络侧，则用户不需要参与计算，即 $O_{comp}^{UE}=0$ 。增大用户和网络侧的计算能力也可以直接降低响应时延，但是会带来能耗的增加。

▪ 服务能效

服务能效是单位能量消耗下所处理的比特数（bit/J），计算如下：

$$EE = \frac{D0}{E} = \frac{D0}{E_{comp}^{NET} + E_{comp}^{UE} + E_{comm}}$$

其中， $D0$ 为 AI 模型原始输入数据的大小（bit）， E 为系统总能耗（J），包含网络侧计算能耗 E_{comp}^{NET} 、用户侧计算能耗 E_{comp}^{UE} 和网络与用户之间的通信能耗 E_{comm} 。计算表达式分别为

$$E_{comp}^{NET} = \frac{O_{comp}^{NET}}{EE_{comp}^{NET}}, \quad E_{comp}^{UE} = \frac{O_{comp}^{UE}}{EE_{comp}^{UE}}, \quad E_{comm} = \frac{D}{EE_{comm}}$$

其中，网络和用户侧的计算能效 EE_{comp}^{NET} 和 EE_{comp}^{UE} （FLOP/J）由其采用的芯片决定，可以通过硬件实验测试或者芯片的浮点运算能力和功率估计得到。如计算芯片的浮点运算能力为 F ，功率为 P_{comp} ，此时计算能效可以表示为 $EE_{comp} = \frac{F}{P_{comp}} (\frac{FLOP/s}{J/s} = FLOP/J)$ 。通信能效由网络的通信传输速率和通信总功率 P_{comm} 确定：

$$EE_{comm} = \frac{R}{P_{comm}} (\frac{bit/s}{J/s} = bit/J)。$$

根据得到的系统能耗，我们可以进一步根据不同地区的能量转换率，计算等效的二氧化碳当量，即

$$CO_2eq = E \times ECR$$

其中， ECR 是每千瓦时能量消耗等效的二氧化碳排放量（kg CO_2 /kWh），与各地区的发电方式相关。

▪ 服务覆盖

服务覆盖定义为在给定的链路预算下，满足服务准确度的区域面积占比，即 $\frac{Area(R \geq R0, a \geq a0)}{Total Area}$ 。对于充分训练的模型，其准确度即是在训练数据分布下的准确度。服务覆盖需要通过系统仿真评估。对仿真区域内的位置，首先通过链路预算确定连接可以覆盖的区域，然后在该区域内收集数据，进行模型训练。一方面，对于给定的模型，调整数据收集的区域，直到训练模型的准确度满足要求，该区域即是模型所能覆盖的区域。另一方面，针对区域内的数据，调整模型大小并进行训练，直到满足准确度要求，该模型即是能覆盖该区域的模型。

由于覆盖能力由模型大小和数据决定，通常较小的模型只能覆盖较小的区域，但是可以通过分布部署多个小模型，以实现更大范围的覆盖。模型的大小和部署密度需要相应的计算能力和密度支持。因此，一方面需要根据服务覆盖的要求，配置相应的模型和计算资源；另一方面，针对多个模型分布式覆盖的场景，可以根据覆盖范围动态选择最佳的服务区域。

▪ 服务密度

服务密度为单位区域内的服务速率，由区域传输容量和推理数据密度决定：

$$C_{\text{area}} = \min \{ C_{\text{comm}}^{\text{area}}, C_{\text{comp}}^{\text{area}} \}$$

其中，区域传输容量可以计算为

$$C_{\text{comm}}^{\text{area}} = \rho_{\text{comm}} \times W \times SE_{\text{avg}}$$

其中， ρ_{comm} 是传输点密度 ($1/\text{km}^2$)， W 是系统带宽 (Hz)， SE_{avg} 是平均频谱效率 (bps/Hz)。区域推理密度可以计算为

$$C_{\text{comp}}^{\text{area}} = \rho_{\text{comp}} \times \frac{F_{\text{comp}}}{O_{\text{comp}}} \times D_0$$

其中， ρ_{comp} 是模型 / 计算点密度 ($1/\text{km}^2$)， $\frac{F_{\text{comp}}}{O_{\text{comp}}}$ 是推理服务的平均吞吐， D_0 是处理的数据量。

3.2.3.2 AI 模型训练服务

网络提供的 AI 模型训练服务包括如图 3-7 所示的分布式模型训练和集中式模型训练两种方式。在集中式模型训练中，网络首先从用户收集数据，然后在网络侧进

行模型训练。在分布式模型训练中，1) 用户基于本地数据，进行有限次迭代的模型训练；2) 用户将模型参数编码后发送给网络；3) 网络对多个用户的模型参数进行聚合，如平均操作，得到全局模型；4) 网络将全局模型发送给用户；5) 用户将本地模型更新为全局模型。重复上述步骤直到模型收敛或达到要求的准确度或迭代次数。上述步骤涉及到多轮的模型训练和传输，影响用户 AI 模型训练服务的关键性能指标，可以分别如下评估。

▪ 服务响应时延

由于集中式训练和分布式训练的流程不同，相应的服务响应时延也不同。

▷ 集中式训练

集中式训练的时延包括收集数据所需要的传输时间和网络进行 AI 模型训练所需要的计算时间，即 $T = T_{\text{comm}}^{\text{cent}} + T_{\text{comp}}^{\text{cent}}$ 。假设每个用户能够享受到的数据传输速率为 R ，则传输时延为 $T_{\text{comm}}^{\text{cent}} = \frac{D}{R} \times N_1$ ，其中 D 为每个样本数据的大小 (bit)， N_1 为用户样本数据的个数。网络进行训练的时延可以计算为

$$T_{\text{comp}}^{\text{cent}} = \frac{O_{\text{comp}}^1}{F_{\text{NET}}^{\text{comp}}} \times N_{\text{sample}} \times M_{\text{epoch}}$$

其中， O_{comp}^1 为一个样本前向和反向计算一次所需要的计算量， N_{sample} 为所有用户样本数据的总数， M_{epoch} 为模型训练迭代相关的次数，即所有训练样本使用的次数。

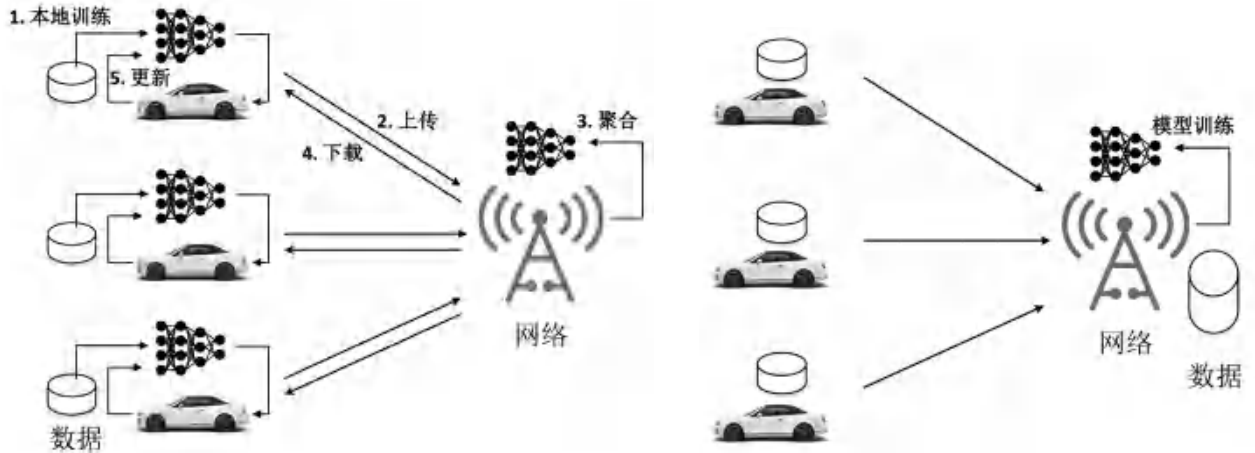


图 3-7 左：分布式模型训练 右：集中式模型训练

▷ 分布式训练

在分布式训练过程中，服务时延包括用户本地训练的时间和与网络传输模型的时间。当考虑基于联邦学习的模型训练时，由于需要迭代更新，性能指标的计算涉及到用户本地进行模型训练迭代的次数和模型融合的次数。由于模型融合和更新的复杂度远低于模型训练，这里忽略网络进行模型融合和用户进行模型更新的时延。服务响应时延计算为

$$T = (T_{comm}^{dist} + T_{comp}^{dist}) \times M_g$$

其中， T_{comm}^{dist} 为用户上传和下载模型的时间， $T_{comm}^{dist} = \frac{D_m}{R_d} + \frac{D_m}{R_u}$ ， D_m 为模型的大小（bit），即模型训练更新的参数数量乘以每个参数的量化比特数， R_d 和 R_u 分别为下行和上行的传输速率； T_{comp}^{dist} 为用户本地训练的时间，即

$$T_{comp}^{dist} = \frac{O_{comp}^1}{F_{UE_{comp}}} \times N1 \times M_{epoch}^1$$

其中， M_{epoch}^1 为用户本地训练迭代相关的次数，即本地数据训练使用的次数。分布式可以充分利用各用户的计算能力，通过本地的多次迭代，降低通信轮数等，从而降低训练时延，加速 AI 训练。

▪ 服务能效

与服务响应时延类似，这里针对集中式和分布式分别描述服务能效的评估方法。

▷ 集中式训练

集中式训练的能耗包括收集数据所需要的通信能耗和网络进行 AI 模型训练所需要的计算能耗。服务能效计算为

$$EE^{cent} = \frac{D0 \times N_{sample}}{E_{comm}^{cent} + E_{comp}^{cent}}$$

其中，通信能耗 E_{comm}^{cent} 为 $E_{comm}^{cent} = \frac{D0 \times N_{sample}}{EE_{comm}}$

计算能耗 E_{comp}^{cent} 为 $E_{comp}^{cent} = \frac{O_{comp}^1}{EE_{NET_{comp}}} \times N_{sample} \times M_{epoch}$

▷ 分布式训练

分布式训练的能耗包括用户本地训练的计算能耗和与网络传输模型的通信能耗。忽略网络进行模型融合和用户进行模型更新的能耗，服务能效计算为

$$EE^{dist} = \frac{D0 \times N_{sample}}{(E_{comm}^{dist} + E_{comp}^{dist}) \times M_g}$$

其中， E_{comm}^{dist} 为一轮模型上传和下载的能耗，与一轮中参与的用户数 K 有关，即

$$E_{comm}^{dist} = \frac{D_m}{EE_{comm}} \times K$$

用户本地训练的能耗为

$$E_{comp}^{dist} = \frac{O_{comp}^1}{EE_{UE_{comp}}} \times N1 \times M_{epoch}^1 \times K$$

分布式模型训练中的参数，如本地训练迭代的次数 M_{epoch}^1 、一轮中参与的用户数和模型量化精度及模型的大小 D_m 均会影响模型训练的收敛速度、性能和能效，需要根据系统的资源和各项性能指标进行优化配置。

▪ 服务覆盖

服务覆盖定义为在给定链路预算下，满足服务准确度的区域面积占比。对于 AI 模型训练服务，需要收集覆盖区域内足够的数据或者模型。对于集中式或分布式训练模式，在给定的训练服务时间内，评估满足准确度要求的区域。

▪ 服务密度

服务密度为单位区域内的服务速率，由区域传输容量和训练数据密度决定。对于集中式训练模式，服务密度主要取决于数据收集的速率和网络侧训练服务器的密度，即

$$C^{arca} = \min \{C_{comm}^{arca}, C_{comp}^{arca}\}$$

区域训练密度可以计算为

$$C_{comp}^{arca} = \rho_{comp} \times \frac{F_{comp}}{O_{comp}} \times \frac{D0}{M_{epoch}}$$

其中， $\frac{F_{comp}}{O_{comp}}$ 是训练服务的平均吞吐。

对于分布式训练模式，服务密度主要取决于模型交互的速率，即

$$C^{area} = \rho_{comm} \times W \times SE_{avg}$$

3.2.3.3 AI 数据服务

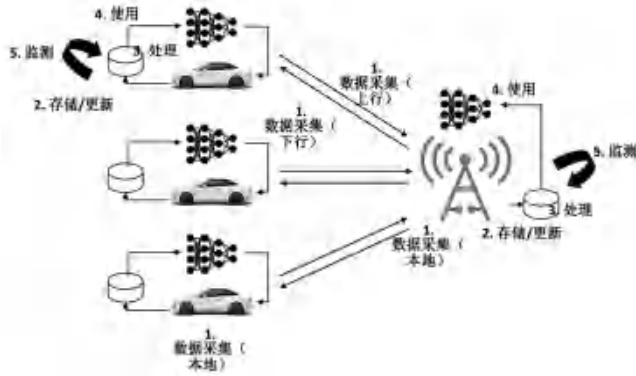


图 3-8 AI 数据服务流程

以通信系统无线空口侧的 AI 数据服务为例，所涉及到的流程如图 3.8 所示，既涉及到用户和网络侧的数据采集、数据存储、数据处理、数据监测等，也涉及到用户与网络间的数据传输。具体来说：1）数据采集，包括上行数据采集、下行数据采集、端侧和网络侧的本地数据采集，是获取数据的主要过程；2）数据存储和更新，主要在端侧和网络侧把采集的数据进行存储，需要综合考虑存储能力和开销；3）数据处理，主要是对数据进行筛选、清洗、分类、标签等；4）数据使用，把数据用于 AI 模型训练、推理和更新等；5）数据监测，通过对数据的分析、特征提取、评测，结合 AI 模型的推理结果，实时监测和更新数据。

但是用于通信系统中的信道信息数据、调度数据、网规网优数据、反馈信息数据等普遍存在数据维度不统一、数据概率分布不可得、数据标签不易分类等特点，导致较难有公认统一的方法对 6G AlaaS 提供的 AI 数据服务进行评估。为了可以直观有效地评价 AI 数据服务的性能指标，需要使用数据扩展、数据特征提取等对数据服务进行处理，然后基于处理后的数据进行性能指标的映射和计算，最终开展 AI 数据服务的性能指标评估。

AI 数据服务的关键性能指标具体评估方法如下。

▪ 服务完整度

服务完整度定义为网络向用户提供的 AI 数据服务中的无损数据所占比例：

$$r_{\text{complet}} = \frac{N_{\text{Lossless}}}{N_{\text{Total}}},$$

其中， N_{Total} 和 N_{Lossless} 分别表示数据样本总数和无损数据样本数。对于 AI 模型训练，有大量的缺失或错误信息的数据样本会显著影响模型收敛的性能，因此需要尽可能地提升训练数据集中高质量数据的占比。

▪ 服务冗余度

服务冗余度是对于网络向用户提供的 AI 数据服务中的冗余数据量的刻画，一般有两种定义方式：一是以冗余信息量的方式定义，二是以冗余信息占比的方式定义。

对于冗余信息量的定义方式，以平均互信息作为度量指标为例，指标单位为 bit，指标值越小说明冗余信息量越少。其定义如下：

$$R_{\text{redund}} = \frac{1}{N_{\text{Total}}^2} \sum_{i,j \in [1, N_{\text{total}}]} I(x_i, x_j),$$

其中， $I(x_i, x_j)$ 表示数据样本 x_i 和 x_j 的互信息，其计算方式为：

$$I(x, y) = \iint P(x, y) \log \frac{P(x, y)}{P(x)P(y)},$$

其中， $P(x)$ 和 $P(y)$ 分别表示 x 和 y 的概率密度函数， $P(x, y)$ 表示 x 和 y 的联合概率密度函数。

对于冗余信息占比的定义方式，以余弦相似度作为度量指标为例，指标单位为 %。其定义为当存在任意两个样本的余弦相似度小于给定阈值时，该两个样本中存在一样本为冗余样本，具体如下：

$$r_{\text{redund}} = \frac{\sum_{i=1}^{N_{\text{total}}} 1(\forall_j \in [1, N_{\text{total}}] \cos \theta_{ij} \leq \varepsilon)}{N_{\text{Total}}},$$

其中， ε 表示给定阈值， $\cos \theta_{ij}$ 表示数据样本 x_i 和 x_j 的余弦相似度，其计算方式为：

$$\cos \theta_{ij} = \frac{\langle x_i, x_j \rangle}{|x_i| |x_j|}$$

3.2.4 性能指标评估示例

针对 AlaaS 的 AI 模型推理和训练服务，这里分别给出具体的性能评估示例。其中，通信参数配置如表 3-10 所示，网络 and 用户的计算芯片和能力不同，具体计算能力如表 3-11 所示。

载波频率 (GHz)	通信带宽 (MHz)	通信传输速率 (Gbps)	功率 (W)	通信能效 (Mb/J)
3.5	100	1.95	636	3.07

表 3-10 通信参数

节点算力配置	浮点运算速度 (TFLOPs)	功率 (W)	计算能效 (GFLOPs/W)
网络	19.5	400	48.75
用户	1.4	30	46.67

表 3-11 节点计算能力参数

3.2.4.1 AI 模型推理服务

以 VGG-16 模型为例，将模型分割为两部分，分别部署在用户和网络侧。VGG-16 模型参数如表 3-12 所示：

模型	输入	模型大小	模型运算量	服务准确度	数据集大小
VGG-16	224x224 pixels	138 M	31 GFLOPs	71.59%	4.80 Mb

表 3-12 VGG-16 模型参数

将 VGG-16 模型划分为 6 个模块，7 个分割点（从左到右依次为标记为 0,1, ……，6），各模块输出的数据维度如下图 3-9 所示。

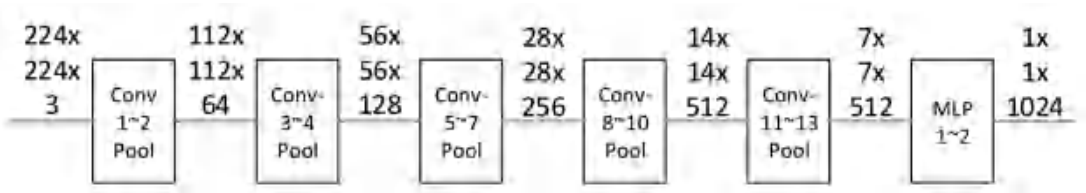


图 3-9 VGG-16 模型分割示意图

当选定分割点后，分割点左侧模块部署在用户侧，右侧模块部署在网络侧。例如，当选择分割点 0 时，此时整个 AI 模型全部部署在网络，用户直接将数据传输给网络推理。当模型部署在网络侧和用户侧时，根据分割方式不同，AI 服务的性能指标各有差异，如表 3-13 所示。

分割点	用户侧： 运算 (GFLOPs)/ 时延 (ms)/ 能耗 (J)	通信： 数据量 (Mb)/ 时延 (ms)/ 能耗 (J)	网络侧： 运算 (GFLOPs)/ 时延 (ms)/ 能耗 (J)	服务响应 时延 (ms)	系统能效 (Mb/J)	总能耗 (J)
0	0/0/0	4.8/2.46/1.57	31/1.59/0.64	4.05	2.18	2.20
1	3.9/2.79/0.08	24.0/12.31/7.83	27.1/1.39/0.56	16.48	0.57	8.47
2	9.4/6.71/0.20	12.0/6.15/3.91	21.6/1.11/0.44	13.98	1.05	4.56
3	18.7/13.36/0.40	6.4/3.28/2.08	12.3/0.63/0.25	17.27	1.75	2.74
4	27.9/19.93/0.60	4.0/2.05/1.30	3.1/0.16/0.06	22.14	2.44	1.96
5	30.7/21.93/0.66	0.8/0.41/0.26	0.3/0.02/0.01	22.35	5.19	0.92
6	31/22.14/0.66	0/0/0	0/0/0	22.14	7.23	0.66

表 3- 13 不同模型分割情况下的性能指标

3.2.4.2 AI 模型训练服务

以 VGG-16 模型为例，涉及到网络 and 终端联合分布式训练服务，用户的算力和通信相关参数见表 3-10 和表 3-11。VGG-16 模型进行分布式训练时的参数如表 3-14 所示，每次迭代所有用户都参与模型训练。分布式训练的次数为 200。相应的性能指标如表 3-15 所示。

VGG-16 模型	输入	模型大小	模型运算量	模型准确度	样本数	训练轮数	激活终端数 / 终端数
分布式	224x224 pixels	138 M	93 GFLOPs	71.59%	1.28	200	100/100

表 3- 14 VGG-16 模型分布式训练参数

VGG-16 模型	单样本单轮训 练时间 (ms)	单样本单轮训 练能耗 (J)	训练能效 (Mb/ J)	单轮训练能耗 (J)	总能耗 (J)	二氧化碳当量 (g)
分布式	66	1.99	2.41	2550857.14	538977336.3	105100.58

表 3- 15 VGG-16 模型分布式训练的关键性能指标

4

第四章 6G AI 即服务 (AlaaS) 的技术优势、产业价值与潜在挑战

4.1 技术优势

4.1.1 时延优势

时延优势包括两个方面，一是资源调度和服务数据传输时延的优势，二是 AI 服务部署时间的优势。

对于 6G 网络本身，相比于 MEC 和 Cloud AI，6G 网络可以更加灵活地选择提供 AI 服务的节点并分配通信资源减少传输时延；另一方面 AI 服务的部署通常是昂贵和耗费时间的，针对一个应用场景去训练和测试一套 AI 模型所花费的时间可能是以周来计算，通过 6G 网络提供 AlaaS，预训练的相似模型可以使用户快速的使用 AI 服务，缩短 AI 服务部署时间。

4.1.2 算力调度优势

AI 在设计和优化 6G 网络架构、协议和运营方面都发挥关键作用。超高可靠性是 6G 网络关键性能指标，而传统 MEC AI 或 Cloud AI 模型的性能极大地依赖于海量预定义的数据集和大量的算力资源，从而给 6G 网络的性能保障带来极大挑战。与任何其他“即服务”解决方案一样，AlaaS 将 AI 能力作为一种服务赋能 6G 网络，AlaaS 组织开发、部署或使用数据分析模型的过程，而 6G 用户无需学习复杂的算法或技术。然后，6G 用户可以专注于训练和配置 AI 模型，从而追求他们的核心能力，而不必担心安装、维护和相关的管理，极大减少了实施 AI 应用程序时的算力资源配置与调度。此外，AlaaS 概念化、设置和维护底层硬件和软件基础设施，用户从资源硬件配置管理中被解放出来，因为提供者拥有开发和维护高效 AI 基础设施所需的专业知识，并且还能够部署昂贵的专用算力基础设施硬件（GPU 或 TPU）进行有效管理配置。

4.1.3 成本优势

AlaaS 提供第三方提供商托管的“即服务”包使 6G 网络中的用户可以使用以较低代价即可使用 AI 服务。6G 网络复杂多变，建立面向 6G 网络的高度智能能力需要大

量投资和专业知识。并且在最终部署前，开发和测试 AI 模型都需要大量的算力资源和时间。因此，在 6G 网络中引入 AlaaS，用户可以通过 API 和工具直接使用 AI 的功能，而无需编写任何复杂的代码，或进行任何复杂的训练，这极大的降低了开发和测试 AI 模型的人力成本、算力成本和时间成本。此外，AlaaS 具有高度的灵活性和适应性，并且采用按使用量付费模式，允许 6G 使用为其所使用的 AI 服务付费，而无需从整个 AlaaS 包中为不需要的 AI 功能付费。

4.2 产业价值

4.2.1 对网络自身的价值

从 5G 开始，人工智能在移动通信网络中逐渐得到了广泛的应用，比如从网管级别的网络配置优化到网元级别的资源调度优化，甚至空口物理层的智能化，此外终端侧的智能化应用也越来越多，但以“打补丁”和“外挂”的应用模式阻碍了 AI 应用效果的发挥。目前网络自治水平不高（自动驾驶网络等级约为 2.2 级），而 6G 网络需实现高水平自治，此外 6G 在网络性能之外还提出了更高的安全可信要求，这使得网络的管理面临着极大挑战。因此，需要内生的 AI 为网络自身提供高性能、高效的服务，服务于网络自身的高水平自治、内生安全可信。

AlaaS 可以服务网络自身，支持实现对运营商和用户意图的感知和实现，实现网络的自我设计、自我实施、自我优化、自我演进，最终实现网络的高水平自治。AlaaS 支撑网络运维更程度的自动化，可以进一步降低人力参与程度，提高资源利用率，实现降本增效。

此外未来网络将承载更多样化的业务，服务更多的应用场景，承载更多类型的数据，因此网络将面临大量新的、复杂的攻击方式。面向网络的 AlaaS 可以实现自主检测威胁、自主防御或协助防御等，可以更加快速地进行网络防护的响应，在降低人力成本的同时，减少由决策延迟带来的损失，为用户提供持续可信的服务。

对于网络自身而言，AlaaS 的服务对象主要是网络运营商，AI 服务可以提供给网络自身，实现高水平的网络自治和内生安全可信，并进一步提高网络性能并降低运维成本，有极大的应用价值。

4.2.2 2B 产业价值

从 5G 网络智能化实践中，外挂式 AI 存在诸多短板和性能不足，但由于 5G 架构、协议功能和流程已经定型，只能在现有架构方案上做增量迭代，效果有限。同时，面向 6G，智能的需求场景将更加广泛，除服务于网络自身的高水平自治、性能优化、为 2C 客户提供极致业务体验外，还将助力千行百业的数智化转型，实现“随时随地”智能化能力的按需供应，这将极大丰富智能应用场景。相比云服务供应商，6G 网络需提供实时性更高、性能更优的智能化能力服务，同时提供行业间的联邦智能，实现跨域的智慧融合和共享。

对于 2B 产业，AlaaS 的服务对象可以是各行业的企业、学校、研究机构等，利用 6G 网络的连结性更高效地提供 AI 服务，为各垂直行业提供降本增效的能力，并进一步促进其产业升级，有极大的应用价值。2B 客户可以从自身的生产数据中挖掘出有用的模型，帮助自身优化生产，方便复制当前的生产模式，或者通过分享模型获得收益。运营商构建平台，获得平台收益。例如，一个工厂在生产过程中产生了大量数据，用这些在平台上训练了一个生产模型或者平台根据这个模型对工厂的生产调度进行优化，生成一个新的模型。当企业在一个平台上将这个模型设置为私有服务，这些模型会被用在这个工厂本身或者被这个企业作为模板复用在新的工厂，方便扩大生产。当企业将这个模型设置为公有服务，就可以将这个模型开放给合作伙伴，用于同类型工厂的生产。在这个例子上，运营商需要运营一个平台（提供算力训练模型，提供模型存储空间，提供服务开放接口，作为模型中间商参与多方的合作），收取平台服务费。工厂提供数据，得到模型，并且分享卖模型的收益。同时也方便自己复制生产流程扩大生产。

AlaaS 服务 2B 产业的范围十分广阔，可以包括金融领域、政府领域、工业领域、医疗领域、教育领域、交通领域、互联网领域等。

以工业领域和教育领域为例：在工业界中，AlaaS 有着极高的产业价值。当前工业界存在大量人工进行工件检

测的场景，这一方面导致工业产品质量检测的时间和经济成本较高，另一方面存在由检测人员工作状态起伏导致检测准确率不稳定的情况。同时，由于传统检测设备精度的限制，高精度工件的一些微小缺陷并不能被发现 [18]。因此，在工业互联网中由无线网络提供 AlaaS，主要通过 AI 图像相关技术自动化地识别工业产品的质量，通过强化学习相关技术触发工业机器人的及时处理，可以在提高质检准确性的同时，极大地降低用工成本。另外，针对每一种工业检测场景，可能都需要机器学习算法的调优，这会存在较长的部署周期。因此，期望 6G 网络能凭借大规模分布式训练、整合、推理，能够快速提供针对性的 AI 服务。

在教育界中，AlaaS 同样也有着极高的产业价值。当前学校在教学过程中已经开始采用各种电子设备和网络，而 AlaaS 可能给教育模式带来变革。在越来越倡导个性化教学的当下，由老师一对一的交流、发现和制定学习方法会耗费大量的时间成本，给老师带来较重的负担。通过 6G 网络收集学生学习过程中由电子设备产生的信息，可以感知学习情境、识别学生特征，通过动态学习诊断、反馈与资源推荐的自适应学习机制，可以适应学生动态变化的学习需求 [19]，从而使教师能够清晰、快速地获取学生动态，根据学生的需求精准地调整教学方法，更好地支持个性化的教学。

4.2.3 2C 产业价值

随着智能终端硬件的不断发展，终端开始存储大量数据，终端的计算能力也越来越强，考虑到数据隐私需求，需要 6G 网络和终端的算力、通信连接和算法模型等资源，以及算力卸载、模型编排等功能，为 2C 客户提供极致业务体验和高价值新型业务。在 2C 场景下，用户和运营商同时提供训练数据，同时使用模型并且获得好处，用户的体验提升，运营商节约了资源，提升服务质量。数字运营企业收取数据收益，分享模型收益。

面向 2C 产业，AlaaS 可以扩展到智能手机、个人电脑、车载终端、智能穿戴设备等多种类终端，简化用户的操作，丰富用户的生活。6G 网络的连结性为 2C 的 AlaaS 带来了极大地提升，如网络的智能感知可以为 PC 端用户带来更流畅的电话视频体验，终端能力受限时提供视频和语音智能修补技术来提高视频通话的质量；在车联网场景下，6G 网络可以提供车载终端 AI 服务，如微软小冰可以

为用户打造智能语音助手，并且可以提供拥有感情的互动。AR/VR 的持续发展也是 6G AlaaS 的 2C 业务的一片蓝海，单个个体不同穿戴设备的协同感知，不同个体之间、个体和环境之间的交互都需要 AI 服务在网络中的实现，结合数字孪生技术，拥有构建虚拟世界，重塑人们生活的能力。

AlaaS 的 2C 产业涵盖了人们日常生活中的各类终端和应用场景，涉及到现代数字生活的方方面面，服务的对象也可以是每一个人，其产业前景和增益都是不可估量的。对于 2C，其分为两个方向，一方面是单个用户的 AI 服务，面向单个用户的 AI 服务涉及到终端和网络的联动，以进行模型的训练和使用。例如，一个自媒体工作者，拍了多段素材，AI 服务平台根据他以往的编辑风格（用户自己提供的数据）、观众反应（自媒体平台提供的数据）生成自动剪辑 AI 模型，缩短工作时长提升点击率。又例如，AI 服务平台根据用户终端的历史（运动轨迹，流量起伏），网络本身的运行状态（拥塞情况，时延，抖动）生成一个联合模型，在终端侧满足节能需求，在网络侧满足提升资源利用率的需求。

另一方面是面向多个类似用户的 AI 服务，例如演唱会，AI 服务平台会根据这些用户的位置，演唱会观众的特殊网络行为生成服务模型，网络据此进行带宽资源的调度，或者根据各个用户的需求编排带宽资源，甚至派出无人机基站满足用户需求。模型也可以复用到其他演唱会。当然，演唱会之类的场景下 AI 服务也可能会结合沉浸式媒体提供服务。又例如，智能交通场景，大量的汽车堵在一条路上，将涉及两个模型，一个是交通疏导指引生成模型，为这些汽车生成交通疏导指引，进行交通疏导。一个是在此过程中进行网络资源调度的模型。这些面向多个用户的应用场景需要与大量不同的数字运营企业协作。此外在通感一体的 6G 网络中，运营商通过无线侧感知物理世界，收集数据，也可以作为训练 AI 模型输入数据，帮助部分企业减少传感器相关基础设施的投入。

4.3 潜在挑战

1、投入产出评估体系

一种技术或架构能否被广泛应用，在于是否能提供

较好的投入产出比。因此 AlaaS 的重要挑战在于需要新的能被客户和 AI 服务提供者接受的评估体系。其中的参数例如收益比，即为训练和维护、运行算法需要消耗的功耗，占用的算力，投入的人力等支出与模型应用后所带来的收益之间的关系。但是某个 AI 模型往往只针对某种特定需求，因此建立评估体系可能较为复杂，或者需要各界较大的投入。

2、算法模型共享

算法模型共享的挑战来自技术和经济两个方面，技术上接口可能难以统一，可能重复以往某某 aaS 的困境。例如 IaaS 也没有作为通用服务，在各个 IaaS 平台上搭起的环境应用，难以移植到其他平台上。应用平台拥有自身的设计和发展技术路线导致应用平台本身架构之间存在差异，某个平台上表现优秀的模型，在另一平台上可能因资源调度或其他问题而变差。经济上模型使用的计费可能难以统一，各个公司的训练成本不一，计费难以统一，此外大公司训练成本会摊薄，初创公司等训练成本较高，造成赢者通吃，破坏算法模型提供生态。同时模型的准确性和应用场景密切相关。许多同类应用场景通常具有一定程度相异的影响因素，例如自动驾驶模型，北京的路面和伦敦的路面摩擦力不一致导致自动驾驶行为偏差，可能导致重复训练同一模型的需求。提高了模型训练总体成本。

3、指标评估体系与异构资源优化

当前，部分 QoAIS 指标尚无成熟的量化评估方式，因此在 AlaaS 平台建设初期可能无法在完善这些指标的量化评估标准。如可解释性，其由于一些算法本身的技术架构和特性令其可解释性较差，此外还有模型的泛化性、可重用性等，学界对此虽已有相关研究，但形成统一的各方认可的评估标准仍然需要较多的工作。因此，在起始阶段设计出足够开放包容的评估体系或架构，以便后续逐步引入上述指标，是当前我们面临的重要挑战。

在 5G AI 中，网络一般作为 Over-the-Top 厂商向用户提供 AI 服务的管道，网络只需要配置通信资源，来满足连接的性能需求。与此不同，6G 原生 AI 通过网络内的通信、计算、模型等资源，向用户提供 AI 服务。6G 网络不仅需要关注通信的性能指标，还需要考虑端到端的 AlaaS 性能指标，并通过配置通信、计算等多种类型的资源，为用户提升可控的 AI 服务。6G AlaaS 需要一套完整的性能指标和评估体系，并建立与多样资源的映射，为网络设计提供指导。

4、模型维护挑战

客户的需求和应用环境是不断变化的，正在使用的模型可能无法应对新的需求，因此，AlaaS 系统需要具备维护模型的能力，评估旧模型，生成新模型，更新模型到应用平台上，对系统的新要求导致更大的投入和技术挑战。

5、如何保障用户使用 AI 服务的安全性和隐私性？

网络在引入 AI 服务后，用户对 AI 服务安全性和隐私性上存在不同的需求选项，如何打破传统通信服务中 QoS 体系和安全体系分开独立设计的模式，使这种需求的差异性得到更好的满足？

6、网络如何提供可信的 AlaaS，以便于其应用推广？

当前 AI 的可信是 AI 在各行各业应用落地和推广的关键挑战，6G 网络智能普惠也需解决 AI 的可信问题，包括可解释性、鲁棒性、再现性等，还有隐私保护、安全性、透明性、审计与问责、伦理与法律、环保等，但是这些内涵是否都要纳入 6G AlaaS 的范畴？6G AlaaS 是否包括其他的可信内涵？AI 可信对 6G 网络设计会有哪些影响？

缩略语简表

英文缩写	英文全称	中文解释
AI	Artificial Intelligence	人工智能
AlaaS	AI as a service	AI 即服务
3GPP	3rd Generation Partnership Project	第三代合作伙伴计划
ETSI	European Telecommunications Standards Institute	欧洲电信标准协会
NWDAF	Network Data Analytics Function	网络数据分析功能
ML	Machine learning	机器学习
RRM	Radio Resource Management	无线资源管理
MEC	Mobile Edge Computing	移动边缘计算
CDN	Content Delivery Network	内容分发网络
gNB	gNodeB	5G 基站
UE	User Equipment	用户设备
MBR	Maximum Bit Rate	最大比特率
GBR	Guaranteed Bit Rate	保证比特速率
AR	Augmented Reality	增强现实
VR	Virtual Reality	虚拟现实
V2X	Vehicle-to-Everything	车联网
QoAIS	Quality of Artificial Intelligence Service	人工智能服务质量
NLP	Natural Language Processing	自然语言处理
CV	Computer Vision	计算机视觉
KPI	Key Performance Indicator	关键性能指标

参考文献

- [1]Sharma, Robin. "Artificial Intelligence in Agriculture: A Review." 2021 5th International Conference on Intelligent Computing and Control Systems (ICICCS). IEEE, 2021.
- [2]Bannerjee, Gouravmoy, et al. "Artificial intelligence in agriculture: A literature survey." International Journal of Scientific Research in Computer Science Applications and Management Studies 7.3 (2018): 1-6.
- [3]Badue, Claudine, et al. "Self-driving cars: A survey." Expert Systems with Applications 165 (2021): 113816.
- [4]Claussmann, Laurène, et al. "A study on ai-based approaches for high-level decision making in highway autonomous driving." 2017 IEEE International Conference on Systems, Man, and Cybernetics (SMC). IEEE, 2017.
- [5]Shaheen, Mohammed Yousef. "Applications of Artificial Intelligence (AI) in healthcare: A review." ScienceOpen Preprints (2021).
- [6]Väänänen, Antti, et al. "AI in healthcare: A narrative review." F1000Research 10.6 (2021): 6.
- [7]Cao, Longbing. "AI in finance: A review." Available at SSRN 3647625 (2020).
- [8]Königstorfer, Florian, and Stefan Thalmann. "Applications of Artificial Intelligence in commercial banks—A research agenda for behavioral finance." Journal of behavioral and experimental finance 27 (2020): 100352.
- [9]Yigitcanlar, Tan, et al. "Contributions and risks of artificial intelligence (AI) in building smarter cities: Insights from a systematic review of the literature." Energies 13.6 (2020): 1473.
- [10]Policarpo, Lucas Micol, et al. "Machine learning through the lens of e-commerce initiatives: An up-to-date systematic literature review." Computer Science Review 41 (2021): 100414.
- [11]Khoali, Mohamed, Abdelhak Tali, and Yassin Laaziz. "A Survey of Artificial Intelligence-Based E-Commerce Recommendation System." Emerging Trends in ICT for Sustainable Development. Springer, Cham, 2021. 99-108.
- [12] 华为技术有限公司, 《6G: 无线通信新征程》白皮书, <https://www.huawei.com/cn/technology-insights/future-technologies/6g-white-paper>
- [13]https://baike.baidu.com/item/%E5%9F%8E%E5%B8%82/33549?fromModule=lemma_search-box
- [14]<https://baike.baidu.com/item/%E5%86%9C%E4%B8%A3/330015>
- [15] 华为技术有限公司, 智能世界 2030, 2021.
- [16] 刘光毅, 邓娟, 郑青碧, 李刚, 孙欣, 黄宇红. 6G 智慧内生: 技术挑战、架构和关键特征 [J]. 移动通信, 2021, 45(04): 68-78.
- [17]Wen C K, Shih W T, Jin S. Deep learning for massive MIMO CSI feedback[J]. IEEE Wireless Communications Letters, 2018, 7(5).
- [18]<https://www.paddlepaddle.org.cn/support/news?action=detail&id=2104>, 2020-07-08.
- [19] 顾小清, 人工智能何以促进未来教育发展, 光明日报, 2022-07-27.

贡献单位

序号	主要贡献单位
1	华为技术有限公司
2	中国移动通信有限公司
3	中国电信股份有限公司
4	中国联合网络通信集团有限公司
5	中兴通讯股份有限公司
6	中信科移动通信技术股份有限公司
7	中国信息通信研究院
8	紫金山实验室
9	上海诺基亚贝尔股份有限公司
10	维沃移动通信有限公司
11	亚信科技（中国）有限公司

6G



微 信 公 众 号

联 系 方 式

邮箱: imt2030@caict.ac.cn

COPYRIGHT@2023 IMT-2030 (6G) PROMOTION GROUP

ALL RIGHTS RESERVED

系统推动中国6G研发与国际合作的重要平台，代表中国6G科技和产业最前沿进展。推进组聚合中国产学研用力量，广泛开展6G愿景需求、关键技术、标准研制、试验验证等工作，深入研究对经济社会的影响，积极推动国际交流与合作。