



6G NETWORK AI CONCEPT AND TERMINOLOGY

摘要

这份白皮书定义了与6G网络人工智能（AI）相关的概念和术语。它还提供了网络AI级别和类别的定义，考虑了6G中网络-AI集成的不同级别和方面，以及网络AI使能的新服务（例如AlaaS）。通过阐明基本概念并为网络AI定义术语，这份白皮书旨在促进6GANA工作组在网络AI场景、需求、架构、算法和流程（例如网络管理和控制）方面的进一步讨论和共识建设。

内容

1 背景.....	4
2 概念和术语的定义.....	5
2.1 网络AI概念与术语.....	5
2.2 AI服务质量 : QoAIS ..	7
2.3 知识、模型与算法.....	10
2.4 相关概念说明.....	10
2.4.1 网络智能化.....	10
2.4.2 原生AI.....	10
3 网络人工智能等级和类别定义	12
3.1 网络人工智能等级定义.....	12
3.2 S0 - AI4NET ..	13
3.3 S1 - 连接4AI类别.....	14
3.4 S2 - 计算融合类别.....	16
3.5 S3 - 数据服务类别.....	18
3.6 S4 - 算法融合类别.....	19
3.7 S5 - 编排和管理服务类	21
别	
4 结论	24
参考文献	25
缩写和缩略语	26

作者

贡献者

程辉彭，刘哲，王飞，王军，邓娟娟，李刚，孙俊帅，李文静，李鹏，冯蕾，周凡琴，李杨，谢峰，康洪辉，王达，边森，段晓岩，艾明阳阳，马木雷，宫晨宇，张凯斌，温海波，段辰，顾芳芳，袁言南，丛伟伟，董新宇，陈敬然，杨婷婷，宁嘉红，徐夏，李鹏宇，王恒，黄冰明，温福喜，肖永晓，梁成超

组织

华为 中国移动 北京邮电大学 中兴 亚信 工业和信息化部移动通信技术股份有限公司 上海科技大学 诺基亚 上海贝尔

vivo

OPPO

大连海事大学

中国电信

中国联通

清华大学

华中科技大学

重庆邮电大学
电信

英特尔

北京科技大学

电子科技大学
中国

刘若峰 阮海军 张刚 冯双秦 刘一杰

1 背景

6g将由人工智能催化。 近年来人工智能（AI）在许多行业的成熟和广泛采用，激发了在6G中对网络智能的AI应用。

根据移动通信的历史，新的典型用例和服务需求会促使新一代移动通信系统出现，反之亦然。类似地，通信中6G和人工智能的出现将促进新的信息通信技术和消费者服务的发展。例如，6G将成为从数字化向智能化演进的B2B产业的新基础设施，以及满足用户对更高性能、智能和个性化需求的新B2C应用（如元宇宙和触觉互联网）的新基础设施。正是在这个背景下，网络人工智能被提出。

目前，业界已就网络AI的方向和趋势达成共识，即在网络中融入基于AI的智能连接、计算、数据处理和分析，走向融合的内生网络智能。然而，对于网络AI的内涵以及网络与AI的融合程度尚无统一认识。相关概念和术语也未能清晰界定。

从我们目前所处的位置出发，本白皮书定义了与网络AI相关的基本概念和术语。它还分析了网络AI的不同选项及其关键方面，包括基于AI的计算、数据处理和分析，以及新服务（如AIaaS）等，并将它们定义为网络AI的级别和类别。本白皮书旨在促进6GANA工作组及行业在网络AI研究中的进一步讨论和进展。

2 概念和术语的定义

2.1 网络人工智能的概念和术语

6G的一个重要愿景是实现包容性智能和智能连接。因此，除了连接基础设施，6G还应支持原生AI。例如，它应该与AI应用深度融合，以满足连接、计算、数据和算法需求——这被认为是6G架构演进的主要驱动力之一。在这个方向上的研究引发了许多讨论，并引入了一系列基本概念和术语。相关概念和术语的定义及内涵如下。

□ **网络人工智能 (AI4NET)** 人工智能用于提升网络性能、效率和用户体验。AI4NET 利用人工智能来优化传统算法（例如空中接口的信道编码和调制）、网络功能（例如移动性和会话管理）以及网络管理（例如资源编排和网络规划）。

□ **人工智能网络 (NET4AI)** 网络为人工智能提供各种支持，例如提高人工智能模型训练和推理的效率和实时性，或改善人工智能的数据安全和隐私保护。NET4AI将网络的范围从连接扩展到计算、数据和算法。

□ **人工智能即服务 (AlaaS)** 人工智能应用，包括AI4NET和新的AI服务，构建在网络基础设施内，并作为服务提供。它们可由运营商或第三方部署。

□ **云人工智能** AI 在云端运行。从这个意义上说，云 AI 与网络架构解耦，但它需要底层网络将数据传输到云端，云端作为数据处理、模型训练和推理的主要智能中心。

□ **网络AI**：网络提供了一种完全分布式¹ 人工智能环境，涵盖人工智能基础设施、人工智能工作流逻辑、数据和模型服务。

从概念上讲，网络AI包括AI4NET、NET4AI和AlaaS，其中NET4AI支持以下内容：AI4NET、新的AI服务以及AlaaS服务。网络AI和云AI可以独立开发或共同开发。

网络AI与AI4NET/NET4AI的关系

通常的理解是，在AI4NET中，AI是一个网络优化器，例如，它增强了网络性能和运维（O&M）效率。在NET4AI中，AI是一种在网络中运行或由网络支持的应用（例如，机器视觉）。网络AI支持两者。也就是说，从网络AI的角度来看，AI4NET和NET4AI是两种不同的场景，网络AI架构需要提供各种能力和优化来支持两者。

除了支持AI4NET和NET4AI，网络AI还需要考虑AI能力的运行，即AlaaS。通过在6G中原生支持网络AI，传统的通信基于生态系统将被转变为多方协作的智能生态系统，使商业和技术合作更加开放、灵活、可信和简化。

2.2 人工智能服务质量：QoAIS

与主要关注连接相关性能指标的传统QoS不同，QoAIS需要从连接、计算、算法等多个维度评估网络AI的QoS

¹ “distributed”描述了网络人工智能的一般含义，即ue和ne具有一定的网络人工智能能力，但这并不意味着网络中运行的每个具体人工智能应用都是分布式的。对于是否使用“distributed”这个词还没有达成共识。

和数据。因此，从广义上讲，QoAIS指标框架将包括性能、开销、安全、隐私、自主性等。除了连接的质量服务（QoS），QoAIS还将包括以下内容：

- 计算：人工智能模型训练和推理的计算功耗与效率，数据预处理等。
- 算法：人工智能/机器学习模型的性能、泛化性、鲁棒性、可解释性、模型训练/推理所需时间等。
- 数据：平衡性、完整性、准确性、动态分配和样本空间准备时间

人工智能服务与QoAIS之间存在一一对应关系。从类别上看，人工智能服务分为人工智能数据类别、人工智能训练类别、人工智能推理类别和人工智能验证类别。对于每种人工智能服务的类别，可以从多个维度设计评估指标，包括性能、开销、安全、隐私和自主性。每个维度可以进一步扩展。以下是一个示例。

人工智能服务类型	维度	QoS 指标
训练	性能	性能界限
		训练时间
		泛化性（数据分布的变化）
		可重用性（包括可迁移性）
		鲁棒性（数据噪声）
		可解释性
		损失函数和一致性之间 优化目标
		公平性
	顶部	存储开销
		计算开销
		传输开销
		能源消耗
	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
	自主性	完全自主
		部分可控自主性 interface)
		完全可控的自主性

QoAIS是网络人工智能编排和管理系统及控制功能的重要输入。网络人工智能编排和管理系统需要分解顶层QoAIS并将其映射到数据、算法、计算和连接上。

QoAIS可以包含AI应用的服务体验指标。以信道压缩为例，可选用归一化均方误差(NMSE)或余弦相似度作为衡量信道恢复精度的KPI；可选用链路/系统级指标（例如误码率或吞吐量）作为反映信道反馈精度对系统性能影响的KPI。QoAIS还可以包含通用评估指标，例如AI服务可用性和响应时间（从用户发起请求到AI服务第一个响应的时间），这些指标与AI服务类型无关。

2.3 知识、模型和算法

□ 知识

知识 是人们在实践中认识客观世界（包括人类自身）所获得的东西，是数据/信息处理和提取的结果。

数据 是可以识别的符号（包括数字、字符、单词、声音、图像和视频），用于记录客观事物。它为客观事物提供表示，但不提供判断或解释。数据对于客观事物的形成至关重要。**信息**，反映客观世界中事物的特点。与数据和信息相比，**知识** 是由处理、抽象、分析、提取和总结信息形成的。知识反映了事物的本质。

数据、信息和知识对应着感知和理解客观事物的三个不同阶段。

□ 模型

A **模型** 要么是一个由字母、数字和其他数学符号组成的方程/不等式，要么是一个描述对象特征和内部关系的数学结构表达式（通常是一个图表、图像和图表）。

一个优化模型由一个目标函数和包含变量的约束组成，用于通过寻求目标函数的极值以及获得极值的变量值来求解优化问题。

在网络AI的语境中，模型指的是机器学习模型。对于机器学习模型，输入是样本数据，输出是期望结果。它设置一个目标函数（通常是为了最小化模型的错误率）和约束，具有大量可调节的参数，这些参数通过模型训练来学习输入和输出之间复杂且通常难以理解的关系。

□ 算法

在广义上，一个 **算法** 指的是完成一项任务的方法和步骤。在数学中，算法是指根据特定规则解决某一类问题的具体有限步骤。在计算机科学中，算法是指使用计算机解决某一类问题的方法和步骤。

算法的特点是确定性、有穷性、有效性、正确性、连续性和逻辑性/通用性。

算法是解决模型的路径或方法。以机器学习为例。机器学习模型中有大量未知参数，但在模型训练中应用合适的算法，可以将这些参数调整以获得最优或局部最优的机器学习模型，然后使用该模型对新输入数据进行推理

□ □算力/□算力

计算能力（也称为**计算力**）指的是数据处理能力。在本文档中，它指的是节点通过数据处理实现特定结果的能力，此类能力包括计算能力和存储能力。

□ 智能

智能是一个总称 **智慧性** 和 **能力**。**智慧**涵盖了感觉、记忆和思考的过程，这导致了语言和动作的产生；而**能力**它负责表达这种语言和动作，即说话和行动。两者统称为智能。也就是说，智能指的是感受、记忆、回忆、思考、说话和行动的整个过程，即 **智慧** 和 **能力**。

在广义上，智力是指有效达成目标所需的知识和技能，包括自然智能（生物智能）和人工智能。

智能化

智能化是一个通过信息通信技术（ICT）、大数据、物联网（IoT）和人工智能等现代技术，使事物具备能够满足各种人类需求的特性的过程。

□ 人工智能

人工智能（**人工智能**）是机器利用知识模仿人类执行某些动作的过程。人工智能是一门技术科学，研究和发​​展模拟、扩展和扩展人类智能的理论、方法、技术和应用系统。通常，人工智能分为弱人工智能和强人工智能。前者使机器能够观察和感知，甚至在一定程度上理解和推断。后者使机器能够适应、像人类一样思考和解决以前从未遇到过的问题。目前的研究主要集中在弱人工智能上。

利用知识的过程包括表示、获取、转移和利用知识的方法。

众所周知，人工智能是由三大要素推动的：数据、算法和计算。数据奠定基础（机器学习的原材料），算法开辟道路（处理），计算构建基础设施（动力）。在这个过程中，需要进一步研究如何结合和利用知识。

机器学习

机器学习是一个研究计算机/机器如何模仿人类学习方式以获取新知识/技能，并利用/重构知识以实现更高性能的学科。它也是一种运用算法解析并持续从数据中学习，最终判断和预测世界发生什么的科技。

机器学习的一般过程是：收集数据>从数据中识别特征>建立初始模型>用数据训练模型使模型有效>用新输入数据对模型进行推理/预测。

机器学习是实现人工智能的重要手段。根据基本原理或类型，机器学习可分为符号学习、连接学习、统计学习和深度学习。根据学习模式，机器学习可分为监督学习、无监督学习、半监督学习、强化学习、迁移学习、深度学习等。

□ 知识图谱

知识图谱是一种可视化或映射知识领域的方式。它展示了知识的发展过程和结构关系。它使用可视化技术来描述知识资源和载体，并挖掘、分析、构建、绘制和展示知识与知识之间的关系。

知识图谱的主要目标在于描述现实世界中各种实体和概念及其关系，从而形成一个庞大的语义网络图。在该图中，节点代表实体、概念或值，边由属性或关系组成。

知识图谱属于人工智能的范围，是知识工程为实现在计算机系统中实现人工智能而进行知识表示的重要方式。图2-1显示了相关概念之间的关系。

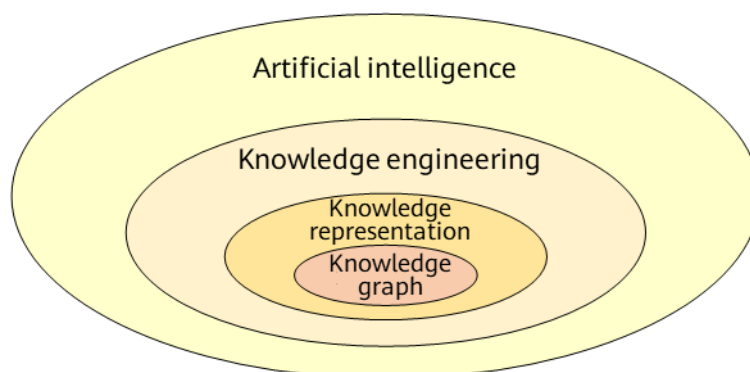


图2-1 AI与知识图谱之间的关系

□ 集中处理

集中处理 指所有由中央信息中心处理的信息。

□ 分布式处理

分布式处理 涉及多个计算节点（位于不同地方、提供不同功能或具有不同数据），通过通信网络连接，每个节点承担同一任务的不同部分，在控制中心的管理和协调下同时运行，以与其他计算节点共同完成同一任务。

□ 分布式机器学习

分布式机器学习 涉及多个计算节点（也称为 工人）从事机器学习以提升性能、保护隐私并支持更大规模的训练数据和更大的模型。

面对大数据和大型模型带来的挑战，训练大规模机器学习（尤其是大规模深度学习）模型对计算能力和存储容量提出了新的要求：由于计算复杂度过高，在单个节点、处理器或计算集群中执行训练任务耗时过长，具有高度并行性的处理器或计算集群成为执行训练任务的自然选择。同样，当在单个节点中存储大量数据不切实际时，分布式存储是唯一的出路。正是在这种背景下，分布式机器学习被提出。

分布式机器学习涉及以下关键任务：数据和模型划分（如何将它们分配到各个节点）、单节点机器学习模型算法与优化（如何在每个节点上优化模型）、节点间通信（如何在不同节点间传输消息），以及模型聚合（如何将每个节点的模型聚合到最终的模型）。

在并行性方面，任务可以通过以下方法进行处理：数据并行、模型并行和混合并行。

典型的分布式机器学习框架包括联邦学习、群学习（swarm learning）和集成学习（ensemble learning）。

2.4 相关概念说明

2.4.1 网络智能化

网络智能化 指人工智能技术与通信网络中的硬件、软件、系统和流程的深度融合。人工智能技术使通信网络的操作过程智能化，以更低的成本获得更高的质量和效率。此外，人工智能技术将革新网络技术和系统，实现敏捷的服务创新，并引领具有原生人工智能的网络。

2.4.2 本地AI

原生AI 指在6G中通过架构和功能设计原生支持人工智能，而非通过附加功能支持人工智能。原生人工智能受以下因素驱动：

□ **高层网络自主性**：网络需要提供本地人工智能功能来检测和满足操作员和用户的意图，这将反过来导致网络自我优化、自我进化，最终实现高度自治。

□ **包容性智能**：行业希望通过智能数字化转型，随时随地提供智能能力。与云服务提供商相比，这些智能服务将更具实时性并带来更高性能。此外，行业间联邦智能的可用性将实现跨域融合与智能共享。

□ **提供新的增值服务与终极服务体验**：用户设备拥有大量数据且计算能力不断增强。考虑到对数据隐私的要求，需要原生智能来促进网络设备间的协作，并为B2C客户提供终极服务体验和新的增值服务。

□ **网络安全与可信性**：未来网络将支持更多样化的服务，承载更多类型的数据并服务于更多用例/场景，这将使网络面临大量意外和复杂的攻击。为此，基于人工智能的安全能力需要嵌入到6G网络的每个过程中，旨在主动和协同地检测和防御威胁。

除了满足基本的通信连接要求外，6G网络还必须综合考虑计算、数据及模型/算法等众多方面。也就是说，6G网络必须在架构上设计原生AI，以满足多样化服务场景和网络自主性的要求，包括用于网络优化和用户体验（例如AI赋能新6G无线技术）以及来自第三方的新的AI服务。

原生AI的含义：在6G网络中，原生AI提供实时、高效且智能的服务和能力，以满足高级网络自主性、行业用户的包容性智能、终极用户体验和原生网络安全的需求。它还为AI工作流的整个生命周期提供完整的运行环境，包括在6G网络架构中的数据收集、数据预处理、模型训练、模型推理和模型评估，使其能够深度融合AI服务所需的计算、数据、算法、连接、网络功能、协议和流程。

插件AI：采用插件式设计的AI应用通常涉及补丁安装，因此面临以下挑战：

□ 缺乏统一的标准化框架，进而导致人工智能应用缺乏有效的验证或保障方法。人工智能应用的事后验证使得端到端（E2E）过程变得繁琐复杂，需要大量人工干预，并对运行网络产生重大影响。结果，真正将人工智能应用于运行网络变得困难。

□ 预验证、在线评估和自动闭环优化的困难。在插件AI模式中，需要从网络收集大量数据并发送到插件AI

用于人工智能模型训练和推理的函数，而集中式数据收集和处理极其困难。这，加上高传输开销，导致了人工智能模型训练和推理中各种问题，包括长迭代周期、高开销、收敛慢和泛化能力差。

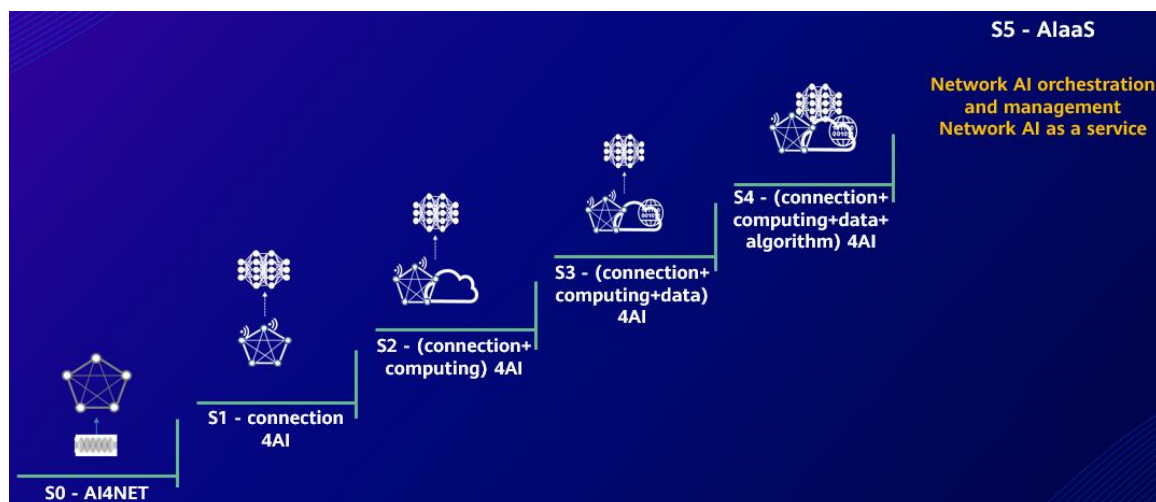
□ 在使用附加人工智能时，计算、数据、模型和连接属于不同的领域。跨系统协作只能通过管理平面实现，这将导致秒级甚至分钟级的延迟，因此无法有效保障服务质量。

由于上述问题和挑战，业内普遍认为6G网络必须采用原生AI设计。

3 网络人工智能级别和类别定义

3.1 网络人工智能级别定义

网络人工智能研究的本质是研究网络与人工智能的融合。因此，有必要从一开始就确定这种融合的选项和发展/升级路径。下列图表从网络的角度说明了网络-人工智能融合的层次。层次越高，融合越紧密、越全面，意味着需要在6G网络的范围内设计更多领域。在图中，S0表示AI4NET，S1至S4表示NET4AI，S5表示AlaaS。



□ **S0 - AI4NET**: 在这个层次上，人工智能主要被用作网络优化器，不一定影响原始网络架构。例如，基于数据的AI模型被用于替换网络中的传统数值算法以优化网络性能和用户体验，或实现智能运维。AI4NET已在5G中研究和应用。随着以深度学习为代表的人工智能技术的成熟以及融合连接和计算的新基础设施的出现，相关应用将更加多元化和成熟，其带来的演进将帮助我们获得更多收益并衍生更多场景来支持全智网络。

□ **S1 - 连接4AI**: 连接服务是一项基础服务，需要分析AI（一项特殊服务）在连接或网络方面是否有特殊要求。例如，与新AI服务相比，5G支持连接QoS服务可能会对可达性、计算速度、吞吐量、延迟、可靠性、安全性和隐私提出新的要求。研究人员必须研究如何通过6G网络更好地满足这些新要求。

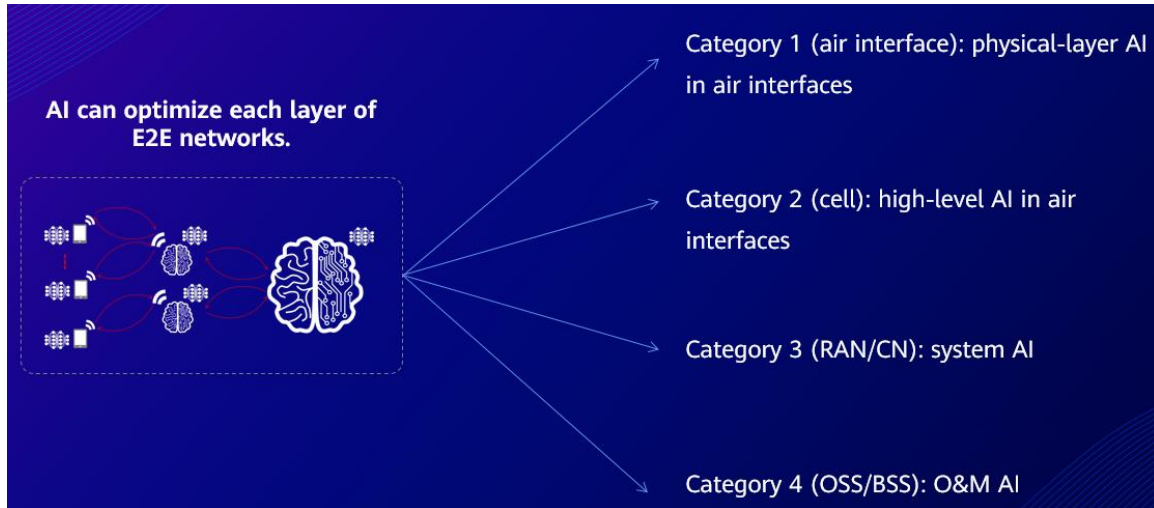
□ **S2 - (连接+计算) 4AI**: 6G将成为融合连接和计算的新基础设施，满足人工智能的连接和计算需求。此类需求可能会推动6G架构中的原生网计算融合或云网设备融合。

□ **S3 - (连接+计算+数据) 4AI**: 6G网络将为人工智能的连接、计算和数据需求提供数据服务。这些需求可能会推动6G架构中的原生网络-计算-数据集成，从而产生一个安全可靠的综合数据服务。

□ **S4 - (连接+计算+数据+算法) 4AI**: 除了满足人工智能的连接、计算和数据需求外，6G网络在一定程度上可以感知人工智能模型。对于不同层次的人工智能模型，6G架构将提供一个全新的原生范式——以适应性强的方式优化和扩展特定的AI模型。

□ **S5 - AlaaS:** AlaaS 协调和管理网络AI的连接、计算、数据和AI工作流程，并可在网络基础设施中构建AI应用。

3.2 S0 - AI4NET的分类



类别1：无线接口中的物理层AI

在这个类别中，人工智能/机器学习应用于物理层模块，如信道建模和估计、信道编码、调制、MIMO和波形设计，以提取时间域、频域和空间域中无线信道的特征。例如，通过利用神经网络，人工智能/机器学习学习无线信道的时间相关性。使用此类相关性训练的模型将能够提供更准确的信道信息。人工智能/机器学习还可以用作解码器。一些研究表明，神经网络解码器可以降低复杂性并更好地补偿非线性。人工智能/机器学习已广泛应用于MIMO系统。例如，通过全网状网络，它可以映射每个天线集和频段中的信道到另一个天线集和频段中的信道。也就是说，在频率复用（FDD）系统中，下行链路信道信息可以通过上行链路信道检测直接获得。人工智能/机器学习的另一个应用是压缩信道状态信息（CSI），以降低CSI报告开销。

我们可以得出结论，人工智能/机器学习为不同的物理层功能提供了一个通用优化模块，从而提高了物理层的适应性和灵活性。它提供了一种进一步改善无线链路性能和探索潜在收益的新方法。

类别2：空中接口中的高级人工智能

在这个类别中，人工智能/机器学习被应用于高级多用户处理场景，包括空中接口的功率控制、服务质量管理、资源分配以及自适应调制和编码（AMC）。资源分配是基站MAC层的一个重要功能。可分配的资源包括接入机会、传输机会、功率和频谱。一些研究表明，人工智能/机器学习可以优化资源分配算法以提高资源效率。大多数传统AMC方案是被动式的，它们仅根据接收器的反馈来调整调制和编码方案。相比之下，人工智能/机器学习可以优化AMC，基于更广泛的学习解选择经验来做出更好的选择。这些应用的本质是基于人工智能/机器学习进行自主和累积学习，以优化相关的调度算法。通过这种方式，基站变得更智能，能够根据学习和经验做出更准确的决定。这样的变化可以转化为在蜂窝级控制和调度中的显著性能提升。

因此，通过广泛将人工智能/机器学习方法应用于无线空中接口的物理层、介质访问控制层和相关协议信令，我们未来或许可以用人工智能重构无线通信空中接口。

类别3：系统AI

在这种类别中，人工智能/机器学习被应用于无线接入网（RAN）/核心网（CN）系统架构。在RAN中，基于基站之间传输的切换奖励信息，人工智能/机器学习帮助基站持续学习和优化切换。在CN中，3GPP定义了网络数据分析功能（NWDAF）来收集和处理人工智能/机器学习所需的数据，并应用和部署人工智能/机器学习模型。因此，人工智能/机器学习可以合成网络切片以实现异构网络，例如集成地面和非地面网络。通过协调复杂的多层异构网络，人工智能/机器学习为用户提供最佳覆盖。

类别4：运维AI

在此类别中，人工智能/机器学习被应用于网络管理和运维系统。网络管理和运维贯穿于网络和服务开发阶段，包括规划、建设、维护、优化和运营（即网络管理的整个生命周期）。运维人工智能利用以下功能的人工智能技术，进行全生命周期的网络运维和管理：

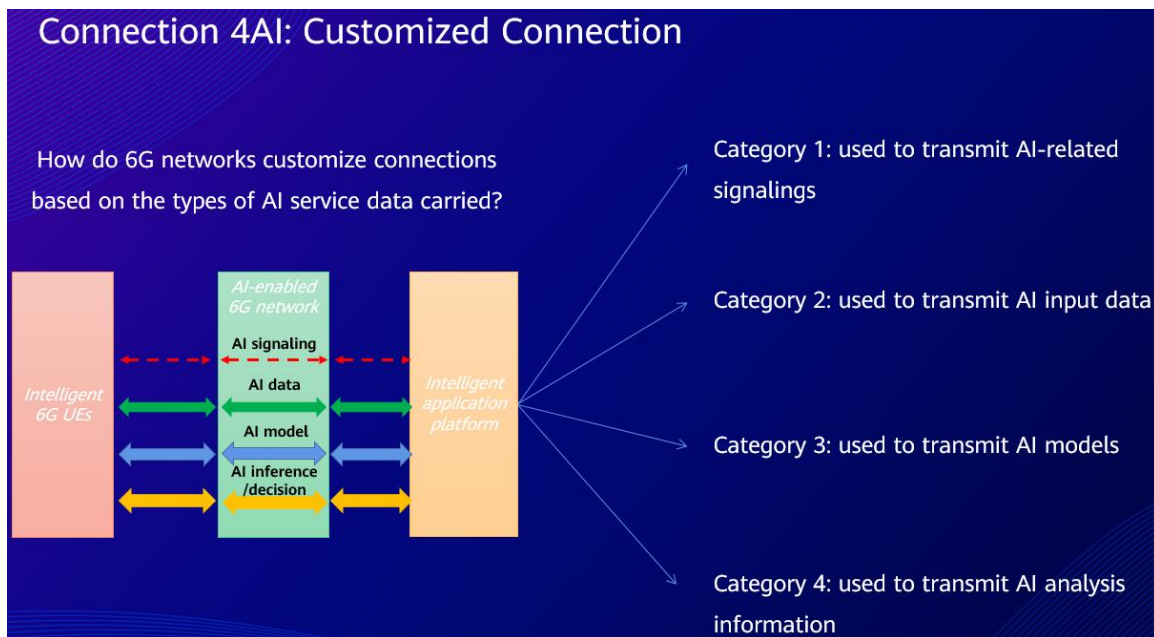
- 设计和编排功能：为了适应6G用例的多样性以及不断变化的服务需求，采用智能感知自动调查资源，并快速完成服务功能、网络能力、资源关联和接口调用的设计。根据设计结果，服务和网络资源将被自动组合，生命周期流程将被自动编排，实现灵活的服务提供和网络资源调度。
- 资源管理功能：此功能涉及大量应用程序，例如网络资源数据管理、资源注册/调度/分配/检查/变更和端到端网络资源拓扑视图。数据服务能力被统一封装并开放。例如，基于AI的图像识别使资源管理智能化，大大降低人工成本。
- 故障管理功能：使用AI进行集中式网络监控，包括端到端网络和服务监控以及闭环故障管理。这些网络监控能力被开放以驱动生态系统创新。
- 绩效管理功能：AI用于对网络和服务质量进行端到端分析，这对于各种分析应用至关重要，例如容量、质量、效率和效益分析。
- 网络规划与优化功能：基于网络数据中间平台，该功能专注于无线网络规划与质量优化。它利用强化学习等技术实现多目标/多参数联合优化。分析与优化能力被开放以实现智能闭环优化。
- 运维调度功能：它实现了运维人员、运维任务、网络割接、网络运维及相关流程的集中调度和管理。采用机器人流程自动化（RPA）等技术实现自动工单派发和知识推送。进度通过统一流程引擎可视化。

3.3 S1 - 连接类别 4AI

连接4AI主要需要回答以下问题：

- 如何根据所承载的AI服务数据类型定制6G网络连接？
- 6g网络如何提供人工智能所需的网络服务？

定制连接

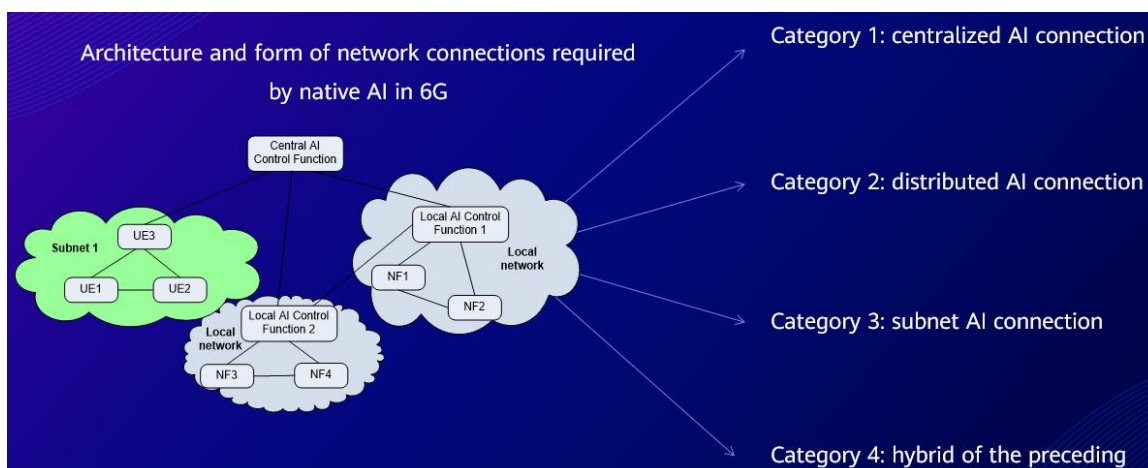


基于连接所承载的AI服务数据类型，定制连接可进一步分为以下类别：

- 类别1：用于传输与AI相关的信令，例如AI分析信息以及相应的计算请求/响应消息。与AI相关的信令可以通过以下任一方式传输：a) 作为用户面数据传输；b) 与非接入层（NAS）信令耦合/融合。例如，NAS信令用于网络中的原生AI控制，而服务/应用相关的AI信令被放置在一个容器中，并通过NAS信令透明传输。
- 类别2：用于传输AI输入数据。
- 类别3：用于传输AI模型。
- 第4类：用于传输AI分析信息，包括中间分析信息（当多个实体执行联合分析时）。

前面的类别可能对连接QoS和网络适应性提出新的要求。即使在同一类别（例如，第4类）中，人工智能服务/应用也可能对连接QoS、适应性等有不同的要求。考虑到人工智能服务场景和连接性能要求，3GPP TS 22.261（基于TR 22.874的结论）也对5G系统的KPI要求进行了分析，并得出了一些结果，主要涉及对吞吐量、延迟和可靠性等传统连接性能指标的影响。那么，3GPP从功能和性能的角度分析了5G系统如何支持不同的人工智能服务/应用，但如何在网络中设计原生人工智能仍然没有答案。

定制网络

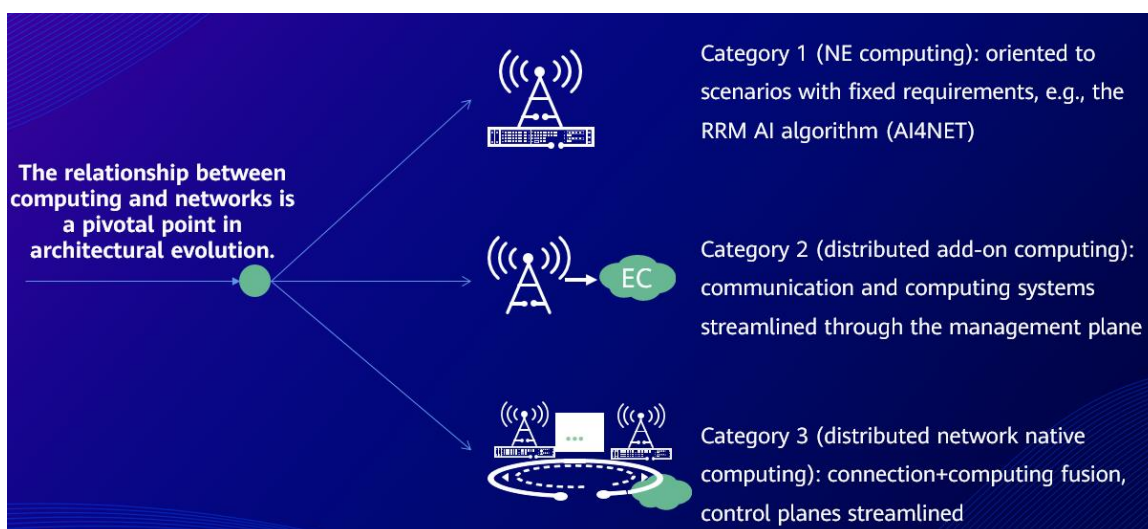


基于原生人工智能在6G中对网络连接架构和形式的要求，定制网络可以进一步分为以下类别：

- 类别1：集中式AI连接。部署一个中央AI控制功能实体来控制AI策略，包括AI分析信息收集、决策和交付。
- 类别2：分布式AI连接，例如边缘AI。类似于多接入边缘计算（MEC），每个本地AI控制功能实体可以管理和控制一个或多个功能实体（用户设备、网络功能、应用功能等）。
- 类别3：子网AI连接，例如，在多个UE之间或在一个虚拟网络（VN）内部。
- 第4类：前几类的混合。

前面的网络形式可以满足不同的人工智能服务/应用程序或控制场景的要求。

3.4 S2 - 计算融合分类



在未来网络架构中，计算将分布在基础设施中，包括中心云、边缘云、网络设备和甚至用户终端。计算和人工智能

算法或基于其构建的功能应用不仅将为网络或设备服务，以提升性能和优化网络运维，同时也会通过统一接口向上层应用开放。计算与网络需要相互感知，以实现资源的最优利用和最佳用户体验。在融入网络的过程中，计算可能逐渐向以下三个类别演进。

类别1：NE计算

在这个类别中，计算通常以专用计算资源的形式服务网络元素（例如，移动通信网络中的基站或 CN 中的设备），旨在实现网络功能或网络元素内的计算。通常，可用的资源有限，主要用于定制 AI 应用服务（即 AI4NET），如无线资源管理、信道估计和波束成形，这些服务旨在提高通信性能或优化网络运维。这一类别的计算通常来自使用通用处理器或可编程组件形成的计算和存储单元。它可以通过接口在运营商的网络管理平台上呈现。此外，管理平面接口可用于在指定的计算单元上加载、更新或删除 AI 算法，使管理平面具有管理和控制计算和算法的能力。与边缘计算（EC）和云计算相比，这种计算能力非常有限，因此无法满足涉及大规模计算和训练的 AI 应用需求。这使得它难以服务第三方应用。

类别2：分布式附加计算

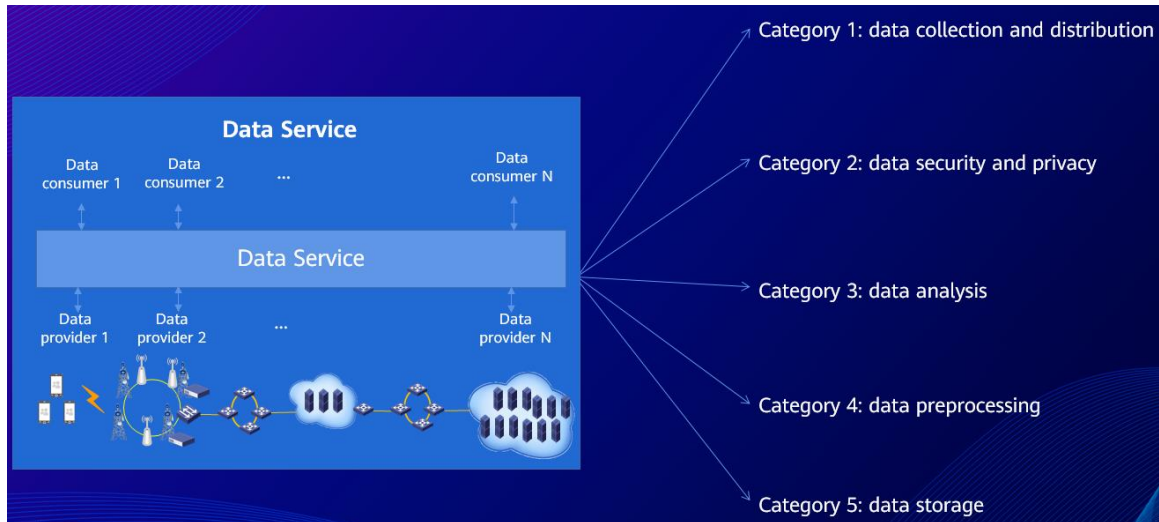
这种类别通常以分布式EC或MEC的形式存在。它源于云计算，从集中的数据中心向下移动到接入网络的边缘，更接近终端用户。附加计算主要由通用CPU提供，也可以由高性能GPU和可编程加速卡提供。相对丰富的计算能力支持网络优化，并有潜力解锁对计算量和延迟提出更高标准的行业应用。附加计算将计算服务分布在更靠近用户的网络边缘，以降低延迟，减少网络资源消耗，从而更好地服务于行业应用，特别是在视频加速、自动驾驶网络和AR/VR等低延迟和高带宽场景中，以及服务于非实时无线协议处理和网络优化的网络应用。通信网络功能对网络管理平台开放，分布式附加计算也在其中呈现。因此，在考虑网络信息分布式计算资源后，AI和其他服务可以被部署或调整。这些服务通过管理平面部署，但管理平面不够动态。因此，网络 and 计算无法在控制平面上统一，因此无法及时响应用户移动和网络变化。网络和服务连接相互独立，属于附加模型。因此，有时无法实现最佳资源利用。

类别3：分布式网络原生计算

在新网络架构中，NE除了具备控制和转发能力外，还拥有计算能力。同时部署了计算节点。这种计算-网络集成所生成的计算能力被称为 *网络原生计算*。在网络设计的早期阶段，计算被视为网络的基本要素。它广泛分布于云、边缘、设备和中间NE中，并集成到网络中。基于计算和连接开发出来的计算、连接及其他服务均作为基本网络服务提供。网络原生计算可以促进原生智能的开发和部署，更好地支持集成感知、通信和计算能力的泛在基站和UE，实现大规模智能分布式协作服务，最大化网络中通信和计算的有效性，同时适应数据分发并保护数据隐私。在新网络架构中，NE和计算单元的控制平面被简化，以弥补类别2（分布式附加计算）的缺点，并更快地响应移动性和网络变化。网络原生计算可以

也促进未来适用型智能应用的生成和发展，例如沉浸式云XR、全息通信、感官互联、智能交互、通信感知和数字孪生。

3.5 S3 - 数据服务类别

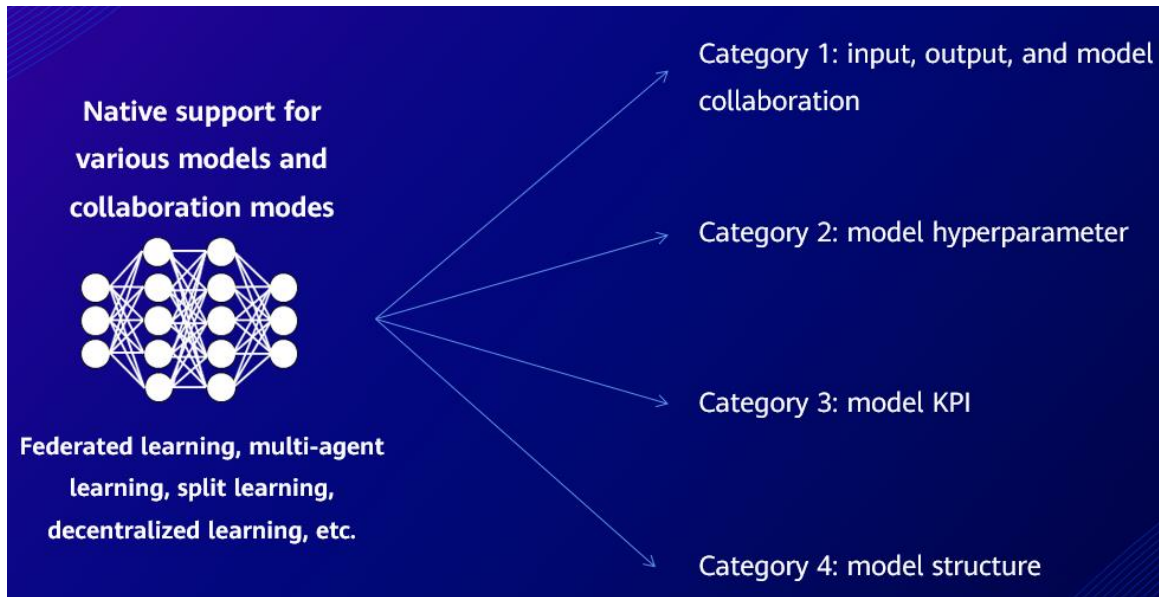


数据是6G的核心生产要素之一。在5G网络中，大多数数据来自通信网络运营和订阅。随着6G将服务从通信扩展到传感、计算和人工智能，数据的范围和类型将增长。作为数据提供者和数据消费者之间的抽象功能，数据服务将数据消费者与数据提供者解耦。当存在多个数据提供者或消费者时，它们通过可重用性帮助维护数据完整性和提高效率。6G数据服务旨在高效地支持端到端数据的收集、传输、存储和共享，因此它们将要解决的最大问题包括：寻找一种以方便、高效和安全的方式为内部或外部网络功能提供数据的方法，以及寻找一种在遵守隐私和安全法律法规的同时减少数据获取困难的方法，以提高数据服务效率和数据消费体验。

根据其潜在功能，数据服务可以分为五类：

- 第一类：数据收集和分发。此服务为数据生产者和消费者提供发布和订阅机制，使基本数据收集和分发比以往任何时候都更加高效。
- 二级类别：数据安全和隐私。安全和隐私保护技术用于为用户提供高质量可信的数据服务。此类服务将保护用户和网络隐私，并通过防篡改和可追溯性保持数据安全。
- 类别3：数据分析。模型、算法、知识和计算相结合，提供统计数据、预测信息、网络异常分析以及优化建议，为内部和外部网络功能的数据消费提供便利。
- 第4类：数据预处理。收集到的数据通过常用工具进行预处理，例如格式转换、去噪和特征提取，以满足智能应用的要求。
- 类别 5：数据存储。此服务存储和检索收集的数据，以支持其他数据服务，如数据安全和隐私、数据分析以及数据预处理。

3.6 S4 - 算法融合的分类



类别1：输入、输出和模型协作

用于模型训练的样本数据集和用于推理的特定任务数据都是 **输入**。在协同式AI操作中，前一次操作的结果也被用作下一次操作的输入。**输出**指AI模型训练过程中每一步的结果，包括协作过程中生成的中间结果，需要发送到一个协作节点，以及最终结果。

定义人工智能模型的输入和输出方式与模型的类型和功能密切相关。可以通过以下方式提升人工智能赋能功能的智能性：

□ 对于决策型人工智能模型，需要定口两种口入和一种口出：

- 第一种输入是 **通用预测**，其中包括服务预测、位置预测、负载预测和用户行为预测。每个具体功能将使用一种或多种通用预测作为输入判断的重要基础。
- 第二种输入是 **个性化数据**。一个应用程序在不同情况下可能需要不同的个性化参数。
- **个性化输出**指人工智能模型分析的结果。需要为每个具体功能点进行定义，以便能够更快地做出适当、准确的决定。

□ 对于非决策式人工智能模型，数据处理的部分或全部步骤被人工智能模型推理所替代。在特定场景下，

- 输入需要给出约束条件和预期效果的个性化定义。基于这些定义，人工智能模型可以执行最优数据处理。
- 输出主要是优化数据处理的结果，因此不需要特别考虑。

AI模型的协作模式包括联邦学习、多智能体学习、模型分割、迁移学习和群集学习。6G网络预计感知AI模型的输入、输出和协作模式，以便为特定的AI操作调整资源。以模型分割模式为例，用户设备将某一层的中间结果发送给网络。由于其能够感知此类中间结果以及AI操作的协作模式（即模型分割），网络向用户设备推荐最合适的分割点。

ue或应用服务器，同时考虑网络本身的状况以及其与ue的通信。这样，模型分段模式可以高效运行。

二级分类：模型超参数

在机器学习中，*模型超参数* 参考模型的的外部配置，主要用于优化和调整模型。通常，在训练开始前手动进行调整为进行。主要的超参数包括学习率、批次大小、优化算法、迭代次数、隐藏层数量、隐藏层神经元数量和激活函数的选择。

学习率 (LR) 指在优化算法中更新网络权重的幅度。通常，以下类型的 LR 是可用的：常数、衰减、动量以及自适应。将要使用的 LR 类型由优化算法决定。值得指出的是，高 LR 可能导致模型收敛困难并导致持续的震荡损失，而低 LR 可能因为模型收敛非常缓慢而延长训练周期。选择一个合适的 LR 对于加速模型收敛、防止局部最优、减少迭代次数以及，更重要的是，提高模型精度至关重要。

另一个重要的超参数是 *批处理大小*，指的是用于训练神经网络模型时发送给模型样本的数量。批处理大小不仅决定了模型的优化程度和速度，而且直接影响内存资源的使用。如果批处理大小太小，梯度可能会持续变化，减慢模型收敛。如果批处理大小太大，可能会出现内存不足，软件程序可能会失败。

考虑到它们对模型性能的直接影响，超参数至关重要。正确调整超参数的设置有助于最大化模型性能并更好地训练模型，从而提高资源利用率。对于6G网络，我们可以预测模型的超参数，通过第三方AI来促进模型训练并最大化模型性能。进一步地，通过开放资源和模型，6G网络可以协助第三方应用程序进行模型训练。更具体地说，通过资源开放，OTT将获得足够的计算、存储和通信资源，以在6G网络中训练模型。通过模型开放，OTT将获得预训练模型，并且只需要微调模型进行训练。

第3类：模型KPI

模型KPIs 包含模型的性能指标及其对通信网络的要求。

□ 在机器学习中，准确率、精确率、召回率和灵敏度等性能指标是衡量模型的关键，也是模型训练的最终目标。

- *准确率* 是通过将正确分类的模型数量除以使用测试集分类的总模型数量来计算的。
- *精准* 和 *召回率* 是两个评估指标，但它们通常一起使用。*精准* 是通过将正确分类的正样本数量除以分类器对所有正样本进行分类的总数量来计算的。*回忆* 是一个分类器正确分类的正样本数量与正样本总数之比。

□ 在移动通信系统中，人工智能/机器学习模型正越来越多地取代传统算法（如语音识别、图像识别和视频处理）在移动设备（如智能手机、汽车和机器人）中，以实现新应用。为了满足人工智能/机器学习需求，6G网络也必须满足一些关键性能指标（KPI）。另一个不可忽视的事实是存在无数种模型和人工智能操作方法。不同的人工智能操作方法和模型大小将对KPI提出不同的要求。例如，在联邦学习架构中，6G网络需要保证一组联邦学习节点的整体QoS，即防止由于这些节点的通信和计算能力不同而导致的低迭代效率。在KPI保证方面，模型分割和分割点调整也可能有效。例如，为了满足不同的KPI要求，我们可以调整传输的中间参数大小

在用户设备（UE）和流媒体服务器（OTT）之间。通常情况下，6G系统将至少支持以下人工智能（AI）/机器学习（ML）操作：

- 节点间分配AI/ML
- 分发和共享 AI/ML 模型与数据
- 分布式/联邦学习

为了支持上述的 AI/ML 操作并确保实时模型训练/推理，通信网络对模型传输、上传和下载有很高的要求。目前，3GPP TS22.261 已经定义了通信 KPI 的要求，包括推理、模型下载以及 UEs 与网络服务器/应用程序之间的联邦学习。

第四类：模型结构

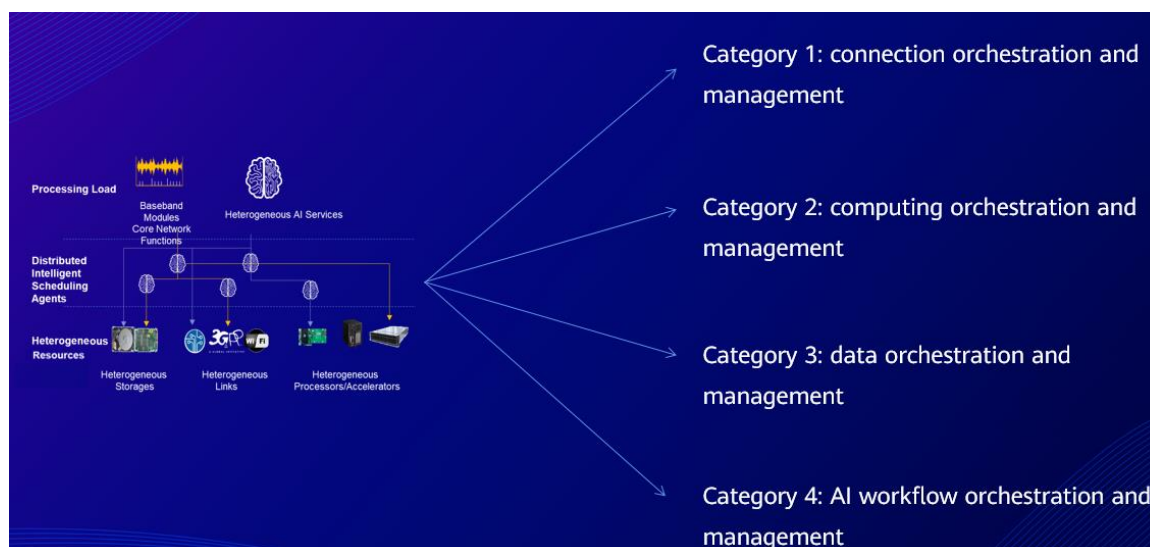
作为机器学习最重要的部分，深度学习涉及以下类型的神经网络模型：多层感知器（MLP）、卷积神经网络（CNN）和循环神经网络（RNN）。MLP是最常见的模型，也称为全连接神经网络——任意两个相邻层之间的节点都是连接的。但是，更多连接意味着更多模型参数，因此需要占用更多的内存和计算资源。

卷积神经网络通常是包含卷积层、池化层和全连接层的向前传播神经网络。它们通过反向传播算法进行训练。卷积神经网络具有三种结构特征：局部连接、权值共享和池化。这些特征使得卷积神经网络在一定程度上对平移、缩放和旋转具有不变性。与向前传播神经网络相比，卷积神经网络的参数更少。

循环神经网络是一种短期记忆网络。在循环神经网络中，一个神经元可以从自身和其他神经元接收信息，形成循环的网络结构。与前馈神经网络相比，循环神经网络在结构上更类似于生物神经网络。循环神经网络已广泛应用于语音识别、语言建模和自然语言生成等任务。

感知人工智能应用中涉及的模型结构被认为是6G网络为人工智能应用提供的最根本的内置支持。通过感知模型的结构，网络可以全面感知人工智能模型，因此能够为人工智能应用提供全面的支持，包括灵活分配资源和协助模型计算。例如，如果模型是MLP模型，网络会分配更多的通信资源给它，因为MLP模型涉及广泛的参数。然而，如果模型是CNN模型，网络只会分配少量的传输资源给它，因为CNN模型的参数要少得多。

3.7 S5 - 配置和管理服务的类别



弹性网络AI部署主要涉及构建相关的编排和管理平台能力，提供按需服务（连接、计算、数据、算法等），并自动化服务部署、测试、管理及运营。网络AI任务主要包括以下类别的编排和管理服务：

类别1：连接编排和管理

对于网络AI任务，连接编排的一个重要目标是自动执行网络请求，最小化应用交付或人工干预，并在满足AI服务的QoS要求的同时，最大化网络资源效率。基于网络能力开放和软件定义网络（SDN）等底层能力，该类别的编排和管理连接。为了实现更高的效率，连接编排和管理必须具备一定的网络感知能力，并能够利用AI算法来保持最佳网络性能。

二级类别：计算编排和管理

计算编排和管理提供最佳的计算资源分配和网络连接。这一类别可以最好地满足网络AI需求，并协调整个网络上的异构资源。一方面，它通过网络分发服务节点的计算和存储信息。另一方面，它感知网络相关信息（例如，路径和延迟）。

在服务本机人工智能网络时，计算编排和管理将涉及边缘的动态、异构和分布式资源；因此，这一类别必须具有以下特征：

- 资源标识：使用统一的资源标识系统，识别来自不同方的各种计算、存储和网络资源。这将促进资源信息的分配和关联。
- 计算感知、建模和评估：评估、测量和建模网络中无处不在的计算资源的状态和分布。这将为基础计算资源发现、交易和调度奠定基础。
- 多方异构资源集成：通过网络控制平面，分发各方的资源信息，并将其与网络资源信息结合，以获得所有资源的全局视图。
- 轻量化：面对网络边缘的动态和复杂环境，利用轻量级资源技术支持实时服务迁移。

类别3：数据编排和管理

□ 分布式数据存储：由于网络模型是去中心化的，网络边缘的大多数人工智能模型采用分布式计算进行任务训练和推理，其数据也分布在不同的存储中。6G网络将致力于以更高的效率和准确性以分布式方式存储模型数据。

□ 数据安全和隐私：如今，数据被视为继土地、劳动力、资本和技术之后的第五大生产要素。在6G时代，边缘产生的数据量将呈数量级增长。大量个人数据将被用于模型训练和提供服务。寻找保障数据安全和隐私的方法是研究者们的重要方向。

类别4：AI 工作流编排和管理

在网络AI训练期间，可能有数百个任务和数千个节点同时运行。因此，有必要将这种复杂的AI服务分解为多个工作流或任务，以降低复杂性。AI工作流编排和管理是一种针对特定环境和服务的精细化编排和管理服务，它能够进行特征提取，

模型训练，模型分割，边缘模型部署，模型推理，和量化感知模型压缩。

4 结论

网络AI涉及跨技术领域的连接、计算、算法和数据的融合与创新。它是在网络架构层面的重大变革。6G网络的范畴将不仅限于连接服务；相反，它将涵盖原生计算、数据、AI以及更多。这将引领通信网络进入一个全新的领域，在那里由于跨技术碰撞而涌现出大量的新思想、概念、术语和服务。本白皮书试图根据行业当前的思考与共识，总结关键部分并解释其定义与内涵。我们还分析了网络AI层级与类别的定义，包括不同类别中服务的描述、潜在服务场景与需求，以及关键研究方向，旨在加速形成行业在网络AI方面的共识。

参考文献

- [1] 6GANA, “6G网络人工智能联盟白皮书”
- [2] 6GANA, “6G网络人工智能联盟：从云人工智能到网络人工智能”
- [3] 华为, “6G, 下一个地平线：从连接人与物到连接智能”
- [4] IMT-2030 (6G) 推广组, “6G 视角和候选技术”
- [5] ITU FG-NET- 2030, “网络2030：迈向2030年及未来的技术、应用和市场驱动因素蓝图”
- [6] IMT-2020, “5G 应用创新发展战略白皮书”
- [7] 六边形-X, “6G愿景、应用场景和关键社会价值”
- [8] 杨, Y, “多层计算网络用于智能物联网,” Nat Electron 2, 4–5 (2019).

缩写和缩略语

3GPP	第三代伙伴关系计划
6GANA	网络AI 6G联盟
人工智能	人工智能
AI4NET	网络AI
AlaaS	人工智能服务
AMC	自适应调制和编码
AR	增强现实
CN	核心网络
CNN	卷积神经网络
CSI	信道状态信息
CT	通信技术
端到端	端到端
EC	边缘计算
FDD	频率分割双工
物联网	物联网
IT	信息技术
KPI	关键绩效指标
LR	学习率
MAC	媒体访问控制
MEC	多访问边缘计算
MIMO	多输入多输出
ML	机器学习
MLP	多层感知机
NAS	非接入层
NE	网络元素
NET4AI	人工智能网络
NMSE	均方误差
NWDAF	网络数据分析功能
运维	运行与维护
OTT	夸张
QoAIS	网络AI的服务质量

RAN 无线接入网
RNN 循环神经网络
RPA 机器人流程自动化
UE 用户设备
VN 虚拟网络
VR 虚拟现实
XR 扩展现实



6GANA

