oppo

意图通信

面向 6G 时代的端网智能体意图交互



前言

随着移动通信网络的每一次代际跃迁，网络连接从“人与人”扩展至“万物智联”；业务场景从语音、视频延伸至自动驾驶、虚拟现实和智能体协作；用户对网络的需求，正在从“连接质量”升级为“综合业务体验”。游戏玩家期待的不是“时延低于 20ms”的技术指标，而是 “流畅不抖动”的游戏体验；高清直播从业者需要的不是“上行带宽高于 3Mbps”的参数配置，而是“画面清晰不卡顿”的播出效果。

在传统通信模式中，端到端的交互基于 3GPP 定义的精密信令体系与标准化流程，虽为跨厂商设备之间的无缝协同奠定了基础，却也付出了一定的产业代价。一项新能力的引入往往需要历经数年的标准讨论与冻结周期，芯片、终端及网络设备各方需投入巨额研发资源，方能完成协议实现与联调。传统模式下，网络只能机械地响应预设的指令参数，却无法感知用户行为背后真实的业务期望与体验诉求。与此同时，终端被淹没于海量网络信令之中，解析这些繁杂指令，不仅推高了芯片和终端厂商的技术准入门槛，更持续消耗着算力与电池寿命。而用户最终所获得的，却依然是一个无法“理解”其真实意图的哑管道。

意图通信正是为打破这一困局而生。

其核心理念简洁而深刻：让用户表达“想要什么”，而非“如何实现” 。在意图驱动的范式下，用户或应用只需以自然语言或结构化方式描述其期望（例如“为这场 4K 直播提供流畅保障”或“为本次游戏对战提供稳定网络服务”），网络自动理解这些需求，并通过意图解析、任务规划和服务调用来满足它们，整个过程无需用户关心底层的信令交互与资源调度。

为实现这一理念，终端和网络需要构建一个全新的“意图层”，同时终端也需要完成从“信令执行者”到“意图表达者”的角色跃迁。在终端侧，用户或应用的原始诉求被转化为标准化的意图结构，通过轻量化的协议通道上行至网络，免去了终端对底层信令的解析负担；在网络侧，智能体对接收到的意图进行解析、验证与计划生成。标准化的意图结构确保网络能够准确“读懂”用户所需，动态的工具调用与服务编排则将高层意图转化为可执行的网络动作，闭环监控与自主调整持续保障意图的履行质量，运营商治理机制则确保整个意图驱动的过程始终处于可控、可信的边界之内。

端侧表达意图，网侧兑现意图。这一端网协同的新范式，正在将连接能力从“可售卖的资源”升级为“可承诺的体验”。对运营商而言，意图通信打开了从“管道提供商”迈向“体验运营商”的战略通道，

当网络能够理解并保障用户的具体业务意图，连接服务便不再只是按流量或时长计费的商品，而是可以按“体验等级”差异化定价的高价值服务，这将催生全新的商业模式与收入增长空间。对设备厂商与芯片厂商而言，终端信令负担的显著减轻意味着开发周期缩短、入网成本降低，新品上市速度得以加快，从而在激烈的市场竞争中赢得先机。对最终用户而言，意图通信意味着用户可以直接表达自己的真实需求而无需理解底层的运行机制。对产业生态而言，意图层作为意图通信的桥梁，为第三方应用提供了开放、易用的交互模式，使得千行百业的创新应用能够便捷地获取差异化的网络服务。

本白皮书旨在系统阐述意图通信的技术体系，包括意图的核心概念、端网智能体意图通信架构、端网智能体示例实现，以及未来挑战。我们期望通过这份白皮书，为产业界提供关于意图通信的完整技术图景，凝聚共识，推动这一变革性理念从标准走向现实。

从“指令驱动”到“意图驱动”，不仅是技术的演进，更是通信网络与人类交互的范式转换。当网络能够理解并主动满足用户的真实意图，通信将不再是一项需要用户去“适应”的技术，而成为一种“自然而然”的服务。这正是 6G 时代赋予我们的机遇与使命。

目录

1、现有通信模式 5

1.1 现有端网交互模式介绍 5

1.2 现有通信模式的限制 7

2、意图通信模式 9

2.1 意图的定义 9

2.2 意图通信模式介绍 13

2.3 意图通信的动机与价值 14

3、端网智能体意图通信架构 16

3.1 整体架构 16

3.2 端侧智能体 17

3.3 网络智能体 21

3.4 端网智能体交互 25

3.5 端网智能体意图通信安全 31

4、端网智能体意图通信示例实现 35

5、意图通信的未来展望 37

5.1 标准演进 37

5.2 商用部署 37

5.3 关键挑战 38

结语 40

1、现有通信模式

1.1 现有端网交互模式介绍

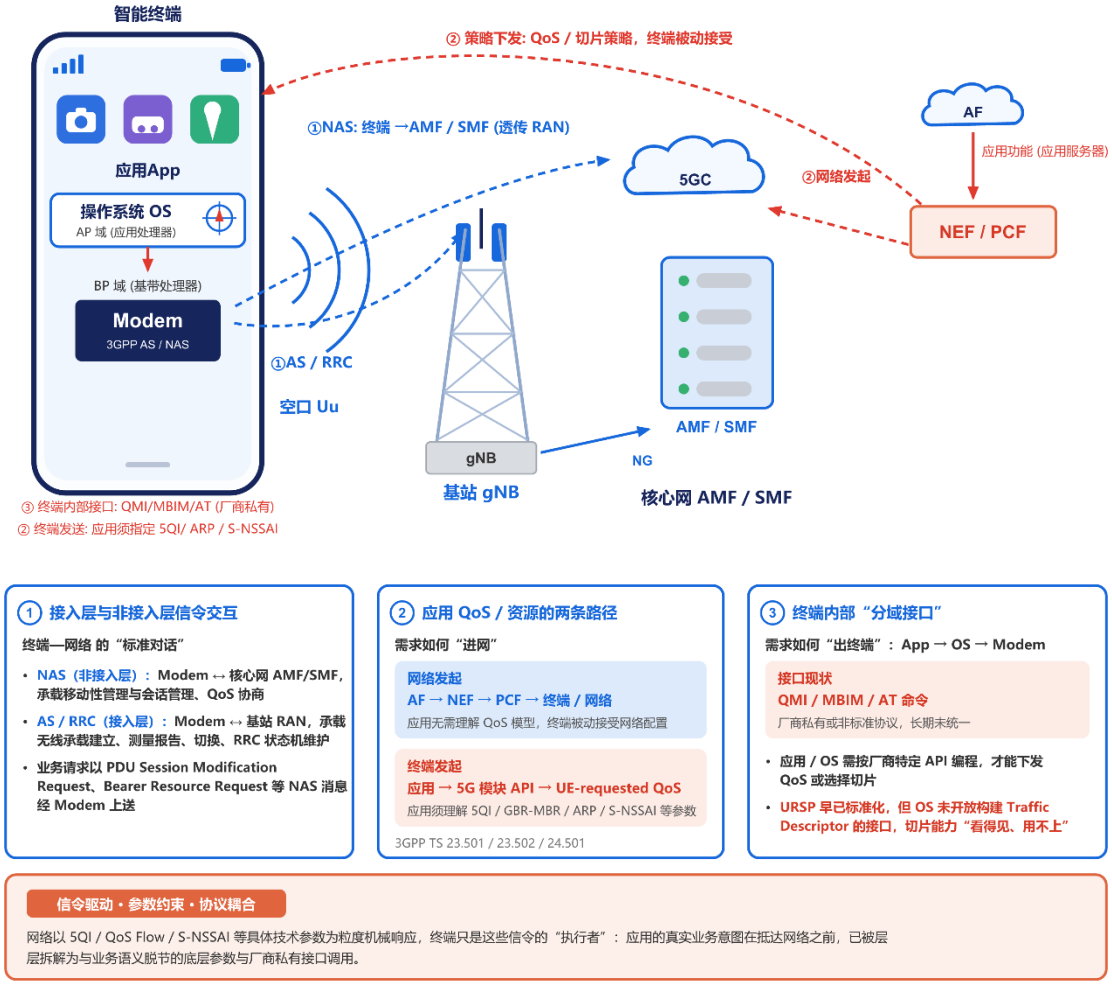


图 1.1-1 现有端网交互模式

现有端网（终端—网络）之间的交互，建立在一套由第三代合作伙伴计划(3rd Generation Partnership Project, 3GPP) 定义的、高度精密且强约定的信令与接口体系之上。终端侧的业务需求，必须被转换为严格定义的信令、参数与流程，方能送达网络并获得满足；网络侧也只能依据预设的标准化流程对终端请求做出响应。从终端应用的原始诉求到达网络，中间实际跨越了「应用层 — 操作系统 (Operating System， OS) — 基带调制解调器（Modem）— 空口 — 基站— 核心网」多个协议层次，其交互模式可归纳为以下三类:

（1）接入层与非接入层的信令交互（终端—网络的“标准对话”）

3GPP 将终端与网络的协议栈分为接入层（AS，Access Stratum）与非接入层（NAS，Non-Access Stratum）。

- 非接入层（NAS）：承载终端与核心网控制面网元（例如 5GC 中的 AMF/SMF）之间的交互，负责例如移动性管理（注册、寻呼、切换）与会话管理（PDU 会话建立/修改/释放、QoS 协商）等功能。终端发出的业务请求，首先以 NAS 消息经 Modem 发送至核心网。
- 接入层（RRC 等）：承载终端与基站之间的交互，负责无线承载建立、测量报告、切换、RRC 状态机维护（IDLE / CONNECTED / INACTIVE）等。核心网授权后的 QoS / 资源，最终由 RAN 通过 RRC 信令配置和空口承载（Data Radio Bearer，DRB）落地。

这套「终端说标准语、网络按标准答」的机制，是跨厂商、跨代际设备能够无缝互操作的根基，但也意味着任何端网交互都被牢牢绑定在协议参数的细粒度上。

（2）应用 QoS / 资源的两条传递路径（应用需求如何“进网”）

终端应用的差异化网络需求，落地为网络保障有两条主要路径：

- 网络发起：由应用功能（Application Function，AF）经网络能力开放网元（Network Exposure Function，NEF）与策略控制网元（Policy Control Function，PCF）将 QoS 策略或切片策略下发给终端与网络（如 VoLTE/IMS、运营商赞助业务）。应用自身无需理解网络的 QoS 模型，终端被动接受网络配置。此类方式多用于运营商可完全掌控的业务。
- 终端发起：当应用需要特定保障（如低时延、大上行），它由终端内的 5G 模块 API 指定 QoS 等级，触发终端向网络发起 QoS 修改或专用承载建立请求。这种方式赋予应用最大灵活性，但要求应用“理解”网络的 QoS 模型（5QI、GBR/MBR、ARP 等具体参数），并由网络依据签约与策略授权。

（3）终端内部的“分域接口”：App → OS → Modem（应用需求如何“出终端”）

智能终端在硬件与软件上通常分为应用处理器（Application Processor，AP，运行操作系统与应用）与基带处理器（Baseband Processor, BP/Modem，运行 3GPP 协议栈）两个域。应用层的业务需求要变成空口上的网络动作，必须先穿越 AP 与 Modem 之间的终端内部接口。这一接口历史上并未完全标准化，不同芯片与终端厂商采用各自私有或部分标准化的控制协议——如高通 QMI（Qualcomm MSM Interface）、MBIM（Mobile Broadband Interface Model，USB-IF 标

准）、以及历史最悠久的 AT 命令。换言之，今天应用的需求要在终端内部先经厂商私有通道翻译成 Modem 能懂的指令，再经 NAS/RRC 变成网络配置。

现有端网交互是一种「信令驱动、参数约束、协议耦合」的范式——网络以具体标准参数为粒度机械响应，终端是这些信令的「执行者」，而应用的真实业务需求在抵达网络之前，已被层层拆解为与业务语义脱节的底层参数与厂商私有接口调用。这正是意图通信所要超越的起点。

1.2 现有通信模式的限制

（1）终端协议栈负担沉重，开发与测试成本高

在传统模式下，终端厂商必须完整实现 AS/NAS 全套 3GPP 协议栈，并持续解析海量与其自身业务无关的网络信令。这带来三重成本：

- 研发与联调成本：芯片、终端厂商需在协议实现、互操作性测试、入网认证上投入巨额资源；任何新特性的引入都要求终端侧同步开发、集成与回归测试。
- 算力与功耗成本：终端需常驻监听与解析信令，持续消耗 CPU、NPU 与电池——而用户既无感知，也未从中获得任何体验收益。
- 碎片化风险：厂商私有实现（QMI / AT 命令扩展等）导致跨厂商、跨版本行为不一致，互通问题频频出现。

（2）终端内部接口割裂，应用适配与能力落地困难

App→Modem 的内部接口缺乏统一标准，应用开发者需针对不同芯片平台（高通 / 联发科 / 海思等）、不同操作系统版本分别适配私有 API；例如，应用/OS 需按厂商特定的 API 编程才能下发 QoS 或选择切片；即便 3GPP 已定义了 UE 路由选择策略（UE Route Selection Policy，URSP，用于把应用业务流映射到 PDU 会话参数，包括切片标识 S-NSSAI、DNN 与接入类型等），主流操作系统长期缺乏对应用暴露传递业务流描述符（Traffic Descriptor， TD）的标准接口，导致切片能力到终端最后一公里迟迟难以落地。网络侧的大量先进能力难以被第三方应用便捷调用，能力开放名难符实。

（3）需求变更受制于标准冻结，产品迭代严重滞后于用户体验演进

现有模式的根本瓶颈在于：任何新业务体验要获得网络保障，都必须先等标准定义出对应参数 → 芯片实现 → 终端集成 → 网络升级，跨越数年。5G 定义的大量先进功能（URLLC、切片、UE 发起 QoS 等）之所以普及缓慢，很大程度上受制于这一漫长而昂贵的开发周期。而当用户的真实意图已从「连接质量」升级为「综合业务体验」——游戏流畅、直播不卡顿、高精度 3D 地图下载、VR 渲染等融合性需求时，网络却依旧只能被动处理单一、预先定义好的指令，无法跟上多样化、个性化用户意图的演进速度。

2、意图通信模式

2.1 意图的定义

2.1.1 用户的意图

用户意图（User Intent）指用户通过自然语言、行为或多模态输入，向操作系统或智能体系统所表达的、希望被理解并满足的真实需求或目的。

（1）用户意图多维输入形态与感知途径

用户表达意图的方式已打破传统的单一文本框或固定控件点击方式，形成了覆盖全场景的多模态交互与上下文感知体系：

- 自然语言输入（Natural Language Inputs）：包括语音对话、文本框输入等直观表达（如“帮我查一下今天的天气”或“订一张明天去上海的机票”）。
- 交互行为与动作输入（Multimodal Behavioral Inputs）：包括屏幕按压力度、圈搜文字或图片、拖拽地址信息、手势、眼神注意力等交互行为（如圈选一段地址直接触发导航意图）。
- 情景与环境状态感知（Contextual & Environmental Perception）：结合用户个人习惯画像、实时地理位置、设备使用状态、历史行为规律等（如根据历史使用习惯，识别出属于主播、游戏等用户标签），被动感知并推导用户意图。

（2）用户意图的表现形态分类

根据表达的直接程度，用户意图通常细分为两类：

- **显性意图（Explicit Intent）**：用户有明确、直白的指令性请求（如“把这张照片发送给张三”）。系统直接进行意图分类与槽位抽取（或 Code 转化）后即可分发执行。
- **隐性意图（Implicit Intent）**：用户未明确下达指令，但系统根据设备多维感知、上下文环境与历史习惯主动识别出的需求（如到达机场自动弹出登机牌，或根据出行时间主动推测行程计划）。

终端智能体在获取用户的显性意图或隐形意图后，判断实现该用户意图是否需要 3GPP 通信网络所提供的服务进行满足。对于与 3GPP 服务相关的用户意图，终端智能体将该用户意图转换为标准化的意图结构或称为结构化的意图。

2.1.2 结构化的意图

在人际交谈中，人的意图可以借助高度灵活的自然语言加以表达，人们通过讨论与澄清，能够理解某一方的具体意图。然而，在通信系统中，若允许完全不受限制的自然语言交互，大语言模型在进行意图理解时容易产生幻觉，无法正确理解请求者的意图以及准确满足其意图。为此，有必要在意图通信模式下，对意图本身进行适度限定。

在自然语言人机交互领域，为使机器能够准确理解人的意图，提示词模板（Prompt Template）是一种被广泛采用的技术。提示词模板是为引导大语言模型完成特定任务而预先设计的结构化提示框架，具有可复用性与参数化特征。其核心构成要素包括角色设定、任务描述、上下文信息、变量占位符、约束条件、示例示范以及输出格式规范；使用者仅需替换相应变量，即可生成适配具体场景的完整提示词。在意图通信系统中，以结构化意图进行终端和网络之间的意图通信，与提示词模版具备类似的设计理念。

结构化的意图可以包括如下的标准化域。

表 2.1.2-1 结构化意图

参数名称	解释
Intent ID	意图唯一标识。贯穿意图的整个生命周期，用于在意图通信流程中唯一定位该意图。
Intent Correlation ID	意图关联标识。用于关联多个相关的意图，帮助意图处理方结合之前的上下文信息来统一理解和处理新的意图。
Issuer	意图发起方。标识是谁提出了该意图。
Intent priority	意图优先级。表示该意图的重要程度。当网络拥塞或资源不足时，网络根据该优先级等级决定优先处理还是拒绝该意图。
Intent type	意图类型。意图的分类，例如该意图为通信连接相关的意图还是业务类相关的意图。
Intent description	意图详细描述。可来源于用户的原始意图，也可用于携带更详尽地补充说明需求，该部分的表达方式不限，可以采用自然语言以及非标准化的参数进行填写，本字段不能覆盖其他字段的语义。
Object	作用对象。意图具体生效的实体，例如特定的 UE ID、APP ID。
Constraint	约束条件。意图生效的限制条件，比如实现意图有效的时间段或空间区域范围。
Target	期望目标。意图希望达到的具体指标，该目标应为可执行、可量化或可验证的目标。

结构化意图的价值主要体现在以下三个层面。

第一，结构化意图仅规定结构格式，而不限定各域内部的具体表达方式。其在具备必要约束的同时保留了灵活性，能够满足用户多样化的意图需求，提升用户意图满足的稳定性与可扩展性。

第二，通过标准化设计，结构化的意图显著降低了意图构建过程中的重复成本，在多任务、多用户以及大规模部署场景中有效提升了意图生成效率。

第三，借助显式约束与格式控制机制，结构化意图增强了意图处理的可控性、可解释性与可评估性，从而有效抑制大模型在意图解析中产生的歧义与幻觉，提高意图解析与满足的确定性。

以下是一些典型的结构化意图的示例。

1、连接保障类意图-游戏/直播

Intent:

Intent ID: intent-1

Issuer: UE ID (e.g., GPSI)

Intent priority: High

Intent type: “通信连接保障意图”

Intent description: 用户想要玩游戏/做直播，检测到当前应用的持续卡顿，需要保障用户当前的游戏/做直播流畅。

Object:

    UE ID

    APP ID

Constraint:

    Duration: xx time – xx time

Target: 与当前业务的 QoE 相比，QoE 提升 20%

2、感知类意图-高清地图

Intent:

Intent ID: intent-2

Issuer: UE ID

Intent priority: Middle

Intent type: “感知意图”

Intent description: 用户想要停车，需要为该用户生成附近一定范围内的车位地图，方便用户寻找停车位。

Object: UE ID

Constraint: 用户当前位置周围 2km。

Target: 地图空间分辨率不低于 0.5m/像素, 地图生成时延不超过 10s。

3、计算类意图-图片渲染

Intent:

Intent ID: intent-3

Issuer: UE ID

Intent priority: High

Intent type: “计算意图”

Intent description: 用户想要渲染一张 Find X10 产品宣传图片。

Object: UE ID

Constraint: Duration: xx time - xx time

Target: 图片分辨率为 2K, 端侧接收图片的时间不超过 10s。

2.2 意图通信模式介绍

意图通信是一种以业务目标而非实现手段为交互对象的端网通信模式。现有模式下，终端向网络提交的是“如何做”，即 5QI 取值、GBR/MBR 门限、S-NSSAI 标识等一组标准化参数；意图通信模式下，终端提交的是“要什么”，即以结构化意图承载的业务目标与约束。例如在一场 4K 直播期间保障画面连续、不卡顿，至于采用何种切片、何种调度策略来达成该目标，由网络自行决策。两种模式的关系如图 2.2-1 所示。

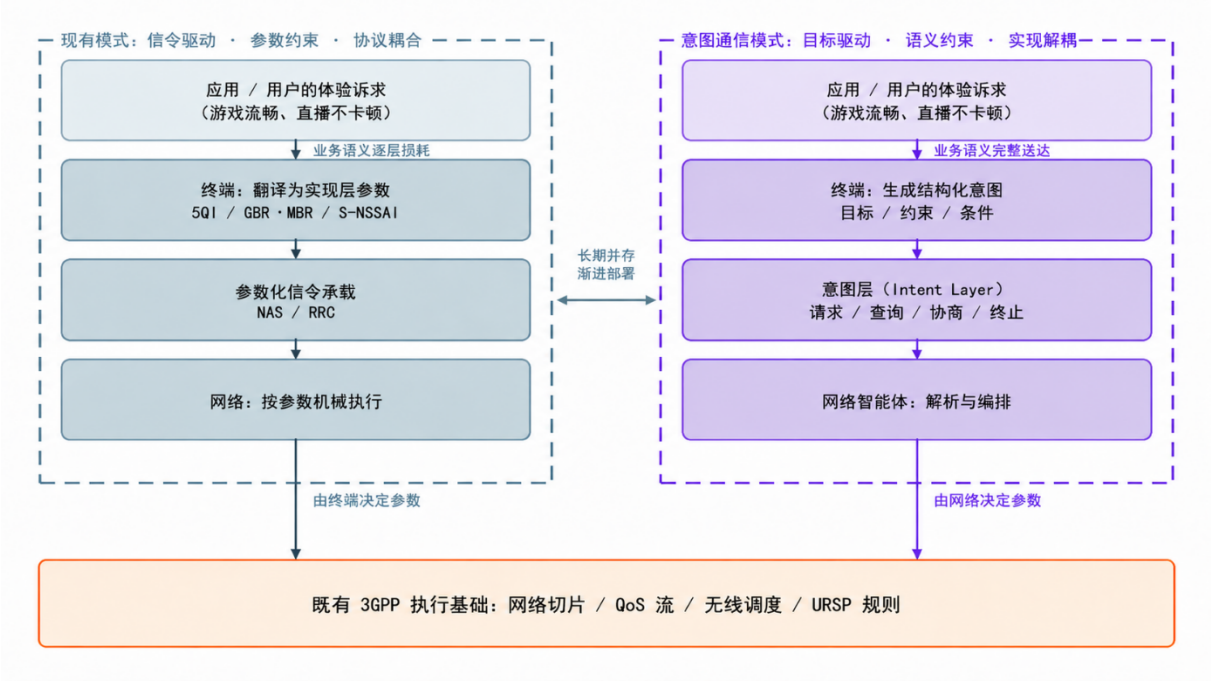


图 2.2-1 现有模式与意图通信模式的关系

意图通信并未取消对实现手段的定义，而是将实现手段的决策权由终端侧移交至网络侧。意图层传递的意图，最终仍需通过 NAS/RRC 信令、QoS 流、网络切片与 URSP 规则等既有机制或新定义的机制满足。因此，意图通信是在现有 3GPP 体系之上叠加一层语义编排能力，而非废除既有的信令体系，可与传统交互方式长期并存、渐进部署。

这一移交带来的直接变化是：终端应用与 OS 层不再需要掌握网络的参数实现逻辑，也不必为每一项新能力同步开发参数映射逻辑；网络接收到的是携带完整业务语义的原始诉求，因而具备在多用户、多业务之间进行全局权衡的依据；端网之间的契约对象，也从参数是否被正确配置，转变为意图目标是否被真实达成。

以实时对战游戏与高清直播推流为例。前者可以由 OS 智能体感知到操作响应抖动加剧，自动生成本局对战期间保障操作响应稳定的意图，网络据此自主选择调度优先级调整或低时延切片接入等手段；后者在弱覆盖场所开播时声明可以接受的画质下限，当网络判断上行资源无法完全满足原始目标时，可反馈一个可承诺的次优目标，再由 OS 智能体结合用户偏好确认或修改。端网之间由此形成意图的双向协商，而不是单向拒绝。

2.3 意图通信的动机与价值

意图通信的提出，有效地解决了 1.2 节中现有模式的弊端，还可以有效避免业务语义在逐层翻译中损耗，网络面对相同标记的数据流无从判断孰轻孰重。同时，运营商从资源计费走向体验计费，需要一个可声明、可核验、可归责的契约载体。因此，与现有模式相比，意图通信在多个方面发生了变化，如表 2.3-1 所示。

表 2.3-1 现有模式与意图通信模式的对比

对比维度	现有模式	意图通信模式
交互对象	5QI、GBR/MBR、S-NSSAI 等实现参数	以结构化意图承载的业务目标与约束
参数决策方	终端翻译并指定	网络解析意图后自主决定
业务语义	逐层拆解，到达网络时已流失	完整送达，可作为全局权衡的依据
能力演进	新功能需新参数，依赖标准、芯片、终端、网络的完整链条	扩展意图取值与网络内部能力即可
责任边界	参数是否被正确配置	目标是否被真实达成
交互形态	单向请求，网络接受或拒绝	双向闭环，支持协商、修改、查询与终止
终端负担	需理解并维护完整的参数与信令流程	仅需准确表达意图，信令解析负担显著下降

产业各方由此获得的收益各有侧重，如表 2.3-2 所示。

表 2.3-2 意图通信对产业各方的价值

参与方	主要收益
用户	无需理解网络术语即可获得与使用场景相匹配的服务保障，交互方式从用户适应网络转向网络适应用户
应用开发者	直接以业务语言声明需求，摆脱对多个芯片平台私有接口的重复适配
终端与芯片厂商	协议栈负担与信令解析开销显著下降，研发资源从协议适配转向体验创新
运营商	网络首次具备理解业务价值的能力，可面向不同场景提供分级的体验承诺
设备商与标准组织	标准工作的重心从穷举参数转向定义语义框架与履约接口，标准资产的复用性提升

上述优势的成立，前提是意图以结构化形式表达。完全自由的自然语言虽然表达灵活，却无法保证网络侧解析的确定性。结构化意图以标准化域约束表达边界，既抑制了解析过程中的歧义与幻觉，又使体验目标可度量、可核验、可留痕，从而支撑意图通信从技术闭环走向商业闭环。

3、端网智能体意图通信架构

3.1 整体架构

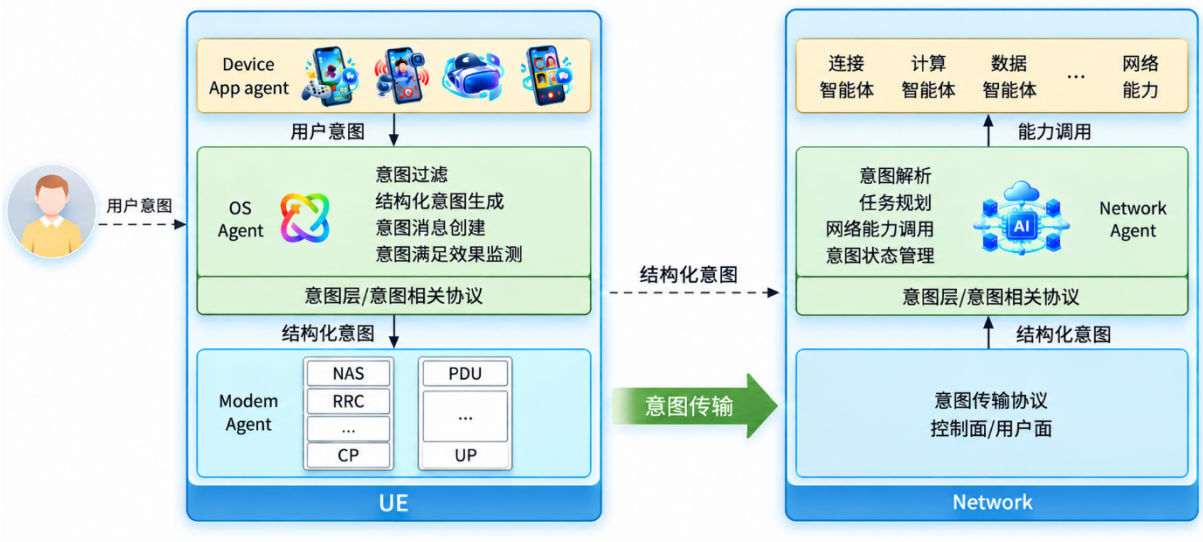


图 3.1-1 意图通信整体架构

终端与网络基于意图进行通信的整体架构如图 3.1-1 所示。从终端角度，用户设备（User Equipment, UE）可划分为 App 层、OS 层与 Modem 层。随着端侧人工智能技术的持续演进及端侧算力的不断提升，终端各层均可通过智能体（Agent）实现能力增强，进而分别演化为应用智能体（App Agent）、操作系统智能体（OS Agent）与基带智能体（Modem Agent）。上述三层智能体协同工作，完成终端侧从用户意图生成到结构化意图转化与发送的全过程。

从终端侧来看，智能体是实现用户意图感知与结构化意图生成的基础。OS 智能体需理解用户以自然语言输入的意图，或与应用智能体协同，实时、准确地感知用户使用应用过程中的体验，在用户无感知的前提下自动实时生成用户意图。OS 智能体对所获取的用户意图进行分析过滤，对于需要 3GPP 通信系统提供保障的意图，OS 智能体生成结构化意图，并将其发送至 Modem。Modem 通过控制面或用户面完成意图的传输。

从网络角度，网络具备更强大的算力，可通过网络智能体对 UE 侧发送的结构化意图进行精确解析，生成任务规划，并确定网络为满足该意图所需调用的能力与服务。鉴于 UE 侧发送的意图具有多样性，网络可部署一个通用智能体与多种专用智能体。通用智能体实现意图解析，任务规划，专用智能体基于专用大模型进行定制化开发，实现特定任务的执行。网络侧多智能体协同配合，可全面满足终端所发送的多样化与复杂化意图。

为支持意图的全生命周期管理，并支持结构化意图生成、智能体能力获取、意图请求、意图协商、意图状态查询、意图终止等功能，需定义意图层（Intent Layer）。该层作为端侧智能体与网络智能体之间端到端的协议层，其语义定义与承载方式解耦，既可承载于 Modem 的控制面协议栈，也可承载于用户面协议栈。

3.2 端侧智能体

3.2.1 操作系统智能体



图 3.2.1-1 OS 智能体架构与意图处理过程

操作系统智能体位于终端操作系统层，是端侧意图链路的起点与中枢。

（1）架构位置

从终端分层看，OS 智能体上接应用层的各类 App 智能体，下接基带层的 Modem 智能体。OS 智能体支持意图层，实现与网络智能体对等交互。它是终端内唯一同时具备应用上下文、设备与网络状态、用户长期偏好以及对外网络接口能力的智能体实体。App 智能体掌握单一业务的局部视图，Modem 智能体掌握传输协议栈与空口的局部视图，而 OS 智能体掌握终端整体视图。这一位置决定了它在端网意图通信中的端侧核心角色。

（2）在意图通信中的角色定位

与传统操作系统仅提供能力调用接口、由应用自行组装网络请求不同，OS 智能体在意图通信中承担五项相互衔接的职责。

第一，意图感知。 用户意图存在两条入口：一是显式入口，用户以自然语言、语音或系统交互直接表达诉求，由 OS 智能体完成语义理解；二是隐式入口，OS 智能体通过应用智能体开放的接口实时感知用户使用应用过程中的业务体验，并结合时间、位置、日程、电量、网络状态等系统上下文，在用户无感知的前提下自动推导出用户意图。两条入口互为补充，使用户意图既可以"说出来"，也可以"被感知到"。

第二，意图判定。OS 智能体对获取到的意图进行分析，判定其中哪些需要由 3GPP 通信系统提供保障。判定依据包括：业务目标是否依赖端到端网络服务，本地资源、本地可执行的手段是否已足以达成目标，网络侧是否已声明支持对应的意图类型（见 3.4.1 节意图能力开放）。判定为不需要网络服务的意图在终端内部闭环消化，不产生端网交互，以避免无效的信令与资源开销。

第三，意图生成与结构化。 对于需要网络服务的意图，OS 智能体依据标准化意图的定义生成结构化意图。明确本次请求的意图类型，对象作用实体，意图保障的起止时间、适用区域与可接受的功耗代价，期望达成的可度量结果。OS 智能体完成结构化意图生成后，将意图下发至 Modem。

第四，用户确认与治理。 意图的生成与发送涉及用户权益，OS 智能体需承担治理职责。涉及套餐变更、增值服务或隐私数据采集的意图，须以用户可理解的方式显式征询同意。对并发的多条意图进行去重、合并与优先级仲裁，避免多头申请与意图冲突。遵循最小必要与本地优先原则，控制上报意图的范围与频度，并保留可审计的操作记录。用户可在任意时刻查询、修改或撤销已发出的意图。

第五，响应解析与生命周期管理。 OS 智能体解析网络侧回复的意图受理、协商、履行与终止结果，并驱动相应的后续动作。当网络因资源约束无法完全满足原始目标而反馈一个可承诺的次优目标时（例如保障 1080p 而非 4K），OS 智能体可结合用户偏好进行确认或修改，形成双向协商而非单向拒绝。OS 智能体在保障周期内按约定节奏查询意图履行状态，在意图到期或业务结束时主动发起终止，从而辅助网络释放意图保障相关资源。最终效果以用户可感知的方式呈现，完成从意图表达到体验兑现的闭环。

上述五项职责在运行态上构成"感知—判定—结构化—闭环"的四阶段链路。第一阶段持续采集应用体验指标与系统上下文，并在场景发生切换时重新评估；第二阶段判定是否需要网络介入，不需要则转入本地优化并继续监测；第三阶段按网络侧下发的意图格式生成结构化意图，经用户确认后下

发 Modem；第四阶段跟踪网络受理与履行结果，按协商结果调整目标，直至意图达成或终止。四个阶段循环往复，使终端侧始终保持"意图—体验"的一致性。

3.2.2 基带智能体

基带智能体位于终端基带芯片层，是端侧意图链路中面向无线接入与 3GPP 协议栈的关键节点。在前述整体架构中，Modem 智能体负责承接已生成的结构化意图，完成其在控制面或用户面上进行可靠传输。此外，Modem 智能体还支持在无线侧提供与意图目标相匹配的协议栈自适应能力。换言之，OS 智能体回答“需要向网络表达什么”，Modem 智能体回答“如何将意图送达网络，并在空口侧支撑该意图的达成”。

（1）架构位置

从终端分层看，Modem 智能体上接 OS 智能体，下接 NAS/RRC/SDAP/PDCP/RLC/MAC/PHY 等协议栈，侧接核心网智能体与无线智能体。结构化意图经 Modem 智能体进入 3GPP 系统，网络履行结果亦由此返回 OS 智能体。

（2）在意图通信中的角色定位

与传统 Modem 主要按 NAS/RRC 等信令被动执行网络配置不同，Modem 智能体在意图通信中承担三类相互衔接的职责。

第一，意图承载与传递。接收 OS 智能体下发的结构化意图并发送至网络，接受网络回传的意图受理、协商、履行与终止等反馈，并上传给 OS 智能体。

第二，协议栈侧的上下文自适应。结合业务约束、信道状况与 Modem 能力，在网络已配置或许可的参数区间内智能调整空口相关行为，使时延、稳定性、能效等 KPI 向意图目标逼近。

第三，与无线侧智能体的协同。在本地优化不足以兑现意图，或意图依赖网络侧资源与策略时，将终端侧意图转化为可被基站理解的交互，并在需要时触发相应的 RRC/MAC 过程，推动空口按需服务业务目标。

（3）双环工作机制（本地行动环 / 网络协同环）

为实现上述职责，Modem 智能体采用“本地行动环 + 网络协同环”的双环架构，两环并行、按需切换，如图 3.2.2-1 所示。

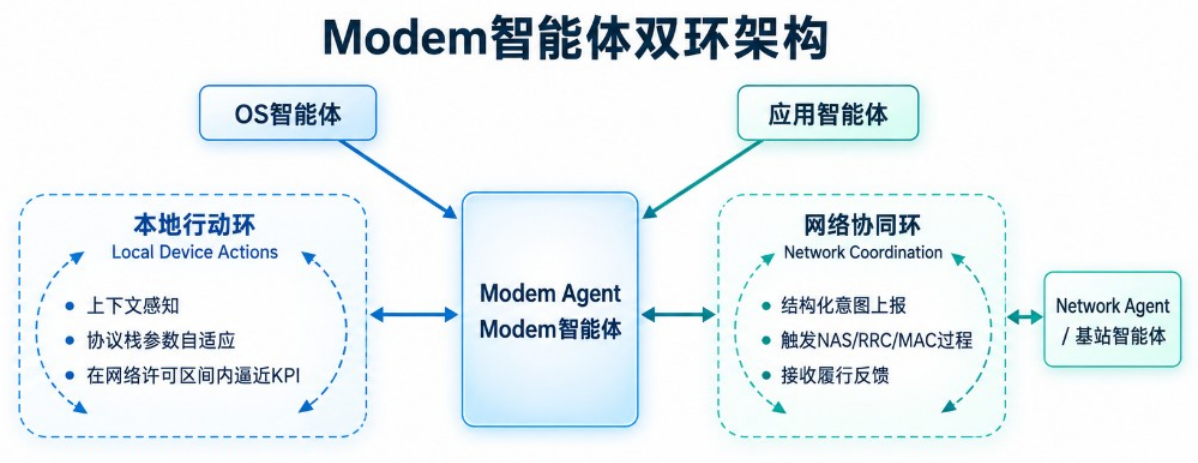


图 3.2.2-1 Modem 智能体双环架构（本地行动环 / 网络协同环）

左侧本地行动环强调上下文感知与协议栈参数自适应，核心原则是在网络许可的参数区间内逼近体验 KPI，而不是绕过网络配置；右侧网络协同环负责结构化意图的上报/查询/协商/终止，必要时触发 NAS/RRC/MAC 过程，并接收满足、降级或拒绝等意图履行反馈。能本地兑现的优先走本地行动环，必须由网络授权或调度配合的走网络协同环。

本地行动环（Local Device Actions）： Modem 智能体持续感知业务时效、无线环境、缓冲与功耗等上下文，在网络配置给出的参数范围内自主优化收发与栈内行为（例如测量与上报节奏、分组处理与缓冲策略、节能态切换等），以更低时延、更高稳定性和更合理能耗逼近意图目标。

网络协同环（Network Coordination）： 当本地优化不足以达成目标，或目标本身依赖网络侧保障时，Modem 智能体与核心网智能体、无线智能体交互，上报结构化意图并触发必要的网络过程；网络返回的履行状态再经 Modem 回传 OS，形成可协商、可降级、可终止的闭环。

（4）与 OS 智能体、网络侧的分工协同

在端侧内部，OS 智能体是意图生成与治理的主体，Modem 智能体是意图出站与无线执行的主体。OS 生成结构化意图后发送至 Modem；Modem 无需重复理解用户自然语言或应用业务语义，而聚焦传输可达性、协议可行性与空口效能。在端网之间，意图请求、查询、修改、协商与终止均可由 OS 驱动、经 Modem 收发。

3.3 网络智能体

3.3.1 核心网智能体

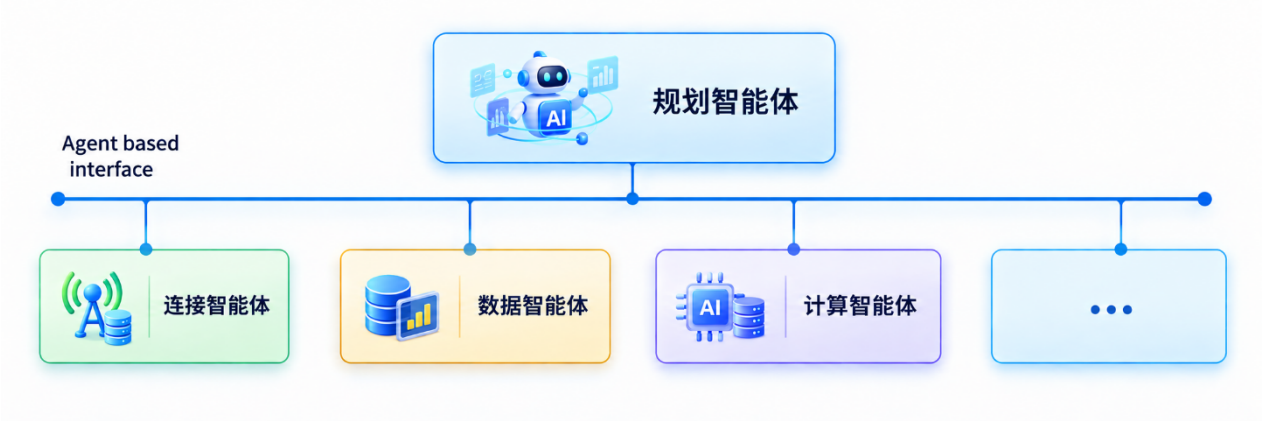


图 3.3.1-1 核心网智能体架构

智能体按照能力划分，可以分为通用智能体和专用智能体。

通用智能体是面向多领域、多类型任务构建的智能体系统，其不局限于单一应用场景，而是能够根据用户意图动态理解任务、分解任务，并自主选择专用智能体或工具完成相应操作。

专用智能体是针对特定领域、业务场景或任务流程进行专业化设计和优化的智能体系统。其通常以通用大模型为基础，并进一步集成领域知识库、专业工具、业务规则、标准化工作流以及特定的评价和约束机制，使智能体能够按照领域专家的方法完成特定任务。相较于通用智能体，专用智能体的任务边界更加明确，在特定领域通常具有更高的专业性、结果一致性和过程可控性。

随着网络业务类型和用户需求不断丰富，用户对网络的需求正在由传统的连接参数配置逐步演进为更加高层、更加复杂的意图表达。例如，“保障游戏流畅运行”“生成高清三维地图”“完成高质量图像渲染”等意图，其背后可能涉及通信连接、数据获取、网络算力等一种或多种网络能力。单一智能体难以同时兼顾任务理解的通用性和任务执行的专业性。因此，为满足多样化、复杂化的用户意图，网络侧可以构建由规划智能体和专用智能体组成的分层智能体体系。其中，规划智能体承担通用智能体的角色，负责意图理解和全局任务编排；专用智能体针对不同类型的网络基础资源和专业能力，负责具体任务的执行。

规划智能体（Planning Agent）主要负责用户意图解析、任务分解以及跨智能体任务编排。当接收到终端发送的意图后，规划智能体首先对意图的目标、约束条件等内容进行分析，并将意图转换

为网络能够理解和执行的任务。对于仅涉及单一网络能力的简单意图，规划智能体可以直接将任务分配给相应的专用智能体；对于涉及多种网络能力的复杂意图，则可以进一步将其拆解为多个相互关联的子任务，选择并调用相应的专用智能体。各专用智能体完成任务后，将任务执行状态和结果反馈给规划智能体，由规划智能体进行结果汇总、关联分析和一致性判断，最终形成面向用户意图的整体满足结果。

专用智能体（Specialized Agent）负责将规划智能体下发的任务进一步转换为具体的网络执行策略。按照网络所提供的基础资源类型，专用智能体可以大致划分为**连接智能体（Connectivity Agent）**、**计算智能体（Computing Agent）**和**数据智能体（Data Agent）**。

连接智能体主要负责处理与通信连接相关的任务和意图。对于云游戏、视频直播、XR、实时会议等业务而言，“游戏运行流畅”“直播不卡顿”“视频会议清晰稳定”等高层体验意图的满足依赖于终端与网络之间通信连接能力的持续保障。连接智能体可以根据业务特征和用户体验目标，结合无线接入、传输和核心网等网络资源状态形成相应的连接保障策略，满足用户连接相关的意图。

数据智能体主要负责处理与数据收集、管理和开放相关的任务和意图。随着网络逐步融合通信、感知和智能能力，数据正在成为网络提供智能服务的重要基础资源。例如，对于“生成一个高清三维地图”的意图，需要获取满足空间范围、分辨率、精度和时效性要求的环境感知数据。数据智能体可以根据任务目标确定所需要的数据类型、数据质量和数据范围，并进一步协调终端、网络节点或其他数据源完成数据采集、聚合、筛选和结果生成，因此，数据智能体承担了从高层的任务和意图向具体数据资源和数据获取任务转换的重要功能。

计算智能体主要负责处理与网络算力相关的任务和意图。随着 AI 推理、图形渲染、和沉浸式业务等计算密集型应用不断发展，与网络算力相关的意图不断增多。网络具备对计算资源进行管理和调度的能力。例如，对于“渲染一幅高质量图像”意图，计算智能体可以根据任务规模、模型类型、端到端时延要求以及网络中可用的 CPU、GPU、NPU 等算力资源，将计算意图转换为具体的计算资源需求，并完成计算节点选择、资源分配，从而利用网络所管理的算力资源完成相应的计算任务和意图。

3.3.2 无线智能体



图 3.3.2-1 无线智能体架构

无线智能体（RAN Agent）是部署在 RAN 域的智能体，负责处理来自端侧智能体、其他 RAN 或核心网的意图，也可反向生成面向端侧智能体、其他 RAN 或核心网的意图。

（1）RAN Agent 所扮演的角色

UE Intent 的接收与解析器：接收 UE 上报的高层业务意图，把 UE 自然语言/高层抽象目标翻译成 RAN 域可理解、可执行的无线层子意图、QoS 目标以及资源约束；

RAN 域意图执行者与分解器：将解析后的 UE 意图拆解为调度、功率控制、移动性、切片以及接入等 RAN 本地控制目标；

UE 与 RAN/核心网之间的意图协商中介：做可行性校验、冲突检测、和端侧智能体双向协商，返回意图满足状态；

闭环监控与反馈代理：持续监测无线 KPI 以及 UE 状态，评估 UE 意图达成度，动态调整资源，向 UE 返回意图执行报告；

跨域协同代理：和核心网侧 Agent、其他 RAN Agent 协同，端到端保障 UE 意图。

（2）UE 意图解析

语义/模型解析：基于 3GPP 意图信息模型，解析 UE 高层目标、约束、优先级以及上下文，区分不可违背目标和尽量满足目标；

意图转译：把业务意图映射为 RAN 域可量化子目标，如目标 BLER、上行/下行时延、速率区间、UE 最大发射功率以及移动性阈值；

上下文补全：结合当前小区负载、干扰、UE 位置、信道质量以及切片资源，补全 UE 意图缺少的无线上下文；

可行性评估：判断当前无线条件下，UE 意图能否满足；如果不能满足，则识别可降级方案。

（3）意图分解与 RAN 本地执行

将解析完成的 UE 子意图下发给 RAN 底层控制模块（调度、功率控制、移动性、切片准入），转化为底层无线配置动作，从而把意图变成具体控制动作。

（4）意图生命周期监控与报告生成

实时采集无线 KPI、UE 状态，持续评估 UE 意图满足程度；

生成意图达成报告，反馈给端侧智能体；

当信道/负载突变，意图无法维持时，触发重协商，通知端侧智能体更新意图。

（5）辅助能力

意图缓存：维护每个 UE 当前生效的意图以及历史意图的上下文；

策略校验：确保解析后的 UE 意图不违反运营商 RAN 策略、切片配额以及安全/隐私规则；

多 Agent 协同：和邻区 RAN Agent、核心网 Agent 协同处理跨小区以及端到端 UE 意图。

3.4 端网智能体交互

3.4.1 端网智能体意图交互流程

一次完整的端网智能体意图交互流程由五个环节构成，形成闭环。分别为意图能力开放、意图请求、意图协商、意图状态查询和意图终止。各环节相互衔接、逐级推进，共同支撑端侧智能体与网络智能体之间意图的高效表达、准确解析、协同协商、状态跟踪与有序终止。

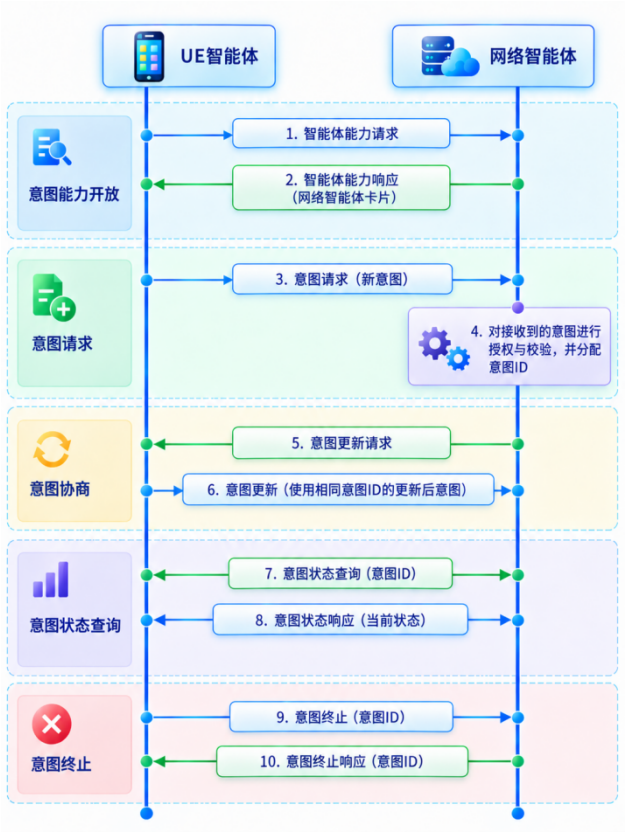


图 3.4.1-1 端网智能体意图交互流程

意图能力开放：在端侧智能体向网络智能体发送意图之前，网络智能体需首先向端侧智能体开放其所能支持的、且已授权端侧智能体可请求的意图类型，以及网络智能体所能支持的结构化意图格式。进一步地，网络智能体还可以明确各结构化域内所支持的语义、参数类型及其取值范围等信息。通过上述能力开放机制，可确保端侧智能体仅在网络能力范围内发送意图，避免不必要的意图发送。通过网络限定范围的方式，保障网络对端侧意图解析和满足的准确性。

意图请求：端侧智能体在接收到网络智能体所发送的、其支持的意图类型信息后，依据自身产生的用户意图，判断该用户意图是否能够通过网络智能体所支持的意图类型予以满足。若可以满足，则端侧智能体根据网络侧所发送的、基于该意图类型的意图格式，生成结构化意图，并向网络智能体发送初始意图请求。

意图协商：网络智能体接收到端侧智能体发送的意图请求后，对意图进行格式检查、授权以及解析。若网络智能体认为所接收到的意图内容不够明确，意图内容可能与其他意图之间引发意图冲突，或当前网络资源无法完全满足意图目标，则网络智能体可与端侧智能体之间开展意图协商。在此过程中，端侧智能体可补充额外信息，或修改该意图内容，从而使网络智能体能够接受并执行该意图；网络智能体亦可反馈一个可承诺的次优目标，由端侧智能体结合用户偏好确认或修改。

意图状态查询：当网络开始执行意图满足时，端侧智能体可向网络实时查询网络对于该意图的执行状态，例如该意图是正在执行中、挂起，还是已完成等。进一步地，端网智能体之间也可以通过意图状态查询，获取端网两侧对于意图保障状态的检测，从而判断意图保障是否达标。由此，端网两侧可以通过意图状态查询，实现对意图保障状态全面、准确的了解。

意图终止：当业务结束或意图保障到期时，端侧智能体向网络智能体发送意图终止请求。网络智能体收到意图终止请求后，终止该意图的保障，并释放相关网络资源，从而完成一个完整的意图生命周期。

3.4.2 端网智能体交互协议

3.4.2.1 智能体协议进展概况

自 2024 年底以来，面向 AI 智能体互操作的协议在开源社区与传统标准化组织两条战线上同步升温，整体呈现**“开源社区快速迭代、事实规范先行；标准化组织稳步吸纳”**的双轨格局。其代表性进展分别如表 3.4.2.1-1、表 3.4.2.1-2 所示。

表 3.4.2.1-1 主流开源智能体协议

协议	发起方与 发布时间	研究重点
MCP ***Model Context Protocol***	Anthropic； 2024 年 11 月	聚焦大模型/智能体与其运行环境（外部资源、数据库、工具）之间的连接接口，解决工具链高度碎片化、每接入一个工具均需定制开发的问题。[1]
A2A ***Agent2Agent***	Google； 2025 年 4 月	面向彼此独立、内部实现不透明的智能体，以 Agent Card 声明能力、以 Task 为协同单元完成通信与互操作，不要求暴露内部状态、记忆或工具链，仅通过外部标准化接口完成智能体间协同。[2]
ANP ***Agent Network Protocol***	ANP 开源社区； 2025 年 2 月	以去中心化身份（Decentralized Identifier，DID）和现有互联网基础设施构建开放协作网络；并采用模块化设计，可按需选取组件或组合协议栈。[3]
AGNTCY	Cisco、 LangChain； 2025 年 3 月	一个集通信协议与互操作框架于一体的开源框架，主要定义智能体注册与能力发现、智能体间通信及基于智能体网关的高效消息分发，提供身份、发现、消息（如 SLIM）与可观测等组件。[4]

与开源社区“快速迭代、事实先行”的路径并行，3GPP、IETF、ITU-T、ETSI 等标准化组织也分别从自身职责边界出发启动了对智能体协议的研究，总体呈现出“不重复造轮子、主动复用开源成果，聚焦架构锚定、信任治理与运营商网络场景落地”的特征。

表 3.4.2.1-2 主要标准化组织智能体协议

标准化组织	研究重点
IETF	**AgentProto BoF**：研究由智能体对话管理、智能体发现、安全等定义良好的模块组合而成的协议框架，优先评估对 HTTP、TLS、OAuth 等既有协议的复用，而非“大而全”的单体协议。[5] **CATALIST（Coordinating Agent To Agent List of efforts）BoF**：协调型 BoF，统筹梳理 IETF 内外各类智能体通信议题（架构、协议、安全、管理、发现），协调各方努力、避免议题重叠与社区碎片化。[6] **DMSC（Dynamic Multi-agent Secured Collaboration）BoF**：面向运营商与企业跨管理域场景，研究以智能体网关为核心的安全协作基础设施，覆盖跨域信任建立、能力注册与发现、策略交换与可观测。[7]
3GPP	3GPP CT 工作组开展面向 6G 的 AI 协议研究 [8]，并结合 6G AI 智能体网络架构新需求 [9]；覆盖 AI 候选协议与设计标准、AI 服务开放、AI 能力/服务发现、授权、调用以及考虑和现有网络的兼容性和互操作性研究。
ETSI	ETSI 下设多个行业规范组和技术委员会，分别专注于网络和系统中人工智能的不同方面。例如，ETSI ISG ENI [10]已发布了多项关于 AI 智能体的成果，涵盖网络切片智能体、移动核心网智能体、多智能体协作框架以及智能体接口和协议。
ITU-T	ITU-T 通过多个 SG 研究组从网络架构、安全信任和应用互操作三个维度进行智能体标准化。例如，SG13 已发布 Y.3193 建议[11]，定义编排智能体与执行智能体双层模型及参考点。SG17 于 2025 年新设 Q16/17，目前拥有 25 项 agentic AI 工作项，覆盖自治智能体安全框架、智能体身份管理与信任框架等。

3.4.2.2 端网智能体协议的潜在设计要素

如图 3.4.2.2-1 所述，MCP、A2A、ANP 等开源协议与 IETF、ITU-T 等标准化组织的在研工作，主要面向互联网的通用智能体交互协议研究，相关机制要在 6G 网络中规模化落地，还需结合电信网络的架构实际进行电信级增强。综合 6G 网络架构的研究进展、对主流开源智能体协议的对比分析，以及 IETF 等标准化组织的同步活动，可将端网智能体交互协议的潜在设计要素归纳为如下几个方面：



图 3.4.2.2-1 端网智能体协议的潜在设计要素

（1）意图处理与管理：支持对意图的高效处理与语义解析，使网络能够准确理解智能体或用户表达的目标与约束，并设计专用的端网意图交互数据结构与消息类型。对应用户的复杂意图，还需要设计意图拆解和意图信息关联等机制。

（2）任务协同：面向长周期、多阶段、高动态的意图通信任务，设计有状态的任务数据结构，需要设计任务从创建、解析、调度、执行、挂起、恢复、降级到完成/失败/取消的状态机设计与触发转换条件，并能够实时记录执行进度、中间结果、上下文信息与异常数据。

（3）多模态数据传输：支持基于意图的多模态数据传输机制，需要建立面向文本、语音、图像、视频等多模态信息的传输通道。智能体交互流量往往由多个不同模态的数据单元构成，不同数据单元在数据体量、传输模式以及时延、可靠性、优先级等 QoS 需求上差异显著。

（4） 智能体身份管理和发现：在 6G 网络中为每个 AI 智能体提供唯一的身份标识，并支持其安全注册，构成智能体之间一切交互的信任前提。目前在开源协议和相关标准化组织中已经有智能体身份的研究：

- A2A：基于 Agent Card 的“自描述+发现”机制。每个智能体以 JSON 格式的 Agent Card 对外声明自身身份、能力、技能、交互模式与服务访问端点，客户端通过获取 Agent Card 完成智能体发现与交互参数配置，实现轻量、松耦合的身份与能力通告。
- ANP： ANP 协议采用去中心化标识 DID，不依赖中心化权威机构，而是基于 DNS、HTTPS、Web 服务器等现有互联网基础设施实现智能体身份的自主掌控、认证与加密通信，契合开放、去中心化的“智能体互联网”愿景。
- IETF：IETF 在研的智能体名称服务 ANS（Agent Name Service）借鉴 DNS 的分层命名与解析思想，提供一套协议中立的注册解析框架，将智能体名称映射到其身份、能力与寻址等元数据，支撑跨平台、跨协议的智能体发现机制 [12]。

总体而言，智能体的命名与发现目前仍由开源协议平台、3GPP 与 IETF 等多方并行探讨，尚无单一标准化组织确立明确归属与统一方案，仍存在进一步收敛的空间。

（5） 协议效率：在满足功能需求的同时，将资源开销与运行性能作为协议设计的内生约束。连接按需建立与释放，通信类型（请求/响应、流式、异步等）与底层传输机制按交互特征匹配。同时协议应原生内建可观测与可追溯能力，使运营商能够掌握协议实际运行状态、支撑风险防控。

（6） AI 交互的容错处理：以“流程可追踪、失败可区分、故障可恢复”为原则，提供过程级关联标识，将同一流程的多次协议交互串联起来，并通过结构化差错信息区分不同失败原因，支撑差异化的诊断、重试与恢复。

3.5 端网智能体意图通信安全

3.5.1 端网智能体意图通信安全核心问题

端网智能体意图通信安全的核心，可以概括为一个基本目标：**如何让网络可信地执行来自可信智能体的可信意图**。围绕这一目标，需要解决三个相互关联的基础问题：

（1） 终端如何产生可被网络验证的意图？

端侧智能体生成的意图不能是网络无法约束的任意自然语言指令，而应具备网络能够识别和验证的结构与语义。

（2）网络如何验证意图来源是否真实可信？

网络不仅需要确认意图来自一个已经完成网络认证的 UE，还需要进一步识别具体由哪个端侧智能体发起，并验证该智能体的身份、可信状态及其所拥有的权限。

（3）网络如何判断该意图是否可以执行？

网络需要判断意图的格式是否合法、语义是否明确，并能够将其映射到网络可理解和可执行的操作对象。并且网络需结合对意图的授权、当前网络状态、以及运营商策略等信息判断该意图是否允许执行。

然而，现有 5G 移动通信网络安全机制在面向端网智能体的意图通信场景中存在如下局限：

一、现有信任体系主要停留在 UE 级，难以识别和认证 UE 内部的多个智能体

现有移动通信网络主要识别“哪个 UE 接入网络”，而无法进一步区分“UE 中的哪个智能体正在使用网络能力”。未来一个 UE 上可能同时运行多个由不同厂商或应用提供商部署的智能体，其能力范围和安全属性可能存在差异。

如果网络仍然只对 UE 整体建立信任，则 UE 内部的恶意智能体可能冒用合法智能体的身份或权限调用网络能力，造成错误计费、业务隔离失效、网络资源滥用，甚至以合法 UE 或合法智能体的身份发起恶意或高风险意图。

二、现有授权机制难以满足意图级动态授权需求

现有网络授权通常围绕 UE 订阅信息、用户签约关系、运营商策略以及适用法规等信息，对服务请求、UE 角色和 UE 能力等进行授权判断。而在意图通信场景中，授权对象进一步细化为“某个智能体在特定上下文中是否可以发起某类意图，以及该意图可以调用哪些网络能力”。端侧智能体的技能、能力以及可信信息还可能由终端厂商或应用提供商提供，传统的基于运营商订阅关系的授权机制难以完整表达这些信息。

与此同时，网络需要根据结构化意图动态完成任务规划、策略生成和资源编排，因此网络必须在执行之前明确判断：该智能体是否有权提出这一意图、该意图是否可以调用相应网络能力、允许使用哪些资源以及可以访问哪些数据。

3.5.2 端网智能体意图通信安全体系

针对上述问题，端网智能体意图通信需要形成覆盖“认证—授权—传输—执行—隐私—审计”的端到端安全体系，主要包括以下六方面能力。



图 3.5.2-1 端网智能体意图通信安全体系

（1）智能体身份管理与认证

建立智能体级身份管理和通用认证机制，使网络能够唯一识别发起意图的 UE AI agent，并验证其身份真实性、可信状态以及身份凭证的有效性。在存在多个智能体的 UE 中，应能够区分不同智能体的安全上下文，防止身份冒用和权限继承。

（2）面向意图的细粒度授权

建立智能体级、意图级的授权管理机制。网络不仅需要判断智能体是否具备访问网络的基本权限，还需要进一步判断其是否有权发起特定类型的意图请求、调用特定网络能力、使用特定资源以及访问特定数据，从而确保意图的生成和执行始终处于明确的授权边界之内。

（3）网络内部安全防线

即使智能体已经通过认证和授权，网络仍应对意图驱动的操作建立内部隔离和约束机制。例如，可通过安全网关、策略控制、执行沙箱以及最小权限等机制限制意图可影响的网络功能和资源范围，避免异常或恶意意图扩散并影响网络整体运行。

（4）智能体意图通信保护

对端网智能体间传输的意图、控制信令及相关业务数据提供机密性保护、完整性保护和抗重放保护，防止意图在传输过程中被窃听、篡改、伪造或重复使用，并确保网络执行的是智能体实际发出的有效意图。

（5）隐私保护

对意图通信过程中涉及的用户和智能体隐私信息实施必要保护，包括长期标识、位置、能力信息、用户偏好以及业务数据等。网络应遵循数据最小化和按需披露原则，仅向完成特定任务所必需的实体提供必要信息，降低敏感信息暴露风险。

（6）意图执行的安全监测与审计

网络需要对意图从接收、验证、解析、授权、规划到执行结果的全过程进行关联追踪和安全监测，并保留必要的审计信息。当出现异常行为、越权操作或执行结果偏离预期时，能够进行检测、阻断、追溯和责任定位，从而实现对智能体行为和网络执行过程的持续安全治理。



图 3.5.2-2 从可信智能体到可信意图、可信执行的安全链条

上述六项能力共同形成从 “可信智能体”到“可信意图”，再到“可信执行”的完整安全链条，并覆盖端网智能体交互全流程。其中，智能体身份认证在意图能力开放前建立信任基础；意图级授权在意图请求与执行阶段明确权限边界；流程中涉及到的端网智能体间的信令与数据交互需实施相应的传输保护与隐私保护，保障交互信息的机密与完整性以及敏感信息的合理使用；安全监测与审计持续覆盖意图请求、规划、执行和终止等关键阶段，从而实现意图通信全过程的安全可控与可追溯。意图通信安全需要与网络智能化能力同步演进，使智能体能够“安全地使用网络能力”，网络能够“可信地执行智能体意图”，从而为端网智能体意图通信的规模化应用提供安全基础。

4、端网智能体意图通信示例实现

基于上述架构与交互流程，我们在商用终端（OPPO Find N6）与运营商商用网络上实现了端网智能体意图通信示例，本示例验证了应用智能体、OS 智能体、Modem 智能体与网络智能体协同完成的意图闭环，覆盖游戏对局与直播开播两类典型业务场景。其中，游戏对局属于时延敏感型业务，直播开播则属于大上行带宽型业务。二者均需依托网络侧高价值资源，方能保障其高质量的用户体验。示例实现功能如下：

- 应用智能体向 OS 智能体提供直播上行码率、游戏链路时延等体验指标；
- OS 智能体完成场景识别，实时多模态感知用户体验的变化，在体验劣化时生成结构化意图，经用户确认后通过 A2A 协议上报给网络智能体；
- 保障期间由 Modem 智能体与 OS 智能体共同完成指标检测并持续核验意图履行效果，业务结束或保障到期后自动向网络智能体发起意图终止；
- 网络智能体受理意图（开始保障、查询保障、终止保障），调用相应网络能力予以满足。

本示例基于对现有开源 A2A 协议的增强，在端侧实现了场景感知、实时多模态用户体验感知，以及根据感知结果实现了结构化意图的生成与交互，以快速验证整体流程的可行性。该方案通过定义网络智能体 Agent Card，实现网络智能体能力的对外开放。本示例将意图传输承载于用户面。

需要说明的是，本演示方案尚不具备 3.4.2 节所述的电信级身份注册、多模态意图传输，以及复杂意图拆解等能力。开源协议若要在 6G 网络中实现规模化落地，仍需结合电信网络架构进行针对性增强。

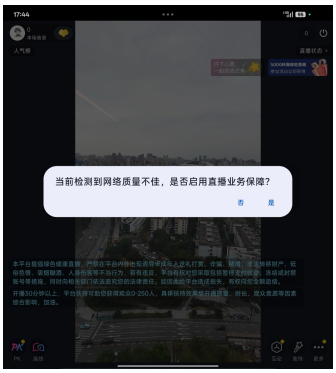
（1）游戏场景

端侧在对局过程中监测链路时延，检测到持续劣化后弹出保障确认提示；用户确认后意图被网络受理并生效，时延由 423ms 降至 43ms；保障到期后意图终止，端侧给出本次保障的最终结果。

（2）直播场景

端侧在开播后监测上行推流码率，识别到码率不足即弹出保障确认提示；用户确认后意图被网络受理并生效，上行带宽由 0.82Mbps 提升至 3.93Mbps；直播结束后意图终止，端侧给出本次保障的最终结果。

表 4-1 游戏与直播场景的示例运行效果

场景	游戏	直播
保障弹窗		
保障成功		
保障结束		

5、意图通信的未来展望

意图通信从理念提出到规模商用，仍需跨越标准定义、技术验证与产业落地之间的多重鸿沟。展望未来，其演进路径可从标准制定、商用部署与关键挑战三个维度加以审视。

5.1 标准演进

意图通信的标准化进程正在 3GPP 框架下系统推进。

在需求定义与方案研究方面，3GPP Rel-20 已正式启动 6G 研究。其中，SA1 工作组围绕 6G 场景用例与业务需求的研究项目已于 2026 年第一季度正式结项，相关成果已写入技术报告 TR 22.870。在此基础上，SA2 工作组在“AI for 6G 架构”关键议题下，已形成涵盖意图处理架构、意图传输、意图结构与描述、意图处理与执行、闭环运营及运营商自治控制等方面的候选解决方案。CT 工作组已同步启动面向 AI Agent 和意图的 Stage 3 协议方案研究，推动意图通信从架构设计走向协议实现。

3GPP Rel-21 作为 6G 的首个正式标准版本，有望完成意图的结构化定义、意图交互、处理流程等关键接口的标准化，为多厂商互操作奠定基础。

5.2 商用部署

意图通信的商用部署不仅是一次技术升级，更是一场围绕“体验价值”的产业重构。

对运营商而言，意图通信推动商业模式从“售卖管道资源”转向“保障体验等级”。运营商可按游戏、直播、工业控制等场景提供差异化体验承诺，并据此分层定价，收入结构从“流量驱动”迈向“价值驱动”。

对终端与芯片厂商而言，终端不再需要深度解析网络信令，只需将用户诉求转化为标准化意图，对 3GPP 协议栈的依赖度显著降低。研发资源可从“协议适配”转向“体验创新”，芯片设计走向轻量化意图处理，功耗与成本进一步优化。终端厂商的角色也从“被动适配者”转向“主动定义者”。

对垂直行业而言，无需理解网络切片、QoS 流等复杂概念，只需以业务语言表达需求，网络即可自动完成资源编排与保障，大幅降低获取高质量网络服务的门槛。

对最终用户而言，网络将根据使用场景自动匹配最优服务保障方案，用户不再需要理解技术术语或在网络间手动切换。通信从“用户适应网络”走向“网络适应用户”。

对应用开发者而言，当网络能力以意图形式开放，应用开发者可直接在代码中声明意图需求，无需关心底层实现，意图层由此成为能力开放与价值创新的枢纽。

5.3 关键挑战

意图通信从愿景走向规模商用，仍面临一系列需要产业界协同攻克的关键课题。这些课题既是当下研究的重点，也勾勒出未来演进的清晰路径。

方向一：意图语义的标准化与一致性。 自然语言与结构化表达均存在理解差异的空间，如何在“表达灵活性”与“语义确定性”之间找到平衡，是 3GPP 标准化工作的核心课题。未来重点在于构建可共享的意图语义模型，确保多厂商对同一意图的理解高度一致，为规模互操作奠定基础。

方向二：多智能体的协同与治理。意图通信架构下，多个网络智能体将同时为相同或不同用户处理意图请求，网络资源的共享特性使多智能体之间的协调成为关键。未来需要在网络侧构建多智能体协作的基础设施，通过规划智能体统筹意图分解与任务协调，并建立智能体间的通信、注册、认证与发现机制；在终端侧探索 OS 智能体与 App 智能体、Modem 智能体的分层协作模式；这一机制的系统性成熟，将推动意图通信从单点智能走向系统智能，从端网分立走向端网协同。

方向三：AI 决策的可解释性与可控性。 意图解析、计划生成与动态调整高度依赖 AI 模型，构建可解释、可审计、可干预的 AI 决策体系，是意图通信从实验室走向生产环境的必经之路。运营商需要能够理解 AI 决策依据并在必要时人工介入，这一能力的标准化与产品化将是未来重点投入方向。

方向四：安全与隐私保护。 意图可能携带用户敏感信息，其传输与处理面临安全风险。构建涵盖意图加密传输、合法性校验、行为审计与异常检测的端到端安全防护体系，是意图通信赢得用户信任、实现规模商用的前提。

方向五：商业模式的验证与生态培育。 技术可行性与商业可行性之间仍需验证。产业界需要找到更多具有说服力的早期商用场景，通过试点验证意图通信的商业价值，带动多方跟进部署，加入意图生态。

方向六：标准与产业的节奏协同。 标准制定需要兼顾技术成熟度与产业期待，商用部署也不能坐等标准完全冻结。如何在标准进展与产业预商用之间找到可行节奏，考验着标准组织、芯片商、终端商、网络设备商与运营商之间的协同智慧。这一课题的推进，将为意图通信的规模化落地铺平道路。

结语

意图通信正沿着“标准定义规则、产业验证可行、商用兑现价值”的路径稳步前行。从“传数据”到“传意图”的理念跃迁，到“体验可承诺、服务可差异、价值可度量”的商业愿景。意图通信正在将通信网络从被动的数据传输管道，重塑为能主动理解并满足用户真实意图的智能服务系统。

这一转型的成功，不仅取决于技术在标准制定与产品实现层面的成熟，更取决于产业界能否协同跨越语义歧义、智能体协作、AI 可解释性、安全隐私、商业模式验证与标准产业协同等六道鸿沟。唯有如此，意图通信才能真正从技术构想走向规模商用，将通信产业从“连接驱动”推向“价值驱动”的新时代。

参考文献：

[1] Anthropic, Model Context Protocol (MCP) Documentation: https://modelcontextprotocol.io/

[2] A2A Community / Linux Foundation, A2A Protocol Specification (latest): https://a2a-protocol.org/latest/

[3] ANP Open Source Community, AgentNetworkProtocol (ANP), GitHub README: https://github.com/agent-network-protocol/AgentNetworkProtocol/blob/main/README.md

[4] AGNTCY, Internet of Agents: https://agntcy.org/

[5] IETF, Minutes of the Agent Communication Protocols (AgentProto) BoF, IETF 126, 2026-07: https://datatracker.ietf.org/doc/minutes-126-agentproto-202607230700

[6] IETF, Coordinating Agent To Agent List of efforts (CATALIST), BoF Request, IETF Datatracker: https://datatracker.ietf.org/doc/bofreq-farrel-coordinating-agent-to-agent-list-of-efforts-catalist/

[7] IETF, Dynamic Multi-agent Secured Collaboration (DMSC), BoF Request, IETF Datatracker: https://datatracker.ietf.org/doc/bofreq-wang-dynamic-multi-agent-secured-collaboration-dmsc/

[8] 3GPP, TR 29.832, Study on the Protocol for Artificial Intelligence in 6G (Rel-20, CT3): https://portal.3gpp.org/desktopmodules/Specifications/SpecificationDetails.aspx?specificationId=5548

[9] 3GPP, TR 23.801-01, Study on Architecture for 6G System; Stage 2 (Rel-20, SA2): https://portal.3gpp.org/desktopmodules/Specifications/SpecificationDetails.aspx?specificationId=4423

[10] ETSI, ISG ENI (Experiential Networked Intelligence): https://www.etsi.org/technical-groups/eni/

[11] ITU-T Y.3193 (04/2026), Distributed intelligence collaboration in IMT-2020 and beyond: https://www.itu.int/epublications/en/publication/itu-t-y-3193-2026-04-distributed-intelligence-collaboration-in-imt-2020-and-beyond

[12] IETF, "Agent Name Service (ANS): A Universal Directory for Secure AI Agent Discovery and Interoperability", Internet-Draft draft-narajala-ans-00. https://datatracker.ietf.org/doc/html/draft-narajala-ans-00