



XXXX

基于可解释性AI的6G O-RAN研究综述

廖勇, 杨善云

(重庆大学微电子与通信工程学院, 重庆 400044)

摘要: 面向第六代移动通信技术 (Sixth Generation Mobile Communication, 6G) 的开放式无线接入网络 (Open Radio Access Network, O-RAN) 智能自治演进中, 高度依赖的人工智能 (Artificial Intelligence, AI) 黑盒特性与透明决策需求存在矛盾。可解释人工智能 (eXplainable AI, XAI) 深度融合是打破黑盒、实现透明决策及构建可信自治网络的核心。本文综述XAI在6G O-RAN中的研究, 提出多层级融合部署框架与应用路径: 阐述O-RAN架构与XAI部署逻辑; 围绕无线资源管理、网络切片运维、网络安全增强、智能协作与零接触运维, 分析XAI实现路径与典型应用; 探讨解释计算开销与实时性矛盾、评估标准匮乏、接口协同限制及隐私风险等技术挑战; 展望轻量化XAI算法、神经符号因果推理模型、闭环自治机制及跨学科方法融合等趋势, 为构建智能、透明、安全的6G可信自治网络提供参考。

关键词: 6G; O-RAN; 可解释性; 人工智能; 网络智能化; 透明化决策; 可信自治网络

中图分类号: A

文献标志码:

doi: 10.11959/j.issn.1000-0801.

A Survey on 6G O-RAN Based on Explainable AI

LIAO Yong, YANG Shanyun

School of Microelectronics and Communication Engineering, Chongqing University, Chongqing 400044, China

Abstract: As the Open Radio Access Network (O-RAN) for Sixth Generation mobile communication technology (6G) evolves toward intelligent autonomy, the "black-box" nature of highly relied-upon Artificial Intelligence (AI) models conflicts with the requirement for transparent and trustworthy decision-making. The deep integration of eXplainable AI (XAI) is essential to demystify this black-box, achieve transparent decision-making, and construct a trustworthy autonomous network. This paper provides a comprehensive review of XAI research in 6G O-RAN and proposes a multi-layer integrated deployment framework and roadmap: we expound on the O-RAN architecture and XAI deployment logic; analyze implementation paths and typical applications across radio resource management, network slicing, network security enhancement, intelligent collaboration, and zero-touch operation; discuss technical challenges, including the trade-off between explanation computational overhead and real-time requirements, the lack of unified evaluation standards, interface-induced coordination constraints, and data privacy risks; and explore future trends, such as lightweight XAI algorithms, neuro-symbolic causal reasoning models, closed-loop autonomous mecha-

收稿日期: XXXX-XX-XX; 修回日期: XXXX-XX-XX

基金项目: 重庆市自然科学基金 (CSTB2023NSCQ-MSX0025)



nisms, and interdisciplinary integration, offering insights for constructing intelligent, transparent, and secure 6G trustworthy autonomous networks.

Key words: 6G, O-RAN, XAI, Network Intelligence, Transparent Decision-Making, Trustworthy Autonomous Networks

1 引言

第六代移动通信技术 (Sixth Generation Mobile Communication, 6G) 与人工智能 (Artificial Intelligence, AI) 的深度融合是网络演进的核心方向。6G 将赋能智能电网 2.0、移动全息及工业 5.0 等扩展现实 (Extended Reality, XR) 应用^[1], 推动网络从传统数据管道跃升为兼具感知与决策的“万物智联”生态系统^[2-5]。

然而, 海量数据与高速率对网络资源调度、安全及决策透明度提出了严峻挑战。传统开放无线接入网络 (Open Radio Access Network, O-RAN) 与黑盒 AI 模型的结合存在不可靠、难运维的瓶颈。可解释性人工智能 (eXplainable AI, XAI) 通过其内在可解释性 (透明结构) 与事后解释性 (外部特征解析) 两大范式, 有效打破了系统算法间的信任壁垒, 为构建透明、可信且高度自治的 6G O-RAN 闭环控制系统提供了核心保障^{[6][7]}。

当前, XAI 在通信领域的应用尚处于起步期, 缺乏 6G O-RAN 统一架构下的系统化整合^{[8][9]}。尽管 AI-RAN 联盟已提出“AI-for-/and-/on-RAN”框架, 现有研究多局限于单点场景。例如, 利用 Shapley 加性解释 (SHapley Additive exPlanations, SHAP) 保障增强型超可靠低时延通信 (eXtended Ultra-Reliable Low-Latency Communication, xURLLC), 以及通过 EXPLORA 框架解释深度强化学习 (Deep Reinforcement Learning, DRL) 的资源分配策略^{[10][11]}。

基于上述背景, 本文系统综述了 XAI 在 6G O-RAN 中的最新进展: 首先梳理 XAI 的嵌入框架

与演进逻辑; 其次归纳其在资源管理、可信运维、网络安全等核心场景的典型应用; 最后剖析现有技术挑战并展望未来趋势。

2 XAI 在 O-RAN 架构中的部署框架

O-RAN 架构包括不同层级, 底层有近实时无线接入网智能控制器 (Near-Real-Time RAN Intelligent Controller Near-RT RIC, Near-RT RIC), 上层包括服务管理与编排 (Service Management and Orchestration, SMO) 非实时 (Non-Real-Time RAN Intelligent Controller, Non-RT RIC)^[12]。不同层级之间依靠标准化开放接口, 比如 A1、E2、O1 完成信息传递, XAI 作用于网络智能化帮助层, 分层部署在不同网络组件内, 为各个组件功能输出透明决策支持, 同时完成异常检测。O-RAN 基本架构及 XAI 部署框架如图 1 所示。

2.1 O-RAN 架构

如图 1 所示, O-RAN 架构主要分为开放式中央单元 (Open Centralized Unit, O-CU)、开放式分布单元 (Open Distributed Unit, O-DU)、开放式无线单元 (Open Radio Unit, O-RU)、Near-RT RIC 和 Non-RT RIC 五部分。

O-CU 是一个逻辑节点, 主要负责处理非实时且对时延要求相对较低的高层协议栈; O-DU 是连接 O-CU 和 O-RU 的中间逻辑节点, 负责处理对时延敏感的底层协议功能, 而 O-RU 是最靠近天线的逻辑节点, 负责将数字信号转换为无线电波。O-DU 部署位置靠近无线接入点的边缘节点, 承担低延迟基带信号处理工作, 也完成无线资源分配任务, O-RU 负责发送和接收无线信号,

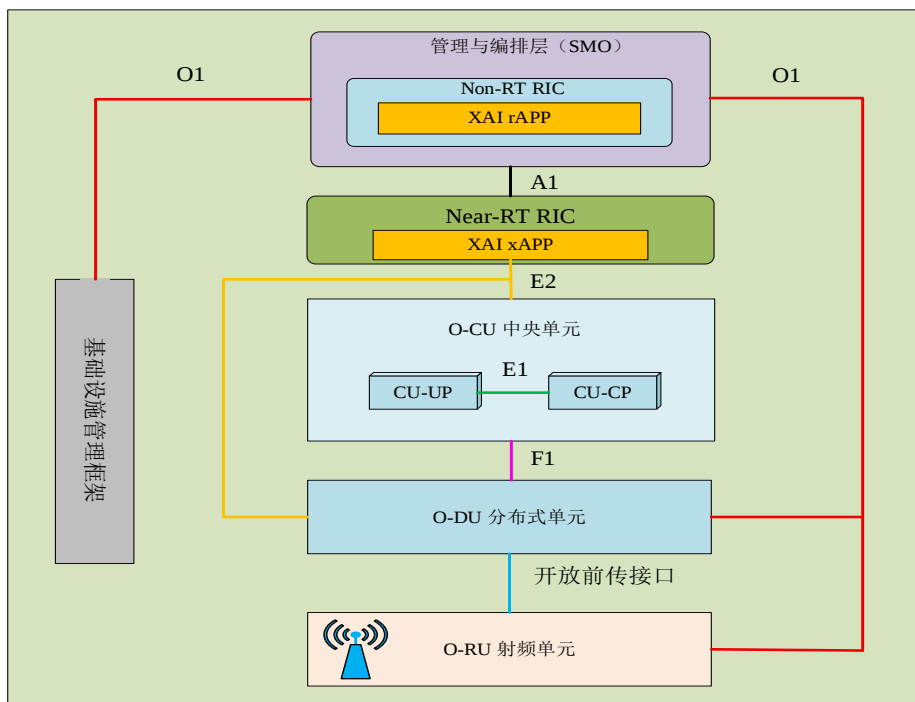


图1 O-RAN基本架构及XAI部署框架示意图

O-CU一般部署在中央云或者边缘云，处理非实时类基带功能，具体内容包括无线资源调度、大规模多输入多输出信号处理。6G网络环境下，上述底层网元会生成数量庞大的遥测数据流，这些数据流可实时输出信道状态信息、无线测量报告、传感器数据等内容，成为XAI算法的原始输入，引入注意力机制或特征归因方法，比如层次相关传播，可对这些输入特征做解释，也能为底层网络的透明化整理出充足的数据支撑。

RICs包含支持非实时功能(>1s)的Non-RT RIC和近实时控制的Near-RT RIC。Non-RT RIC部署于服务管理与编排(Service Management and Orchestration, SMO)平台内，通过其框架提供的R1接口为rApp提供基础服务，并利用O1接口进行数据监控以执行智能RAN优化^[12-14]。基于ML/DL推理生成的智能策略经由A1接口下发至Near-RT RIC，实现闭环指引。作为核心自动化平台，SMO通过O1、A1及O2接口负责全网元的编排与管理，在简化网络复杂性的同时显著

提升了运维效率。

Near RT RIC负责控制和优化，通过精准的数据监控和在E2接口上的操作，以近实时的尺度(从10毫秒到100毫秒)实时监控O-RAN节点(CU和DU)及其资源。它托管了多个近实时无线接入网应用程序(Near-Real-Time RAN Application, xApp)，这些应用可以通过E2接口收集近实时信息，并根据通过A1接口接收到的非实时RIC的策略提供增值服务。xApps包括移动性管理、资源管理、频谱管理等。

O-RAN架构中相关术语及其部署位置如表1所示。

2.2 XAI部署逻辑

在6G O-RAN背景下，应用场景与XAI技术分类通过需求驱动的逻辑紧密关联^[15-19]。XAI在不同网络组件中的部署，遵循着基于动态挑战和机制演进梯度的递进逻辑，如图1所示，它主要以xApp或rApp的形式部署在Near-RT RIC和Non-RT RIC中。整体部署主要有事后解释型、解



表1 O-RAN 架构中相关术语及其部署位置

术语	核心功能	部署位置/所属层级
O-CU	处理非实时的高层协议栈，负责无线资源调度和信号处理。	中央云或边缘云
O-DU	处理时延敏感的底层协议功能，承担低延迟基带处理与无线资源分配。	靠近无线接入点的边缘节点
O-RU	负责数字信号与无线电波的转换、底层信号收发及生成海量遥测数据流。	最靠近天线的节点
SMO	智能自动化管理平台，负责整个RAN域的编排、管理、自动化以及降低网络复杂性。	管理与编排层
Non-RT RIC	位于SMO内，执行时间粒度大于1秒的非实时控制，基于AI/ML生成并下发智能RAN策略。	位于SMO框架内
Near-RT RIC	负责10至100毫秒的近实时控制与优化，实时监控与调度底层的O-RAN节点资源。	RAN控制层
rApp	运行于Non-RT RIC框架的应用程序，用于获取数据并执行非实时的智能RAN优化与策略触发。	运行于Non-RT RIC框架
xApp	运行于Near-RT RIC的独立应用，根据上层策略执行移动性、频谱等近实时资源管理。	托管于Near-RT RIC平台

释引导学习型、神经符号推理型三个类型。

事后解释型是一种被动诊断算法，其主要强调外部诊断，部署于O-CU/O-DU/O-RU以及Near-RT RIC层。其优势在于它可以在不改变原生黑盒决策模型的前提下，让其网络安全性能增强，无线资源管理延迟降低。

解释引导学习型是一种动态优化算法，其技术将解释性融入模型的训练闭环。它主要通过SMO/Non-RT RIC层强大的算力和开放接口来对全局进行动态优化。这样的机制极其适合网络切片运维和多节点的智能协作场景。

神经符号推理型是一种主动自治算法，该机制主要部署于SMO层。针对复杂的全局无线资源管理和零接触运维中的约束导向问题其通过实现深度学习和逻辑规则的内在融合来提供严格的安全与物理边界保障。

上述机制，落地应用均离不开O1、A1、E2接口直间的数据交互。事后解释的低延迟审计反馈需要通过E2接口实现；解释引导学习的策略意图需要通过A1接口传递；神经符号推理需要通过O1、O2接口作为全局输入。这种接口于算法类型的相关匹配，可以优化系统总体开销、适

用不同实时性需求。

虽然O-RAN的分层解耦与开放接口架构为XAI的提供了较为灵活的部署空间，但仍然面临一些有待解决的问题。首先，需要面对的就是在多供应商环境下，不同厂商提供的AI/ML模型在底层架构、输入特征空间以及输出动作空间上存在巨大差异，由于没有相关标准这就导致了其难以在SMO层进行全局维度的对齐与聚合，进而阻碍了全网视角的因果推理。其次，现有的E2和A1接口协议主要针对精简的控制信令和策略意图设计，并未预留大规模传输XAI解释结果的信道容量。XAI的解释结果通常是极其庞大的数据。如果将这种庞大的数据高频地塞入E2接口，可能会直接导致控制信道拥塞，从而破坏Near-RT RIC的实时性。此外，目前的E2SM（E2 Service Model，E2SM）规范定义了如何上报性能指标或网络接口数据，但尚未正式定义用于承载复杂AI解释的标准数据结构格式。

XAI在O-RAN中的部署位置以及接口关系如表2所示。

表2 O-RAN 架构下不同 XAI 机制的部署与应用特征对比

XAI 机制类型	算法性质	部署架构层级	关键依赖接口
事后解释型	被动诊断算法	Near-RT RIC, O-CU / O-DU / O-RU	E2 接口
解释引导学习型	动态优化算法	SMO / Non-RT RIC	A1 接口
神经符号推理型	主动自治算法	SMO 层	O1 接口

3 XAI 在 O-RAN 中的应用场景

面对 6G O-RAN 的动态挑战, 可解释人工智能基于机制演进梯度, 以需求驱动深度赋能无线资源管理、切片运维、安全增强、智能协作及零接触运维等方面。具体而言, 事后解释型 XAI 侧重外部诊断, 利用 SHAP 实现安全与资源管理中的黑盒审计与异常检测; 解释引导学习型 XAI 融入训练闭环, 通过 SHAP 调整联邦学习损失函数以折中性能与透明度; 神经符号推理型 XAI 则融合内在逻辑, 引入 TANGO 规则规避资源分配等约束导向下的局部最优。该体系推动网络由被动诊断向主动自治演进, 全面提升 O-RAN 生态的透明度与可信度, 未来亟需进一步优化开销以适配实时性要求。

3.1 无线资源管理的透明化优化

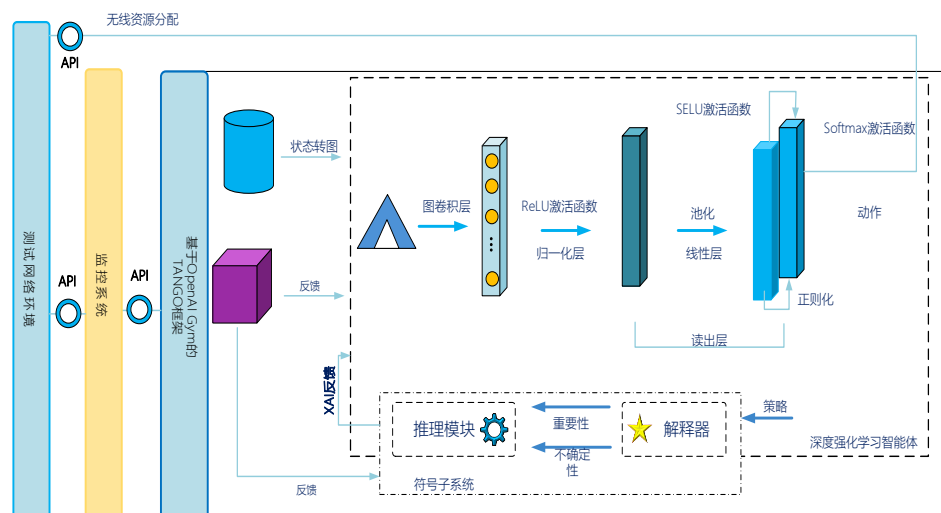
本节将梳理 XAI 在无线资源管理领域的三个典型应用场景: 服务于下一代基站资源分配的 TANGO 框架、用于毫米波波束对齐的波束对齐引擎 (Beam Alignment Engine, BAE), 以及用于信道估计的基于梯度的可解释方案 (Gradient-based XAI scheme for Channel Estimation, GRACE)。同时, 横向对比这三种分别代表内在神经符号推理、事后实例置信度评估及事后梯度反向传播的 XAI 技术流派, 深入剖析其在 O-RAN 资源管理场景中的性能演进与开销权衡^[20-23]。

Rezazadeh 等人^[24]提出了一种新型图强化学习框架 TANGO, 如图 2 所示。该框架内嵌符号子系统以实现内在可解释性, 其核心机制为: 贝叶斯图神经网络 (Graph Neural Network, GNN) 解释器生成输出后, 依据符号推理规则定期将其转换为逻辑图强化学习 (Graph Reinforcement Learning, GRL) 的奖励函数, 进而指导学习过程。在真实 gNodeB 测试床的无线资源分配任务中, TANGO 将问题建模为马尔可夫决策过程, 其状态空间和动作空间分别对应网络拓扑/负载信

息与物理资源块分配决策。其贝叶斯 GNN 解释器利用变分贝叶斯推理近似后验分布, 计算节点和边的掩码以评估特征重要性与不确定性; 推理器则根据预定义规则 (如对高不确定性决策施加惩罚) 进行奖励整形, 确保决策满足领域约束并避免局部最优。从技术演进角度看, TANGO 代表了向内在可解释性迈进的前沿方向, 其优势在于将符号逻辑闭环嵌入奖励函数, 实现逻辑自治; 但变分推理带来的高计算复杂度使其难以适应极低延迟的 O-RU 底层调度, 更适合部署于 Non-RT RIC 或高规格 Near-RT RIC 平台。

波束对齐作为无线资源管理的核心子任务, 常面临高开销与噪声干扰的挑战。Khan 等人^[25]提出了一种适用于毫米波多输入多输出系统的鲁棒性可解释深度学习波束对齐引擎。该框架依据 DFT 码本的接收信号强度指示测量值, 利用卷积神经网络 (Convolutional Neural Network, CNN) 预判最优窄波束, 从而压缩码本扫描开销。具体流程为: 基站发送宽波束导频, 用户设备上报 RSSI, CNN 以此作为输入进行分类预测。同时, BAE 引入深度 K 近邻算法 (Deep k-Nearest Neighbors, DkNN) 评估网络内部表示, 输出人类可读的置信度指标, 并检测分布外输入以提升模型抗噪与对抗攻击的能力。该方案可作为 xApp 部署于 O-RAN 的 Near-RT RIC。

信道估计的准确性直接关系到资源分配的效率。Gizzini 等人^[26]提出了一种基于梯度的可解释信道估计方案。针对前馈神经网络, GRACE 引入层级相关传播技术, 通过梯度反向传播处理分类模型输入, 精准识别对决策贡献最高的相关特征。在 6G O-RAN 架构下, GRACE 可作为 rApp 嵌入 Non-RT RIC。相比前述两种方案, GRACE 代表的基于梯度事后解释路线在保真度与计算复杂度间取得了优异平衡, 彻底消除了外部扰动偏差, 其计算开销显著低于 TANGO 与 BAE, 是边缘资源受限场景的更优选择。

图2 新型图强化学习TANGO框架^[12]

综上所述，XAI在无线资源管理领域的应用有效推动了模型从黑盒状态向透明决策的转变。TANGO通过神经符号逻辑互补加速收敛并提升鲁棒性；BAE借助DkNN量化置信度以应对噪声干扰；GRACE则依赖内部梯度过滤实现高效无偏的信道估计。这些方案有效降低了资源分配的不确定性，增强了O-RAN系统的可审计能力与信任度。XAI在无线资源管理场景中的典型应用如表3所示。

3.2 网络切片的智能可信运维

本节将梳理现有研究，聚焦XAI在切片准入控制、资源分配、QoS优化及联邦学习等场景的关键应用。我们将首先分析一般AI/ML模型在切片管理中的局限性，例如Phyu等人^[15]的调研与

Rahman等人^[18]的QoS框架，进而横向对比严格意义上的XAI方案在同一业务场景下的演进脉络与技术优劣势，例如Roy等人^[16-17]提出的FL-XAI模型。

Phyu等人^[27]系统调研了机器学习（Machine Learning, ML）在网络切片中的应用，其划分的层级结构涵盖了准入控制、资源分配、故障管理等多个运行环节。例如，在准入控制中常采用随机森林或神经网络预估切片请求的可行性及潜在冲突；在资源分配环节，则依托强化学习动态调整物理资源块与带宽以适配QoS需求。该调研明确了ML在O-RAN的Non-RT RIC中实现零接触运维的价值，但同时指出，ML模型的“可解释性不足”是当前亟待解决的核心痛点。尽管该研

表3 XAI在无线资源管理场景中的典型应用

典型场景	XAI技术	贡献	局限
TANGO框架	贝叶斯GNN解释器	处理DRL黑盒决策问题，提高资源分配环节的透明度，优化资源分配的可靠性，强化相关环节的信任度，加快DRL模型的收敛速度，规避结果落入局部最优区间，保证所有执行流程符合对应领域的约束条件。	计算复杂度出现上升，对应变分贝叶斯推理的开销同步提高，解释一致性的稳定状态可受规则定义方式干扰。
BAE引擎	DkNN	处理波束对齐过程中存在的高开销与噪声干扰，优化系统鲁棒性，完善决策判断维度，压缩码本扫描环节的资源消耗，强化异常检测环节的鲁棒性。	DkNN完成邻域搜索的过程会增加计算开销，处理分布外输入时，还可能引入额外的延迟。
GRACE方案	LRP	平衡信道估计环节性能与复杂度之间的关系，优化模型可解释性，优化模型鲁棒性，消除基于扰动生成的偏差，完成高效率输入分类。	阐释计算开销上升相关内容，对应梯度传播环节，分析不同信道模型条件下，解释方面存在的一致性问题。

究探讨了引入 SHAP 或 LIME 量化特征贡献、以及结合联邦学习等开放性方向,但其主要聚焦于一般 AI/ML 模型在提升系统性能上的潜力,旨在揭示黑盒问题,并未提出具备实质解释机制的 XAI 架构。

在 O-RAN 的具体场景中, Rahman 等人^[28]提出了一种基于标准 AI/ML 框架与 Non-RT RIC 的切片 QoS 优化方案,如图 4 所示。该框架部署于 SMO 层,通过 rApp 监测延迟、吞吐量等关键性能指标,并利用 ML 模型动态调整资源。其完整工作流程为:首先经 O1 接口收集实时数据;其次调用深度 Q 网络(Deep Q-Network, DQN)或长短期记忆网络(Long Short-Term Memory, LSTM)预测 QoS 违规风险;最后在 Non-RT RIC 内完成策略优化以实现闭环自动化。然而,该框架中的 DQN/LSTM 模型仅输出资源调度动作,缺乏显式的内在因果逻辑,属于典型的黑盒优化。虽预留了接入注意力机制等解释模块的空

间,但当前状态进一步凸显了在实际部署中引入专属 XAI 机制以打破透明度瓶颈的迫切需求。

为突破上述一般 AI/ML 方法面临的信任危机,后续研究引入了不同的 XAI 解释器(如 SHAP 与 LIME),展现了截然不同的技术特征。Roy 等人^[29]探讨了联邦学习(Federated Learning, FL)应用于 6G O-RAN 切片时性能与可解释性的制约关系,提出了一种解释引导的内嵌式 FL 方法。在该方案中,作为 FL 客户端的 gNB 在本地训练深度神经网络以预测资源需求,服务器负责参数聚合。同时,解释器利用 SHAP 值量化信噪比、流量负载等特征的重要性,并将结果反馈至训练环节以调整损失函数。这种解释引导的奖励整形促使模型优先生成可解释决策。此方案有效结合了 FL 的分布式优势与 XAI 的透明性,高度适用于 O-RAN 的 xApp 部署。

此外, Roy 等人提出的联合可解释性与敏感性感知的联邦深度学习框架是该领域的另一典型

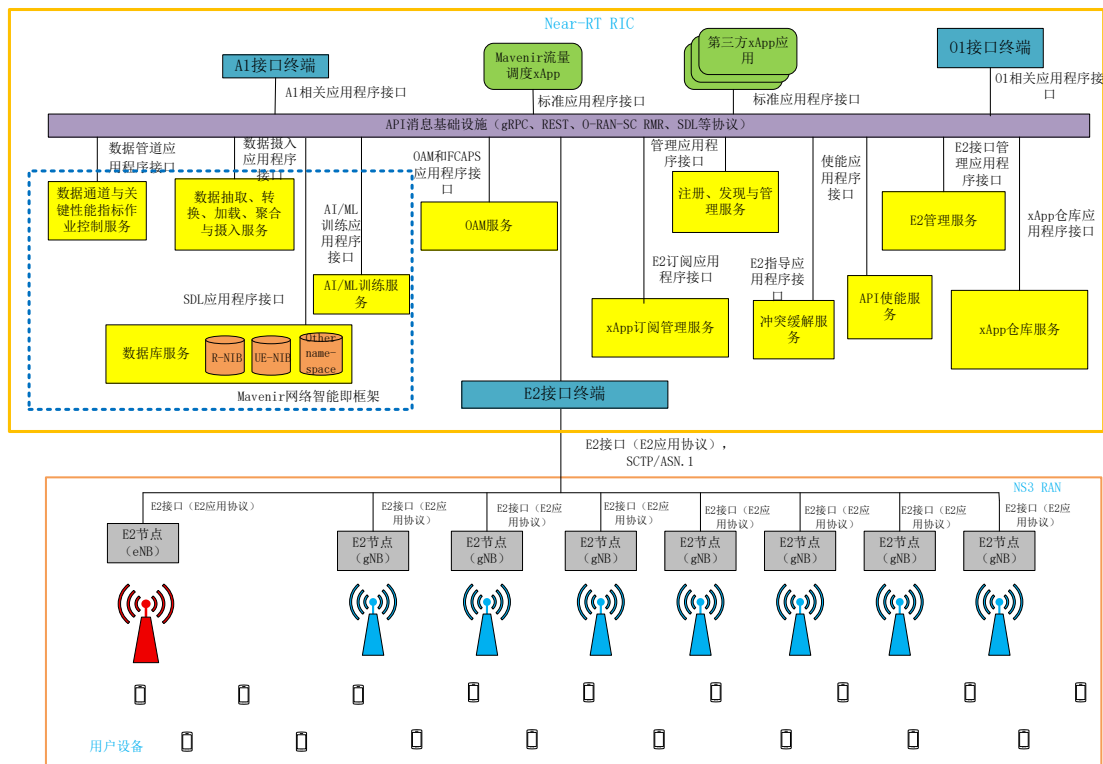


图 4 O-RAN QoS 优化框架^[28]



案例。该框架在预测切片丢包率的同时，同步优化模型依托LIME实现的可解释性与敏感性。客户端运行DNN处理本地数据，服务器则收集解释反馈并执行基于梯度范数的敏感性聚合操作，防止高敏感参数主导模型。该框架同样可作为rApp嵌入6G O-RAN，为可信切片运维提供坚实支撑。

综上所述，现有研究勾勒了XAI在网络切片运维中的演进路径：从Phyu^[27]对ML基础框架的问题揭示，到Rahman^[28]围绕O-RAN集成的黑盒实践，再到Roy等人^[29-30]将FL与XAI深度融合以权衡性能与透明度。XAI的引入有效降低了切片故障风险，提升了QoS的可审计性。XAI在智能可信运维场景中的典型应用如表4所示。

3.3 增强网络安全的韧性

本节聚焦XAI在O-RAN网络安全中的应用，深入剖析其典型案例，并横向对比事后特征归因、大语言模型语义辅助、内在注意力机制与量子增强等不同XAI技术流派在响应延迟、计算开销及可解释性方面的权衡，以直观展现各方案的部署可行性^[31]。

Basaran与Dressler提出的XAIomaly框架^[10]是XAI应用于6G O-RAN的典型案例。该框架针对极高可靠低延迟通信（extreme Ultra-Reliable Low-Latency Communication, xURLLC）业务流设计，旨在提供可解释且可信的异常检测。

XAIomaly集成了SHAP（SHapley Additive ex-Planations）方法，通过计算流量模式、延迟指标及接口信令等输入特征对模型输出的贡献度，明确异常判定的依据。在O-RAN的Near-RT RIC部署中，该框架依托深度学习模型监控实时数据流，随后调用SHAP生成特征重要性热图，精准定位引发异常的关键因素，如突发丢包或未授权访问。这一机制不仅提升了检测准确率，还支持运维人员审计决策过程，有效解决了黑盒模型的信任缺失问题。其核心创新在于将XAI嵌入O-RAN的E2接口，实现了从检测到解释的完整闭环，为零接触运维提供了有力支撑。

另一核心案例是Chatzimiltis等人提出的AI-on-RAN框架^[32]。该框架通过整合大语言模型（Large Language Model, LLM），将XAI输出的特征重要性转化为自然语言告警，拓展了网络安全运维的可操作性。在其内部，XAI利用LIME（Local Interpretable Model-agnostic Explanations）或SHAP对异常检测模型进行决策分析。随后，基于GPT变体的LLM将抽象特征转换为人类易读的描述。此流程作为rApp嵌入Non-RT RIC，并依托A1接口引导Near-RT RIC的响应策略。该框架有效衔接了技术解释与运维实践，缩短了响应时间，提升了网络在应对DDoS、接口劫持等多模态威胁时的自身韧性。

Porambage等人提出的XcARet架构^[33]强调了

表4 XAI在智能可信运维场景中的典型应用

典型场景	XAI技术	贡献	局限
切片准入控制、资源分配	SHAP值或LIME	处理ML模型存在的可解释性不足问题，搭建系统化参考框架，平衡模型性能和信息透明度，支持隐私保护，也能完成跨域协调。	未直接涉及6G O-RAN，扩展性需验证；额外解释计算开销
O-RAN QoS优化框架	注意机制	处理QoS违规与资源利用存在的相关问题，搭建闭环自动化运行模式，提高运维过程的信息透明度。	计算开销和数据隐私挑战；解释标准化缺失
FL性能-可解释性权衡方案	SHAP值	解决FL在切片中的性能-透明权衡问题，提升QoS满足率和可解释性	准确率轻微牺牲（2%-5%）；解释计算增加开销
联合可解释性和敏感性感知FL	LIME、梯度范数	处理丢包预测环节存在的透明性与鲁棒性问题，可提高解释一致性与对抗攻击的抵抗能力。	敏感性-aware聚合增加复杂度；多方法解释可能不一致

XAI在绿色安全 O-RAN 中的作用。针对 6G 开放接口,该架构采用注意力机制和特征归因方法解释资源消耗异常并检测能源相关攻击(如伪造负载引发的异常高功耗)。通过分析功率指标和流量日志等模型输入,XAI生成解释报告以支撑可持续运维。Kaur 与 Gupta^[34]则聚焦 6G IoT 环境,提出结合 SHAP 与 LIME 的 XGBoost 架构以增强 AI 安全性,如图 5 所示。该方案先训练 XGBoost 模型进行入侵检测,再利用 SHAP 实现设备 ID、数据包大小等特征的重要性可视化,并结合 LIME 的局部解释辅助 IoT 节点异常行为的审计。此外,Mehta 等人^[35]提出一种量子辅助 XAI 驱动的深度学习框架,用于连接自治电动车辆的虚假数据注入检测。在 6G 底层网络环境下,该框架结合量子增强神经网络与 XAI 识别篡改车辆传感器数据等注入行为,并生成因果解释链以支持实时响应。通过对上述方案的横向比较可知:为克服事后解释的高延迟瓶颈,XcARet 引入内置的注意力机制,在保证透明度的同时有效降低了安全防护对网络设备造成的能源与计算开销;而量子辅助 XAI 则提供了一条在算力维度突破经典计算瓶颈的新路径,在应对未知恶意流量攻击时,展现出远超传统机器学习模型的泛化能力与毫秒级响应速度。

现有公开案例可呈现 XAI 在网络安全韧性增强领域的多元价值,XAIomaly 和 AI-on-RAN 围绕实时解释和自然语言转化展开研究,XcARet 聚焦绿色可持续性,XGBoost-SHAP 和量子 XAI 将应用范围扩展至 IoT 和 CAEV 场景,XAI 降低误报风险,提高 O-RAN 系统的可审计性和响应速度,支撑可信自治网络建立。

综合来看,在 O-RAN 的安全防护架构中,XAI 方案的选型必须与部署位置的时间尺度严格匹配:以 SHAP 和 LLM 为代表的高算力、高语义解释方案更适合部署在 Non-RT RIC 用于离线的长期威胁溯源与审计;而轻量级的内在注意力机制或量子增强方案,才真正具备下沉至 Near-RT RIC 进行近实时动态防御的潜力。

XAI 在网络安全场景中的典型应用如表 5 所示。

3.4 跨越智能协作与零接触运维

本小节聚焦 XAI 在智能协作零接触运维中的应用,逐一展开典型案例,包括 XRL 为基础的资源管理和 RAN 切片、AI-and-RAN 工作负载编排、EXPLORA 框架的通用解释、语义感知 DRL 在协作优化中发挥的作用。

要想实现跨层级协同应建立“微观解释感知”和“宏观意图导向”的映射机制。具体而

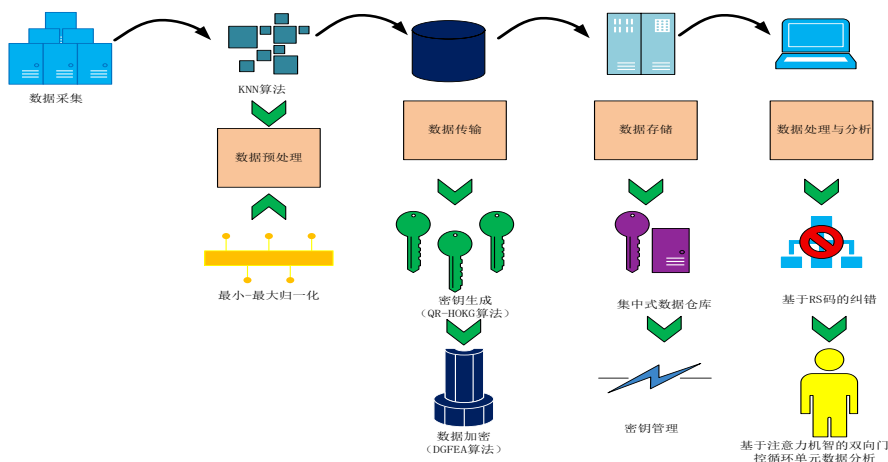


图5 XGBoost与SHAP的IoT安全增强^[34]



表5 XAI在网络安全场景中的典型应用

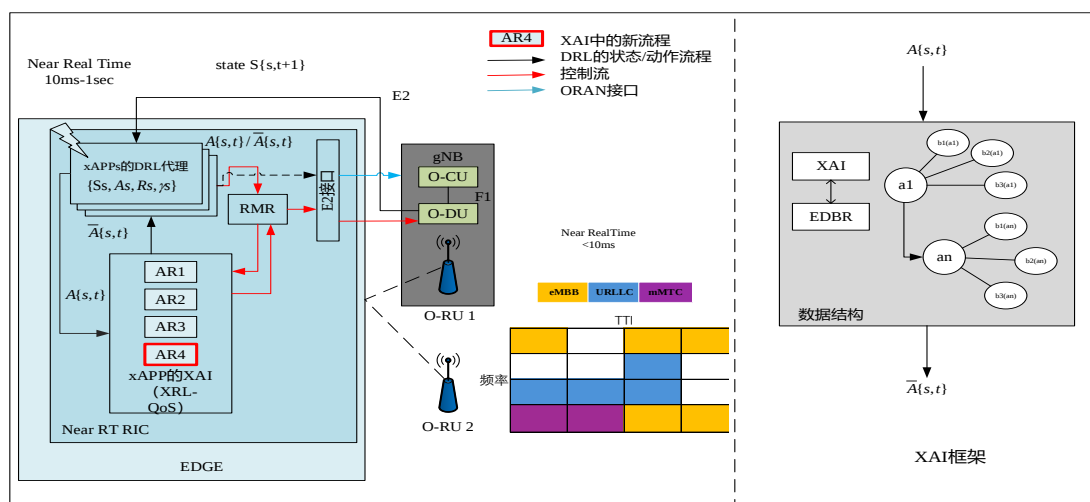
典型场景	XAI技术	贡献	局限
xURLLC异常检测	SHAP值	处理异常检测环节中黑盒模型信任不足的相关问题，提高检测环节的精度，强化审计方面的能力，可支撑实时闭环运行。	解释生成时间虽毫秒级，但可能影响极端实时性；解释一致性问题
异常解释	LIME或SHAP值、LLM	解决响应延迟和运维效率问题，桥接技术解释与实践；降低误报率	LLM融合增加计算开销；多模态威胁下解释一致性挑战
XcARet架构	注意机制和特征attribution	解决能源相关攻击问题，支持可持续运维和异常审计	计算开销影响实时性；绿色指标解释标准化缺失
IoT安全增强	SHAP值和LIME	解决IoT入侵检测的透明问题，提升检测精度和可信度	多方法（SHAP与LIME）可能产生矛盾解释；额外开销
在CAEV中的FDI检测	量子增强神经网络结合XAI	解决FDI检测延迟问题，提升决策透明度和响应速度	量子计算开销高；解释深度与一致性需优化

言，底层Near-RT RIC通过E2接口捕获毫秒级的无线资源波动，从而生成局部解释，然后经过语义抽象，映射为Non-RT RIC可理解的SLA性能画像。这种协同依赖于A1接口下发的“解释性策略”与O1接口上报的全局遥测数据，形成从底层感知到高层决策的闭环。

Mhatre等提出了一种应用于6G O-RAN资源管理与切片的可解释强化学习（eXplainable Reinforcement Learning, XRL）框架，如图6所示^[35]。该框架融合强化学习与注意力机制，通过分析“状态-动作”对的贡献度实现任务卸载决策的全程溯源。建模过程将O-RAN环境表征为多智能体马尔可夫决策过程（Markov Decision Process, MAPD），由代表O-DU/O-CU的智能体根据负

载、时延与能耗指标生成策略。随后，XAI模块通过类SHAP归因法量化特征重要性，解释为何优先选择边缘而非云端处理。相关解释以xApp形式集成于Near-RT RIC以支持实时调整。

此外，Shah等人提出了一种面向零接触运维的主动式AI-and-RAN工作负载编排框架^[36]。该框架构建了“AI异常检测—XAI归因分析—LLM报告生成”的自动化闭环流水线。具体而言，AI模型首先监测O-RAN组件的负载状况；随后，XAI模块基于LIME算法定位异常根因；最后，基于BERT变体的LLM将定位结果转化为自然语言操作指令。该框架集成于SMO层，展示了“从归因到意图”的纵向交互逻辑：Non-RT RIC层的rApp利用O1接口获取长周期全局日志以进

图6 XRL框架结构^[35]

行异常建模, XAI 归因结果并不直接干预底层, 而是作为解释性约束的策略意图经由 AI 接口下发。LLM 在此作为语义转换桥梁, 将复杂的特征归因转化为可执行指令, 显著压缩了人工干预空间。Fiandrino 等提出的 EXPLORA 框架^[11]是一个通用 XAI 平台, 扩展应用范围覆盖智能协作和零接触运维, 该框架面向 O-RAN 开发, 调用集成梯度方法完成 AI/ML 模型决策的解释工作, 比如协作场景中, 该框架可以说明决策优先选择边缘卸载的原因, 梯度分析结果显示, 边缘节点能耗贡献最低。零接触场景下, EXPLORA 可以 rApp 形式部署, 支持自动化报告生成, Guo 和 Wang 进行相关研究^[37], 引入 XAI 引导的语义感知 DRL, 对应 xTL 框架, 研究最初围绕城市交通灯控制展开, 研究提出的思路可应用于 6G O-RAN 协作, 比如提取语义以压缩通信开销。框架借助 XAI 完成无关语义特征过滤, 通过注意力机制完成协作决策解释, 应用于 O-RAN 可优化多代理间信息交换。

综上所述, XAI 在智能协作与运维场景中展现出核心价值: XRL 与 EXPLORA 实现了卸载透明化与通用解释^{[11][38]}, AI-and-RAN 构建了自动化运维流水线^[36], 而 xTL 则提供了语义优化的可行路径^[23]。XAI 有效增强了协作鲁棒性与运维自治性, 为 6G 网络端到端智能化奠定了基础。实现 XAI 跨层级协同的有效路径在于: 利用 Near-RT RIC 进行近实时局部解释, 经由 O1/A1 接口实现解释语义的上报与全局意图融合, 最终在 SMO

层确立全网一致的自治逻辑。

XAI 在跨越智能协作与零接触运维场景中典型应用如表 6 所示。

4 技术挑战与发展趋势

4.1 技术挑战

XAI 应用于 6G O-RAN 存在开发空间, 同时需要突破多项技术障碍, 这些障碍会直接作用于系统的实时性、可靠性、安全性。

1) 复杂的采样与特征计算, 引入的显著延迟是 XAI 在 6G O-RAN 近实时高动态场景落地的核心瓶颈。当前主流 XAI 方法大多涉及复杂的采样与特征计算, 会引入显著的额外延迟^[41-44]。在 O-RAN 的 Near-RT RIC 中, 控制回路的响应时间被严格限制在 10ms 以内。然而, 完成一次详细的特征重要性分析往往需要消耗数十毫秒甚至更长。这种延迟在高动态场景中尤为致命: 由于无线信道的相干时间极短, 数十毫秒的计算延迟会导致解释生成的速度严重滞后于物理环境的剧烈变化, 使解释结果失去指导实时决策的有效性。例如, Rezazadeh 等提出的 TANGO 框架虽通过符号推理提升了可解释性, 但其变分贝叶斯计算大幅增加了训练与推理环节的时间成本^[24]。因此, 如何在保持现有解释深度的前提下, 有效压缩算法复杂度以适配 6G 的严格实时性标准, 是当前落地应用需要解决的首要瓶颈。

2) 不同 XAI 方法处理同一模型, 解释结果缺少一致性, 也没有统一可信度评估标准, 生成

表 6 XAI 在跨越智能协作与零接触运维场景中典型应用

典型场景	XAI 技术	贡献	局限
资源管理和 RAN 切片	注意-based 解释器、SHAP-like attribution	解决任务卸载决策不明朗问题, 提升协作信任度和切片效率	解释计算增加延迟 (约 10ms 开销); 跨代理解释标准化问题
AI-and-RAN 工作负载编排框架	LIME、LLM	解决运维干预和响应时间问题, 实现自动化管道和自治水平提升	XAI 引入额外延迟; 报告生成的一致性挑战
EXPLORA 框架	集成梯度	解决协作决策透明问题, 支持自动化报告和跨层解释	解释计算开销影响实时性; 不同方法不一致
xTL 框架	注意机制	解决通信开销和协作效率问题, 提供语义级解释	语义提取增加复杂度; 解释标准化缺失



的解释内容完全不同,运维决策会陷入混乱。举例来说,SHAP计算全局特征重要性的结果,同LIME局部线性近似结果放在一起,处理深度强化学习这类复杂模型时,二者结论经常无法匹配^{[45][46]}。运维人员无法判断哪一种解释更贴合模型的真实运行机理,网络切片运维场景或是安全增强场景中,这类不一致性会扩大操作风险,XAIomaly框架给出的异常归因结论,可能和GRACE梯度过滤输出的结果发生冲突^{[10][26]},当前学术界工业界,还没有统一指标模式,可用来量化XAI的输出质量,涉及解释保真度、鲁棒性、稳定性这类维度都没有统一规则,XAI无法进入标准化部署阶段。

3) 现有研究没有围绕多层级的解释展开,难以支撑实现跨层级协作。跨层级解释的协同机制受O-RAN接口与协议隔阂约束,O-RAN采用分层架构,包括Non-RT RIC和Near-RT RIC,解释信息需要在A1、E2等接口之间完成传递与融合^[47-49]。现有研究大多只围绕单一层级的解释展开,智能协作场景中,边缘节点这类底层模块的决策逻辑,上层模块如SMO无法完整读取。XRL框架的卸载解释需要从边缘节点传输到SMO层,现有技术缺少高效标准化协议支撑这类跨层级语义交互,会出现信息丢失或同步延迟,零接触运维发展过程中,这类通信瓶颈会拉低闭环自治的实现效率。

4) XAI在提高解释透明度的同时会带来数据隐私泄露风险增大和安全防御要求提高等方面的问题。XAI提高透明度过程中,它输出的详细解释内容,可能在无意识状态下泄露敏感数据或模型结构。联邦学习运行场景中,攻击者可借助高精度特征归因,推断用户行为模式,反演训练数据^[3],AI-on-RAN框架借用LLM生成自然语言报告,配合相关工作完善网络安全防护时,报告内容提高方案可操作性,也会给逆向工程攻击创造入口^[3]。O-RAN具备开放接口特性,开放接口

本身允许更多外部主体接入网络,如何给出足够决策透明度,同时搭建有效隐私屏障阻挡对抗性攻击,是实现高可靠6G网络必须解决的内在冲突。

4.2 发展趋势

为应对前述挑战,6G O-RAN中XAI的未来演进将聚焦高效融合与闭环自治,推动其由辅助诊断工具向网络核心控制引擎跃升。具体发展方向集中于以下四点:

1) 面向边缘计算与实时响应的轻量化XAI设计。针对O-DU与Near-RT RIC等边缘节点算力受限的挑战,需开发低复杂度解释机制。除通过模型蒸馏或近似计算简化SHAP等高算力事后归因算法外,未来研究将从基于实例或梯度的局部解释(如BAE引擎、EXPLORA框架),进一步向融合内嵌注意力机制的内在可解释性方向演进,以彻底消除复杂事后解释器带来的延迟瓶颈。

2) 神经符号推理与深度学习深度融合。纯数据驱动模型往往缺乏严密逻辑,而神经符号推理能显著增强解释的因果性。未来应将符号推理与深度学习紧密结合,例如,TANGO框架通过融合贝叶斯GNN与规则推理展示了领域知识嵌入黑盒模型的路径^[24]。该方向将在网络切片与协作决策领域延伸,引入因果推断替代传统相关性分析,挖掘决策深层成因,彻底打破黑盒模型的信任壁垒。

3) 驱动“检测-解释-优化”闭环自治机制。XAI将深度嵌入O-RAN控制闭环,实现全流程自动化运作。EXPLORA框架已展示通用解释在零接触运维中的潜力^[21],未来系统可进一步整合LLM,生成可执行的代码或配置指令以构建自愈合管道。此类方向已在AI-and-RAN工作负载编排领域展开探索^[6],将强力驱动网络零接触管理的落地。

4) 适配异构环境的跨学科理论融合。为应

对 6G 网络的高度复杂性, XAI 需吸纳博弈论、量子计算与控制论等跨领域方法^[50-53]。具体而言, 在网络安全领域, 量子辅助 XAI 提升了 FDI 检测速度与透明度, 并增强了应对对抗性数据投毒攻击的鲁棒性^[32]; 在资源管理环节, 引入语义感知 DRL 实现了多智能体通信开销压缩与语义级解释强化^[23]。跨学科融合不仅将强化 AI 在空天地一体化等异构网络中的泛化能力, 更将助力业界制定统一的电信级 XAI 量化评估标准, 加速可信 6G 网络的规模化部署。

5 结束语

本文系统梳理了 XAI 与 6G O-RAN 的融合路径和应用场景。首先介绍了 XAI 在 O-RAN 中的部署框架与逻辑, 这是两者深度融合的基础。接下来分析了 XAI 在 O-RAN 四大核心领域的应用: 第一类应用是无线资源管理, 现有 TANGO、BAE、GRACE 三类方案可提高资源分配和信道估计效率; 第二类应用是网络切片运维, 可完成透明审计; 第三类应用是网络安全, 现有 XAIomaly 框架和 AI-on-RAN 框架可强化异常检测和响应; 第四类应用是智能协作与零接触运维, 现有 XRL 和 EXPLORA 可支持任务卸载和闭环自治。这些应用展现了 XAI 在提高透明度、降低运维成本、增强鲁棒性方面的优势。

XAI 虽然已经取得初步成效, 但仍然需要平衡解释效率和实时性, 且在一致性评估、跨层级协同、隐私安全等方面存在挑战, 这些问题在 Near-RT RIC 毫秒级响应和联邦学习场景中更为突出。本文通过对现有研究的梳理揭示了一个根本性的范式转变: 即 XAI 必须从单纯的事后诊断工具转变为嵌入智能闭环的内在可解释性设计原则。

为呼应这一范式转变, 后续研究方向应紧扣内在可解释与闭环控制集中发力: 一是搭建面向智能闭环的 XAI 标准化框架, 打破现有接口壁垒

以实现跨层级的解释协同; 二是整合跨领域技术路径, 如探索量子计算辅助 XAI 与语义感知的融合, 以突破实时性算力瓶颈并构建原生可解释模型; 三是在真实应用场景中完成端到端的自治闭环实验验证。上述方向的探索将推动 XAI 在 6G 异构环境中落地, 适配包括扩展现实与工业 5.0 在内的各类新兴应用, 最终助力智能、透明、自治的下一代无线网络逐步成型。

参考文献:

- [1] BRIK B, BOUTIBA K, DAHMANI A, et al. Explainable AI in 6G O-RAN: A tutorial and survey on architecture, use cases, challenges, and future research[J]. IEEE Communications Surveys & Tutorials, 2025, 27(5): 2826-2859.
- [2] WANG C X. On the road to 6G: Visions, requirements, key technologies and testbeds[J]. IEEE Communications Surveys & Tutorials, 2023, 25(2): 905-974.
- [3] ZHOU D, SHENG M, LI J, et al. Aerospace integrated networks innovation for empowering 6G: A survey and future challenges[J]. IEEE Communications Surveys & Tutorials, 2023, 25(2): 975-1019.
- [4] MAO B, LIU J, WU Y, et al. Security and privacy on 6G network edge: A survey[J]. IEEE Communications Surveys & Tutorials, 2023, 25(2): 1095-1127.
- [5] CORONADO E, et al. Zero touch management: A survey of network automation solutions for 5G and 6G networks[J]. IEEE Communications Surveys & Tutorials, 2022, 24(4): 2535-2578.
- [6] 王思野, 高萃雨, 赵中原, 等. 6G 与 XAI 的融合[J]. 移动通信, 2024, 48(8): 111-117.
- [7] 杨志航. 算法透明实现的另一种可能: 可解释人工智能[J]. 行政法学研究, 2024(3): 154-163.
- [8] 杜人淮, 马会君. 人工智能的一体化发展及其水平评价研究[J]. 贵州大学学报(自然科学版), 2025, 42(6): 1-11.
- [9] 包义明, 林阳, 屠要峰. 人工智能技术与应用前沿[J]. 中兴通讯技术, 2025, 31(4): 64-74.
- [10] BASARAN T, DRESSLER F. XAIomaly: Explainable, interpretable and trustworthy AI for xURLLC in 6G Open-RAN[C]// 2024 3rd International Conference on 6G Networking (6GNet). Paris, France: IEEE, 2024: 93-101.
- [11] FIANDRINO C, BONATI L, D'ORO S, et al. EXPLORA: AI/ML EXPLAINability for the open RAN[J]. Proceedings of the ACM on Networking, 2023, 1(CoNEXT3): 19.



- [12] 解宇瑄, 李响, 李婷, 等. 6G无线接入网通算智融合关键技术与标准化思考[J/OL]. 中兴通讯技术, 1-13[2026-04-15].
- [13] 索士强, 许盛浩, 曾婷, 等. 面向6G的通感空口融合无线接入网方案[J]. 电信科学, 2022, 38(9): 71-76.
- [14] 王晴天, 刘洋, 刘海涛, 等. 面向6G的网络智能化研究[J]. 电信科学, 2022, 38(9): 151-160.
- [15] FIANDRINO C, ATTANASIO G, FIORE M, et al. Toward native explainable and robust AI in 6G networks: Current state, challenges and road ahead[J]. *Computer Communications*, 2022, 193: 47-52.
- [16] JUNEJO S, SHAIKH F K, CHOWDHRY B S, et al. Role of AI and Open RAN in 6G networks: Performance impact and key technologies[C]//2024 IEEE International Conference on Advanced Telecommunication and Networking Technologies (ATNT). Johor Bahru, Malaysia: IEEE, 2024: 1-4.
- [17] AL KASSIR H, MASOOD A, KAZMI S H A, et al. Automated end-to-end AI/ML lifecycles for radio management in 6G networks[C]//2024 IEEE Future Networks World Forum (FNWF). Dubai, United Arab Emirates: IEEE, 2024: 365-371.
- [18] KHAN N, COLERI S, ABDALLAH A, et al. Explainable and robust artificial intelligence for trustworthy resource management in 6G networks[J]. *IEEE Communications Magazine*, 2024, 62(4): 50-56.
- [19] LIATSAS L, KIBALYA G M, ANTONOPOULOS A. XAI-driven model design for resource utilization forecasting in cloud-native 6G networks[C]//2024 IEEE International Mediterranean Conference on Communications and Networking (MeditCom). Madrid, Spain: IEEE, 2024: 566-571.
- [20] O-RAN Alliance. O-RAN: Towards an open and smart RAN [R]. Alfter, Germany: O-RAN Alliance, 2018.
- [21] ABETA S, KAWAHARA T, UMESH A, et al. O-RAN alliance standardization trends[J]. *NTT DOCOMO Technical Journal*, 2019, 15(1): 38-45.
- [22] QIN P, FU Y, ZHANG J, et al. DRL-based resource allocation and trajectory planning for NOMA-enabled multiUAV collaborative caching 6G network[J]. *IEEE Transactions on Vehicular Technology*, 2024, 73(6): 8750-8764.
- [23] WU M, GUO K, LI X, et al. Optimization design in RIS-assisted integrated satellite-UAV-served 6G IoT: A deep reinforcement learning approach[J]. *IEEE Internet Things Magazine*, 2024, 7(1): 12-18.
- [24] REZAZADEH F, ROY S, CHERGUI H, et al. Toward explainable reasoning in 6G: A proof of concept study on radio resource allocation[J]. *IEEE Open Journal of the Communications Society*, 2024, 5: 6239-6260.
- [25] KHAN M, ABDALLAH A, CELIK A, et al. Explainable and robust millimeter wave beam alignment for AI-native 6G networks[C]//ICC 2025 - IEEE International Conference on Communications. Montreal, QC, Canada: IEEE, 2025: 753-758.
- [26] GIZZINI A K, MEDJAHDI Y, MABROUK M B. GRACE: Gradient-based XAI scheme for channel estimation in wireless communications[C]//2024 IEEE International Mediterranean Conference on Communications and Networking (MeditCom). Madrid, Spain: IEEE, 2024: 572-577.
- [27] PHYU H P, NABOULSI D, STANICA R. Machine learning in network slicing—A survey[J]. *IEEE Access*, 2023, 11: 39123-39153.
- [28] RAHMAN M H, KARIM F, ZULKARNAIN, et al. Leveraging O-RAN SC AI/ML framework & non-RT RIC for AI-driven network slice QoS optimization[C]//2024 IEEE Globecom Workshops (GC Wkshps). Cape Town, South Africa: IEEE, 2024: 1-6.
- [29] ROY S, CHERGUI H, VERIKOUKIS C. Toward bridging the FL performance-explainability tradeoff: A trustworthy 6G RAN slicing use-case[J]. *IEEE Transactions on Vehicular Technology*, 2024, 73(7): 10529-10538.
- [30] ROY S, REZAZADEH F, CHERGUI H, et al. Joint explainability and sensitivity-aware federated deep learning for transparent 6G RAN slicing[C]//ICC 2023 - IEEE International Conference on Communications. Rome, Italy: IEEE, 2023: 1238-1243.
- [31] 刘杨, 彭木根. 6G内生安全: 体系结构与关键技术[J]. 电信科学, 2020, 36(1): 11-20.
- [32] CHATZIMILTIS S, SHOJAFAR M, MASHHADI M B, et al. AI-on-RAN for cyber defense: An XAI-LLM framework for interpretable anomaly detection[J]. *IEEE Transactions on Network Science and Engineering*, 2025.
- [33] PORAMBAGE P, PINOLA J, RUMESH Y, et al. XcARet: XAI based green security architecture for resilient open radio access networks in 6G[C]//2023 Joint European Conference on Networks and Communications & 6G Summit (EuCNC/6G Summit). Gothenburg, Sweden: IEEE, 2023: 699-704.
- [34] KAUR N, GUPTA L. Enhancing IoT security in 6G environment with transparent AI: Leveraging XGBoost, SHAP and LIME[C]//2024 IEEE 10th International Conference on Network Softwarization (NetSoft). Saint Louis, MO, USA: IEEE, 2024: 180-184.
- [35] MEHTA D, JAMEEL F, ALI A, et al. Quantum-assisted XAI-driven DL framework for FDI detection in connected autonomous electric vehicles underlying 6G[C]//2025 5th Intelligent Cybersecurity Conference (ICSC). Tampa, FL, USA: IEEE, 2025: 349-354.
- [36] SHAH S D A, HAFEEZ M, SALAMA A, et al. Proactive AI-

- and-RAN workload orchestration in O-RAN architectures for 6G networks[J]. IEEE Open Journal of the Communications Society, 2025, 6: 7939-7954.
- [37] GUO J, WANG S. xTL: Reducing communication overhead with XAI-guided, semantic-aware DRL for urban traffic light control[C]//2025 IEEE Intelligent Vehicles Symposium (IV). Cluj-Napoca, Romania: IEEE, 2025: 2033-2038.
- [38] MHATRE S, ADELANTADO F, RAMANTAS K, et al. Enhancing AI transparency: XRL-based resource management and RAN slicing for 6G O-RAN architecture[C]//ICC 2025 - IEEE International Conference on Communications. Montreal, QC, Canada: IEEE, 2025: 5755-5760.
- [39] ŞENGÖZ N, YİĞİT T. Towards third generation AI: Explainable and interpretable AI[C]//2022 7th International Conference on Computer Science and Engineering (UBMK). Diyarbakir, Turkey: IEEE, 2022: 523-526.
- [40] SUN H, ZHANG P, WANG Y, et al. Advancing 6G: Survey for explainable AI on communications and network slicing[J]. IEEE Open Journal of the Communications Society, 2025, 6: 1372-1412.
- [41] 耿光磊, 高博, 熊轲, 等. 联邦学习赋能 6G 网络综述[J]. 物联网学报, 2023, 7(2): 50-66.
- [42] 王首峰, 郭建超, 边森. 原生 AI 融合网络数字孪生赋能下一代无线网络自治[J]. 中兴通讯技术, 2025, 31(5): 30-36.
- [43] 张彤, 任奕璟, 闫实, 等. 人工智能驱动的 6G 网络: 智慧内生[J]. 电信科学, 2020, 36(9): 14-22.
- [44] ABOUELMAATI D, ESFAHANI A, GOUDAZI S, et al. Empowering Next-Gen Networks: AI-Driven Autonomy in O-RAN and SON Architectures[J]. IEEE Access, 2025, 13: 186903-186936.
- [45] ANTWARG L, MILLER R M, SHAPIRA B, et al. Explaining Anomalies Detected by Autoencoders using Shapley Additive Explanations[J]. Expert Systems with Applications, 2021, 186: 115736.
- [46] BLANCO L, OVIEDO J A, GARCÍA-AVILÉS D, et al. UNITY-6G: UNified archITecture for Open RAN enabled distributed, scalable and sustainability enhanced 6G networks[C]//2025 Joint European Conference on Networks and Communications & 6G Summit (EuCNC/6G Summit). Poznan, Poland: IEEE, 2025: 440-445.
- [47] CORONA J, RODRIGUES P, ALMEIDA L, et al. From black box to transparency: Consistency and cost within XAI[C]//2024 IEEE Globecom Workshops (GC Wkshps). Cape Town, South Africa: IEEE, 2024: 1-6.
- [48] LIANG X, WANG Q, AL-TAHMEESSCHI A, et al. Energy consumption of machine learning enhanced Open RAN: A comprehensive review[J]. IEEE Access, 2024, 12: 81889-81910.
- [49] RAGO A, FASCISTA A, PIRO G, et al. 6GxAID: Integrating 6G networks and eXplainable artificial intelligence for drone-based assistance in emergency situations[C]//2025 25th Anniversary International Conference on Transparent Optical Networks (ICTON). Barcelona, Spain: IEEE, 2025: 1-4.
- [50] RENDA A, DUCANGE P, GALLO G, et al. XAI models for quality of experience prediction in wireless networks[C]//2021 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE). Luxembourg, Luxembourg: IEEE, 2021: 1-6.
- [51] 黄海松, 朱云伟, 张国章. 人机协作技术的发展与应用研究综述[J]. 贵州大学学报(自然科学版), 2025, 42(4): 1-8.
- [52] 姚丽茜, 陈建辉, 车艳, 等. 基于区块链和 5G 网络的物联网云联健康管理系统研究[J]. 贵州大学学报(自然科学版), 2022, 39(3): 85-92.
- 廖勇 (1982-), 男, 四川自贡人, 博士, 副研究员, CCF 杰出会员, 主研 6G 通信、人工智能技术及其应用;
- 杨善云 (2005-), 男, 重庆合川人, 本科生, 电子信息工程专业。