6G 网络原生 AI 技术需求白皮书

6GANA TG1

2022/1/20

摘要

新一代 6G 移动新系统，旨在支撑和实现未来 6G 时代的万物智连、万务智联和智能普惠，未来它将成为支撑和赋能各行各业更强大的智能基础设施和平台。伴随着各级消费者和各行各业数字化、智能化应用的发展和深入，业界对 6G 新网络原生/内生 AI 能力和服务的需求日臻强烈。本白皮书将基于未来 6G 新业态、新行业场景用例、新技术等方面需求和趋势，系统地分析提炼原生/内生 AI 设计的技术需求和影响（例如：在能力，服务，架构，标准化等方面），从而有力地支撑 6G 新网络原生/内生 AI 相关技术方案的设计实现和最终落地。

参与本白皮书撰写的单位，包括但不限于：中兴通讯，中信科移动，华为，中国移动，中国电信，中国联通，上海科技大学，重庆邮电大学，上海诺基亚贝尔，维沃，紫金山实验室，亚信，爱立信，大连海事大学，鹏城实验室。谨此衷心地感谢和致敬各个单位的贡献和输出。

目录

摘要 2

目录 2

1. 前言 5

2. AI 与 5G 通信系统的结合应用 5

2.1 AI For 5G 发展现状 6

2.2 5G For AI 发展现状 6

2.3 “5G 网络 AI”发展现状 7

2.4 AI 和 5G 移动系统结合应用面临的挑战 11

3. AI 与 6G 新系统内生融合机理 11

3.1 必要性分析 12

3.2 可行性分析 12

3.3 增益性分析 13

3.4 新特征新范式分析 14

4. AI 和 6G 新系统内生融合的技术需求 15

4.1 能力需求（功能，性能等方面） 15

4.1.1 算力能力需求 15

4.1.2 算法能力需求 16

4.1.3 数据能力需求 17

4.1.4 其它能力需求 18

4.2 服务需求（对内，对外等方面） 19

4.2.1 算力服务需求 19

4.2.2 算法服务需求 20

4.2.3 数据服务需求 21

4.2.4 其它服务需求 23

4.3 架构需求（产品，部署等方面）.....24
4.3.1 算力架构方面.....24
4.3.2 算法架构方面.....27
4.3.3 数据架构方面.....28
4.3.4 其它架构方面.....31
5. 技术需求和原则总结.....31
6. 参考文献.....32
7. 附录.....33
定义和缩写.....33

文档作者列表：

贡献者	单位
杨立，谢峰，康红辉，薛妍，王梦涵，牛娇红	中兴通讯股份有限公司
艾明，段小嫣，孙万飞，舒敏	中信科移动
彭程晖，刘哲，王君，王飞	华为技术有限公司
李刚，温子睿	中国移动
夏旭，于梦晗，王恒，齐文	中国电信
黄兵明，廖军	中国联通
杨旸，吴连涛，李凯，巩宸宇，马牧雷	上海科技大学
梁承超，唐伦，柴蓉、王国仲	重庆邮电大学
沈钢，叶晨晖，张凯宾，顾方方	上海诺基亚贝尔
袁雁南，孙布勒	维沃移动通信有限公司
李兰兰，尤建洁	紫金山实验室
欧阳晔，王达，柏杨，赵燕	广州亚信技术有限公司
苏苓，　郝丹丹	爱立信
杨婷婷，宁嘉鸿，崔正琦	大连海事大学/鹏城实验室

1.前言

原生/内生 AI 被业界认为是未来 6G 移动新系统的核心架构特征之一【1】。AI 科技和应用发展日新月异，其相关的科学理念、范式模式、算法模型、方法手段等，不仅要能更紧密、更深层次地内嵌融合到 6G 新系统的架构、网元和功能流程之中，它们还要能凭借未来 6G 更强大泛在的电信基础设施平台之优势，更大地彰显出 AI 效能增益和价值，全面助力实现未来 6G 时代“万物智连”、“万务智联”、“智能普惠”等美好愿景。本白皮书将先简要介绍原生/内生 AI 的相关背景和动机，再力求全面系统地分析阐述：6G 移动新系统在原生/内生 AI 的设计过程中，所涉及到的诸多技术需求（例如：功能、性能、服务、架构等方面），全面涵盖 AI 算力，AI 算法，AI 数据三大 AI 基本元素；进一步地，基于最新的研究业态进展，我们还将继续分析汇炼原生/内生 AI 对 6G 移动新系统架构和标准化等方面的综合影响和需求。

2.AI 与 5G 通信系统的结合应用

AI（特别是机器学习 ML 和深度学习 DL）与移动通信系统的结合应用，始于第五代移动通信系统 5GS（由 3GPP 标准化定义），但 5GS 系统在设计之初（我们称之为原生阶段），并没有充分地考虑 AI 业务应用和其相关能力服务，例如：如何基于 AI 新范式去改善优化某种通信类工作机制。虽然 5GS 移动系统已有着向云原生化 Cloud Native，和 IT 软件定义、服务虚拟化方向演进的趋势（例如：已支持服务化的核心网 SBA CN 和核心网云化部署等功能），同时 O-RAN、OpenRAN 等联盟组织，也在积极地推动无线网络的开放化、云化、虚拟化等工作，但受限于技术成熟度、安全性、系统运维复杂度等多方面的条件制约，5GS 无线接入侧 NG-RAN 子系统总体上仍然保留着过去传统“烟囱式 CT 化基站架构”和相对固化的 RAN 协议栈模型。伴随着 AI 功能和业务的逐步渗入且价值彰显，5GS 只能在既定的系统架构和协议栈体系中，通过在核心网侧引入新逻辑功能节点 NWDAF 和各式各样模块级“外挂叠加式”的 AI 功能，来进一步增强优化系统自身各方面的性能和对外服务能力。每个“外挂叠加式”的 AI 功能模块，基本都是针对已识别且特定的通信类问题（例如：切片质量保障、用户 QoE 优化、移动性预测、故障定位、网规网优运维等），主要旨在提升 5GS 系统性能、优化无线传输效率和简化网络管控运维等目的。

2.1 AI For 5G 发展现状

大量实践已证明：AI/ML/DL 等方法是针对传统无线通信中多维复杂且计算密集型问题求解的可行高效手段。当前，AI/ML/DL 已在 5GS 系统内的多个层面和多业务领域，进行了初步的应用尝试和价值验证，例如：通过智能化平台辅助简化 5GS 系统运维、优化服务场景识别、网络异常检测、故障根因分析和提升系统节能效果，通过智能化模块去优化网络的策略资源部署和参数设值的精准度，通过智能压缩 CSI 反馈信息资源开销和智能调制编码去提升空口的资源利用率，通过智能模式去识别预测网络流量分布、用户轨迹行为，以提升用户业务体验等【2】。上述这些实例应用都可被称为 AI for Network(AI4NET)，即体现了 AI 作为先进技术手段，对 5GS 系统某些方面赋能提质、优化增效的价值。在 5GS 系统中，各种 AI 算法模型，基本都是针对特定已识别通信问题的应用和改善，即已经过大量专门的线下训练和效果验证，这类似“头痛医头，脚痛医脚”，但它确实可一定程度地提升通信业务性能，降低系统运维成本。总体上，当前的 AI4NET 仍然缺乏系统性、周密性、全局性、可解释性方面的考虑，较大限制了 AI 能力可拓展性、迭代增强性和 AI 模型泛化应用能力等。此外，当前 5GS 系统中的各种 AI 资源和能力（涵盖 AI 算力，AI 算法和 AI 数据方面）并不具备开放和服务化的特征，大部分仅仅限于系统内被利用和应用。

2.2 5G For AI 发展现状

当前，AI 各种移动应用（例如：语音、图像、视频、数据的 AI 识别归类和处理等）主要还是通过终端本地化或集中式云 AI 的服务模式（例如：以亚马逊 AWS、微软 Azure、谷歌云、阿里云为代表的公有云服务体系）提供给终端用户，因此 5GS 系统更多地扮演着底层数据传输管道的角色。5GS 系统把 AI 模型训练和 AI 业务应用相关的数据流，都当成一般的用户业务数据进行传输；云 AI 模型训练所需的大量样本数据，也通常通过应用层数据的形式，在 5GS 系统中端到端的传输流转。另外一方面，3GPP Rel-16 定义了 NWDAF 以及其相应的交互接口，致力于在网络内部实现一些智能化应用以及对内对外的 AI 赋能，进而减轻对传统云 AI 的依赖。上述这些方面可被称为 Network for AI (NET4AI)，即体现了 5GS 系统对 AI 业务和服务的价值。面向 NET4AI 智能类应用，当前 3GPP TS22.261【3】已开始制定一些特定场景下相应的网络 KPI 连接性能要求。然而，在上述 NET4AI 移动应用例子中，上层云 AI 模式所有的操作对于 5GS 系统几乎都是透明的，并没能实现上层 AI 应用和下层网络管道之间的跨层深度融合或配合；同时，5GS 系统内的各种通信、感知和计算资源能力和数据，对上层云 AI 应用服务器或者 NWDAF 等智能网元的开放度和被利用度也不够，也没能做到和上层 AI 业务应用的最佳适配。从 5GS 网络整体看，其内的各个逻辑网元节点所涉及和拥有的算力，算法和数据资源，并没能被外部 AI 功能

实体充分地调度和利用，它们中大部分仅仅受限服务于传统通信业务目的，或时常处于低效闲置无用的状态，例如：基站中大量基带算力和无线数据。未来，如何更大程度且更优地去利用好6G 新网络中的各种算力，算法和数据资源和能力，赋予其更大的业务服务价值，进一步拓展移动运营商的盈利体系，这都需要 6G 移动新系统去原生地实现。

2.3 “5G 网络 AI”发展现状

5G 网络在设计之初，虽然没有原生地考虑 AI，但是 5G 网络在如何与 AI 相结合互利即“5G Network AI”方面，却在不断地探索和演进。在“5G Network AI”探索中，3GPP、ITU-T、ETSI 和 TMF 等标准化组织或者论坛都取得了一定的进展。

3GPP 在“5G Network AI”方面进行了初步的探索，其状况与进展如下：

3GPP Rel-16 在 5GS 网络架构中，引入了新的逻辑功能实体 NWDAF，如图 1 即网络数据分析功能（具体参见 TS 23.288【4】）。NWDAF 通过与 5GC 中其它功能实体(例如：AF、PCF、AMF、SMF 等)交互，提供多种网络数据分析服务，包括：接收核心网各个功能实体 NF 的网络数据分析请求，基于该请求采集相应的网络数据；利用 AI 算法对采集数据进行分析推理，得到网络分析结果信息；再将网络分析结果信息提供给请求的 NF 实体。进而，各个 NF 利用 NWDAF 提供的网络分析结果信息，对 5GS 网络和终端工作状态进行监控，并对通信业务进行闭环控制优化。截至 Rel-17，NWDAF 已支持对网络业务体验、网络性能、切片负载、NF 负载、终端移动性/通信/异常事件、服务质量(QoS)可持续性、用户数据拥塞情况等进行分析。

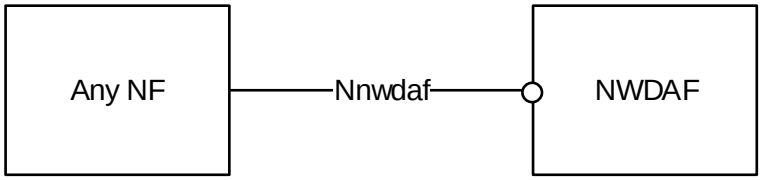

图 1: 基于 NWDAF 的 5G 网络数据分析架构（参考 3GPP TS 23.288）

Rel-17 对 5GS 网络数据分析架构和功能进行了增强，包括：NWDAF 的逻辑功能拆分及逻辑功能间的交互、多 NWDAF 实例如何协作数据训练和模型共享、引入新功能实体以提高数据收集效率、增强实时性。它主要包括以下特征：

-NWDAF 实体部署更加灵活，支持中心式、分布式、中心与分布结合式几种方式；

-支持多个 NWDAF 实体之间的协作(例如：分析聚合、分析转移、共享 AI 数据/模型等)；

-将 NWDAF 功能进一步分解为 MTLF(模型训练逻辑功能)与 AnLF(分析逻辑功能)，其中 MTLF 可向其它 NWDAF 实体提供 ML 模型；

-引入 DCCF(数据采集控制功能)、ADRF(分析数据存储功能)和相关数据采集优化流程；

-支持 NWDAF 实体从 UE 采集数据；

-支持针对边缘计算业务体验和网络性能的分析；

-支持对 UE/会话相关的切片负载、数据/信令的离散分布、WLAN 性能、用户面性能、会话拥塞控制、冗余传输等方面分析。

Rel-17 还对 5GS 系统管理面的智能化，进行了研究和相关标准化工作，如图 2 即引入管理数据分析(MDA: Management Data Analytics)，具体参见 TS 28.104【5】。MDA 采集与网络和服务事件及状态相关的数据，包括：网络性能测量、Trace/MDT/QoE 等报告、警报、配置数据、网络分析数据和 AF 服务体验数据等。MDA 基于特定 AI 算法进行相应的数据分析，生成分析结果报告，基于分析结果报告进行网络管理操作，从而实现网络管理运维的自动化与智能化。

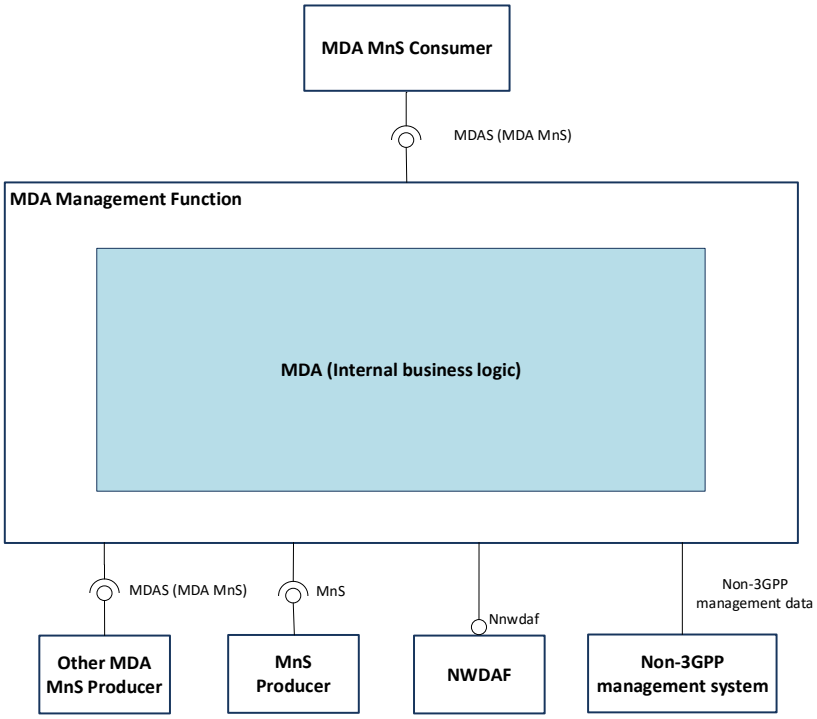

图 2: MDA 功能与服务框架（参考 3GPP TS 28.104）

Rel-17 对无线接入网 RAN 智能化的研究也正在展开，具体可参见 TR 37.817【6】，其潜在应用场景包括：网络节能、负载均衡、终端移动性管理等方面。

Rel-18 继续对 AI/ML 模型在 5GS 系统中传输的性能需求指标进行了研究，具体参见 TS 22.261【3】。5GS 系统将会依据 AI/ML 业务或应用需求，为相关 AI/ML 模型数据的传输提供 QoS 保障。5GS 系统还可对 AI/ML 模型数据的传输状态（例如：传输速率、时延、可靠性）进行监控，并上报给 AI 应用服务器，以供 AI 应用服务器据此监控信息调整 AI 应用层参数。

目前，3GPP SA2 正处于 Rel-18 新立项讨论阶段，网络智能化主要从下面两个角度考虑：

- **AI for Network 方面**：聚焦 5GC 网元相关分析、研究潜在的架构增强、新场景等。例如：研究是否及如何增强 5GC 架构以支持联邦学习和在线学习、UPF 业务数据上报 NWDAF 用于智能分析、UE 是否及如何使用来自 NWDAF 的分析建议、数据收集和数据存储增强、NWDAF 辅助的 URSP 等；

- **Network for AI 方面**：聚焦基于 SA1 Rel-18 AI/ML 模型数据传递性能要求，研究 5GS 辅助的 AI/ML 业务传输；支持 AI/ML 模型分发、传递、训练；用于不同 AI 应用视频/语音识别、机器人控制、汽车等方面。例如：研究支持应用层 AI/ML 可能的架构和功能扩展、研究可能的 QoS 及策略增强、研究 5GS 如何辅助 UE 客户端和 AS 之间的联邦学习等。

ITU-T SG13 也进行了 5G 网络与 AI 结合的相关探索工作。2017 年 11 月 SG13 成立了包括 5G 在内的未来网络机器学习(ML5G)焦点组，2020 年 7 月 ML5G 焦点组结束了第二阶段的工作，并向 SG13 提交了关于 AI 用例、架构框架、智能级别、数据处理、机器学习功能编排器、服务框架等的十项技术规范。此外，ML5G 焦点组还提出一套针对机器学习的管理子系统，针对机器学习全生命周期中各个阶段所需的不同功能，提出了跨多域、多云、不同层级的多层级 ML 工作流。

ETSI 在 ICT 系统与 AI 结合的领域开始相关探索的时间较早，2017 年 2 月即成立了体验式网络智能行业规范小组(Experiential Networked Intelligence Industry Specification Group，ENIISG)。ENIISG 定义了一种用于网络运维、业务编排、网络保障等应用，提供智能化服务的人工智能引擎，其功能架构【7】如下图 3 所示。

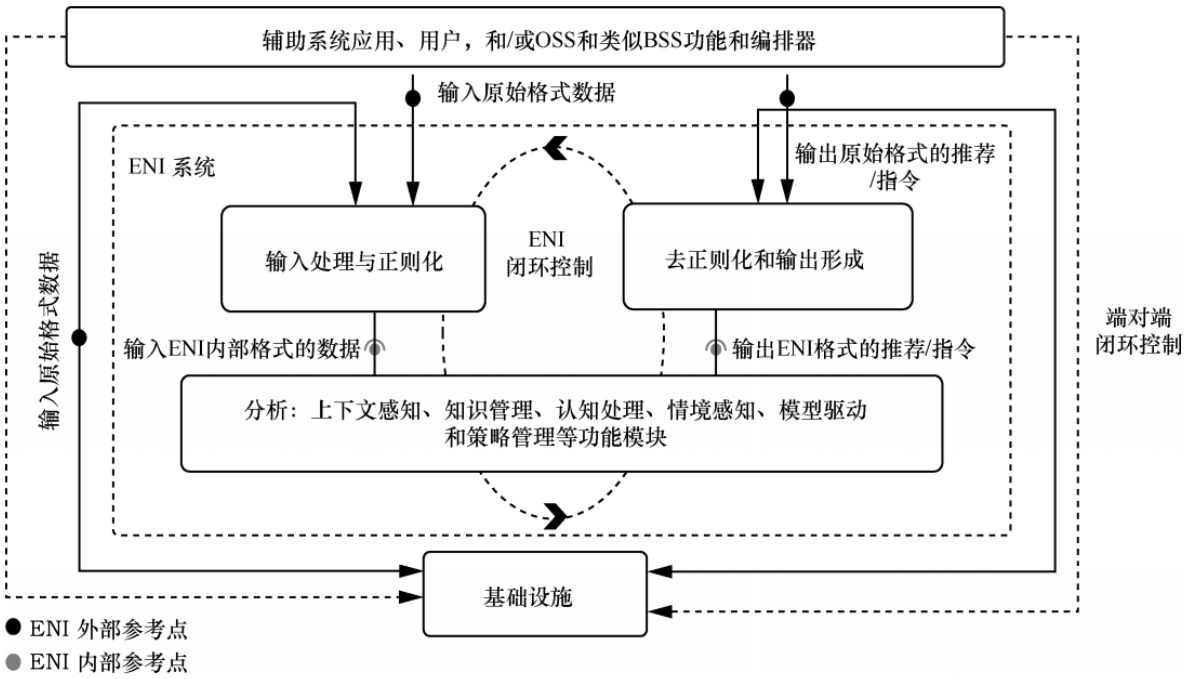

图 3：ENI 功能架构

ENI 系统目前包含：知识管理、模型管理、策略管理等模块，通过对数据进行处理，经过 AI 模块后，可以自动化地为网络提供服务运营和保障，以及提供切片管理和资源编排。目前，ENI 的功能还在不断地演进丰富，例如：支持基于意图驱动的网络等。

TMF 组织在 5G 网络与 AI 结合的相关工作中，主要聚焦于 OSS/BSS 网管方面的探索。当前 TMF 正在开展人工智能与数据分析(AI and Data Analytics，AI＆DA)项目，该项目主要从架构、用例、AI 术语、数据处理、AI 训练等方面进行研究，研究方向与具体内容【8】如下表 1 所示。

表 1：TMF AI&DA 项目主要工作

AI&DA 项目研究方向	具体内容
AI User Story& Use Cases	AI 在网络运维上的用户故事以及相应的用例
Data Model	数据的数字模型化标准，希望统一 AI 数据模型格式
Service Management for AI	基于 AI 管理电信网的整体功能架构，包括各接口和功能模块的描述
AI Data Training Repository	AI 对数据的需求、数据的来源（如 Log）、处理（如数据清洗）、使用等
AI Maturity Model	AI 相关概念澄清、AI 网络应用的分级成熟度等

2.4 AI 和 5G 移动系统结合应用面临的挑战

5GS 网络希望能更好地利用 AI 能力进行自我增强和支持 AI 类业务应用，尤其首先是在核心网侧，进而扩展到无线接入网 RAN 侧。Rel-16 引入的 NWDAF 功能，其主要目的就是为了提升 AI 数据采集和分析能力，例如：NWDAF 可为其他核心网功能 NF 和终端 UE 提供分析结果信息，辅助优化相应的网络业务发放。NWDAF 还支持从 5GS 网络运维网管系统中采集数据，为此 NWDAF 还提供了专门服务，用于相应网络功能的注册和元数据开放。尽管如此，5GS 系统与 AI 的结合应用还面临着如下挑战和缺陷：

• **数据源有限**：NWDAF 实体采集和分析的数据主要是 5GC 核心网功能接收的数据，但并没有充分考虑来自无线侧基础设施、环境、终端和各类传感器的广义数据，因此在 AI 数据样本方面存在不足。

• **传输带宽消耗**：无论集中式或分布式 NWDAF 部署，为了收集 AI 数据，都需要消耗大量的传输带宽资源。当数据源离 NWDAF 实体较远的时候，还会造成数据更新时延的问题。

• **缺少数据隐私保护**：NWDAF 偏集中式数据采集分析，数据源通常主要来自同一业务领域，因此架构设计中数据隐私的保护考虑不足，容易泄露用户隐私。

• **不支持对外 AI 服务**：NWDAF 是 5GC 核心网内部功能，主要用于增强优化 5GS 系统自身，外部 AI 应用不能直接在 5GC 核心网或无线接入网 RAN 的 AI 功能体系中被服务和受益。

• **基础设施利用不充分**：网络切片功能、超高可靠低时延通信 URLLC、 海量机器类通信 mMTC 等 5G 关键功能特性，在架构设计上都只是为了性能、功能和运营角度满足垂直行业需求，但都未专门考虑原生/内生 AI 的支持(如：数据治理和服务、分布式架构等)，网络基础设施中的各种资源（或低效或闲置），并没被业务应用充分地利用和价值转换。

• **数据治理和服务缺失**：AI 不只是涉及到数据采集训练和分析推理两个方面。为了提供 6G 原生/内生 AI 的支持，还需要专门对 AI 数据治理和服务的架构进行系统性设计，而这并不在 5GS 系统当初的考虑范围之内。数据服务的价值未来有待更大地彰显。

3. AI 与 6G 新系统内生融合机理

ODICT 业界普遍认为：原生/内生智能（Native AI）将成为未来 6G 移动新系统的核心特征之一【1】，因此 6G 新系统将从一开始的需求和架构设计阶段，全面充分地考虑如何和 AI 深度融合。较大区别于当下 5GS 系统通过 AI 功能叠加、补丁、外挂等方式的结合应用，原生/内生智

能的目标，将会对 6G 新系统设计带来诸多的深刻影响和挑战，因此本章节将先剖析阐述 6G 原生/内生智能的动机和理由（例如：内生 AI 为何能更好地适配未来 6G 新场景、新用例，如何带来新业态和新价值增益等方面）。

3.1 必要性分析

面向未来移动网络的演进，网络管理运维需要从降本增效的局部智能化运营，向端到端的高水平网络自智自治迈进。但是，现有网络 AI 用例的研发，普遍采用了打补丁，外挂，烟囱叠加等方式进行，缺乏统一的系统框架。大部分 AI 模型应用效果缺乏有效的验证和 QoS 质量保障手段。AI 模型训练学习和分析推理在时间上解耦，AI 模型效果验证只能在事后进行，对现网影响大，无法实现高水平的网络自智自治，无法实现 AI 模型效果预验证、在线评估和全自动的闭环快速优化。此外，AI 模型（再）训练需要大量的样本数据，而集中式采集数据困难，导致网络传输开销大，AI 模型迭代更新的周期较长，训练开销较大、收敛慢、AI 模型泛化性差等。因此 6G 新系统需要进一步改善 AI 模型的训练和应用性能，提升网络自智自治水平。

面向各个垂直行业用户，助力千行百业的数智化转型，探索新的智能化商业模式，提供 6G 新场景和新能力都是内生智能的重要驱动力。在保护垂直行业数据隐私、数据不出园区的前提下，6G 新系统需能提供分布式、区域性算力资源、平台和服务，实现“随时随地”智能化能力的按需灵活供应，实现“以数据为中心的计算”。这相比传统云 AI 服务供应商，可提供更高的实时性、隐私性、性能更优的智能化能力和服务；另外，6G 内生智能还可提供行业间的联邦智能，促进实现跨域、跨行业的智慧融合和数智共享 。

面向未来智能终端的演进，海量终端将产生更大量的数据，终端的计算和智能能力也越来越强，6G 内生智能需能协同好网络 AI 和终端 AI，为 ToC 用户提供极致的业务体验和更高价值的新型 DICT 业务服务 。此外，保障未来网络安全可信也是重要的研究课题，内生智能可促进实现网络内生安全可信，可自主检测和主动防御各种潜在的攻击和威胁等。

3.2 可行性分析

5GS 系统和 AI 的成功结合应用，已一定程度地证明了网络智能化的可行性，本章节将从 AI 三要素的角度，分析未来 6G 新系统原生/内生 AI 深度融合的可行性。

算力：由于时延、可靠性和数据安全隐私保护等方面的要求，在 5G 时代，计算能力和算力资源的下沉部署已成为趋势。到了 6G 时代，数据的连接和计算可能进一步融合，例如：出现通信连接计算融合的双基础设施，这为内生 AI 融合设计提供了计算服务相关的条件基础。

算法：虽然当前将各种数据集中到 AI 云，在 AI 云中进行集中式处理的方式有其优势，但在数据隐私、极致性能和计算能耗等方面，一直存在其难以解决的问题。如果未来 6G 新系统能做到将算法模型和智能能力融合到网络中，无论数据和任务在哪里、数据和任务的智能处理就能在那里，这将成为一种有效替代选项。因此，AI 算法模型将由于数据和计算资源能力的下沉，也将同时被下沉到网络边缘（例如：边缘节点，无线基站）来执行和维护。

数据：传统通信网络主要是数据传输的管道，除了网络自身的管理运营数据，一般自身是不主动产生和处理业务相关的数据。由于数据方面的限制，这可能是 5G 将 AI 的重点放在网络自身性能提升优化和管理运维自动化的原因。到了 6G 时代，由于各种感知技术、行业数字化、边缘计算等走向成熟，6G 新网络本身就是巨大的无线传感器网络，它将具备主动产生和处理海量异构数据的能力，这也就为 6G 原生/内生 AI 融合设计提供了数据服务相关的基础，例如：大量训练样本采集和预处理，数据资产按需流转等。

3.3 增益性分析

“6G 原生/内生智能”将极大强化 AI 相关资源（包含 AI 算力、AI 算法、AI 数据）和 6G 移动新系统之间的耦合，从过去的所谓“通算智三张薄皮”耦合进化升级为“通算智一张厚皮”，从而 6G 新系统内的各种资源的集成复用率、利用效率、综合性能等方面会被提升优化，进而实现更高的系统成本和业务应用之间的性价比，普通消费者也能享受到更泛在、更廉价的算智类业务服务。

相比于传统集中式的云 AI 服务器和边缘智能节点工作模式，原生/内生 AI 模式使得 AI 资源能够更广泛、均衡地、更灵活按需地分布部署在 6G 新系统泛在基础设施平台之中，AI 算智类操作将更贴近数据源、任务源和终端用户，且更能高效适配空口的动态状况（例如：用户环境和信道变化，网络拓扑和资源更新等），因此 6G 新系统（特别是基站）将更易面向用户的动态环境进行快速而精准的闭环优化、更实时地进行策略调整和趋势预测。

在以联邦学习为代表的分布式 AI 机器训练模式下，去集中化的内生 AI 模式更有利于用户数据的隐私保护，分摊数据和算力任务的压力，强化（子）网络本地安全自治。在 AI 相关数据（算法模型、训练样本、基本参数、特征参数等）的采集处理和传输流转方面，内生 AI 将可能依托于 6G 新系统专门的“数据面”和“智能面”等逻辑功能，进行更高效灵活且鲁棒的数据流转和共享，进而带来更低的 AI 数据传输延时，更少的传输资源消耗和系统能耗。

通过将 6G 原生/内生 AI 技术进一步标准化，还可促进 ODICT 异厂家之间的通算智类设备、功能模块和 AI 任务流程的对接协作，甚至带来未来新业态和新商业模式的重构。这还将进一步促

进更多的参与方去联合构建更广泛、安全可信的 AI 智能资源能力服务平台，从而实现 6G“泛在AI”和“智能普惠”的愿景。

3.4 新特征新范式分析

未来 6G 移动新系统要实现原生/内生 AI，将会面临相当大的技术挑战，具体如下：

1）**高度差异化的智能服务质量 QoAIS 需求**，一方面无线网络工作状态波动大，用户和业务需求动态性高，尤其是垂直行业用户的需求变化更大；另一方面当前移动系统尚缺少对 AI 服务质量 QoAIS 系统性的评估和保障体系。

2）**有限的通信和计算资源**，网络边缘节点设备上的算力和智能不足，存储能力弱，无线接入节点与边缘设备之间的传输带宽有限；当前缺少对 AI 服务相关的异构资源系统性管控和调配，比如：算力，数据，连接等方面。

3）**当前缺乏拟真的 AI 训练与验证环境**，一方面拟真环境的实现对同步数据的实时性和数据量要求极大；另外在拟真环境下，为了避免 AI 功能启用后对真实物理网络带来的潜在负面影响，如何实现 AI 模型应用效果的预验证和保障。

4）**面向垂直行业需求**，内生 AI 需能提供 AI 资源、功能、服务的按需灵活编排到端-边-网-云各种节点之中，提供端网紧密协作的分布式 AI 架构和 AI 服务质量评估和保障体系。此外，内生 AI 需使能行业间跨域的联邦学习、知识经验数据共享框架，促进与行业业务逻辑映射的数字孪生网络的融合。

5）**面向未来网络自治自智**，内生 AI 需能实现网络自治自智所需的自发现、自编排、自配置、自优化、效果自评估、多域闭环，AI 效果预验证、在线评估和闭环快速优化，通信-存储-计算多维异构资源融合，更充分高效地利用网络内各种资源，支持安全可信的网络自治自智方案。

6）**面向普通终端用户**，内生 AI 需能对终端数据和算力资源进行智能感知，利用海量终端的大量数据和算力资源，实现端网融合的分布式 AI 架构，并且需对用户数据进行脱敏、关联聚合，保障用户数据隐私安全。

从上述若干技术挑战可看出：6G 原生/内生智能不仅要能实现高水平的网络自治自智，也要能为众多的垂直行业用户和普通终端用户，提供更高质量有保障的 AI 服务。6G 移动新系统至少需具备如下新技术范式特征：

1）端到端 AI 服务质量的全生命周期编排和管控，构建 AI/ML 服务质量评估和保障体系，实现全生命周期编排 AI/ML 算法模型，算力，数据等资源。

2）基于内生 AI 的计算和通信深度融合，在网络内数据、算力、带宽资源和传输时延都受限的场景下，内生 AI 需考虑计算和通信资源的联合编排，6G 将能重构网络架构、协议编排和功能流程，全面适应 6G 空口和网络侧传输特性，优化内生 AI/ML 模型性能。

3）内生 AI 要能与数字孪生融合，内生 AI 可给数字孪生网络提供所需的 AI/ML 模型，实现数据增广和自生成，降低数字孪生网络对物理网络数据采集的需求，数字孪生网络可对 AI/ML 工作流或 AI 模型效果可进行预验证和优化，避免 AI/ML 功能后续引入后对现网的负面影响。

4. AI 和 6G 新系统内生融合的技术需求

4.1 能力需求（功能，性能等方面）

未来，6G 新系统将会是深度融合了“通”、“感”、“算”、“智”、“存”于一体的超级无线基础设施平台，它对内将同时具备“通信”、“感知”、“计算”、“智能”和“存储”等方面更强大的能力，例如：更多的特征功能和更好的 KPI 性能；同时它对外也将能更强地提供“通信”、“感知”、“计算”、“智能”和“存储”等方面的综合业务、服务和应用，例如：面向普通 ToC 消费者，ToB 类用户或第三方客户。因此 6G 新系统既是超级的无线通信传输管道，又会是巨大的分布式雷达传感网络、泛在算力和 AI 智能服务器阵列等。6G 内生智能和上述 6G 系统特征息息相关，下面我们首先将从“算力”，“算法”和“数据”三大主要方面阐述 6G 内生 AI 在基本能力方面的需求。

4.1.1 算力能力需求

算力作为 6G 智能原生平台的新生产力，是支撑一切数据能力和服务的坚实基础。现阶段 5G、大数据、AI 等技术的高速发展，不断推动着数据的爆炸式增长和 AI 算法的复杂程度不断提高，从而带来了对算力规模、算力能力等方面需求的快速提升。对于融合了计算、通信、感知和 AI 能力的新一代 6G 算网融合系统，算力的发现、感知、度量、按需调度利用和最大化开放成为未来移动网络发展的重要趋势之一。

算力资源发现能力：6G 新系统涉及云计算、雾计算、边缘计算等多层次技术之间的相互协作。这些计算技术的能力特点和覆盖范围不同，例如：可以分别针对区域级别、本地级别和设备级别的物联网应用和服务。6G 新网络需具备及时的多层次异构算力发现和注册能力，以支持实

时的数据感知处理、控制执行等基本功能。算力发现注册能力作为 6G 新系统的基础功能，起到桥梁的作用，是算力服务的基础。

算力需求感知能力：随着越来越多数据的产生以及更强大复杂的算力算法运用，6G 新系统的算力感知能力需要变得越来越智能。典型的算力服务需求也从简单的数据感知、收集和表示转向算力服务需求信息的提取和分析。未来，6G 新系统可广泛地应用于环境监测、城市管理、医疗健康等场景。6G 新系统带来的触觉网络将助力对算力需求的实时感知，实现高效的数据处理、信息提取和分析决策。

算力统一度量能力：6G 新系统涉及的算力主要包括：逻辑运算能力（一种通用的基础运算能力，硬件芯片代表是 CPU）、并行计算能力（一种专门为了加速处理图形图像等数据类型统一的高效计算能力，硬件芯片代表是 GPU）和神经网络计算能力（一种用来针对机器学习、神经网络等进行加速的计算能力，硬件芯片代表是 NPU）三大类【9】。对于提供异构算力的各种设备和平台而言（不同厂商芯片，不同计算类型，以及不同用户的计算算力需求），如何对异构算力资源进行统一量化，这是算力调度和使用的基础。这就需要 6G 新系统能提供将各自异构算力资源映射到统一量纲的度量函数或方法。

算力按需调度能力：6G 原生算网融合系统结合了边缘计算、雾计算、网络云化以及智能控制的优势，通过强大的网络连接，实现了更广泛的算力资源的管控和动态按需调度。区别于传统云计算资源纳管采用集中式的资源管理或者集约化的资源提供，在 6G 原生算网融合的资源管理体系中，更多考虑了网络状态（如：空口信息、网络延时、网络损耗等）对于分布式算力资源调度方面的影响，以达到综合性能和资源最优化。因此 6G 新网络的核心能力是提高泛在分布式计算效率，算网融合正是为了提高智能计算、通信和 AI 服务的工作效率。

算力开放交易能力：6G 新系统算网融合下的原生算力，将不再单纯地由移动运营商或者云服务提供商所提供。只要能够贡献闲散算力的节点，比如：基站、手机终端、电脑、游戏机、企业空闲的小型数据中心等，都可成为新模式的算力资源。考虑到安全、经济、高效、可靠等因素，6G 新系统需要定义合理的机制，例如：区块链技术，对泛在算力资源进行管理和开放交易【10】。

4.1.2 算法能力需求

作为 AI 三要素之一的算法（模型）是 6G 新网络实现内生智能和提供各类 AI 服务的决定性因素之一，6G 内生 AI 算法能力需求至少包括如下几个方面：

AI 算法指标需求：目前 AI 算法和模型门类众多，且迭代发展极快，不同 AI 算法模型可以实现的功能性能，所需要的执行时间、运算复杂度等方面差异性较大。从 6G 新网络内生智能和对外 AI 服务质量的角度，6G 需要规范化定义出 AI 算法的指标需求和所包含的指标内容，并面向不同 AI 服务给出各个指标的量化范围。例如：可从 AI 算法模型的增益、运算复杂度、网络种类、执行时间、输入输出需求、数据依赖性和泛化能力等角度，去综合刻画算法能力需求。不易量化的 AI 算法指标可以分级描述。

AI 模型训练需求：通常 AI 模型需要基于大量数据，进行较长时间多次的训练才能收敛。在 6G 内生 智能系统中，特定 AI 模型训练的具体模式，需要根据系统数据收集量、计算资源、隐私保障等客观需求，和系统当下的通信计算等资源状态来进行实时的决策，潜在的训练模式可能包括：离线训练、在线训练、联邦训练等。AI 模型训练要能最优地匹配 6G 系统的实时资源状态。

AI 模型描述与交互需求：6G 新系统中将会有非常多种类的 AI 模型，服务于不同的 AI 用例和业务，它们需要能在不同网络架构的各个网元节点之间进行交互和共享利用。对此，6G 需要定义出 AI 算法模型的描述语言（或方法），并通过统一定义的流程，保障任何 AI 算法模型的可交互与可执行。

AI 算法演进需求：无线通信和 AI 算法是两条飞速发展的快车道，因此 AI 算法模型的架构和能力均需具备前向可演进的能力。AI 算法演进可以是参数层面的更新演进（性能自优化），也可以是模型架构结构层面的革命演进（架构自生长）。

4.1.3 数据能力需求

6G 新网络内生智能的实现主要是基于数据驱动，因此具备完善且高效的数据能力至关重要。从 AI 数据需求的角度出发，6G 至少需要具备以下几个关键能力：

数据收集能力：未来网络中的数据资源非常丰富，包括：基础设施的资源信息，业务支持系统中的客户、伙伴等信息，行业通信系统中的各行各业相关信息，以及终端用户和环境信息等。要想更高效安全的收集这些数据，需要相应的网络数据收集功能去支持。网络数据收集的过程包括：与各种数据源建立安全连接，确定收集范围和方式，把收集到的数据（经过预处理）存储在数据库中，并对数据库进行相关操作维护等。

数据分析能力：面对丰富的数据资源，使用数据挖掘、机器学习等方法对收集来的大量数据进行分析提炼，以求最大化地开发数据的功能，挖掘出关键的信息和知识，给客户提供所需要的数据服务。例如：根据历史数据统计特征实现对网络故障和业务损伤的自主检测，识别网络异常等，并能通过用户数据去预测未来事件，分析用户行为和偏好，以更好地为用户服务。数据分析

和挖掘的算法多种多样，且不同算法面对不同的数据类型也会呈现出不同特点，6G 新网络内生 AI 需要能根据不同的业务数据类型，来选择最合适的分析和挖掘算法。

数据隐私安全保护能力：收集和存储敏感类数据涉及到隐私风险，需要承担隐私保护的责任。数据脱敏是回应隐私关切、实现法律遵从的重要动作，这对于在 6G 新系统中实现安全的 AI 数据服务尤为重要。AI 模型训练和推理过程中的数据脱敏技术一直备受关注，近年来关于 AI 中的差分隐私，同态加密，多方安全等技术领域都有着大量的研究工作。

数据存储能力：未来数据存储由核心向边缘转变，将会分布式地存储在云、边、端各节点中，并且需要支持非结构化、半结构化和结构化的多元数据存储架构。6G 新系统需能将数据按照分层分域的方式进行存储，提供多层次的容错和冗余能力。例如：云端可存储高价值明细数据及必要的轻度汇总数据，而边端可存储低价值明细数据、轻度汇总数据。以边端场景为例，利用本地存储的明细数据完成快速处理，赋能本地应用；处理后的结果数据，会根据存储及使用规范，向云端或其它的边缘节点传输。

数据开放能力：将汇聚后的内外数据通过融合加工、去隐私化处理、标准化地进行封装，形成对外数据服务，以标准化的数据服务方式（包括数据集、AI 模型数据、预测服务等），向 6G 内生 AI 网络及外部第三方用户提供开放服务。数据服务类型可结合 6G 实际应用场景，进行规范化分类，例如：人联网类（用户位置类、用户上网行为类、用户通信掉话率类等）、物联网类（终端类、设备类、车类等）及其它各种分类。

4.1.4 其它能力需求

除了上述 AI 算力，AI 算法和 AI 数据三大基本要素能力方面的需求，6G 内生智能还要能进一步促进 6G 新系统内生安全（主动免疫）的实现，例如：在面对未来更多未知和不确定的业务场景和用户环境下，6G 新系统要能基于内生 AI 能力，实时地感知分析、推理和预测出各种潜在的威胁和风险、实现全面自主免疫、主动防御和多网元节点之间协作联合的安全防御策略等。基于 6G 新网络内生智能的自学习训练和推理，各种新式变种的安全威胁和风险，都将可能被尽早地识别判定出，从而在 6G 全网内实时地做到安全认知和经验同步，形成无死角的联合协同防御并消灭任何的风险后患。

当前，5GS 系统已通过外挂式 AI 的辅助手段，实现了不错的系统节能增益。例如：对某区域 5G 网络，每万站点可实现年节电 2000 万度，显著节省了电费支出，同时网络 KPI 和用户感知也基本保持平稳【11】。尽管如此，未来 6G 新系统仍然有不小的能耗冗余压缩空间。未来，

6G 内生智能还需能进一步促进 6G 新系统实现能效和节电方面持续地提高，和更大的节能增益绿色减排，以有力支撑国家新能源和“双碳”发展策略。

4.2 服务需求（对内，对外等方面）

相对于 6G 新系统的内在 AI 相关能力，6G 内生 AI 新系统对外所能呈现的业务和服务，对于用户和客户更为重要和彰显价值。未来，6G 新系统服务的多主体对象领域从小到大大致可分为：个人（ToC），家庭（ToH），企业（ToB），行业（ToI），社会（ToS）等；每个不同的主体领域或层级，对 6G 内生 AI 业务和服务的需求都会有所差异，例如：AI 性能方面，其背后也会对应着上述 4.1 章节中不同的 AI 能力需求。下面我们还是从算力，算法和数据三大基本要素方面展开阐述。

4.2.1 算力服务需求

AI 密集型计算需求的高速增长，进一步催生了 6G 新系统算力服务的快速发展。未来，6G 新系统需要提供内部和外部 AI 业务应用所需的算力服务、数据服务和算法服务，实现系统算力开放共享和智能 AI 服务。多层算网融合作为未来新一代网络架构，其设计需协同考虑网络和计算融合的演进需求。实现泛在的算力连接和 AI 算力在网络中的全局优化包括：算力的按需灵活调度使用，算力服务的合理分布。多层算网融合的目标是：1. 网络可以感知和管控无处不在的计算服务，而用户无需关心网络中的计算资源状态。2. 算力灵活动态地部署，灵活调度算力到有计算需求的服务上。根据上述总体要求，6G 新系统算力服务主要要能提供以下服务类型。

1.时延敏感型服务：通过 6G 新系统的智能感知和决策，实时为用户 AI 服务提供充沛的算力支持，加速传输和计算效率，从而达到较短时延。相比于移动云计算 MCC 所需要的 100 毫秒量级时延，基于多层算网融合体系的 6G 新系统，可满足 1~10 毫秒量级的超低时延服务要求。应用场景例如：沉浸式虚拟现实 XR。

2.能耗敏感型服务：用户或客户可通过 6G 新系统的算力调配，将高能耗型的 AI 计算任务迁移到 6G 新网络中，进行分布式并行计算，从而避免本地计算带来的巨大能耗。应用场景例如：抢险救援，智能穿戴设备等。

3.隐私敏感型服务：与传统集中式计算相比，6G 新系统内的用户数据信息不再需要经过复杂的核心网到达云端数据中心。用户无需上传私密数据，可通过 6G 新系统算力调度和分配进行安全可靠高隐私的计算。应用场景例如：普惠金融，数据孤岛等。

4.服务体验敏感型任务：由于多层算网融合与用户的近距离优势，可通过 6G 新系统智能感知更准确地预测和判断出用户的计算行为和需求，从而提早合理地调度算力，提供更实时有效的计算服务。应用场景例如：沉浸式虚拟现实 XR、感官互联等。

4.2.2 算法服务需求

6G 新网络通过 AI 算法（模型）来实现对内和对外的内生智能服务，AI 算法品质决定了内生智能应用的质量和性能。6G 新网络对外提供的 AI 算法（模型）服务至少需要支持以下特性：

规范化 AI 算法模型描述和存储：AI 算法模型种类众多，面向不同的场景用例有很大的差异。6G 新系统需要给出规范化的算法模型（服务）描述，在统一的框架下用规范化描述语言来定义算法模型和其相关服务，且该描述是可解释的。每种 AI 算法模型需给出配置参数和配置文件，且支持内容更改。算法描述文件能针对算法的定义、配置、使用、性能、适用场景等给出解释。

AI 算法分类索引和服务接口：对于 AI 算法模型给出分类和索引，通过格式化标准化的接口使用 AI 算法服务，支持关键词等多种方式的检索利用。6G 新系统需支持单个 AI 算法部署和多个 AI 算法堆叠联合部署。在部署 AI 算法模型的时候，能自定义地修改调整算法，但修改后的算法如果要作为 6G 新网络智能内生算法，必须经过严格认证。

AI 算法认证：对外提供服务的 AI 算法模型必须经过认证后，才能作为 6G 新网络智能内生的适用算法。该认证需要能给出测试验证并确认其性能。认证结果信息需包含在算法描述文件中，包括但不限于：应用场景、性能、数据安全性、数据隐私性、资源开销等。只有经过严格认证的 AI 算法模型，才能在 6G 智能内生网络中被检索，部署和使用等。

AI 算法训练：也称作算法建模，根据提供的数据集寻找模型的最优参数。需要在描述文件中提供 AI 算法训练的信息，包括但不限于：训练数据，训练环境，训练资源开销、测试结果信息等。AI 算法训练作为 Network AI 的服务之一，在 6G 新系统分配指定的资源中运行，使用指定的训练数据集，训练过程可监测可控制。

AI 算法复用：6G 新系统如果对每一个不同的 AI 用例请求都独立配置内生 AI 算法或模型，将会产生巨大的成本开销。为此，6G 新系统应支持复用网络中已存在或已训练好的 AI 模型，来应对需求较相似的多个场景用例。具体的复用机制可以按照不同的需求，从 AI 算法模型框架、算法模型超参数，算法模型具体参数等多个级别由浅入深地进行。

AI 算法测试：在测试数据集上进行测试，评估 AI 模型的泛化能力和其它性能。测试指标包括但不限于：准确率、召回率等。6G 新系统需要能提供完整的测试环境和测试结果，当测试结果不合格时，该 AI 模型不能被部署使用。AI 算法测试作为 Network AI 的服务之一，在 6G 新系统分配指定的资源中运行，使用指定的测试数据集，测试过程可监测可控制。

AI 算法推理：算法模型经过认证后，可用于新的数据推理和决策。通过规范化的服务接口提供 AI 算法推理服务，该服务由经认证的用户使用。AI 算法推理的结果可用于进一步修正算法模型。

4.2.3 数据服务需求

未来，6G 新网络的作用之一就是创造一个“智慧泛在互联”的世界，基于无处不在的大数据，将 AI 能力和服务赋予各个领域应用。6G 新网络将在设计之初，就充分考虑与人工智能和大数据技术融合，将 AI 和大数据的应用融入到网络的基因当中，形成一个端到端的完备体系架构。为了实现这一目标，6G 新网络可支持独立的数据面功能，通过构建架构级的数据服务框架，在确保数据安全和隐私的基础上，建立起支撑内生智能的知识图谱，以满足全网中数据采集、机器学习、智能服务和应用赋能的全网全域内生智能需求，进而提供更高的数据传输速率和更可靠的传输链路，满足未来更多种业务类型增长和更严苛 QoS 要求。

AI 应用的本质就是：基于不断增强的算力对大数据中蕴含的价值，进行充分挖掘与持续学习的过程。6G 内生 AI 要以大数据为基础【12】，通过网络和终端泛在地“收集”、“处理”、“传输”和“使用”数据的四大基本步骤，不断推动 6G 新技术和新业务的发展。围绕 AI 数据处理过程，上述四个基本步骤可进一步分解为：数据发现，采集，预处理，传输，流转，训练，处理，分析，推理，决策，挖掘，增广，交易，存储，隐私保护等诸多动作，而这些动作贯穿在大数据处理和应用的流程中，具体如图 4 所示【13】，上述这些数据动作即可对应着各个 AI 数据功能和服务。

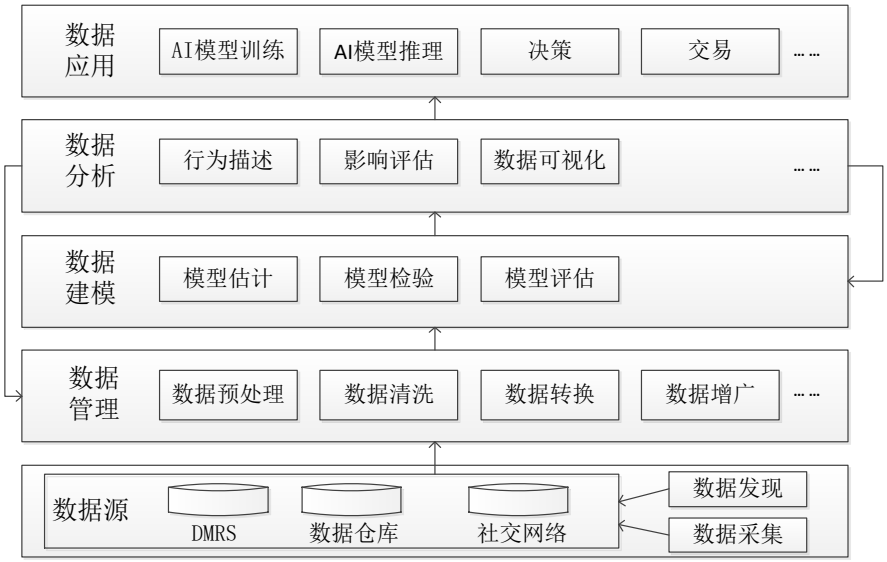

图 4: 大数据处理和应用

在大数据技术的加持之下，人工智能技术取得了快速发展，AI 对数据的质量、规模和个性化等方面的要求也越来越高，需要场景化、领域化的数据。这就需要增强传统大数据流程中各环节的处理能力，比如：在数据采集、标注和特征提取等方面进一步提升【14】。在使用深度学习或机器学习构建 AI 分析模型的过程中，数据集从各种数据来源收集获得，如：文件、数据库、传感器等。但是，收集的原始数据并不能直接用于执行分析过程，需要对原始数据进行预处理。为了能从机器学习和深度学习应用模型中获得更好的结果，数据的格式也必须以适当的方式进行表达和存放。某些特定的机器学习和深度学习模型需要特定格式的信息，例如：随机森林算法不支持空值，因此要想顺利地执行随机森林算法，必须从原始数据集中管理空值。在数据处理过程中，数据增广是智能网规网优常用的技巧之一，主要用于增加训练数据集，弥补采集样本数据不足的问题，让训练数据集尽可能的多样化，使得训练的 AI 模型具有更强的泛化能力【15】。特征提取是从收集到的初始数据开始，构建有信息性和非冗余性的特征值，从而促进后续的机器学习和泛化步骤。当 AI 算法输入数据太大而无法处理且存在冗余时，可将复杂的特征数据转化为一组简化的特征。针对不同的使用场景，从相应的数据集数据中提取特征，经过语法处理和语义分析得到满足对应场景使用主题的特征数据【16】。

基于前述的各个内生 AI 数据功能，6G 新系统还要能对外按需地提供各类 AI 数据服务，这既包含 AI 数据（源样本、数据集、特征集、模型文件、评估效果等）自身，也包含各个 AI 数据功能所对应的数据服务项目。通过合理地布局网络架构中的存储功能模块和数据传输方式，可保证 AI 数据在网络中高可靠低时延的传输，满足 AI 算法模型训练，验证和推理决策等过程的需求。反之，AI 技术以智能方式支持多种数据服务。

4.2.4 其它服务需求

对应于传统通信业务的 QoS 和 QoE 概念，为了能更精准的评判、度量和保障 6G 内生 AI 系统所能提供 AI 业务服务的质量等级，6G 新系统还需能提供完备的智能服务质量，即 QoAIS 评估和保障体系。具体地，QoAIS（Quality of AI Service）是对 AI 业务服务质量，进行系统性评估和保障的一套指标体系【17】。6G 新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服务体系，即 AIaaS。考虑到不同的智能应用场景（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）对 AI 业务服务的质量有着不同的需求，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，以及网络编排管控 AI 各要素（包括算法、算力、数据等）的综合效果。

6G 新网络内生 AI 服务可以大致分为 AI 数据类、AI 训练类、AI 推理类和 AI 验证类，每一类 AI 服务均需要一套自己的 QoAIS。传统通信业务的 QoS 主要考虑了传输时延和吞吐率（MBR、GBR 等）与连接相关的性能指标，而 QoAIS 需要能涵盖 AI 工作性能、开销、安全、隐私和自治等多个方面，需从连接、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，因此 QoAIS 指标体系首先需从内容上进行扩展。表 2 示意了 AI 训练类的一些指标需求。

表 2. AI 训练类服务的 QoAIS 指标体系

AI 服务类型	评估维度	QoAIS 指标
AI 训练类	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

注*：不同类型 AI 服务间的共同评估指标

QoAIS 是 6G 新网络内生 AI 编排管理系统和控制功能的重要输入参数，网络内生 AI 管理编排系统需要对顶层的 QoAIS 指标进行分解，再转化映射到对数据、算法、算力、连接等要素各方面的 QoS 具体要求上。

为了能够最大化有效地利用各行各业的知识、经验和数据等资源，6G 新系统还需能提供单个行业内和跨不同行业间的联邦学习机制和跨行业数据、知识经验等共享的对外服务体系。

4.3 架构需求（产品，部署等方面）

基于前述的 6G 内生 AI 在能力和服务方面的需求，为了能实现和确保满足这些需求，6G 内生 AI 需要具备不同于传统以通信业务为中心的移动系统架构和功能体系，例如：6G 内生 AI 架构将更强调系统功能和数据的自生成、自优化和强利用，实现以数据为中心的计算，突出系统分布式、协作式和以任务/项目为中心的网络操作，任何一个智能网元节点都可成为智能业务服务的发起者或终结者等。下面我们还是从算力，算法和数据三大基本要素方面展开阐述。

4.3.1 算力架构方面

近年来，人工智能技术取得了飞跃式的进步，应用程序逐渐向智能化和复杂化的趋势发展。6G 提出的沉浸式云 XR、感官互联、智慧交互等新型应用场景，对经典移动通信网络也提出了诸多挑战，例如：各种终端设备海量接入，网络规模日益膨胀，这对传统网络的运维、管理是极大挑战。过去计算和通信各行其道，传统移动网络对算力资源和通信资源之间，缺乏有效的管理和融合利用手段；未来外部的安全威胁日益严重，传统移动网络对细粒度安全防护束手无策；未来 6G 各种业务部署要求更迅速敏捷，而传统网络的业务部署、迭代升级耗时耗力。

6G 新网络在构建部署的同时，除了要考虑传统的运维问题，还应该考虑到对边缘算力资源的最大化利用。可以期待的是,未来 6G 新网络应当具备通过分发各个服务节点的算力信息、存储信息、算法信息等，结合网络信息（如路径、时延等），针对用户需求，提供最佳的资源分配和网络连接方案，并实现整网资源最优化使用能力。通过结合多智能体感知，云原生计算，数字孪生等技术，提升网络的通算一体化感知能力. 以上网络面临的本身系统管理扩展，通算一体化调度，以及极致用户体验和安全等方面问题，都迫切要求一个新的 6G 新网络架构出现。

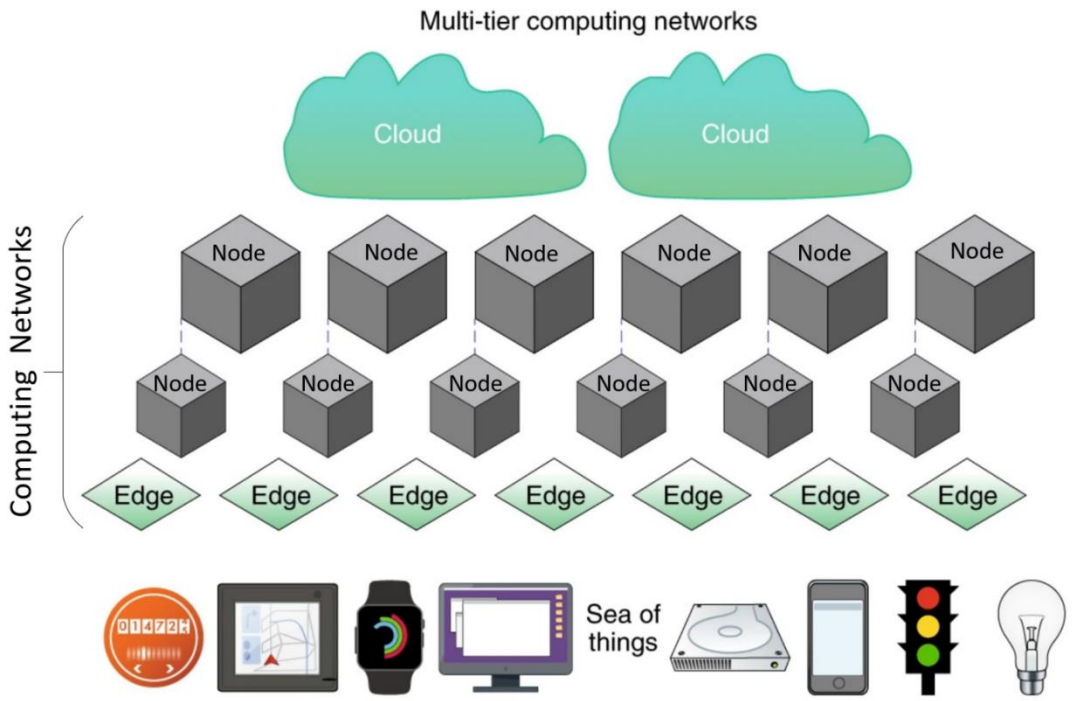

图 5：多层算网融合系统部署架构

如图 5 所示，未来的 6G 新网络将是一个云-网-边-端混合的多层算网融合系统【18】，将采用计算、通信、控制、存储融合度更高的新型架构。算网融合系统是一个多层的网络，其中的计算和通信网元节点可以提供内生在网算力。计算和通信融合可以更好地发挥出计算和通信能力。算网融合可支持各种应用和组件，一方面通过算力调度功能对各种计算资源进行合理地分配，另一方面提供网络算力和通信服务并管理这些服务，包括：控制物理设备以实现特定所需的网元逻辑功能、处理和转发从云层到边缘层的控制命令等。算力架构可以包括：云层、算网融合系统（包含边缘层）和端层，具体如下:

- **端层**是用户终端设备层，即用户任务发起层，通常包括：个人电脑、手机、可穿戴设备和物联网设备等。先进的终端（例如：功能强大的智能手机和嵌入式单板计算机）可以实现本地网关和简单组网功能。

- **边缘层**是算网融合系统的最底层，接受端层发起的任务，根据调度策略进行本地计算或向高层节点转发，这些边缘节点可以是环境中部署的设备。一般认为应在边缘设备附近实现实时控制和操作功能，即：对低延迟有较高要求的操作，通常是在算网融合系统的边缘层实现的。

- **云层**是算网融合系统之上的最顶层，提供与终端用户的应用层接口、控制命令和特定领域的应用。它将网络功能与后端应用集成在一起，实现算网融合系统与上层应用层的互通。

上述多层算网融合系统部署架构，可使我们能够开发出算力与通信融合的各种创新服务和应用。特别地，算力可以由网络中心/高层(即云层、算网融合)迁移到更接近用户的位置(即：边缘层、端层)，以获得以下增益:

(1). **最小化延迟**: 一部分业务应用只涉及基本的控制循环; 换句话说，要能在一定的时间条件下，触发一定的动作。然而，这些控制回路通常对时间非常敏感。在本地或靠近数据源的地方分析并做出快速决策，可以极大地减少那些控制循环的延迟，因为远程的云层可能无法对数据进行实时处理并存在额外的传输延迟。

(2). **提高可靠性**: 网络会将传感器数据用于公共安全或控制关键基础设施。通向远处的云层的上行链路很容易被破坏，因此将本地处理作为备用选项，或完全使用本地处理是很有价值的。例如：在工业控制回路或应急响应系统中，利用边缘计算进行本地处理，可以提高业务安全性和可靠性。

(3). **解决隐私问题**: 一些用户数据是敏感数据，或是一些法律要求不能存储在特定地理边界之外的数据。虽然云服务提供商通常被认为是可信的，但用户无法知晓和控制自己的数据实际存储在哪里，以及谁可以访问它。而本地网关和边缘节点则处于本地操作员的控制之下，可以提供更好的信任。

(4). **节约带宽**: 网络中的上行带宽通常受到限制，将大量数据从边缘设备传输到云服务器并不总是可行的。在本地先执行数据（预）处理，再将聚合和过滤后的数据发送到云端，可以显著减少上行传输带宽的消耗。

(5). 跨云边端协同：系统根据业务逻辑，实际资源等情况，为完成服务而跨云-边-端进行协同调度操作，至少包括：端-端协同，云-边协同，端-云协同等不同方式。在实际生产生活场景下，云边端彼此之间的概念，其实并不存在非常具体绝对的划分，云边端之间的界限可相对模糊，并且可根据实际服务内容有所改变。

在多层算网融合系统部署架构中，网络中的各级算力资源得到系统的管理和有效的调度利用，并能够和通信资源进行联合地调度或独立调度，最大化地满足用户业务/应用对算力和通信性能的不同需求，从而提高用户体验和资源的利用效率。

4.3.2 算法架构方面

通信网络，特别是移动通信网络，由于其具有非常复杂的网络体系、分布式的部署方式、极大规模的节点分布、稳定的系统运行要求等特性，因此针对 6G 新网络原生/内生 AI 算法架构的设计，应当充分考虑通信网络不同特性对 AI 算法与算法架构的需求和影响，从 6G 新系统架构的角度设计 AI 算法，实现网络 AI 系统（network AI system）。6G 新网络的原生/内生 AI 算法架构应当至少满足如下需求：

稳定性（Stability）：由于电信级移动网络需要非常稳定地运行，因此对 AI 算法运行和维护的稳定性也提出了较高的需求，AI 算法架构需要具备较为稳定的结构与系统化设计，并能够提供较强的容错能力。

实时可管可控（Real-time manageable and controllable）：原生/内生 AI 算法架构需支持网络信息的智能化实时监控、实时管理与实时控制。AI 算法架构在设计时，应充分考虑电信网络实时性需求与业务实时性保障。

安全与隐私保护（Safety and privacy）：AI 算法架构应当原生地支持不同算法模块、算力设备之间参数传递的安全和隐私保护。

中心化和分布式混合层级设计（Hybrid and layered design）：电信移动网络在不同层级具有分布式和中心化的多重混合式部署。部署原生 AI 算法在架构层面，应当支持分布式部署、分布式运行、以及多层灵活混合式部署和运行，从而可支持电信网络不同层级的需求。

算网一体化（In-network computing）：随着算网融合概念与研究工作不断地深入，算网一体化协同将会是未来 6G 新网络的重要特性之一。因此，原生/内生 AI 算法应当充分考虑算网一体这一网络特性，在架构设计上体现出算网深度协同的需求。

模型与数据划分（Distribution of model and data）：由于网络中算力设备能力的不同及设备部署方式的差异，分布式算法架构需要充分考虑通信网络 AI 模型、网络数据的划分与分布式训练与存储，从而可更好满足部署需求和算网一体化需求。

模型通信方式（Model exchange）：分布式或混合式 AI 架构需要不同层级/功能/区域间对计算模型的交互，因此 AI 模型传输（通信）的模式的定义与设计，也决定了分布式 AI 算法模型的性能。AI 算法架构根据不同的算法和场景，应当支持同步/异步/混合式模型通信交互方式。

数据与模型双重驱动 (Model and data-driven)：传统知识模型驱动的电信移动网络设计，在多个方面具有较大的优势，可弥补仅依靠数据驱动的机器学习算法所伴随的泛化性、可解释性缺乏等弊端。因此，AI 算法架构应当原生地满足数据与模型双重驱动的需求。

数据异构（Heterogeneous data）：AI 算法架构还需要考虑无线数据的特殊性和异构性。

通信效率（Efficiency）：原生/内生 AI 算法架构在设计时，应充分考虑通信网络效率的需求，需要减低模型进度，压缩模型更新，改进通信调度等。同时，AI 算法模型的收敛性和泛化能力对通信效率也有较大影响，因此在 AI 算法架构设计时也需重点考虑。

可演进与自演进（Evolvable）：AI 算法架构还应当具有较强的可演进性，以充分应对网络技术发展和新业务不断变化。

4.3.3 数据架构方面

数据是 AI 算法模型的基础，对于要支持原生/内生 AI 能力服务的 6G 新网络，AI/ML 相关的数据管理功能，可以被认为是与移动性管理、会话/任务管理重要性相并列的基本控制功能。因此数据管理的逻辑架构对于 6G 新网络架构至关重要，它主要体现在对各个数据功能实体的组织编排方面。

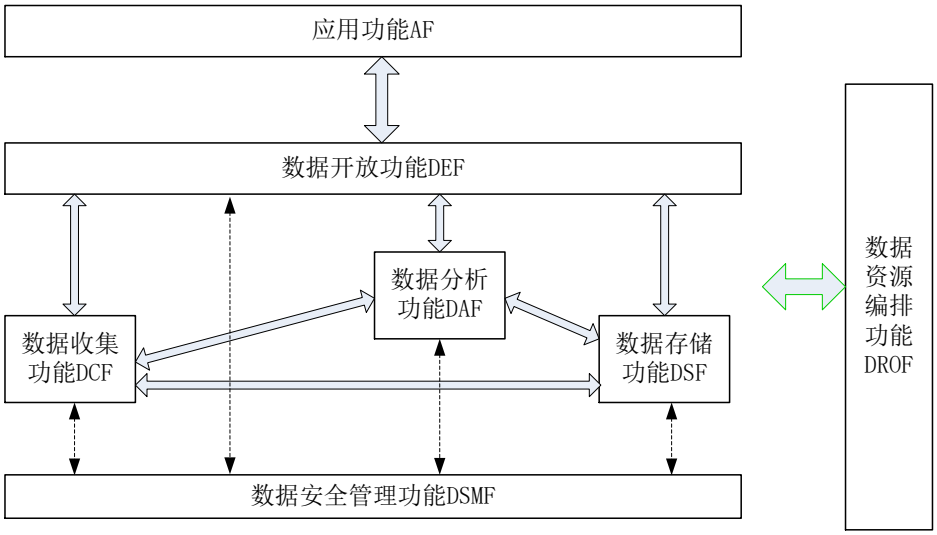

图 6：6G 数据管理逻辑架构示例 1

图 6 是一种可能的 6G 数据管理逻辑架构，其中逻辑功能实体主要包括：

- **数据开放功能 DEF (Data Exposure Function)**：用于向应用功能 AF 开放 6G 网络数据相关服务(DaaS: Data as a Service)。DEF 可基于 AF 提供的数据业务需求和/或服务等级协议

SLA，将 6G 网络数据进行去隐私化、聚合等处理，作为数据服务提供给外部/第三方 AF 或系统内部 AF。

- **数据收集功能 DCF (Data Collection Function)**：用于采集 6G 新系统内的网络、业务和终端数据等，供 DAF 实体使用。

- **数据分析功能 DAF (Data Analytics Function)**：用于基于 DCF 采集的数据，进行 AI/ML 模型训练，进而利用 AI/ML 算法得到和网络、业务和终端相关的分析结果信息。DAF 逻辑功能又可细分为模型训练子功能 MTLF 和分析推理子功能。

- **数据存储功能 DSF (Data Storage Function)**：用于存储 AI/ML 相关的数据。

- **数据安全管理功能 DSMF (Data Security Management Function)**：用于面向 AI/ML 相关数据进行隐私保护和安全管控。DSMF 与用户签约与授权管理功能 USAM (User Subscription and Authorization Management Function）、其它数据管理功能（DCF、DAF、DSF、DEF）之间相互交互，实现对用户数据处理（例如：收集、分析、存储）的授权许可、修改和撤销的检查和事件通知，从而控制其它数据管理功能对用户数据的使用和处理（例如：在用户数据使用授权撤销后，终止对该用户数据采集和分析，删除存储的数据等）。

- **数据资源编排功能 DROF (Data Resource Orchestration Function)**：用于管理（例如：分配、调度、删除）AI/ML 数据处理相关的资源。DROF 本身也可基于 AI/ML 算法（例如：深度强化学习），对这些数据处理资源进行动态智能的编排和管理。

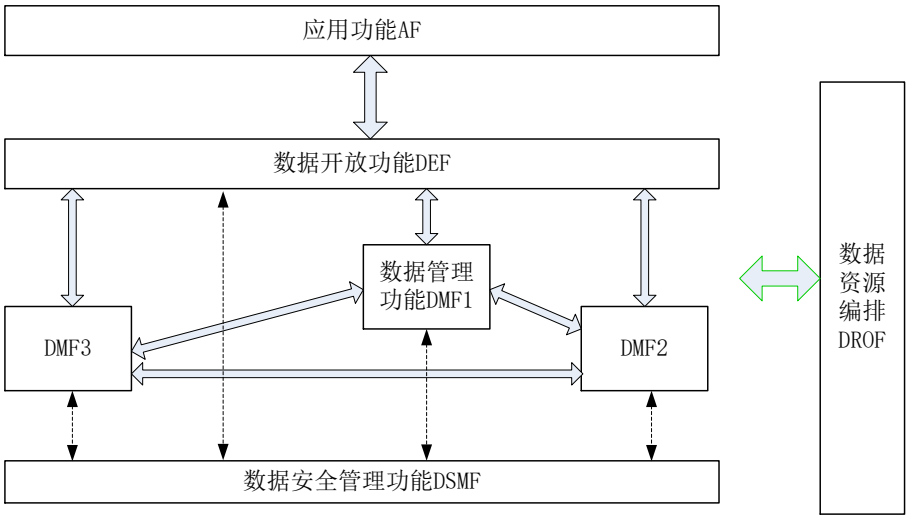

图 7：6G 数据管理逻辑架构示例 2

图 7 是另一种可能的 6G 数据管理逻辑架构。它与图 6 相似，该架构也包含 DEF、DSMF 和 DROF 功能实体。但与图 6 的不同之处在于，该架构定义了数据管理功能 DMF（Data Management Function），DMF 实体包含了数据收集 DCF、分析 DAF 和/或存储 DSF 功能。因此，图 7 中架构示例可视为图 6 中架构的一种变体。

图 6 和图 7 中数据管理相关的功能实体可采用集中式或分布式部署：

- DEF 实体可以部署在云端或边缘网络，用于向云端应用服务器/AF 或边缘网络中的本地应用服务器/AF 提供数据服务。

- DMF 可以部署在云端或边缘网络，亦可在 6G 网络内部分布式部署，从而支持网络中泛在的 AI 能力，以及与其它 DMF 之间的多点协作。

- DCF、DAF 和 DSF 也都可在云端、边缘网络或在 6G 网络内部分布式部署。

- DSMF 可在云端、边缘网络或在 6G 网络内部分布式部署，以满足网络中不同区域对 AI/ML 相关数据的隐私保护与安全管控的不同需求。

- DROF 可在云端、边缘网络或在 6G 网络内部分布式部署，以满足面向数据处理资源管理的实时性和本地化需求等。

上述 DEF、DSMF、DMF、DCF、DAF 和 DSF 都属于控制面功能实体，DEF、DSMF 一般部署在核心网(或与其它核心网功能实体共址)，而 DMF、DCF、DAF 和 DSF 可以部署在核心网或无线接入网（或与其它核心网或无线接入网功能实体共址）。DROF 则属于管理面功能实体。

在上述 6G 数据管理逻辑架构中，“数据”包括但不限于以下类型：

- 用于 AI 分析推理的输入数据，例如：网络数据、业务数据、终端数据；

- AI/ML 模型信息，包括：模型（超）参数、模型描述文件信息（例如：该模型适用于哪种分析类型、该模型的应用条件/状态）等。

- AI 分析结果信息，即基于 AI 输入数据、AI/ML 模型算法得到的分析结果，例如：对网络或终端未来行为、状态的预测信息，或当前行为、状态的统计判断信息等。

云边协同：要能实现数据的分布式协同采集、处理与治理。云端负责统一采集调度入口，边端执行具体采集任务；云端负责统一数据处理入口，与边端协同计算引擎配合，实现协同计算逻辑的调度与控制，边端负责下推算子的执行；云端负责统一数据治理入口，云端负责不同类型表的统一元数据注册与统计分析、规则配置等，边端完成数据治理执行。

4.3.4 其它架构方面

6G 要做到原生/内生 AI，需要主动拥抱比传统通信网络更广阔的生态，因为 AI 三要素算力、算法、数据都很难完全由一方来全部提供，很多 AI 场景用例需要由多方协作来共同完成的，例如：通过 6G 新网络实现多家企业在不直接共享其物联网终端数据的情况下，实现可信的联邦训练等。因而，6G 新网络将 AI 作为其内生能力和服务，需要在算力、算法、数据多个架构层面上，原生支持多方共同协作参与机制，实现多方可以简单的、透明的、公平的、可信的方式，参与到某个具体的 AI 训练、验证或推理过程中，包括但不限于：

1）**算力**：即 6G 新网络的异构算力资源可以是由多方共建、共享、共维护的，用户和客户可以将其拥有的空闲算力资源，基于 6G 内生 AI 架构透明公平地进行分享利用。

2）**算法**：6G 新网络内的 AI 算法，包括网络自用 AI 或第三方 AI 应用，其来源可以是多样的，6G 内生 AI 架构需要支持对相关算法库的高效管理运营，支持对多方 AI 算法的可信验证等，支持多方参与不同类型的分布式 AI 训练和决策。

3）**数据**：6G 新网络 AI 数据源可以来自任何第三方，6G 内生 AI 架构需要原生地支持相关多方数据的安全隐私保护，例如：当用户参与到 6G 某个联邦训练任务时，不会因为训练过程中的梯度同步过程，而造成用户隐私泄露。

5.技术需求和原则总结

本白皮书简要地回顾了 AI 和 5G 移动系统结合应用的历史和现状，分析了未来 AI 和 6G 移动新系统原生内生融合应用的基础和变革趋势。在这条充满挑战的演进发展之路上，业界必须对 6G 原生/内生 AI 在能力，服务，架构技术需求方面，先有着较为完整且统一的认知和共识，进而在此基础之上形成原生/内生 AI 能力，服务，架构体系和具体内容的构建和拓展。6G 原生/内生 AI 的实现，在其能力，服务，架构技术需求方面都会涉及到“算力”，“算法”和“数据”三要素方面，而这些技术影响和需求和今日以云 AI 为代表的集中式 AI 体系有着许多不同，原生/内生 AI 总体呈现出“泛在式”、“分布式”、“本地式”、“协同式”、“鲁棒式”、“隐私式”、“开放式”等特征。除此之外，未来为了能支持可差异化、定制化的 AI 能力和服务应用，6G 原生/内生 AI 还需要支持所谓 QoAIS 类的广义服务质量保障和评估机制，以支撑对应的 AI 商业模式新应用。6G 原生/内生 AI 需能支撑和构建比传统 ICT 网络更广阔的业态，其不仅可拥有更广阔安全可信的“算力”，“算法”和“数据”来源提供者，更能更广、更深、更厚地赋能和造福更多的行业客户和用户。

6.参考文献

【1】IMT-2030 网络组. 6G 网络架构愿景与关键技术展望白皮书[R]，IMT-2030 中国 6G 大会, 2021.

【2】中国移动研究院等. 5G 无线网络智能化研究[R]，CCSA TC5 WG9, 2021.

【3】3GPP. 3GPP TS 22.261: Service requirements for the 5G system(Release 17)[S], 2021.

【4】3GPP. 3GPP TS 23.288: Architecture enhancements for 5G System (5GS) to support network data analytics services(Release 17)[S], 2021.

【5】3GPP. 3GPP TS 28.104: Management and orchestration; Management Data Analytics(Release 17)[S], 2021.

【6】3GPP. 3GPP TR 37.817: Study on enhancement for data collection for NR and ENDC(Release 17)[S], 2021.

【7】欧阳晔, 王立磊, 杨爱东,等. 通信人工智能的下一个十年[J]. 电信科学, 2021, 37(3):36.

【8】程强, 刘姿杉. 电信网络智能化发展现状与未来展望[J]. 信息通信技术与政策, 2020(9):7.

【9】中国联通. 中国联通异构算力统一标识与服务白皮书[R], 2021

【10】中国联通. 中国联通云网融合向算网一体技术演进白皮书[R], 2021

【11】中兴通讯. 中兴通讯自主进化网络白皮书[R], 2020.

【12】Mohammad Sultan Mahmud,Joshua Zhexue Huang, etal. A survey of data partitioning and sampling methods to support big data analysis[J], Big Data Mining and Analytics, vol.3, issue:2, pp85-101, 2020.

【13】彭进香,张莉. 大数据处理技术与应用[M], 清华大学出版社,2020.

【14】Yongyong Ge. A Survey on Big Data in the Age of Artificial Intelligence[C], pp.72-77, 2019 6th International Conference on Information, Cybernetics, and Computational Social Systems (ICCSS)

【15】Jinkang Zhu, Chen Gong, Sihai Zhang, Ming Zhao, Wuyang Zhou. Foundation study on wireless big data: Concept, mining, learning and practices[J], China Communications , vol.15, issue:12, pp.1-15, 2018.

【16】Yuanwei Liu, Suzhi Bi, Zhiyuan Shi, and Lajos Hanzo, When machine learning meets big data a wireless communication perspective[J], IEEE Vehicular Technology Magazine, vol.15, issue:1, pp.63-72, 2020.

【17】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.

【18】Y. Yang, “Multi-tier computing networks for intelligent IoT,” Nat. Electron., vol. 2, pp. 4－5, Jan. 2019.

7.附录

定义和缩写

缩略语	英文全称	中文全称和说明
3GPP	3rd Generation Partnership Project	第 3 代合作伙伴项目
5GC	5G Core	5G 核心网
5GS	5G Mobile System	第 5 代移动通信系统
ADRF	Analytic Data Repository Function	分析数据存储功能
AI	Artificial Intelligence	人工智能
AI4NET	AI for Networking	通过 AI 辅助提升优化网络
AN	Access Network	接入网
AnLF	Analytic Logic Function	分析逻辑功能
CN	Core Network	核心网
CSI	Channel state information	信道状态信息
DaaS	Data as a Service	数据相关服务
DAF	Data Analytics Function	数据分析功能
DCCF	Data Collection Control Function	数据采集控制功能
DCF	Data Collection Function	数据收集功能
DEF	Data Exposure Function	数据开放功能
DL	Deep Learning	深度学习
DMF	Data Management Function	数据管理功能

DROF	Data Resource Orchestration Function	数据资源编排功能
DSF	Data Storage Function	数据存储功能
DSMF	Data Security Management Function	数据安全管理功能
MCC	Mobile Cloud Computing	移动云计算
MDT	Minimize Driving Test	最小化路测
MEC	Mobile Edge Computing	移动边缘计算
ML	Machine Learning	机器学习
mMTC	Massive Machine Type Communication	海量机器类通信
MTLF	Model Training Logic Function	模型训练逻辑功能
NET4AI	Network for AI	通过网络提供 AI 业务服务
NF	Network Function	网络功能
NWDAF	Network Data Analytic Function	网络数据分析功能
OAM	Operation and Maintain	运维网管
ODICT	Operation Data Information Communication Technology	运营数据信息通信技术
QoAIS	Quality of AI Service	智能服务质量
QoE	Quality of Experience	用户业务体验
QoS	Quality of Service	用户业务质量
RAN	Radio Access Network	无线接入网
RL	Reinforcement Learning	强化学习
SBA	Service Based Architecture	基于服务的架构
SLA	Service Level Agreement	服务等级协议
ToB	To Business	面向商业对象
ToC	To Customer	面向普通消费者

URLLC	Ultra-Reliable Low-Latency Communication	超高可靠低时延通信
URSP	UE Route Selection Policy	用户终端路由选择策略
USAM	User Subscription and Authorization Management Function	用户签约与授权管理功能
XR	X Reality	扩展现实

6GANA