



下一代移动通信核心网与AI协同演进方案研究

宋旭光¹, 孙逊¹, 邱巍¹, 马洪源², 周维¹, 冯志强¹, 卢友伟¹

(1. 中国移动通信集团设计院有限公司, 北京 100080;

2. 中国移动通信集团有限公司, 北京 100032)

摘要: 随着5G-A技术规模商用与6G技术的预研推进, 核心网对实时性、灵活性与智能决策能力的需求显著提升, 但现有5G核心网的AI融合模式存在明显局限。围绕“核心网向AI原生架构演进”核心目标, 首先系统剖析现有核心网与AI融合模式的技术瓶颈, 明确“协同不足、耦合严重、数据低效”三大核心问题; 其次介绍了AI智能体、算网协同、统一数据服务等三大关键支撑技术; 最后设计“探索期兼容部署、演进期重构融合、原生期AI内生”三阶段演进方案, 逐步实现从“5G+AI外挂”到“6G原生AI Core”的范式升级, 可支撑运营商在5G-Advanced向6G演进过程中, 逐步构建具备“自感知、自决策、自优化、自修复”能力的智能核心网, 为通感一体、车联网、沉浸式交互等场景提供技术支撑, 推动“万物智联”愿景落地。

关键词: 核心网AI协同; AI智能体; 算网协同; 统一数据服务; 分阶段演进

中图分类号: TN915

文献标志码: A

doi: 10.11959/j.issn.1000-0801.2025281

0 引言

在全球数字化转型与“万物智联”愿景的驱动下, 移动通信核心网正经历从“功能驱动架构”向“智能驱动架构”的范式跃迁。随着5G-Advanced (5G-A) 技术进入规模商用阶段, 其在工业互联网、车联网、沉浸式交互等领域的应用逐步深化, 而6G技术预研已聚焦“空天地一体化”“通感算智融合”等核心方向, 这使得AI技术不仅是核心网的“辅助工具”, 而是预期成为突破性能瓶颈、催生新型服务的“核心引擎”。

从业务需求维度看, toC消费级用户与toB行业级用户的AI化需求呈现“高实时性、强动态性、广适配性”特征。对于消费级场景, AI多模态交互业务需实现近实时响应(如语音-图像联合交互), 以确保多类型信息传递的自然流畅性;

具身智能终端(如家用服务机器人、AR眼镜)需快速感知环境变化并调整通信策略, 避免因时延导致的交互卡顿。对于行业级场景, 工业自动导引车(AGV)依赖确定性通信支撑高精度生产调度, 需避免因网络抖动导致的生产误差; 通感融合业务需整合无线通信(radio)、图像(camera)、激光雷达(LiDAR)等多源数据, 通过协同处理提供全域环境感知能力。这些需求场景均对核心网的实时性、灵活性与智能决策能力提出更高的要求。

为应对上述新需求, 产业界已形成核心网与AI深度融合的演进共识。3GPP在SA2工作组启动“AI Agent通信协议(A2A/MCP)”标准化立项, 明确AI智能体与核心网功能的交互规范; 同时推动NWDAF功能增强, 新增跨域数据分析与能力开放接口。设备商层面, 华为提出



“Agentic Core 架构”，将 AI Agent 作为核心网功能的“原生组件”，实现任务驱动的动态协同；中信科则倡导“算网协同+AI 内生”技术路线，将算力资源纳入核心网统一管控，通过 AI 优化算力与网络资源的联合调度。这些探索共同指向“核心网与 AI 从‘松散协同’走向‘深度内生’”的发展方向。

1 5G 核心网与 AI 融合的现状与局限

当前 5G 核心网存在 3 种引入 AI 的主流模式，分别为：外挂式 AI 模式、网元级 AI 增强模式和初步协同模式，这 3 种模式均存在明显技术局限，难以支撑核心网与 AI 深度协同需求。

外挂式 AI 模式独立部署 AI 模块（如故障分析服务器），依托核心网标准化接口获取网元数据，虽能快速落地故障分类、流量预测、告警压缩等单点功能，但其与核心网的协同深度存在显著局限：一方面，AI 模块无法实时获取网元内部关键状态（如 SMF 的会话资源占用率），仅能基于接口传输的有限数据生成决策（如故障根因识别），导致决策存在滞后性；另一方面，该模式缺乏闭环执行能力，例如，故障处理场景中，AI 模块输出根因后需要人工触发网元配置调整，无法自动联动核心网完成故障修复，且功能覆盖局限于运维领域，无法支撑动态 QoS 调整、算力按需调度等业务层面的智能化需求。

网元级 AI 增强模式通过在 SMF、UPF 等传统核心网元内部嵌入 AI 能力实现局部智能。例如，针对 UPF，重点增加 AI 业务流处理能力，通过嵌入 AI 驱动的业务流分类算法及相关智能硬件配置，实时识别高清视频、工业控制指令等业务类型并匹配差异化转发策略，或通过 AI 流量清洗算法过滤攻击报文，保障数据传输安全。但此模式存在两大核心局限：一是 AI 能力与网元深度耦合，AI 能力迭代往往伴随着网元功能升级，开发与维护成本高；二是不同网元的 AI 能力缺乏协同

机制，无法形成全局智能决策。

初步协同模式以 NWDAF（网络数据分析功能）为核心载体，通过对接自有或第三方 AI 模型，依托 NWDAF 采集多网元数据并输入模型进行分析，尝试突破单网元 AI 的局限，实现跨网元层面的智能决策。但该模式仍受多重制约，一方面，NWDAF 与各网元的数据交互流程固化，无法灵活适配不同 AI 模型的数据需求；另一方面，第三方 AI 模型与核心网元的交互缺乏标准化接口，稳定性与兼容性难以保障。

2 协同演进的关键支撑技术

2.1 AI 智能体技术

在下一代移动通信核心网从“5G+AI”向“原生 AI Core”演进的过程中，AI 智能体（AI Agent）作为具备自主感知、决策与协同能力的核心载体，打破了传统网元级 AI、外挂式 AI 的功能局限，通过标准化协议与分层能力体系，实现核心网“跨域协同、意图驱动、闭环自治”的智能需求。

2.1.1 定义与标准化内涵

AI 智能体并非传统意义上的“算法模块”，而是在通信网络场景下被赋予明确目标导向的“自动化智能实体”。从 3GPP SA1 的定义来看，AI 智能体是“能够与核心网环境（含网元、终端、业务系统）交互获取上下文信息，通过推理、自学完成决策，并自主或协同执行任务以达成预设目标的自动化实体”，核心强调“环境交互—自主决策—目标达成”的闭环能力；ISO-IEC 则进一步细化其技术属性，明确 AI 智能体须具备“感知外部状态变化、动态调整执行策略、兼容多类型接口”的基础能力，且需满足通信网络对“低时延、高可靠”的刚性要求。

在核心网场景中，上述定义的落地体现为两大核心导向：一是“网络适配性”，即 AI 智能体需兼容核心网标准接口（如 N4、N15）与协议栈

(如HTTP、QUIC), 避免改造现有网元架构; 二是“业务针对性”, 需要针对toC消费级业务、toB行业级业务的差异化需求, 动态调整感知维度与决策逻辑, 而非通用化智能实体。

2.1.2 核心技术特征

相较于传统AI模块, AI智能体的核心优势体现在“场景泛化、协同能力、自我演进”三大维度, 面向核心网场景的典型AI Agent架构示意图如图1所示。

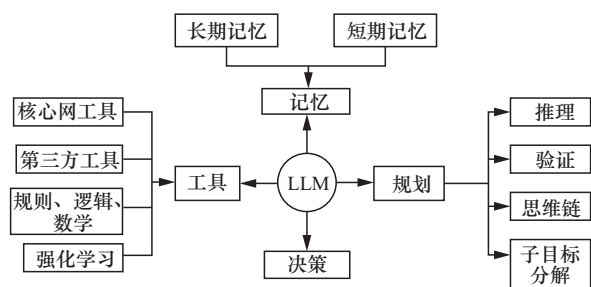


图1 面向核心网场景的典型AI Agent架构示意图

(1) 角色泛化与场景自适应: AI智能体具备“一核多角色”的适配能力, 可根据核心网业务场景动态切换功能定位。在2C场景中, 可作为“体验保障智能体”, 基于用户终端类型、业务偏好, 实时调整QoS参数; 在2B场景中, 可转化为“行业任务智能体”, 例如, 工业场景下的“AGV通信智能体”, 通过感知AGV运动轨迹、车间网络负载, 协同SMF生成专属会话策略, 保障控制指令传输可靠性。这种泛化能力的核心在于“能力包模块化”, 将核心网常用功能封装为可插拔的能力包, 智能体可按需加载, 无需重构底层逻辑。

(2) 工具调用与跨域协同: AI智能体通过标准化协议实现“工具解耦+多智能体协作”, 突破传统AI的单点局限。在工具调用方面, 可依托MCP协议(model context protocol)对接核心网工具与第三方工具, 例如, “算力调度智能体”通过MCP协议向边缘MEC发送算力请求, 获取10TOPS算力资源支撑XR渲染业务。在跨智能体

协同方面, 基于A2A(agent-to-agent)协议, 不同功能的智能体可实现能力发现、安全交互与任务拆解, 例如, “多模态交互智能体”接收用户“高铁会议AI协同保障”意图后, 通过A2A协议联动“接入智能体”完成身份认证、“切片智能体”创建会议专属切片、“体验监测智能体”实时监测业务传输时延, 形成跨域协同闭环。

(3) 记忆管理与自我演进: 一是记忆分层管理。短期记忆存储当前任务上下文(如用户实时意图、网元临时状态), 例如, “故障处理智能体”的短期记忆记录当前告警的网元ID、时间戳; 长期记忆沉淀历史数据(如用户偏好、故障处理经验), 如存储过去6个月某区域的拥塞规律, 为后续调度提供依据; 二是在线自我演进。基于实时业务反馈动态迭代模型, 例如, “QoS优化智能体”若监测到用户AI通话卡顿, 会自动分析卡顿原因, 调整QoS参数生成逻辑, 后续同类场景的策略适配准确率将明显提升; 同时, 通过能力包更新机制, 可快速融入新功能, 无需重新部署智能体。

结合核心网业务复杂度与AI自主化程度, 可以将AI智能体的能力划分为五级, 不同级别对应核心网不同演进阶段的需求, AI智能体能力分级与核心网应用场景对应表见表1。

2.1.3 端到端工作机制

以核心网典型的“AI多模态交互业务”为例, 假如用户发起“AR虚拟试衣+实时AI修图”的业务请求, AI智能体的端到端工作机制可拆解为“感知-分解-执行-反馈”四步, AI智能体端到端工作机制如图2所示。

工作机制具体流程如下。

(1) 意图感知与环境采集: “多模态交互智能体”接收用户多模态输入, 例如, “帮我AR试衣并修图”语音+终端位置信息, 同时通过核心网统一数据面采集环境数据, 包括当前网络负载(如SMF会话资源占用率)、边缘算力状态(如



表1 AI智能体能力分级与核心网应用场景对应表

能力级别	级别名称	核心能力描述	核心网典型应用场景	自主化程度
L1	简单任务执行	基于预设流程执行单一任务，无自主决策能力，需要人工触发输入参数	固定网元的故障告警上报（如UPF端口异常通知）	人工驱动（10%）
L2	任务闭环处理	可自主分析任务需求、执行流程，并验证结果有效性，无需人工干预单步骤	网络流量统计与基础调度（如闲时带宽压缩）	半自主（40%）
L3	用户意图解析	理解模糊用户意图（如保障家庭智能设备通信），拆解子任务并规划执行路径	动态QoS策略生成（如为家庭IoT设备动态分配带宽）	意图驱动（70%）
L4	记忆-意图交互	结合长期记忆（用户历史偏好）理解意图，主动发现潜在需求	个性化网络服务推荐（如通勤时段自动保障语音）	主动交互（90%）
L5	全域自主决策	无需人工干预，自主识别目标、协同跨域资源，应对复杂不确定性场景	无人集群自主组网+故障自愈（如无人机通信恢复）	完全自主（100%）

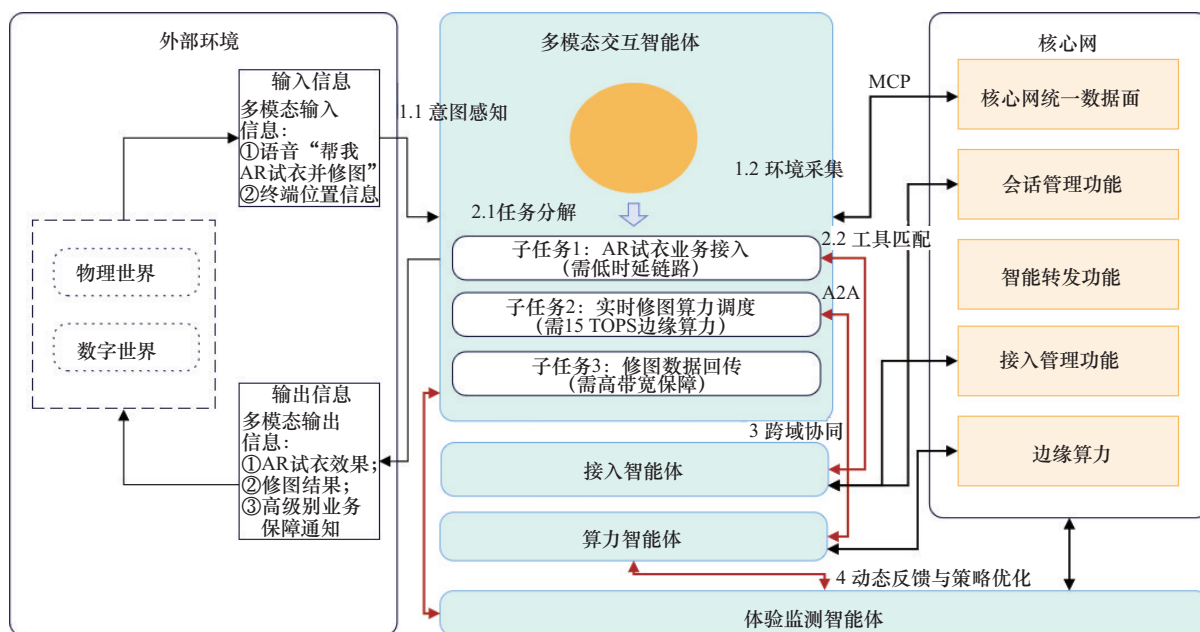


图2 AI智能体端到端工作机制

MEC节点GPU利用率）和终端能力（如AR设备支持的码率），形成完整的任务上下文。

（2）任务分解与工具匹配：智能体通过内置LLM（核心网轻量化大模型）将用户意图拆解为3个子任务：①AR试衣业务接入（需低时延链路）；②实时修图算力调度（需15 TOPS边缘算力）；③修图数据回传（需高带宽保障）。随后，基于A2A协议匹配对应工具，对接“接入智能体”处理AR业务认证；对接“算力智能体”调用算力调度工具申请边缘GPU资源。基于MCP协议联动核心网智能转发功能配置高优先级转发规则。

（3）跨域协同执行：“多模态交互智能体”通过A2A协议发起协同。①“接入智能体”完成用户身份认证后，向核心网会话管理功能请求创建AR业务专属会话；②“算力智能体”通过MCP协议调用算力调度功能分配边缘MEC节点；③“体验监测智能体”实时监测AR画面时延与修图响应速度，并将数据同步至其他智能体。

（4）动态反馈与策略优化：若“体验监测智能体”发现修图响应延迟过高，立即通过A2A协议通知“算力智能体”切换备用MEC节点，同时核心网智能转发功能调整数据转发路径；任务完成后，智能体将用户体验评分（如“修图流

畅”)存入长期记忆,优化后续同类业务的算力调度优先级,实现能力迭代。

2.2 算网协同技术

算网协同技术作为核心网与AI协同演进的关键支撑,其核心价值在于破解传统核心网“算力资源与网络资源割裂管控”的痛点,通过AI驱动实现两者的动态联合优化,构建“算力按需调度、网络随算适配”的一体化资源管控体系。该技术的实现依赖任务控制功能、计算控制功能两大核心组件。

任务控制功能的核心职责是实现AI业务“双重需求”的精准识别与标准化转化,既要感知业务的算力诉求,也要明确网络传输的性能指标,为后续资源匹配提供清晰依据。其关键能力体现在两方面,一是多维度需求拆解,针对不同AI业务场景,拆解出算力与网络的具体参数。在算力维度,须明确推理模型参数规模、算力类型、算力时延阈值;在网络维度,须定义带宽需求、丢包率上限、端到端传输时延等指标。二是标准化格式转化,将拆解后的非结构化需求(例如“保障AR手术指导的实时性”)转化为机器可识别的标准化任务描述格式,包含需求ID、算力参数集、网络QoS指标、业务优先级等字段,为计算控制功能与网络侧提供统一的适配基准。

计算控制功能作为算力资源的“总调度台”,承担全场景算力资源的统一管控与动态适配职责,覆盖终端、边缘、云端三级算力节点,确保算力资源与业务需求的精准匹配。其核心技术能力包括两方面,一是全场景算力池化管理:整合终端算力、边缘算力、云端算力,构建全域算力资源池,并通过实时监测模块获取各节点的运行状态,包括CPU/GPU利用率、内存占用率、算力空闲量及网络可达性。二是AI驱动的动态调度:基于任务控制功能输出的标准化需求,结合算力池实时状态,通过强化学习算法选择最优算力节点。例如,对低时延AI业务,优先匹配终端

或边缘算力以缩短传输距离;对大算力需求业务,则调度云端GPU集群实现规模化算力供给。同时,计算控制功能支持算力资源的弹性扩缩容:当AI业务流量突发时,可自动唤醒备用算力节点并完成资源分配;当业务流量下降时,及时释放冗余算力以降低能耗,实现算力资源的高效利用。

2.3 统一数据服务

作为核心网AI模型训练与推理的关键“数据底座”,统一数据服务框架通过打破多源数据孤岛、优化数据全生命周期处理流程,为AI算法提供高质量、高可用、高安全的数据支撑,从根本上解决传统核心网中“数据分散存储、格式异构、质量参差”导致的AI数据供给瓶颈,是实现核心网与AI深度融合的基础支撑技术。

2.3.1 分布式数据采集与整合

通过“网元级数据代理+标准化接口+全域数据湖”的三层架构,实现核心网多源数据的高效汇聚与统一管理,解决传统“点对点”数据交互导致的格式混乱、获取低效问题。

(1) 网元级数据代理部署。在核心网网络功能侧部署轻量化数据代理模块,实时采集三类关键数据:①用户行为数据,如终端接入时间、业务类型偏好、位置切换记录等;②网络状态数据,如网元CPU/内存负载、链路丢包率、会话建立成功率等;③业务运行数据,如视频业务码率波动、AI多模态交互日志、工业控制指令传输时延等,确保数据采集的全面性与实时性。

(2) 标准化数据整合。将各数据代理采集的异构数据进行统一标准化格式转换,随后将标准化数据加载至分布式数据湖,实现“一次采集、全域共享”,避免多网元数据重复存储与格式适配成本。

(3) 动态数据订阅机制。支持AI模型根据需求动态订阅数据,可通过标准化API发起订阅请求,数据湖直接推送目标数据,无须与核心网网



元单独对接，大幅缩短数据获取时延。

2.3.2 非结构化数据处理

针对核心网 AI 业务（如通感融合、智能交通）产生的图像、激光雷达点云、语音交互记录等非结构化数据，通过“特征提取-向量存储-快速查询”的技术路径，解决传统结构化数据库处理效率低的问题，为 AI 多模态推理提供高效数据支撑。

3 核心网与 AI 协同演进的分阶段方案

核心网与 AI 的深度融合需遵循“从局部适配到全域原生”的递进逻辑，紧密衔接 5G 向 6G 的演进节奏，形成探索期、演进期、原生期 3 个层层递进的发展阶段。各阶段既聚焦解决当前技术瓶颈，又为下一阶段的架构升级铺垫基础，最终实现从“5G+AI 外挂辅助”到“6G 原生 AI Core 自主运行”的范式跨越，支撑“万物智联”场景的全维度需求。

3.1 探索期：兼容式部署，突破单点智能瓶颈

探索期的核心诉求是“在不中断现有 5G 核心网业务的前提下，快速验证 AI 价值”。这一阶段需平衡“智能化突破”与“网络兼容性”的矛盾，避免过度改造现有架构导致的业务风险，同时通过局部智能解决运维效率低、业务体验波动等显性问题。

具体而言，技术落地围绕“轻量化嵌入”展开。一方面，在 AMF、SMF 等传统网元内部嵌入低复杂度 AI 算法，无须重构网元核心逻辑，例如，为 AMF 配置用户接入负载预测模型，通过历史接入数据预判早晚高峰时段的用户规模，动态调整接入资源分配策略；为 SMF 部署会话资源调度算法，结合实时会话请求量与网络负载，优化会话建立优先级，提升单网元的资源利用率。另一方面，以独立虚拟机或容器实例部署外挂式 AI Agent，通过核心网标准接口对接现有网元，实现特定场景的智能化覆盖。例如，“运维

Agent”通过 NWDAF 获取多网元告警数据，自动完成故障分类与根因初步定位，大大缩短故障响应时间；“体验保障 Agent”实时监测 XR、AI 多模态等业务的 QoS 指标，当发现体验波动时，即时推送优化建议至 SMF 调整会话参数。

这一阶段的关键价值在于“以最小改造成本验证 AI 可行性”，通过单点智能的落地积累场景经验与数据基础，为后续架构重构提供实践依据。

3.2 演进期：重构式融合，破解协同与耦合难题

随着探索期对 AI 场景的验证与数据积累，核心网 AI 演进自然进入以“重构式融合”为核心的演进期。此阶段需解决探索期遗留的“功能耦合”“跨域协同不足”等深层问题，推动核心网从“单点智能”向“跨域协同智能”转型，为原生 AI Core 的构建奠定架构基础。

技术路径聚焦两大方向，一是推动传统网元的“无状态化+工具化”重构，打破 AMF、SMF 等网元的功能边界，将强耦合的业务模块拆分为独立的工具化组件。例如，将 AMF 拆分为“移动管理模块”（负责用户位置登记与切换）与“链路管理模块”（负责基站 N2 链路维护），将 SMF 拆分为“会话控制模块”与“策略管理模块”。所有拆分后的组件统一注册至“核心网能力池”，支持根据业务需求动态调用与弹性扩缩容，避免单一模块升级对全网业务的影响。二是重构网络分层架构，新增“AI 数据中台”与“意图解析层”两大核心组件：AI 数据中台整合多网元的结构化数据（如网元负载、会话参数）与通感融合产生的非结构化数据（如路况图像、激光雷达点云），通过统一的数据处理流程为 AI 模型提供高质量数据支撑，破解“数据孤岛”问题；意图解析层则对接用户或行业的模糊需求，将其转化为标准化的算力、网络参数，实现“业务意图”与“资源配置”的精准映射。

通过架构重构与组件新增,演进期逐步消除AI与核心网的协同壁垒,使跨域资源调度效率有效提升,为原生期的全维度智能自治做好技术铺垫。

3.3 原生期:原生式AI架构,实现6G全维度智能自治

当演进期完成架构重构与能力铺垫后,核心网AI将迈入面向6G需求的“原生期”。此阶段的核心目标是构建“AI原生”的核心网架构,称为原生AI Core,使核心网具备“自感知、自决策、自优化、自修复”的全域自主能力,完全适配6G“空天地一体”“通感算智融合”等复杂场景需求。

原生AI Core的技术形态主要体现在3个方面,其一,确立“四层三域”的组网架构,即“服务执行层-网络功能层-服务控制层-网络系统层”的4层逻辑架构,融入卫星通信、地面网络、低空通信“三域”资源。服务执行层实现通信、感知、计算资源的可编程组合,支持通感算一体化业务;网络功能层以AI Agent集群为核心,包含智能控制实体、算网协同实体等原生网元,替代传统功能耦合的网元形态;服务控制层负责跨域任务拆解与Agent协同,实现全局资源的最优调度;网络系统层则提供数字孪生、生成式AI等底层支撑,为决策优化提供仿真环境。其二,核心网元全面升级为“AI原生形态”,例如用户面具备AI驱动的业务流实时分类与动态转发能力,可根据工业控制、多模态交互等业务的差异化需求调整转发策略;卫星核心网的星上处理单元嵌入轻量化AI模型,能自适应太空环境的链路波动,保障空天地业务的连续性。其三,实现端到端的意图驱动闭环,用户或行业只需输入模糊意图,核心网即可自动拆解为“接入认证-算力调度-链路优化-状态监测”等子任务,通过AI Agent协同完成资源配置,并借助数字孪生平台预验证方案可行性;任务执行过程中,实时根据业务反馈优化策略,全程无需人工干预。

原生期的核心网AI彻底摆脱“外挂辅助”的定位,成为6G生态的核心支撑,能够应对空天地网络的异构性、业务需求的多样性,真正实现“万物智联”场景下的高效自治。

4 结束语

本文针对5G-A向6G演进中核心网与AI融合的技术瓶颈及场景需求,开展了系统性研究。首先全面梳理了外挂式AI、网元级AI增强、初步协同三种主流融合模式的核心短板,明确了“协同不足、耦合严重、数据低效”是制约核心网AI深度融合的关键问题,为后续技术突破指明了方向。在此基础上,本文构建了完整的核心技术支撑体系:AI智能体技术通过标准化协同协议与分层记忆学习机制实现跨域协作与自我演进,其分级能力体系可适配不同演进阶段需求;算网协同技术依托任务控制功能需求解析与计算控制功能算力调度双中枢,精准匹配业务的“算力-网络”双重需求,实现全域资源动态优化;统一数据服务框架则通过网元代理采集、标准化整合与向量存储,为AI模型提供高质量、高安全的多源数据支撑,三者共同构成了原生AI Core的技术基石。

基于上述技术储备,本文遵循“最小改造验证-架构重构突破-原生智能落地”的递进逻辑,设计了核心网AI演进的三阶段方案:探索期通过网元级AI轻量化嵌入与外挂式AI Agent,在不中断现有业务的前提下实现单点智能突破并积累场景经验;演进期进一步推动网元“无状态化+工具化”重构,新增AI数据中台与意图解析层,逐步解决协同与耦合难题,为架构升级做好铺垫;原生期则确立“四层三域”组网架构,推动核心网元升级为AI原生形态,最终实现端到端意图驱动闭环,让核心网具备“自感知、自决策、自优化、自修复”的全域自主能力。

**[作者简介]**

宋旭光（1997-），男，中国移动通信集团设计院有限公司工程师，主要研究方向为移动通信核心网、边缘计算。

孙逊（1985-），女，中国移动通信集团设计院有限公司高级工程师，主要研究方向为移动通信核心网络规划咨询与设计、5G/6G新技术策略、行业专网解决方案等。

邱巍（1976-），女，中国移动通信集团设计院高级工程师，主要研究方向为5G/6G核心网、行业专网解决方案等。

马洪源（1984-），男，中国移动通信集团有限公司高级工

程师，主要研究方向为移动通信核心网。

周维（1977-），男，中国移动通信集团设计院有限公司正高级工程师，主要研究方向为2G/3G/4G/5G融合核心网、物联网、边缘计算的工程咨询设计。

冯志强（1981-），男，中国移动通信集团设计院有限公司工程师，主要研究方向为2G/4G/5G/NB融合核心网、物联网、5G专网的工程咨询设计。

卢友伟（1998-），男，中国移动通信集团设计院有限公司工程师，主要研究方向为移动通信核心网、边缘计算。