Grant Agreement No.: 101192650

Type of action: HORIZON JU Research and Innovation

Topic: HORIZON-JU-SNS-2024-STREAM-B-

Call: HORIZON-JU-SNS-2024

unity-6G

D2.2 TRUST MODEL AND TRUST MANAGEMENT APPROACHES

Work package	WP2
Task	T2.4
Due date	31/01/2026
Submission date	31/01/2026
Deliverable lead	LINKS
Version	1.0
Dissemination Level	Public
Editor	Cristian Brunetto (LINKS)
Authors	Cristian Brunetto (LINKS), Maria A. Serrano (NBC), Alice Piemonti (MAR), Gianluca Fontanesi (NokiaBL), Paweł Hatka (PUT), Hanna Bogucka (PUT), Harilaos Koumaras (NCSRD), Stefanos Plastras (NCSRD), Rafael Cavalcanti (KEY), Matteo Pagin (KEY), Farhana Javed (CTTC), Engin Zeydan (CTTC), Josep Mangues (CTTC), Jorge Baranda (CTTC), Cristian Vaca (CTTC)
Internal Reviewers	Maria A. Serrano (NBC), Gianluca Fontanesi (NokiaBL), Adrian Kliks (PUT), Stefanos Plastras (NCSRD)
External Reviewers	Javier Velazquez Martinez (TID), Oriol Font (SRS)

Abstract	This deliverable provides an analysis of trust models and management strategies for 6G networks, focusing on security, reliability, and resilience in highly dynamic and distributed environments. It examines trust relationships among heterogeneous entities, their impact on interoperability, data protection, and service continuity, and proposes guidelines for assessing, establishing, and maintaining trust. The work aligns with UNITY-6G objectives by integrating AI-assisted mechanisms and DLT-based evidence to enable verifiable and adaptive trust management across multi-domain scenarios.
Keywords	Trustworthiness, LoT, DLT, AI, ZTA, O-RAN

Document Revision History

Version	Date	Description of change	List of contributor(s)
V0.1	14/10/2025	1st version of the table of contents	Cristian Brunetto (LINKS)
V0.2	3/11/2025	Inputs start	All the authors
V0.3	8/12/2025	Inputs end	All the authors
V0.4	9/12/2025	Ethical Revision	All the reviewers
V0.5	10/12/2025	Revision started	All the reviewers
V0.6	19/12/2025	Revision end	All the reviewers
V0.7	28/01/2026	Final Version	Engin Zeydan (CTTC), Javier Velazquez Martinez (TID), Cristian Brunetto (LINKS)
V1.0	30/01/2026	Submitted version	

DISCLAIMER

Co-funded by the European Union

6GSNS

Project funded by
Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra
Swiss Confederation
Federal Department of Economic Affairs, Education and Research EAER
State Secretariat for Education, Research and Innovation SERI

The **unity-6G** project has received funding from the Smart Networks and Services Joint Undertaking (SNS JU) under the European Union's Horizon Europe research and innovation programme under Grant Agreement No *101192650*. This work has received funding from the Swiss State Secretariat for Education, Research, and Innovation (SERI).

COPYRIGHT NOTICE

EXECUTIVE SUMMARY

This document addresses trust management in highly heterogeneous, dynamic, and AI-driven environments for future 6G networks. Unlike traditional security paradigms, trust in 6G extends beyond authentication and encryption, handling integrity, resilience, privacy, and accountability across multiple domains and stakeholders. This is essential for ensuring secure and reliable interactions in scenarios where openness, disaggregation, and distributed intelligence are fundamental design principles.

The document begins by outlining the motivation for trust in 6G and its fundamental role for achieving a secure connectivity. It then provides a review of the State of the Art (SotA), including research initiatives, relevant European Union (EU) projects, and standardization initiatives such as National Institute of Standards and Technology (NIST) Zero Trust Architecture (ZTA), European Telecommunications Standards Institute (ETSI) specifications for permissioned distributed ledgers, and Open Radio Access Network (O-RAN) security guidelines. These references establish the baseline for designing trust frameworks aligned with global standards and interoperability requirements.

Building on this foundation, D2.2 introduces conceptual and functional architectures for the UNITY-6G trust domain, detailing design principles, functional blocks, and enabling technologies. Key components include trust adapters, orchestration modules, and a dedicated trust layer leveraging Distributed Ledger Technologies (DLTs), smart contracts (SCs), and blockchain oracles to provide verifiable and auditable trust guarantees. These mechanisms aim to support multi-domain orchestration, Service Level Agreement (SLA) monitoring, and secure collaboration among distributed Artificial Intelligence (AI) agents.

The deliverable also proposes a general trust model, introducing concepts such as Level of Trust (LoT) and AI-assisted trustworthiness evaluation, complemented by Explainable AI (XAI) techniques to ensure transparency and accountability. To demonstrate its applicability, the document explores scenario-specific customizations for Internet of Things (IoT), O-RAN, and Wi-Fi environments, highlighting how DLT-enabled federated learning (FL) and reputation-based mechanisms can enhance trust in distributed AI workflows and multi-vendor ecosystems.

Rather than defining a single architecture, D2.2 presents multiple architectural options and design guidelines, reflecting the complexity and evolving nature of trust in 6G systems. This approach provides the conceptual and technical groundwork for subsequent implementation and validation activities. In doing so, the deliverable contributes directly to Work Package 2 and specifically Task 2.4, which focuses on defining trust models and management strategies for heterogeneous 6G environments.

The insights and frameworks presented in this document will serve as a reference for the future of the project, ensuring that trust management becomes a native capability of the UNITY-6G architecture.

LIST OF FIGURES

LIST OF TABLES

ABBREVIATIONS

3GPP	Third Generation Partnership Project
ACL	Access Control List
AGSP	Aggregator Service Provider
AI	Artificial Intelligence
API	Application Programming Interface
AS	Autonomous System
BFT	Byzantine Fault Tolerance
CLSP	Client Service Providers
CTI	Cyber Threat Intelligence
DID	Decentralized Identifier
DLT	Distributed Ledger Technology
DLT-TF	DLT Trust Functions
DTL	Dynamic Trust Level
DTE	Distributed Trustable AI Engine
DU	O-Distributed Unit
eBPF	extended Berkeley Packet Filter
EU	European Union
EVM	Ethereum Virtual Machine
ETSI	European Telecommunications Standards Institute
FaaS	Function as a Service
FedAvg	Federated Averaging
FG DLT	Focus Group on Distributed Ledger Technology
FL	Federated Learning

IBI	Intent-Based Interface
IDMO	Integrated (Cross-Domain) Management Orchestrator
IMT-2030	International Mobile Telecommunication 2030
IP	Internet Protocol
IPsec	Internet Protocol Security
ISG PDL	Industry Specification Group for Permissioned Distributed Ledgers
ITU-T	International Telecommunications Union - Telecommunications Standardization Sector
LDAP	Lightweight Directory Access Protocol
LIME	Local Interpretable Model-agnostic Explanations
LoT	Level of Trust
LoTw	Level of Trustworthiness
MEC	Multi-Access Edge Computing
ML	Machine Learning
MNO	Mobile Network Operator
mMTC	massive Machine-Type Communications
mTLS	mutual Transport Layer Security
MQTT	Message Queuing Telemetry Transport
MTD	Moving Target Defense
Near-RT RIC	Near-Real-Time RAN Intelligent Controller
NIST	National Institute of Standards and Technology
NMSE	Normalized Mean Squared Error
Non-RT RIC	Non-Real-Time RAN Intelligent Controller
NTN	Non-Terrestrial Network
O-CU	ORAN Central Unit

OIDC	OpenID Connect
O-RAN	Open Radio Access Network (O-RAN)
OSS/BSS	Operations Support Systems / Business Support Systems
PDL	Permissioned Distributed Ledger
PDP	Policy Decision Point
PE	Policy Engine
PEP	Policy Enforcement Point
PoC	Proof of Concept
PoT	Proof of Transit
QoS	Quality of Service
RBAC	Role-Based Access Control
RF	Radio Frequency
RIC	RAN Intelligent Controller
RSN	Robust Security Network
RSNA	Robust Security Network Association
RU	O-Radio Unit
SBA	Service Based Architecture
SC	Smart Contract
SCM	Smart Contract Manager
SCO	Smart Contract Orchestrator
SGX	Intel Software Guard Extensions
SHAP	SHapley Additive exPlanations
SLA	Service Level Agreement
SMO	Service Management and Orchestration

SNS JU	Smart Networks and Services Joint Undertaking
SOAR	Security Orchestration, Automation and Response
SotA	State of the Art
TDX	Intel Trust Domain Extensions
TEE	Trusted Execution Environment
TN	Terrestrial Network
UE	User Equipment
UEBA	User and Entity Behavior Analytics
URLLC	UltraReliable LowLatency Communications
VNF	Virtual Network Function
VPN	Virtual Private Network
WG	Work Group
XAI	Explainable Artificial Intelligence
xApp	NearRT RIC application (ORAN)
rApp	NonRT RIC application (ORAN)
XDP	Express Data Path
ZTA	Zero Trust Architecture
ZTMM	Zero Trust Maturity Model

1 INTRODUCTION

Trust has become a cornerstone for next-generation mobile networks, particularly in the context of 6G. The openness, disaggregation, and AI-native control loops envisioned for 6G introduce a need for trust that goes beyond traditional security paradigms. In highly dynamic, multi-domain environments, trust is not limited to identity verification; it encompasses integrity, reliability, privacy, and resilience across heterogeneous entities and services. This evolution requires systematic approaches to model, evaluate, and enforce trust throughout the entire service lifecycle.

This deliverable addresses these challenges by analysing trust models and management strategies specifically in the context of 6G networks. It explores how trust relationships influence interoperability, data protection, and service continuity, and how Distributed Ledger Technologies (DLTs) and Artificial Intelligence (AI)-assisted mechanisms can provide verifiable, auditable, and adaptive trust guarantees. The document is structured to first review the state of the art (SotA) in trust-related research and standardization, then introduce conceptual and functional architectures for trust domains, and finally present scenario-specific customizations such as DLT-enabled federated learning for Internet of Things (IoT) and Open Radio Access Network (O-RAN) environments. Rather than describing a single solution, the deliverable proposes multiple architectural options and design principles that will inform subsequent development and integration activities.

The deliverable consolidates insights from standards, research initiatives, and emerging technologies to support the objectives of WP2 to enable secure, reliable, and explainable trust management.

The content is organized as follows:

- **Chapter 2** reviews the SotA in trust-related research, relevant European Union (EU) projects, and standardization efforts (National Institute of Standards and Technology (NIST) Zero Trust Architecture (ZTA), European Telecommunications Standards Institute (ETSI) Permissioned Distributed Ledgers (PDL), O-RAN security).
- **Chapter 3** introduces the UNITY-6G trust domain architecture, describing its design principles, functional blocks, and integration of DLT components such as SCs and blockchain oracles.
- **Chapter 4** presents the trust model, outlining general approaches to trustworthiness, Level of Trust (LoT), and AI-assisted mechanisms, followed by scenario-specific customizations for IoT, O-RAN, and Wi-Fi environments.
- **Chapter 5** discusses the road ahead for trust modelling, identifying open challenges and future directions for implementing a unified trust layer in 6G systems.

By providing these analyses and architectural options, the deliverable lays the groundwork for subsequent steps in UNITY-6G, where the proposed concepts will be refined into concrete implementations and validated through experimental scenarios.

2 RELEVANT STATE OF THE ART ON TRUST DOMAIN

2.1 RESEARCH ACTIVITIES IN TRUST DOMAIN

Trust has become a central pillar of both O-RAN and 6G security architectures, driven by the unprecedented disaggregation, multi-vendor openness, and AI-native control loops that define these systems. SotA works fall into at least one of the following categories:

- Trust in network components
- Trust models / ZTA
- Trust for AI/Machine Learning (ML) in RAN Intelligent Controller (RIC) / xApps
- Trust for multi-vendor ecosystems
- Trust for distributed systems (federation, 6G)
- Trust for DLT / blockchain
- Trust for software lifecycle of apps (e.g. xApps/rApps)

Trust in Network Components - Anchoring Security in Disaggregated Infrastructure: Trust in O-RAN infrastructure must be anchored in a Hardware Root of Trust (RoT) , such as a Trusted Platform Module (TPM), to secure Commercial Off-The-Shelf (COTS) servers and white-box radios [1], [39].
Trust Models and ZTA: O-RAN and 6G are adopting ZTA based on the "Never Trust, Always Verify" principle (NIST SP 800-207) [15]. A Service Management and Orchestration (SMO) platform typically acts as the central Policy Decision Point (PDP), utilizing threat intelligence to make access decisions. Distributed Policy Enforcement Points (PEPs) at the O-RAN Central Unit (O-CU), O-Distributed Unit (DU), and RICs (RAN Intelligent Controller) enforce these policies [2][3]. A critical challenge addressed by recent research is the "runtime compliance gap". Standard ZTA verifies trust at entry; however, advanced frameworks propose continuous monitoring (e.g., "Trusted Reporter" in the O-Radio Unit (RU)) to detect compromise after the initial connection is established [1][4].
Trust for AI/ML in RIC / xApps: As the "brain" of the network, the RIC relies on AI/ML, introducing unique "cognitive" trust vulnerabilities. The ecosystem is vulnerable to data poisoning (manipulating training data to implant backdoors) and adversarial evasion (crafting inputs to deceive models). International Telecommunications Union-Telecommunications Standardization Sector (ITU-T) Y.3172 defines the architectural framework for securing these ML pipelines [5]. To establish trust in "black box" models, XAI techniques are being integrated into the security loop. XAI (Explainable Artificial Intelligence) helps operators understand why a model made a specific decision (e.g., a handover), allowing for the detection of anomalies driven by malicious inputs [5]. Trustworthiness at the AI layer is reinforced by recent studies, demonstrating how adversarial machine learning attacks can distort RIC decision logic or corrupt RAN intelligence, while also underscoring the need for rigorous trust validation of AI models and their lifecycle within cyber-range environments [19]. Collectively, these works highlight the critical importance of end-to-end trust modelling, spanning infrastructure, interfaces, AI workflows, and multi-domain orchestration, and reveal substantial gaps that next-generation cyber-ranges must address.

Trust for Multi-Vendor Ecosystems: The shift to a multi-vendor environment requires operators to act as systems integrators, managing trust across diverse supply chains. Global Plugfests and Open Testing and Integration Centers (OTICs) serve as trust-building exercises, validating that vendors correctly implement security specifications (e.g., secure interfaces) and ensuring that "open" interfaces do not rely on proprietary, undocumented extensions [40]. Operators are moving toward a "Zero Trust Supply Chain," requiring rigorous vendor vetting and the use of immutable logs to attribute configuration changes or faults to specific vendors, preventing "blame shifting" during incidents [7].

Trust for Distributed Systems (Federation, 6G): 6G envisions a hyper-distributed "network of networks" requiring federated trust models. Cross-certification and federated identity management allow entities from different domains (e.g., a satellite provider and a terrestrial Mobile Network Operator (MNO)) to verify each other without a single central authority [8]. Privacy-preserving Federated Learning enables models to be trained at the edge (e.g., on User Equipments (UEs)) without sharing raw data. Trustworthy aggregation algorithms are essential to filter out malicious updates from compromised nodes before they corrupt the global model [9]. Complementing this, a broader line of work extends blockchain-supported FL to resource-constrained IoT and emerging 6G environments, with a strong emphasis on lightweight design and alignment with ongoing standardization efforts [17]. Here, FL participation by heterogeneous IoT devices is mediated by DLT adapter, verifier, and aggregator components, where only hashes and compact metadata of model updates are anchored on a permissioned ledger (e.g., an IOTA-based Tangle [67]), preserving energy and bandwidth while still enabling end-to-end auditability and reputation tracking. By mapping this architecture onto Third Generation Partnership Project (3GPP), ETSI, ITU-T, and O-RAN work items, it directly targets trust for distributed systems and federation, trust for DLT/blockchain, and trust in AI workflows across domains. For next-generation cyber-ranges, these contributions offer concrete, standards-aware blueprints to instantiate federated, cross-domain trust experiments that span IoT endpoints, RIC/xApps, and 6G control infrastructures, closing part of the gap between abstract trust models and deployable, deployment-ready implementations.

Trust for DLT/Blockchain: DLT provides the immutable "system of record" necessary for multi-party trust. ETSI Industry Specification Group (ISG) PDL and ITU-T X.1400 series define standards for using DLT in telecommunications to ensure verifiable trust and interoperability [10][11]. DLT is applied to dynamic spectrum sharing, enforcing SLAs via SCs, and managing decentralized identities (Self-Sovereign Identity) for roaming users [12]. Complementary research on blockchain and DLTs introduce mechanisms for decentralized trust orchestration, immutable security logging, and cross-operator policy enforcement suited to multi-tenant and federated O-RAN/6G deployments [18].

Trust for Software Lifecycle of Apps (xApps/rApps): The software-defined nature of O-RAN introduces significant supply chain risks via third-party xApps and rApps. The Software Bill of Materials (SBOM) is a prerequisite for trust, providing a transparent inventory of all components and dependencies to allow for rapid vulnerability scanning [14]. A rigorous "Application Lifecycle Management" process is mandated, including digital code signing by the vendor and mandatory signature validation by the SMO before any xApp can be onboarded or instantiated on a RIC [40]. At the O-RAN level, recent work on blockchain-enabled reputation for federated learning shows how

decentralized trust can be embedded directly into the AI lifecycle of RIC/xApps in multi-vendor, multi-operator environments [1]. In that architecture, FL clients corresponding to different vendors or operators are onboarded via SCs, which then govern the submission of performance metrics and the computation of on-chain reputation scores that weight each client's impact on the aggregated model. By deploying this logic on an efficient Layer-2 Ethereum Virtual Machine (EVM) chain, the framework keeps latency and transaction overhead compatible with O-RAN control loops while providing immutable audit trails for model contributions, thereby addressing trust for AI/ML in RIC/xApps, trust for multi-vendor ecosystems, and trust for DLT/blockchain simultaneously. Such an implementation is particularly suitable for cyber-range scenarios where realistic, reputation-aware FL workflows need to be emulated under different attack and misbehaviour patterns, allowing a systematic assessment of how trust mechanisms affect both model quality and RAN control decisions.

2.1.1 Security and Trust Research in Distributed, AI-Driven 6G RAN Systems

Research activities in the trust domain focus on ensuring secure, reliable and verifiable operation of modern mobile networks, especially as architectures evolve toward increasingly open, virtualized and intelligent systems. The analysed materials show that both AI-driven decision-making and distributed network control introduce new trust challenges, requiring systematic research into robustness, integrity and resilience. As mobile systems progress toward 6G-class architectures, trust becomes essential across radio access, edge processing and cooperative learning environments.

A significant research area concerns trust in AI/ML algorithms integrated into network functions. Machine-learning models supporting tasks such as anomaly detection, resource allocation or interference prediction are exposed to adversarial manipulation. Attackers may poison training data, tamper with model updates, insert adversarial samples during inference or attempt to steal model parameters through accessible interfaces. These risks can degrade prediction accuracy, introduce hidden behaviours or leak sensitive information, underlining the need for robust, trustworthy and continuously monitored AI pipelines [21]. Research directions include secure model training, input validation, adversarial detection mechanisms and protection of ML-based services against model inversion, extraction and membership inference.

Another major research focus arises from FL, which is highlighted as a promising method for distributed intelligence in wireless environments. While FL enhances privacy by keeping data local, it remains vulnerable to attacks on training data, manipulation of model updates, inference leakage and disruption of communication during model exchange. Research efforts therefore investigate secure aggregation, anomaly detection for corrupted participants, differential privacy, homomorphic encryption and robust learning procedures capable of maintaining model quality even in adversarial conditions [20]. These topics are especially relevant for spectrum sensing and cognitive-radio scenarios, where FL-based cooperation directly affects spectrum availability and interference avoidance.

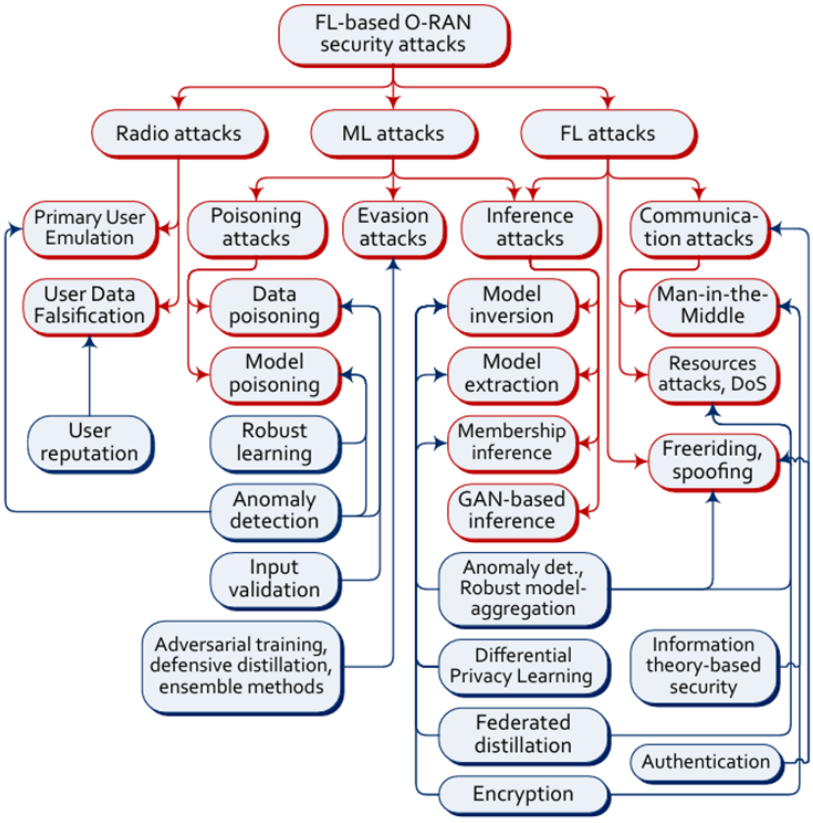

Figure 1 FL-based attacks and countermeasure classification.

At the radio and communication layer, trust is challenged by the openness of the wireless medium. Networks are exposed to jamming, spoofing, eavesdropping, signalling manipulation and energy-draining attacks. Such threats are particularly harmful for Ultra-Reliable Low-Latency Communications (URLLC) and IoT devices that rely on predictable latency and reliable connectivity. Research in this domain explores physical-layer security, lightweight authentication, radio-level anomaly detection and early-warning mechanisms deployed at the edge, leveraging distributed intelligence to detect malicious behaviour closer to its source [21].

Finally, the transition toward O-RAN and Multi-Access Edge Computing (MEC) introduces trust issues in programmable and distributed network infrastructures. Open interfaces and modular xApps/rApps broaden the attack surface and require research into secure Application Programming Interface (API) design, isolation of control logic and continuous verification of software components. Edge computing decentralizes processing, increasing exposure to attacks originating from various parts of the network and demanding stronger integrity protection and trustworthy orchestration mechanisms.

In summary, the development of trustworthy next-generation networks relies on a coherent set of priorities: securing AI/ML and FL pipelines, reinforcing trust at the radio layer, protecting programmable O-RAN interfaces, and securing distributed edge infrastructures. These activities form a foundation for trustworthy next-generation systems, where openness and intelligence must be coupled with mechanisms that ensure integrity, confidentiality and resilience across all layers of the network.

2.2 EU R&D INITIATIVES IN TRUST DOMAIN (EU AND OTHERS)

The Smart Networks and Services Joint Undertaking (SNS JU) 6G-IA Security Working Group's white paper, "Innovative Approaches for 6G Security" [27], outlines the main security challenges for 6G networks and proposes a portfolio of solutions. It presents a broad, forward-looking security vision centred on trustworthiness, privacy, and resilience in future 6G architectures, and serves as a roadmap for identifying gaps in current (5G/B5G) security approaches, while defining research clusters and project guidelines needed to build a secure, privacy-aware, resilient, and scalable 6G ecosystem. This white paper organizes R&D efforts around several pillars:

- New security frameworks, such as AI-driven "zero-trust" architectures and continuous trust evaluation.
- Decentralized and adaptive security solutions targeting the distributed, cloud-edge-IoT continuum typical of 6G.
- Use of advanced technologies (e.g., physical-layer security, post-quantum cryptography, anomaly detection) to strengthen networks against future threats.
- Scalable and zero-touch approaches, automation, orchestration, and "security by design" to enable seamless deployment of secure services across large, dynamic 6G infrastructures.

UNITY-6G operationalizes the 6G-IA pillars by enforcing a zero-trust posture with continuous Level-of-Trust (LoT/LoTw) evaluation from the Data Continuum; coordinating cross-domain, intent-based decisions via the SBMA/IDMO; anchoring identities, policies, and evidence in a DLT/PDL-backed Trust Layer (DIDs/VCs, smart contracts, oracle/adaptor) for verifiable and auditable assurance; and hardening AI/FL with XAI and reputation-aware mechanisms across O-RAN, IoT, and Wi-Fi. This closed loop enables zero-touch adaptation—placement, isolation, and reconfiguration—driven by LoT thresholds and smart-contract enforcement, ensuring consistent, explainable security at scale.

2.2.1 Trust Modelling and Trust Computation Projects

2.2.1.1 iTrust6G Project

The iTrust6G project [13] introduces a dynamic, behaviour-driven, explainable, and evidence-backed trust framework for 6G networks. Unlike traditional static or credential-based models, iTrust6G treats trust as a continuously updated state derived from the real-time behaviour of devices, users, and services. Its trust architecture is designed to ensure that only trustworthy entities gain access, even if they possess valid credentials, and to guarantee that every trust decision is transparent, auditable, and explainable. Figure 2 provides a high-level architecture of iTrust6G. The iTrust6G architecture integrates:

- Context Collection & Behaviour Monitoring
- Dynamic Trust Level Computation
- Explainability Module (XAI)
- Policy Decision & Enforcement Points (PDP/PEP)
- Transparent Evidence Repository (TER)

This creates a full trust loop: Monitor => Compute Trust => Decide Access => Explain => Log => Adapt. In iTrust6G, each entity (device, service, or user) is assigned a Dynamic Trust Level (DTL) based on: (i) behavioural patterns, (ii) contextual information (location, time, network conditions), (iii) Communication profile deviations and (iv) Signal integrity and contextual factors. This allows the network to detect suspicious behaviour even if the entity has valid authentication. Trust is also not computed from a single signal. iTrust6G fuses behavioural analytics, context (location, velocity, topology), User and Entity Behaviour Analytics (UEBA), like anomaly detection and signal integrity cues. This prevents attackers from poisoning just one vector. Every trust score and every access control decision is also paired with an explanation produced by Shapley Additive Explanations (SHAP)/Local Interpretable Model-agnostic Explanations (LIME) like methods or interpretable ML. This makes trust understandable, verifiable, and auditable.

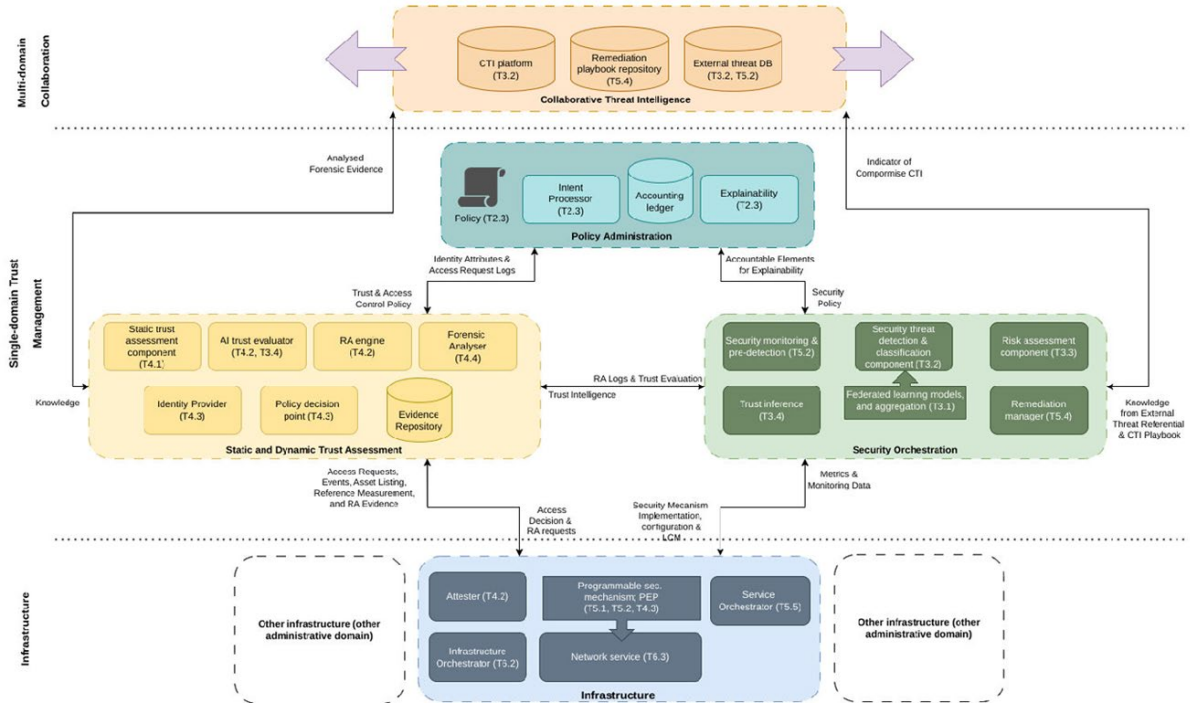

Figure 2 High level iTrust6G architecture [22]

2.2.1.2 ROBUST-6G Project

The ROBUST-6G project [23] models trust as an emergent property of autonomous, explainable, and privacy-preserving 6G network components, called macro-services. Instead of assigning trust only to users or devices, ROBUST-6G distributes trust computation across a network of cooperating agents that continuously collect local evidence, train federated models, and make context-aware trust decisions. Trust in ROBUST-6G is therefore distributed, explainable, and adaptive, designed to increase resilience and autonomy in 6G's heterogeneous, AI-driven environments. The ROBUST-6G project uses federated learning, local sensing, and physical-layer

attributes to compute explainable and privacy-preserving trust decisions. Its trust management framework uses multi-modal evidence, Radio Frequency (RF) sensing, anomaly detection, and reinforcement learning to evaluate trust locally, federated training to align trust globally, and distributed coordination to take trust-driven actions such as migration, isolation, and resilience optimization. Through XAI, privacy protection, and multi-agent autonomy, ROBUST-6G provides a scalable and accountable trust architecture for highly dynamic 6G environments.

2.2.1.3 NATWORK Project

The NATWORK project [24] treats trust as a dynamic, distributed, multi-domain property that continuously adapts to changing conditions in 6G networks. Unlike trust models centred on identity or behaviour alone, NATWORK embeds trust into orchestration logic, service placement, slice management, and network defence across multiple administrative domains. Its trust framework combines AI-driven Moving Target Defence (MTD), cross-domain policy negotiation, and federated threat intelligence to create a resilient and self-adaptive trust environment. NATWORK's design mirrors biological systems: trust is not static, it evolves, mutates, and adapts in real time to preserve system resilience. Trust is computed locally but aligned globally through Cyber Threat Intelligence (CTI) exchange and policy negotiation, while the orchestrator continuously reconfigures network functions and paths to maintain trust under dynamic threats. With MTD as a trust enforcer, federated reinforcement learning for trust optimization, and compliance-aware orchestration, NATWORK delivers a resilient, adaptive, and privacy-preserving trust model suited for multi-domain, heterogeneous 6G deployments. In summary, in NATWORK, the trust is a cross-domain, coordinated process enabled by AI-driven Moving MTD, inference engines, CTI exchange, and policy negotiation, emphasizing dynamic reconfiguration and federated resilience across multiple administrative domains.

2.2.1.4 RIGOUROUS Project

The RIGOUROUS project [25] embeds trust throughout the entire service lifecycle of 6G systems, starting with onboarding, extending through runtime monitoring, and enforced by automated policy actions as given in Figure 3.

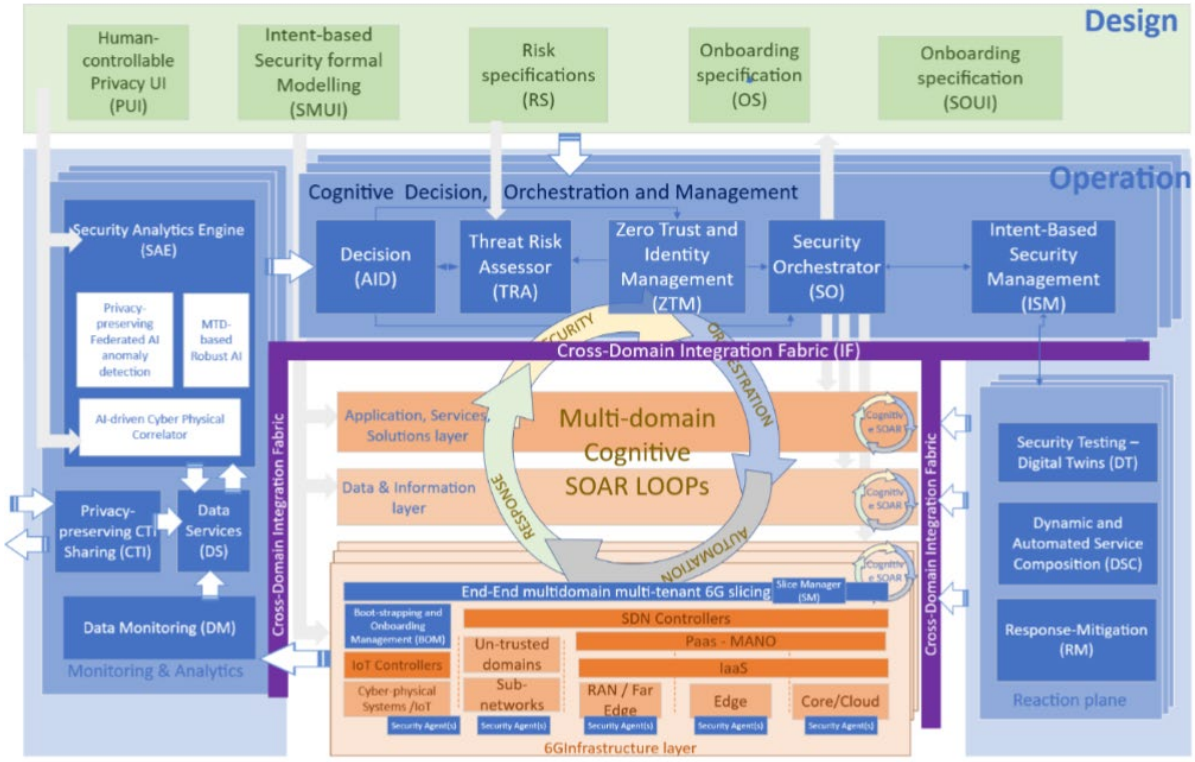

Figure 3 RIGOUROUS Final High-level Functional Architecture (HLFA) [28]

Instead of treating trust as a property of users or devices, RIGOUROUS defines trust as a system-level guarantee derived from privacy compliance, security integrity, and runtime behaviour consistency. Trust is achieved not through static identity verification, but through continuous validation that a service operates exactly as declared in its privacy and security policies. RIGOUROUS implements a privacy-driven trust model, defining trust as continuous compliance with declared privacy, organizational, and regulatory constraints. Trust is maintained through Privacy Manifests at onboarding, real-time Security Orchestration Automation and Response (SOAR)-based monitoring, Digital Twin simulations, and user-configurable privacy preferences. Any trust deviation triggers automated mitigation and is logged for auditability, making trust measurable, enforceable, explainable, and aligned with dynamic multi-domain 6G orchestration.

2.2.2 Trust-Enabling Architectures and System-Level Trust Frameworks

2.2.2.1 SAFE-6G

SAFE-6G [29] is a pioneering SNS JU project that redefines trust for 6G by introducing a comprehensive, user-centric trustworthiness paradigm. Rather than focusing solely on conventional security, SAFE-6G integrates and balances five critical domains, such as safety, security, privacy, resilience, and reliability, treating them as interconnected dimensions of a holistic trust framework tailored for the dynamic, human-centric 6G environment.

At the heart of SAFE-6G is an adaptable architectural framework that leverages advanced, distributed AI and machine learning to achieve cognitive coordination across trust dimensions. This framework interprets user or tenant intents in real time

and dynamically translates them into actionable configurations and policies, making it possible to fine-tune trustworthiness to align with specific requirements, use cases, or regulatory needs. By implementing a zero-touch, end-to-end orchestration of trust across onboarding, operation, and decommissioning, SAFE-6G enables elastic, scalable, and personalized security regimes for 6G systems.

Practical impact comes from the project's ability to deliver unprecedented agility and transparency in trust management. SAFE-6G's intent-based, AI-assisted trust coordination bridges the gap between high-level user expectations and actual network behaviours. This enables networks to continuously adapt their trust posture, ensuring not just robust security but also optimal usability and responsiveness for diverse applications, from the edge-cloud continuum to immersive metaverse scenarios. By placing trust (measured as a Level of Trust) as a core Key Value Indicator (KVI), SAFE-6G provides a blueprint for 6G ecosystems to close existing gaps in usability, resilience, and reliability, ushering in a new era of trustworthy digital infrastructure.

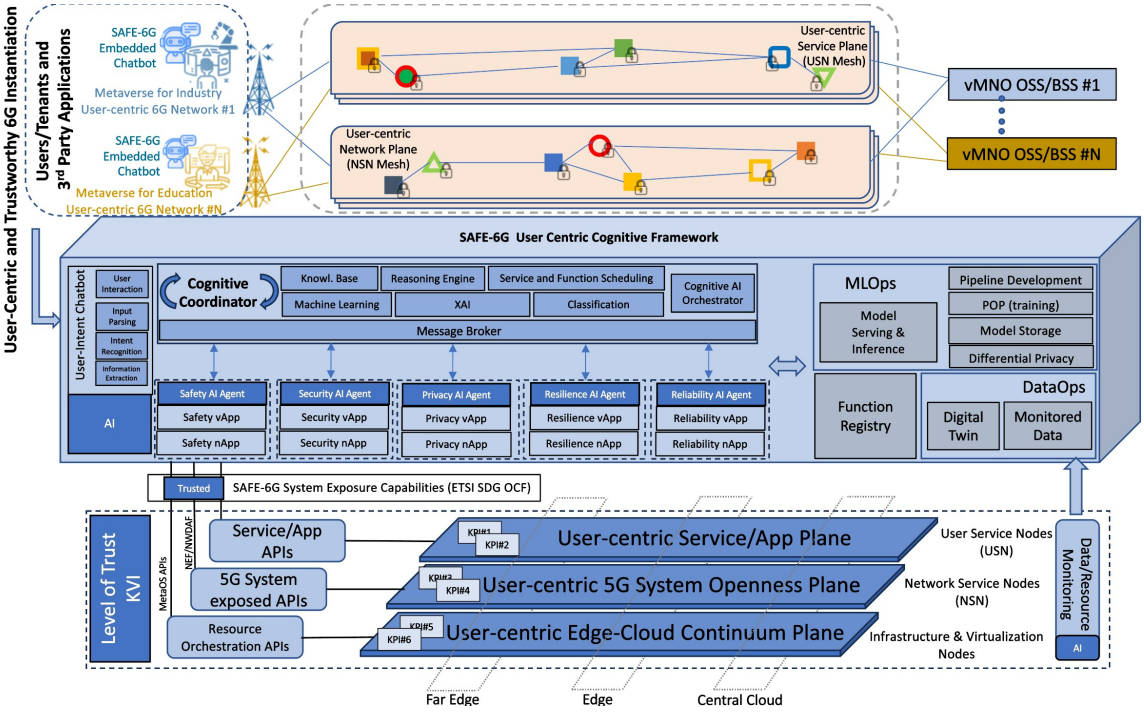

Figure 4 SAFE-6G Reference Architecture

2.2.2.2 MARE Project

MARE (Programmable, Modular and Disaggregated Security Plane for 6G Ecosystems) Project [26] aims to create a reliable 6G service provisioning platform by defining a novel security plane with open, programmable security functions. It focuses on proactively handling attacks and threats in 6G networks through enriched security services, known as Dynamic Open Trust (DOTs). These DOTs are advanced security primitives designed to be modular and programmable, allowing them to be combined into flexible security services that maximize protection across different network environments.

2.2.3 Trust-Enabling Architecture Projects

2.2.3.1 HORSE Project

The HORSE (Holistic, Omnipresent, Resilient Services for future 6G wireless and computing Ecosystems) project [31] provides an architectural blueprint that strongly aligns with the trust management goals addressed in UNITY-6G Task 2.4. Focused on end-to-end security and resilience for 6G networks, HORSE introduces a layered, cloud-native architecture that integrates threat intelligence, autonomous mitigation, and robust policy enforcement across heterogeneous domains. Its design is inherently trust-centric: the Smart Monitoring (SM) component, the Distributed Trustable AI Engine (DTE), the Data Pre-processing and the Detection and Mitigation Engine (DEME) cooperate to deliver accurate, trustworthy detection of anomalies and cyber threats. These modules feed into the Intent-Based Interface (IBI) and the Common Knowledge Base (CKB), where generative AI enriches attack-mitigation knowledge to support reliable and explainable decision-making. This aligns with UNITY-6G's requirement for secure and trustworthy interactions between distributed learning entities and for accurate monitoring data collection. In particular, HORSE anticipates the need for trust management by incorporating mechanisms that verify the reliability, provenance, and robustness of both data and AI-driven decisions. The Reliability, Trust, and Resilience Provisioning (RTR) module ensures that mitigation actions derived from the CKB are validated and transformed into enforceable network strategies, facilitating SLA-compliant and trustworthy orchestration in dynamic multi-domain environments, an objective shared with UNITY-6G's trustworthy orchestrator and trust management module (TMM). Furthermore, HORSE advances the concept of dual-context security by integrating real-time monitoring with Digital Twin-based emulation through the Sandboxing (SAN) and Early Modelling (EM) modules. This approach enables predictive assessment of security actions, supporting proactive trust establishment and continuous validation of the network's behaviour under different threat conditions. By merging these capabilities, HORSE provides a holistic vision of a future 6G architecture where trust is not only managed but continuously optimized through intelligent, distributed, and autonomous mechanisms. As such, HORSE offers valuable architectural insights and practical building blocks that can inform UNITY-6G's trust management design, particularly concerning integrated network composition, secure collaboration among distributed AI components, and the enforcement of trustworthy, intent-driven policies across heterogeneous environments.

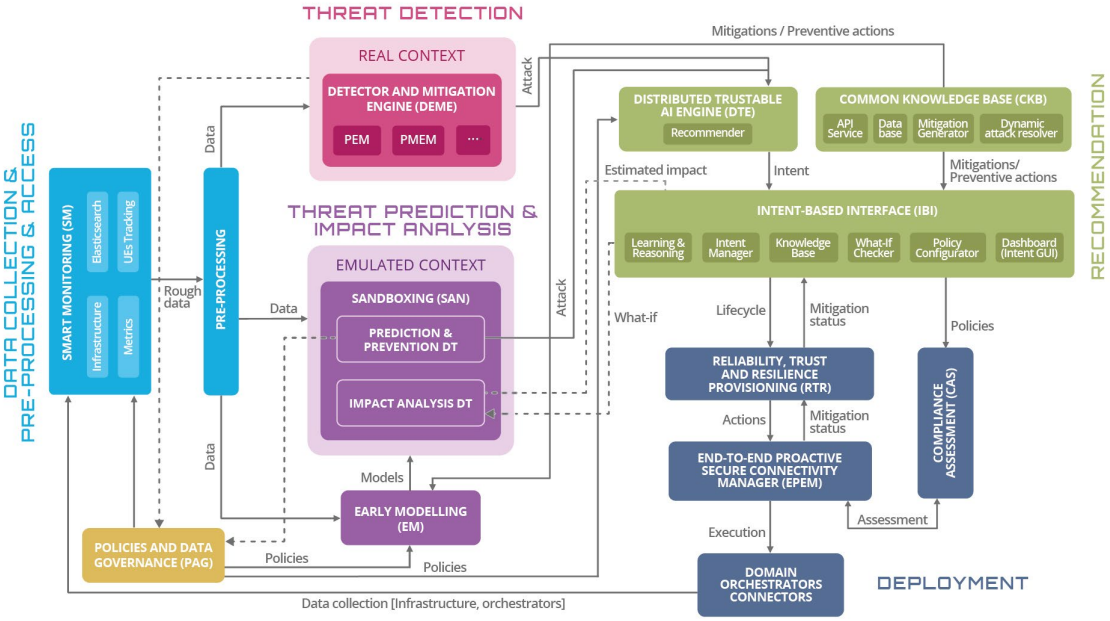

Figure 5 High level System Architecture of HORSE Project

2.2.3.2 6GCLOUD Project

6G-Cloud [32] is an EU SNS JU research project that aims to design, build, and validate an AI-native, cloud-friendly 6G system architecture across the whole “cloud continuum”, from centralized cloud to edge and extreme edge. It focuses on composing 6G network functions (RAN, core, edge) as flexible, service-oriented building blocks that different stakeholders can deploy and orchestrate dynamically in multi-tenant, multi-provider environments. Trust is a core theme: the project explicitly designs business and cloud interfaces that account for security, reliability, *trust*, and energy efficiency, and it explores zero-trust security models, AI-driven threat detection, and trusted routing to make 6G cloud infrastructures more resilient against attacks and misconfigurations. In short, 6G-Cloud is about making future 6G telco clouds not only intelligent and elastic, but also inherently trustworthy and secure for all actors sharing the same cloud-native infrastructure.

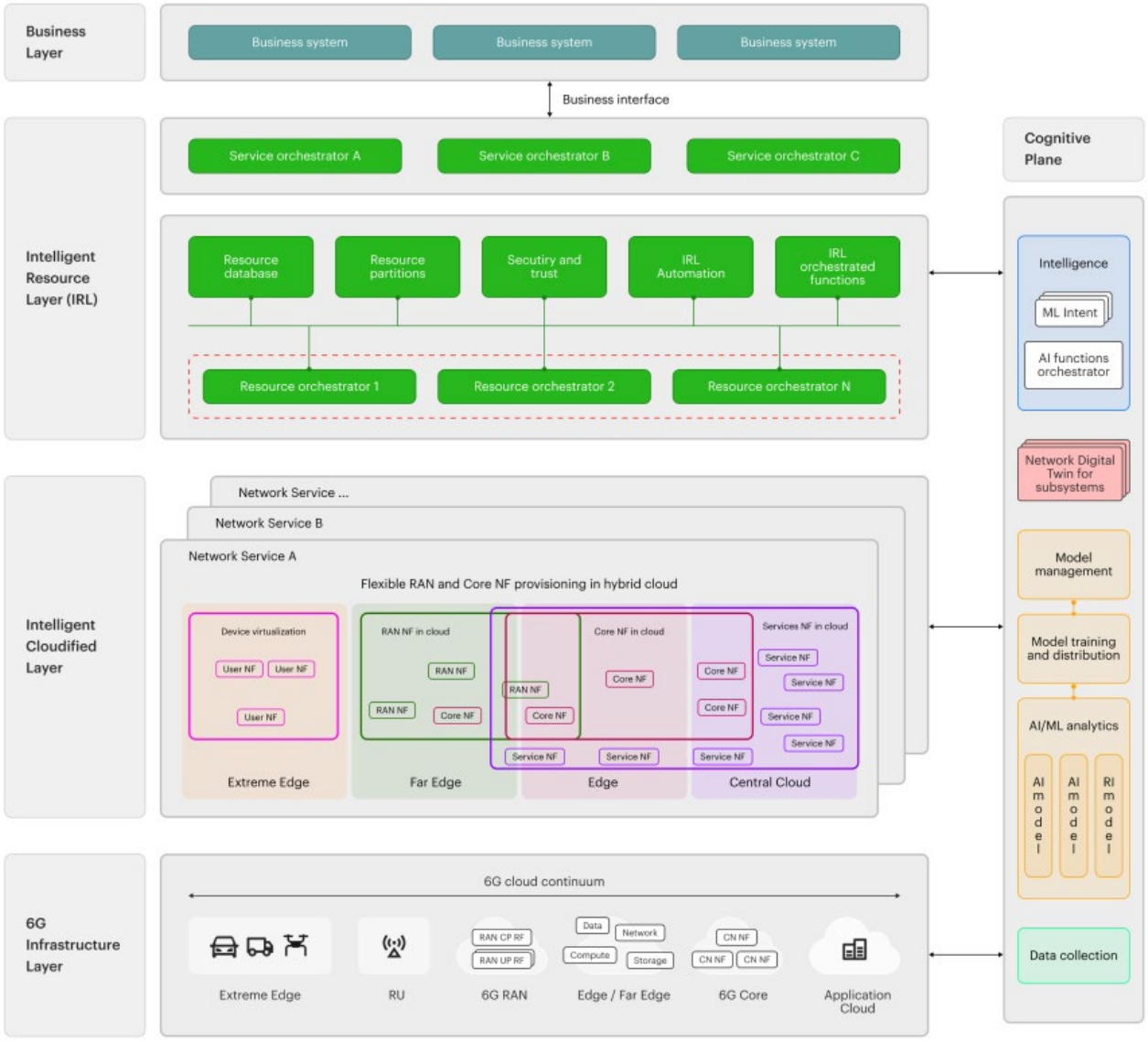

Figure 6 High-level architecture 6GCloud

2.2.3.3 ELASTIC Project

ELASTIC [30] is an EU SNS JU project on “Efficient, portabLe And Secure orchesTration for reliable servICes”, aiming to redesign 6G service orchestration across cloud, fog, and edge using WebAssembly, Function as a Service (FaaS), confidential computing, extended Berkeley Packet Filter (eBPF)/Express Data Path (XDP), and federated learning to make deployments more flexible, secure, and energy-efficient. Within this architecture, the *Trust & Access Control* block is the piece that enforces *who* can do *what* on *which* data and *under what conditions*. It relies on Trusted Execution Environments (TEEs) and remote attestation to verify that workloads really run in approved, protected enclaves before granting access, and then applies access-control policies to guard sensitive data and services across cloud, fog, and edge. In other words, ELASTIC’s Trust & Access Control turns the low-level confidential computing mechanisms into end-to-end assurances that only attested,

policy-compliant components can interact with critical data and functions in the 6G orchestration plane.

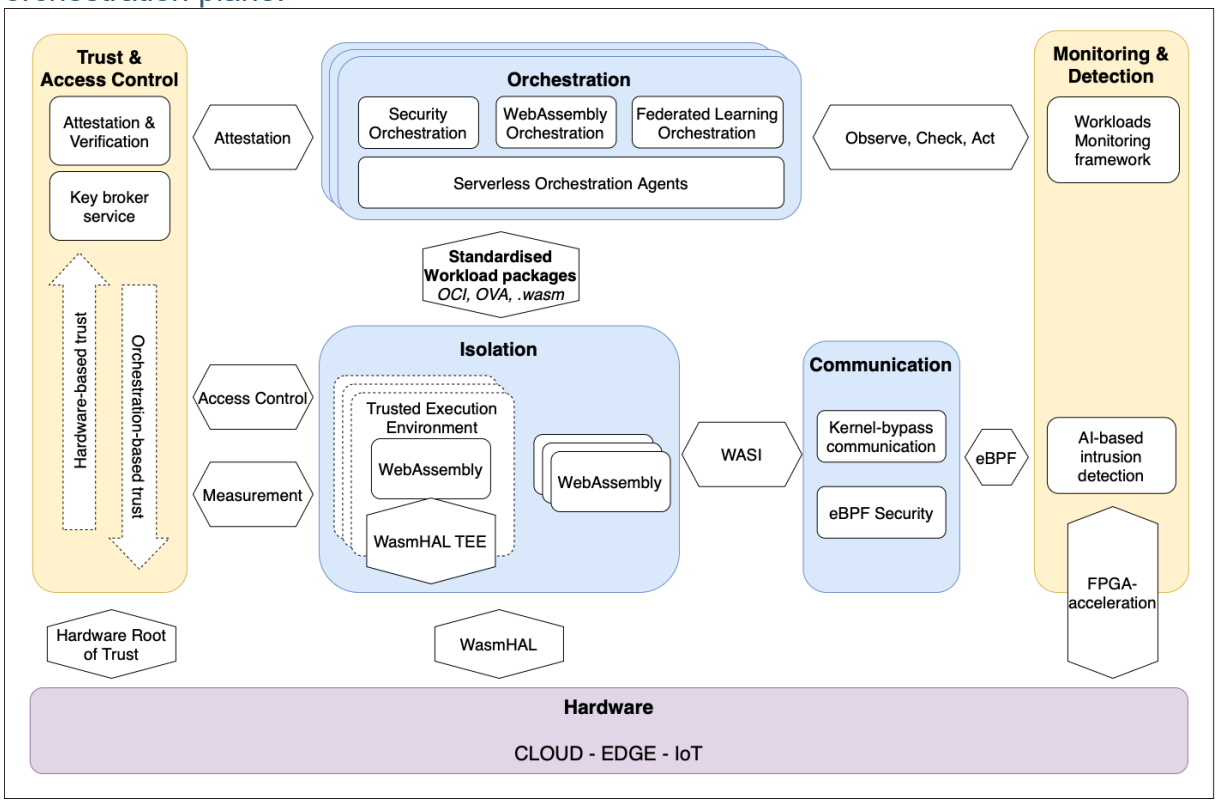

Figure 7 High-level architecture ELASTIC Project

2.2.3.4 SUSTAIN-6G

SUSTAIN-6G (SUSTainability-Advanced and Innovative Networking with 6G) is the SNS JU “Sustainability Lighthouse” project [33] that aims to put sustainability at the core of how 6G is designed, deployed, and used. Rather than treating energy efficiency or carbon impact as afterthoughts, it develops a holistic sustainability framework for 6G that jointly considers environmental, societal, and economic dimensions. The project surveys the evolving 6G technology landscape, maps SotA solutions, and identifies concrete sustainability needs, requirements, and indicators for both networks and vertical sectors. On this basis, SUSTAIN-6G proposes technical and methodological solutions, validates them via proof-of-concepts, and assesses their real sustainability impact. The end goal is to deliver guidelines, best practices, and strategic roadmaps so that future communication systems and vertical applications embed sustainability “by design,” from device to application and across the full lifecycle of network assets. A key feature of the project is that it treats “sustainable 6G” and “6G for sustainability” as two equally important sides of the same coin. On one hand, sustainable 6G refers to designing the network itself, radio, core, computing, and infrastructure, so that it optimizes energy and resource consumption, minimizes environmental impact, and remains economically accessible. On the other hand, 6G for sustainability looks at how 6G can *enable* greener and more sustainable verticals such as energy smart grids, e-health, and smart agriculture, by supporting new use cases and measuring their sustainability benefits. The framework, SUSTAIN-6G is building, is explicitly meant to bridge these aspects; identifying where sustainability needs arise, what challenges 6G and verticals must address, which technological

capabilities are required, and how society and the economy benefit from responsible innovation within planetary boundaries.

2.2.3.5 PRIVATEER

The PRIVATEER project [34] advances a trust-aware, intent-driven orchestration framework for next-generation 5G and 6G networks by combining remote attestation, path validation, trust scoring, and closed-loop automation within the ETSI ZSM architecture. The goal is to ensure that multi-domain network services (Edge-Transport-Core) remain compliant with security, integrity, & confidentiality requirements, even under dynamic or adversarial conditions, through autonomous, trust-driven reconfiguration.

Figure 8 illustrates PRIVATEER's holistic trust architecture for 5G/6G networks, combining attestation, intent-based orchestration, data-driven trust analytics, and secure, DLT-backed trust exposure. The layered architecture shows how trust evidence is captured, processed, quantified, enforced, and finally exposed to external consumers (e.g., operators, service providers, or automated systems).
At the foundation, the Attestation layer provides hardware-rooted and path-rooted trust evidence through microservice attestation (using µProbes and Security Probes operating in confidential computing environments such as Intel SGX/TDX) and Proof of Transit (PoT) nodes that verify the integrity of transport paths. This evidence is supplied to the 5G and cloud-native execution layer, where trusted workloads operate across edge, transport, and core domains, with telemetry and observability supporting broader trust assessment.

Above this, the Data and Security Analytics layer transform raw attestation and network data into actionable intelligence through local and global model updates, adversarial training, and anomaly detection, generating insights into integrity deviations or misbehaviour.

The Intent-Based Networking (IBN) and Orchestration layer operationalizes trust by embedding trust requirements into service intents, templates, and SLA policies; a decision engine and LoT (Level of Trust) controller use continuous trust scores to trigger closed-loop actions such as VNF redeployment or path reconfiguration when trust falls below defined thresholds. At the top, CTI sharing ensures cross-domain trust federation via credential wallets for end users, infrastructure providers, and service providers.

On the right, the DLT-based Trust Exposure layer unifies the architecture by providing immutable, verifiable, and privacy-preserving trust management: off-chain storage retains raw evidence, on-chain components manage identity, secure execution, and trust quantification, while SCs expose trust scores and claims to authorized consumers. Together, these layers implement PRIVATEER's holistic trust model, making trust a measurable, enforceable, and shareable asset that drives autonomous, zero-trust 6G network behaviour.

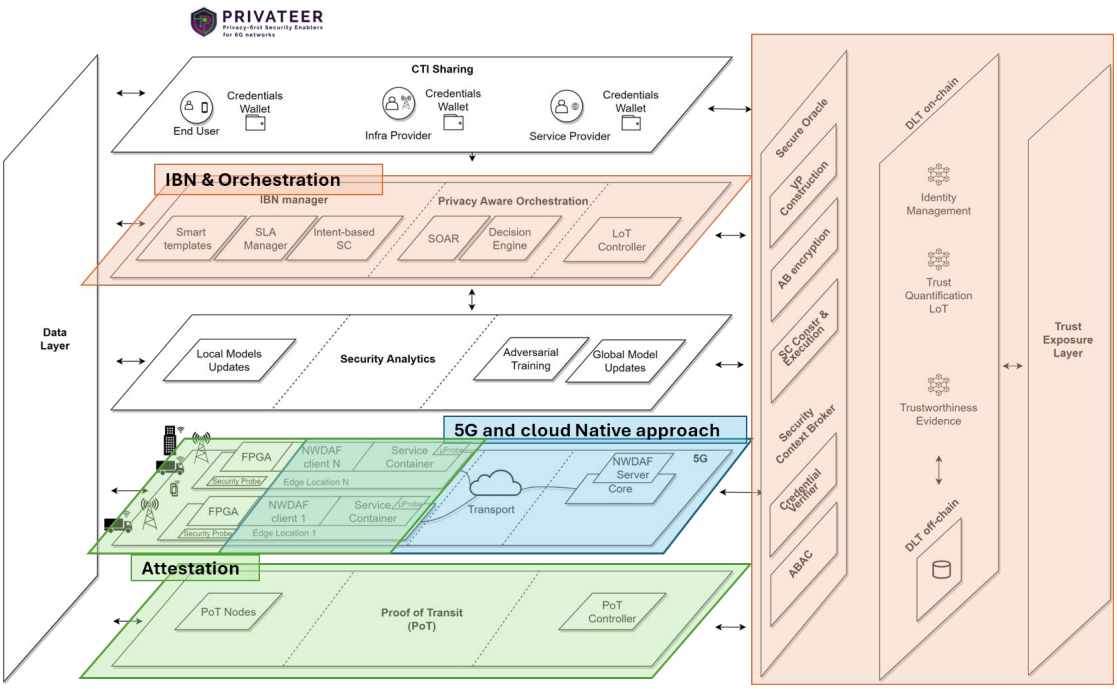

Figure 8 PRIVATEER's holistic trust architecture for 5G/6G networks

2.2.4 Summary Comparisons and UNITY-6G Unique Perspective

UNITY-6G is unique across the SNS trust landscape because it does not introduce yet another standalone trust model, but instead acts as an AI-native, system-level trust integrator and orchestrator that unifies, contextualizes, and operationalizes trust signals produced by all other projects. The UNITY-6G project complements the trust models of above projects by providing a unifying, AI-native trust orchestration fabric that integrates dynamic trust evaluation, distributed intelligence, and cross-domain evidence into a coherent, end-to-end assurance framework.

While *iTrust6G* focuses on entity-centric, behaviour-driven trust computation for access control, UNITY-6G elevates such trust scores into semantic service intents and lifecycle-aware orchestration decisions across the TN-NTN continuum. Compared to *ROBUST-6G*, which distributes trust inference among autonomous macro-services using federated and explainable AI, UNITY-6G provides a higher-level cognitive coordination layer that resolves conflicts between multiple AI agents and aligns their trust outputs with global service objectives and policies.

In contrast to *NATWORK*, which embeds trust into cross-domain orchestration through AI-driven Moving Target Defence and federated resilience, UNITY-6G treats NATWORK-like trust indicators as external, domain-specific signals that are fused with semantic KPIs, digital-twin insights, and intent priorities to drive end-to-end optimization.

Relative to *RIGOUROUS*, where trust is defined as continuous privacy and policy compliance at the service level, UNITY-6G generalizes this notion by integrating compliance-derived trust with performance, resilience, and semantic correctness into a unified decision space. At the system level, UNITY-6G complements *SAFE-6G* by

providing the architectural mechanisms to compute, aggregate, and act upon Level of Trustworthiness (LoTw) KVIs in real time.
Compared to *MARE*, which offers programmable security primitives (DOTs), UNITY-6G orchestrates such primitives coherently across domains using AI-native intent translation. In relation to *HORSE and 6G-CLOUD*, which emphasize trust-aware architectures and digital-twin-based validation, UNITY-6G leverages these capabilities as inputs to a closed-loop, semantic-aware control plane.
Finally, unlike *ELASTIC and PRIVATEER*, which enforce trust at the execution and runtime-assurance layers through confidential computing, attestation, and path validation, UNITY-6G sits above them as the unifying governance and reasoning layer, transforming heterogeneous trust evidence into explainable, conflict-free, and goal-driven orchestration actions, thereby making trust not just measurable or enforceable, but globally intelligible and actionable across the entire 6G ecosystem.

2.3 ANOTHER PROJECT

2.3.1 5GSTAR

The Open-RAN security testbed [65] developed in the framework of project 5GStar funded by the National Centre for Research and Development in Poland provides a fully programmable environment for analysing and mitigating radio-access anomalies under realistic network conditions. It integrates all major O-RAN architectural components, Non-RT RIC, Near-RT RIC, O-DU/O-CU, O-RUs and a 5G Core, linked through standardized open interfaces that allow security-focused xApps and rApps to be dynamically deployed and evaluated.

The platform is used to reproduce and study jamming attacks, including continuous and keyed jamming (Figure 9). Real RF hardware enables over-the-air interference injection, while the testbed collects metrics such as ACK/NACK ratios and transport-block error patterns. Near-RT RIC xApps correlate these measurements to detect abnormal radio behaviour. Mitigation strategies evaluated in the testbed demonstrate up to a 4× reduction in URLLC packet delay, showing the effectiveness of rapid control-loop reactions enabled by O-RAN's programmable structure.

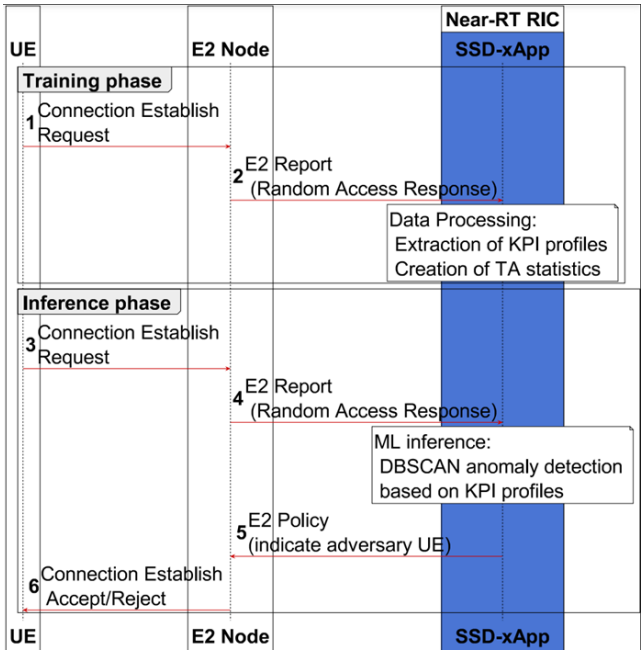

Figure 9 UML sequence diagram for jamming detection and mitigation

The testbed also supports experiments on signalling storms (Figure 10), relevant for dense IoT and mMTC deployments. Timing Advance statistics, connection-attempt patterns and device-level activity are monitored to generate anomaly scores for UEs generating excessive signalling traffic. Implemented xApps achieve over 68% detection accuracy while keeping the false-alarm rate below 0.5%, demonstrating the capability of O-RAN control loops to detect large-scale signalling anomalies with low overhead.

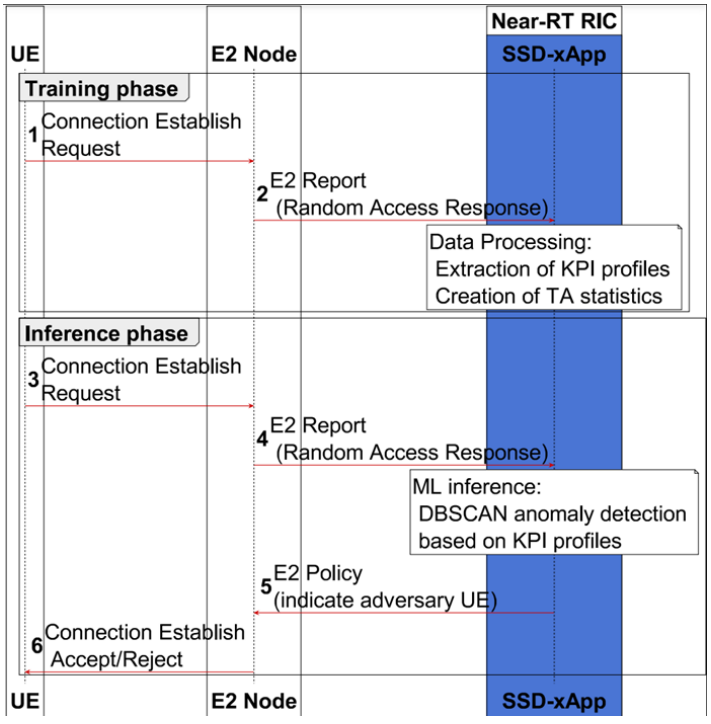

Figure 10 UML sequence diagram for signaling storm detection and mitigation

The platform's programmability and real-hardware integration make it suitable for evaluating interactions between radio-layer disturbances and RIC-driven control logic, enabling the assessment of end-to-end trust mechanisms in open, multi-vendor environments. As such, the testbed (Figure 11) represents a practical basis for validating security and resilience techniques required in future open and intelligent RAN deployments.

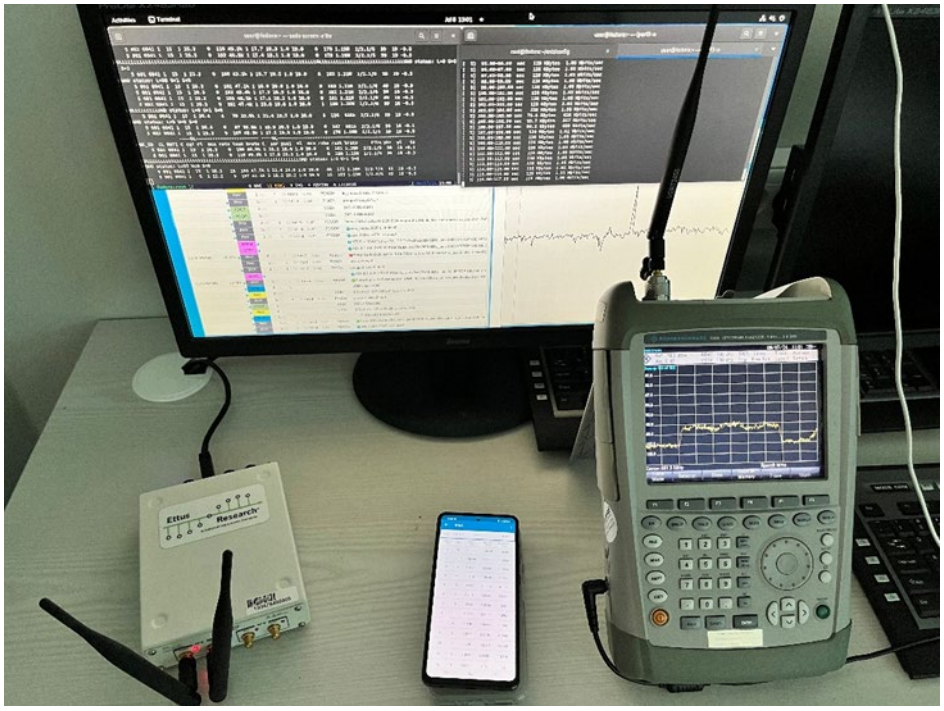

Figure 11 View of the testbed

2.4 STANDARDISATION EFFORTS IN TRUST DOMAIN

2.4.1 NIST Standardization

A ZTA philosophy, which is proposed by NIST, provides the guiding "never trust, always verify" principle. It shifts security from a static, location-based defence to a dynamic, identity-based, and continuously verified posture. NIST SP 800-207 introduces and defines the term "ZTA" as defined by the following seven tenets of zero trust [37]:

1. All data sources and computing services are considered resources
2. All communication is secured regardless of network location
3. Access to individual enterprise [operator] resources is granted on a per-session basis
4. Access to resources is determined by dynamic policy
5. The enterprise [operator] monitors and measures the integrity and security posture of all owned and associated assets
6. All resource authentication and authorization are dynamic and strictly enforced before access is allowed

7. The enterprise [operator] collects information about the current state of assets, network infrastructure and communications and uses it to improve its security posture

Cybersecurity and Infrastructure Security Agency (CISA) Zero Trust Maturity Model (ZTMM) establishes a phased approach to incrementally strive towards a ZTA, as modelled in Figure 12 [36]. The CISA ZTMM identifies 5 pillars for zero trust: Identity, Devices, Networks, Applications & Workloads, and Data. The three cross-cutting functions, Visibility and Analytics, Automation and Orchestration, and Governance, apply to all five pillars

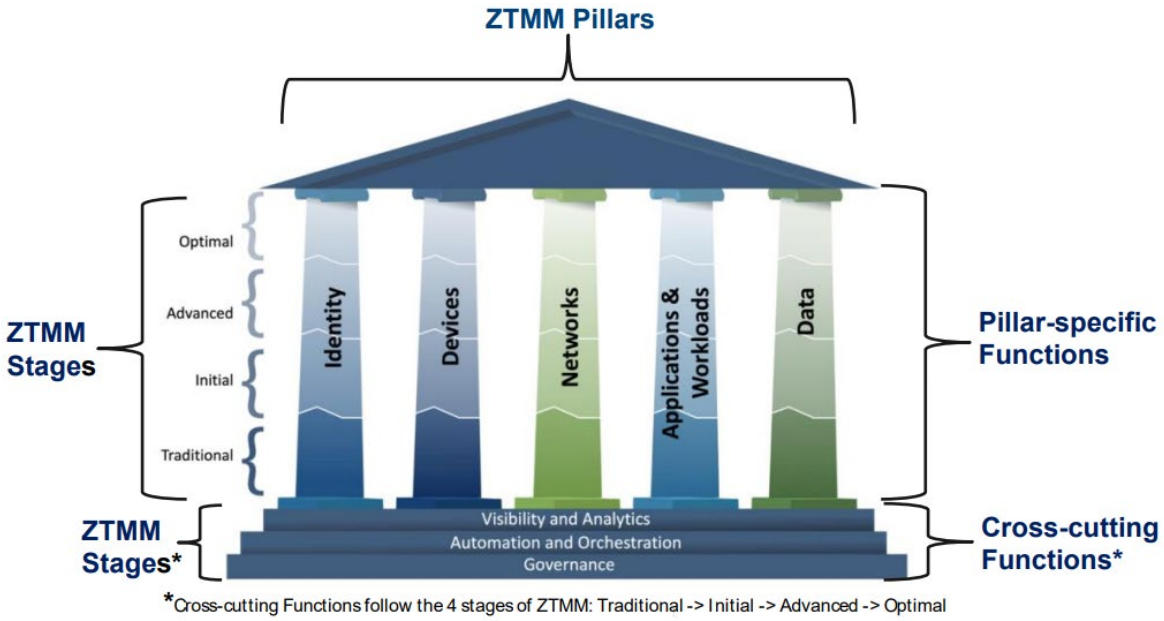

Figure 12 US DHS CISA Zero Trust Maturity Model

The NIST ZTA model is a logical architecture built on two key functions:

1. *Policy Decision Point (PDP)/Policy Engine (PE)*: The "brain" of the ZTA. It ingests data (e.g., from monitoring, identity systems), consults the defined security policies, and makes the decision to grant, deny, or revoke access.
2. *Policy Enforcement Point (PEP):* The "shield" or "gatekeeper." It sits in the data path, requests a decision from the PDP, and then enforces that decision (e.g., by allowing or blocking the connection).

2.4.2 O-RAN Alliance

Trust management has become one of the most critical research and standardization areas within O-RAN, because openness, disaggregation, and multi-vendor interoperability inherently dissolve the traditional perimeter-based trust model that 4G/5G relied upon. O-RAN introduces openness and modularity, any vendor's RIC, DU, or CU may interoperate. That creates new trust boundaries across:

- Components (O-RU, O-DU, O-CU, Near-RT RIC, Non-RT RIC, SMO, O-Cloud)
- Interfaces (A1, E2, O1, O2, Y1, Open Fronthaul)
- Applications (xApps/rApps from third parties)

Hence, trust in O-RAN is not static but it must be established, verified, and continuously monitored among entities that may not share the same administrative control. Work Group (WG)11 is the dedicated security group within the O-RAN Alliance [41]. Its formal mandate is to define, assess, and specify the security requirements for the O-RAN architecture and its interfaces, components, and operational lifecycle. O-RAN WG11 is responsible for defining the O-RAN threat model, specifying security requirements and controls for all O-RAN interfaces and components, and studying advanced topics such as AI/ML security, Zero-Trust, O-Cloud protection, and supply-chain assurance, forming the security backbone of the O-RAN ecosystem. WG11 uses the Microsoft **S**poofing, **T**ampering, **R**epudiation, **I**nformation disclosure, **D**enial of Service, ***E***levation of privilege (STRIDE) model to systematically categorize and analyse threats against every O-RAN component and interface. WG11 translates the core NIST ZTA tenets into tangible principles for the O-RAN architecture:

1. *"All data sources and computing services are resources":* In O-RAN, this means every Network Function (NF), container, xApp, rApp, and data flow is treated as a resource to be protected with its own "micro-perimeter".
2. *"All communication is secured regardless of network location":* This tenet drives the WG11 mandate to secure *all* O-RAN interfaces (A1, E2, O1, O2, etc.) with strong cryptographic protocols like Mutual Transport Layer Security (mTLS) and Internet Protocol Security (IPsec), assuming that the underlying transport network (the O-Cloud) is not trusted.
3. *"Access to... resources is granted on a per-session basis":* This is a direct rejection of the old Virtual Private Network (VPN)/persistent tunnel model. In O-RAN, trust is not static. Access is established using per-session, certificate-based mutual authentication (mTLS) and is authorized via mechanisms like OAuth 2.0.
4. *"Access to resources is determined by dynamic policy":* This is the "brain" of ZTA. Access decisions are not based on static Access Control Lists (ACLs). They are dynamic policies based on the identity of the subject (e.g., an xApp), the resource being accessed, and the real-time security posture of the component. This is the core function of the RICs.
5. *"The enterprise [operator] monitors and measures the integrity and security posture of all... assets":* This tenet mandates continuous monitoring and logging. This work has led to a dedicated "Study on Continuous Security Monitoring" to define what data to collect and how to analyse it for anomalous behaviour

O-RAN ALLIANCE Security Work Group incorporates ZTA in O-RAN security specifications. One core responsibility of WG11 is to ensure multi-vendor O-RAN deployments remain secure even with no implicit trust, it studies how Zero Trust principles can be applied to O-RAN components, focusing on continuous authentication, policy enforcement, and trust brokering [39]. "Study on Zero Trust Architecture for O-RAN," in [39] explicitly applies the principles of NIST Special Publication 800-207 (ZTA) and is guided by the CISA Zero Trust Maturity Model (ZTMM) for its incremental implementation. To protect external and internal interfaces and align with ZTA, O-RAN ALLIANCE has specified the security controls as summarized in Table 1.

Table 1 Specified Security Controls for O-RAN Interfaces [39]

Security Principles	Non-Fronthaul Interfaces						Open Fronthaul Interfaces			
	A1	R1	O1	O2	E2	Y1	C-plane	U-plane	S-plane	M-plane
Confidentiality	TLS	TLS	TLS	TLS	IPsec	TLS		PDCP		TLS/SSH
Integrity	TLS	TLS	TLS	TLS	IPsec	TLS		PDCP		TLS/SSH
Authenticity	mTLS	mTLS	mTLS	mTLS	IPsec	mTLS	802.1X	802.1X	802.1X	mTLS/SSH/802.1X
Authorization	OAuth	OAuth	NACM	OAuth		OAuth	802.1X	802.1X	802.1X	NACM/802.1X
Data Origin Authentication	mTLS	mTLS	mTLS	mTLS	IPsec	mTLS		PDCP		TLS/SSH
Replay Prevention	TLS	TLS	TLS	TLS	IPsec	TLS		PDCP		TLS/SSH

NOTE: 3GPP Access Stratum (AS) Control Plane and User Plane messages that are transported via the Open Fronthaul U-Plane (LLS-UP) are confidentiality and integrity protected by Packet Data Convergence Protocol (PDCP). PDCP security controls remain in place when the message traverses the Open Fronthaul U-Plane.

However, the O-RAN Alliance document in [39] *does not contain the specific mapping* of these PDP and PEP roles to O-RAN's functional components, therefore, current standards-level efforts are described as "conceptual" and are "lacking details regarding the deployment and implementation" of the PDP/PEP. In particular, the document [40] studies security threats, risks, and mitigation strategies for O-RAN systems, providing a comprehensive threat modelling framework developed by O-RAN WG1.1.

The SMO framework is the logical and specified nexus for O-RAN security management. The SMO provides the O1 (management) and O2 (orchestration) interfaces, giving it direct control over the provisioning, configuration, and security posture of all other O-RAN components. The SMO's role is not limited to management; it also serves as the central trust orchestrator for the entire O-RAN domain. The security of control and management planes is orchestrated from this central point of trust, the SMO, which then propagates security controls to all other component connections. The greatest strength of the O-RAN model - its openness to innovation through third-party applications (xApps and rApps) - is also one of its most significant security risks. This model allows third-party, untrusted code to run on the RICs, the "brains" of the network. This also introduces a novel and substantial software supply chain risk that does not exist in monolithic RAN architectures. For example, a seemingly benign "energy saving" rApp could contain a malicious payload to exfiltrate data, or a compromised xApp could be used to launch a denial-of-service attack.

To address this "trust deficit" in third-party code, WG11 created the Application Lifecycle Management (AppLCM) Security framework [41]. The purpose of this framework is to establish provenance (who wrote this app?) and integrity (has it been tampered with?) before the application is onboarded, provisioned, or executed. The AppLCM security framework is a multi-step, trust-gating process with defined roles for the application vendor, the SMO, and the RICs.

2.4.3 ETSI Standardization

ETSI earlier work in NFV security focused on "Trust Domains" and attestation. At the moment, its new frontier is the *ISG PDL.* This group is developing a comprehensive framework to create an open, industrial-grade ecosystem for trust, with a focus on DLT. The ETSI ISG PDL's work, including the group report on "Trust in Telecom System", directly addresses the need for a new trust foundation. The most profound shift in the ETSI PDL model is the move away from centralized trust anchors. In traditional telecom networks, trust is siloed and it exists between a user and their specific operator. This model breaks down in multi-operator scenarios, such as roaming, network slicing, or shared infrastructure. ETSI's solution is to create a "decentralized trust foundation". it can be implemented through specific, foundational technologies, The model on Decentralized Trust Evaluation across task lifecycle stages is built on following key issues [42]:

a. Efficient Trust Data Management using DLT for storage & retrieval.
b. SC-based Trust Enforcement for compliance & adaptation.
c. Enable trust without centralized authorities (important for multi-operator and multi-vendor telecom environments).

ETSI TC ESI and ETSI ISG PDL, ETSI TC DATA have been involved in standardizing DLTs. The ETSI ISG PDL reference architecture in Figure 13 makes this vision concrete by layering permissioned ledgers as a reusable platform beneath telecom and vertical applications. At the bottom, the DLT layer hosts one or more “endorsed DLT types”, while an optional DLT abstraction layer exposes a common set of primitives so that higher layers remain agnostic to the specific ledger technology. On top of this, the platform services layer is split into atomic platform services (e.g., identity, key management, asset registration, notarization) and composite platform services (e.g., reputation systems, SLA enforcement, cross-domain workflows), each with mandatory and optional components. An optional application abstraction layer then provides uniform interfaces to these services, allowing different providers (Entities A and B) to implement their applications using platform services from different vendors without sacrificing interoperability. The architecture explicitly accommodates off-chain storage and external entities, reflecting the need to couple PDL-based trust anchors with existing OSS/BSS systems, data lakes, oracles, and analytics platforms in realistic deployments [42].

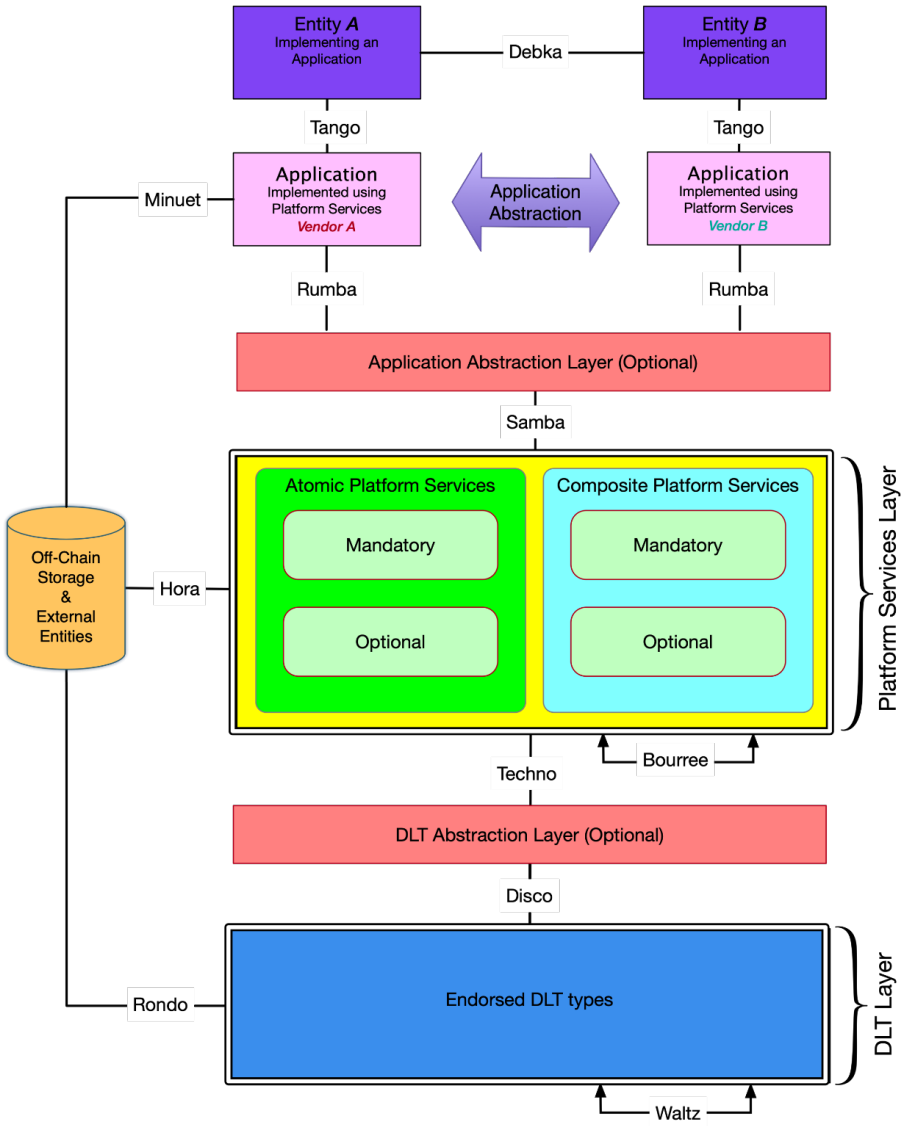

Figure 13 ETSI-ISGPDL Reference Architecture

For telecom operators and 6G research, this architecture is particularly relevant because it cleanly separates DLT technology choices from trust services and applications, enabling standardized, PDL-based trust functions to be reused across O-RAN, slicing, roaming, and federated data-sharing scenarios. In the context of cyber-ranges, it provides a blueprint for instantiating different DLT backends under a common platform-services layer, and for experimenting with alternative trust policies (e.g., reputation, evidence aggregation, cross-operator SLAs) without reengineering the applications themselves.

Building on this reference architecture [42], ETSI GR PDL 004 refines the role of SCs by treating them as engineered artefacts with a well-defined system architecture and lifecycle, from upfront governance to end-of-life termination [42]. The document introduces a three-phase lifecycle Figure 14 in which SCs move from a planning phase, where stakeholders and standardisation bodies define governance, draft human-readable templates, negotiate terms, and compile them into machine-readable specifications that must match the intended contract; through a coding and testing

phase, where software engineers implement the contract, perform code verification, validation, and iterative testing on dedicated testbeds, feeding back whenever test outputs diverge from requirements; and finally into a deployment and execution phase, where validated contracts are deployed on the target PDL, exercised via APIs in production, monitored, debugged if necessary, and ultimately terminated or upgraded according to predefined policies. For telecom and 6G cyber-ranges, this lifecycle model provides a practical checklist for emulating "real" smart-contract development and operations, capturing not only on-chain behaviour, but also governance decisions, testing workflows, and runtime assurance steps that are crucial when SCs encode SLAs, roaming agreements, or cross-operator trust policies.

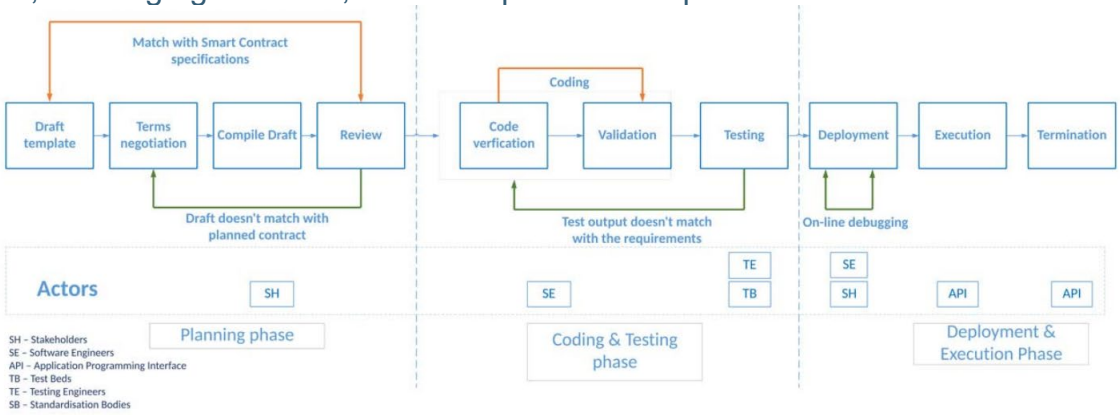

Figure 14 Lifecycle of a Smart Contract

Beyond the engineering lifecycle, ETSI GR PDL 004 also illustrates how SCs can be embedded into end-to-end service delivery and assurance, as shown by the "SCs with Quality of Service (QoS) monitoring" scenario back (see Figure 15). Here, the user interacts with a Decentralized Application (dApp) to request a telecom service whose QoS is governed by an SLA encoded in a SC. The request (1) triggers contract execution on the PDL via the dApp (2-3), with multiple operator PDL nodes and a regulatory-authority node all maintaining a consistent, immutable view of SLA terms and state. In parallel, the service is actually delivered over the underlying network (5), while QoS metrics are collected both at the user side (e.g., in a trusted enclave) and by an independent performance monitor. These measurements are periodically fed back (see Figure 15) into the PDL-resident SCs, which can automatically verify compliance, flag violations, and support regulatory oversight. For telecom operators, regulators, and cyber-range designers, this pattern concretely shows how PDL and SCs can couple live QoS evidence with on-chain SLA logic, enabling realistic experiments on cross-operator assurance, dispute resolution, and trust in performance reporting [42].

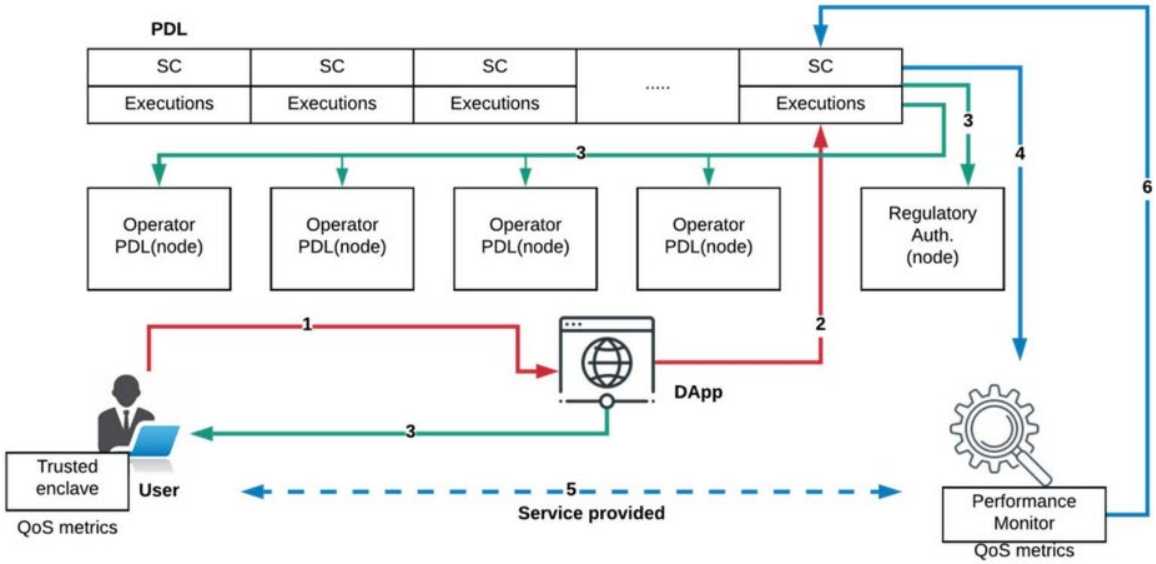

Figure 15 Smart Contract with QoS monitoring

At a more granular level, ETSI GR PDL 004 abstracts the internal operation of these contracts using the "Reference Architecture of a SC without contract chaining (see Figure 16)", which clarifies how off-chain evidence and on-chain state interact [42]. Figure 16 illustrates how external data (e.g., oracles) are ingested via an API, processed through the contract's initialization-execution-termination logic, and finally recorded as state update. External data sources (e.g., QoS oracles, monitoring probes, or OSS/BSS systems) feed measurements or events through an API into the SC, which then progresses through an execution pipeline: initialization, execution of the contract logic, and eventual termination when a timer elapses or predefined conditions are met. Throughout this process, the logic updates the contract's internal state and records the resulting state transitions as transactions on the PDL, producing an immutable history of how inputs, timers, and rules led to specific outcomes. This reference model is directly applicable to telecom-oriented contracts such as SLAs, roaming agreements, or FL reputation updates because it separates oracle design, contract logic, and ledger recording in a way that cyber-ranges can systematically exercise, monitor, and mutate when evaluating different trust policies and failure modes.

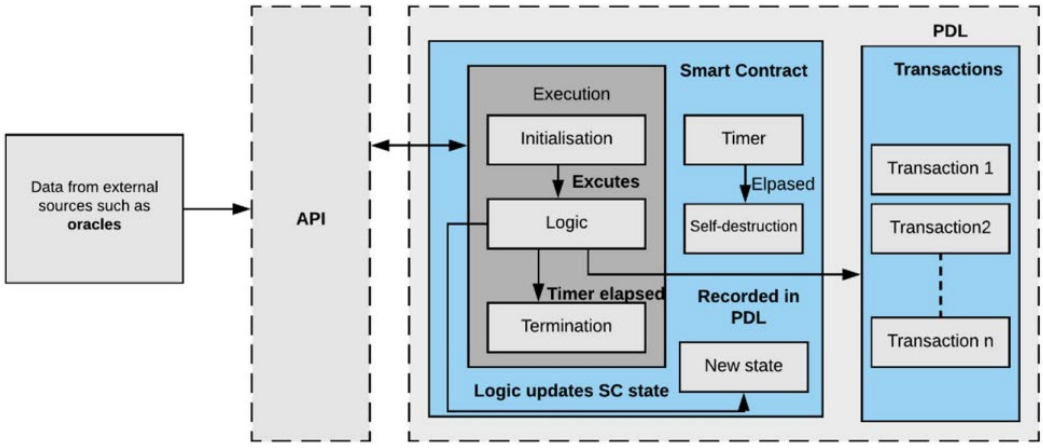

Figure 16 Reference architecture of a SC without contract chaining

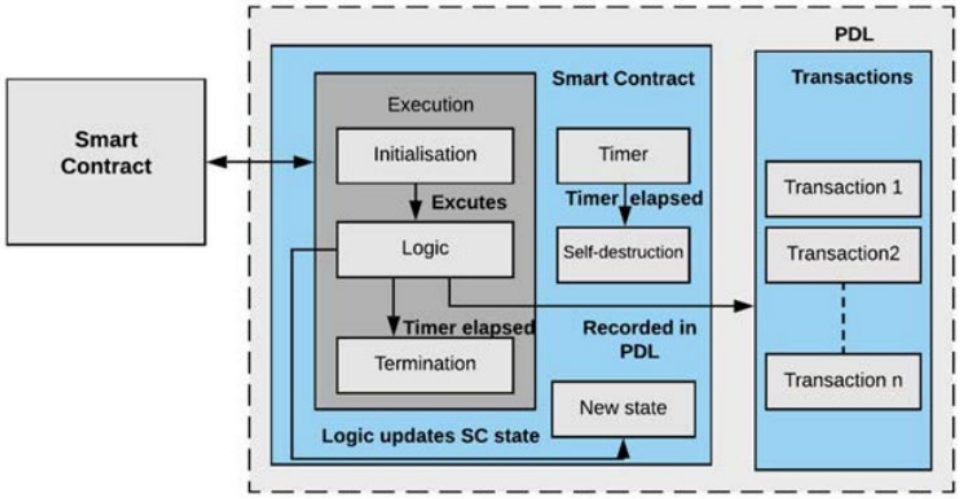

Figure 17 Reference architecture of a SC with contract chaining

From a PDL perspective, [42] the ETSI material captures how 6G is expected to blur the traditional separation between "users" and "providers" [42]. UEs can simultaneously act as *consumers* of network services, *providers* of data/resources, and *participants* in advanced task-collaboration swarms (e.g., sensing, cooperative driving, or edge intelligence), all interacting with the 6G network infrastructure and service platforms. This many-to-many interaction graph motivates the ETSI PDL vision of a shared, permissioned ledger where identities, credentials, reputations, and SLA commitments of UEs and network service providers are recorded and evaluated. In such a setting, SCs and PDL-based trust services can express and enforce multi-party agreements (e.g., data sharing, collaborative inference, resource pooling) while keeping an immutable history of who contributed what, under which policies and conditions precisely the kind of behaviour that cyber-ranges need to emulate when exploring 6G UE collaboration and UE-as-a-provider use cases [44].

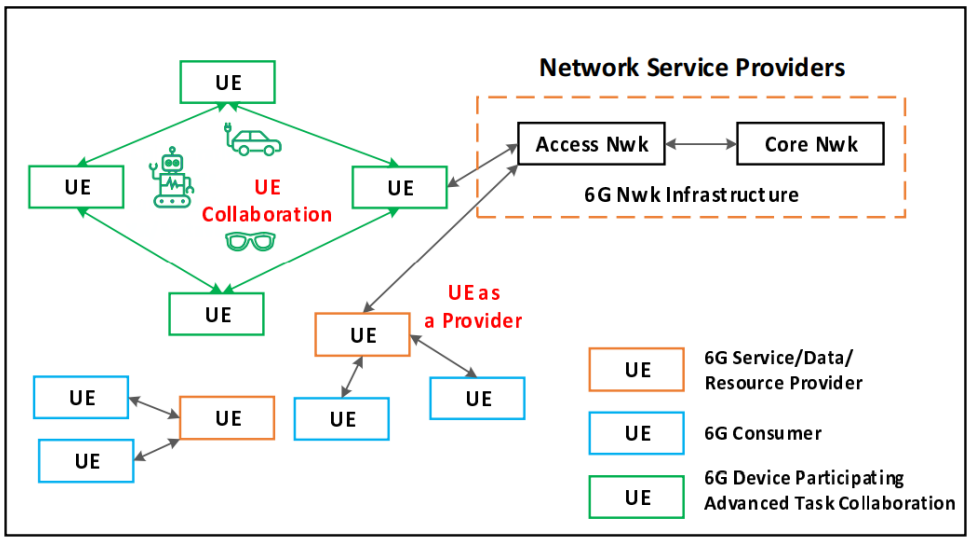

Figure 18 Future 6G where UE can act both as consumer and provider of services.

Figure 19 then zooms in on a single user (User1) and their devices to highlight how context and behaviour dynamically reshape trust over time. Figure 19 illustrates how

mobility and device changes (UE1 to a new location, or UE1 to UE2) alter interactions with edge/core/cloud network functions and lead to changing trust. In Scenario 1, User1 and UE1 move to a new location but continue to access services from the same set of edge/core/cloud network functions, changing the environmental and connectivity conditions under which trust must be assessed [44]. In Scenario 2, the same user accesses the network through a different device (UE2), potentially with different hardware, software, and past behaviour, even though the logical user identity is unchanged. In both cases, NF service producers (NF1…NFX) must adapt their trust decisions based on evolving evidence about user behaviour, device characteristics, and context. For cyber-range experiments, this illustrates that trust models, PDL-anchored records, and smart-contract policies must be context-aware, allowing trust to be recomputed as users move, switch devices, or change how they engage with network services.

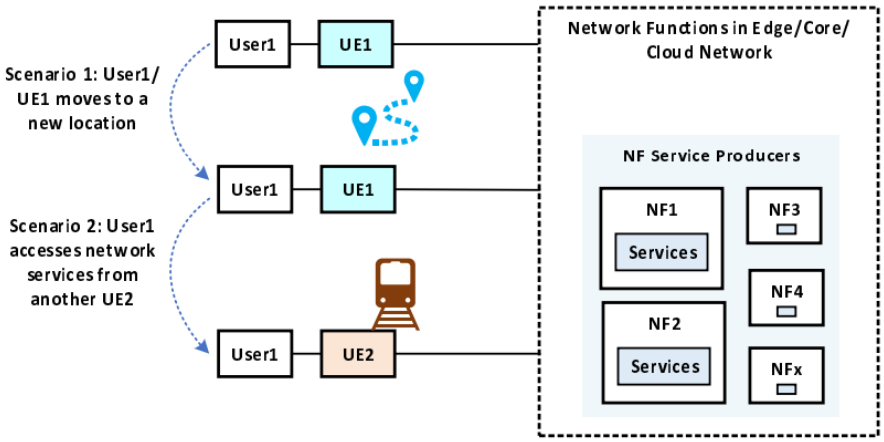

Figure 19 Dynamic context and behaviour of User1/UEs

2.4.4 ITU Standardization

Within the ITU-T standardization environment, trust is formally defined as the "measurable belief and/or confidence which represents accumulated value from history and the expecting value for the future". This definition marks a pivotal shift from "Hard Security," which validates credentials, to "Soft Trust," which evaluates behaviour, competence, and benevolence over time. The distinction is critical: a compromised node may possess valid cryptographic keys (passing security checks) but exhibit anomalous data patterns (failing trust checks) [45]. Figure 20 shows an example architecture diagram from ITU-T Focus Group on Application of Distributed Ledger Technology (FG DLT) [48].

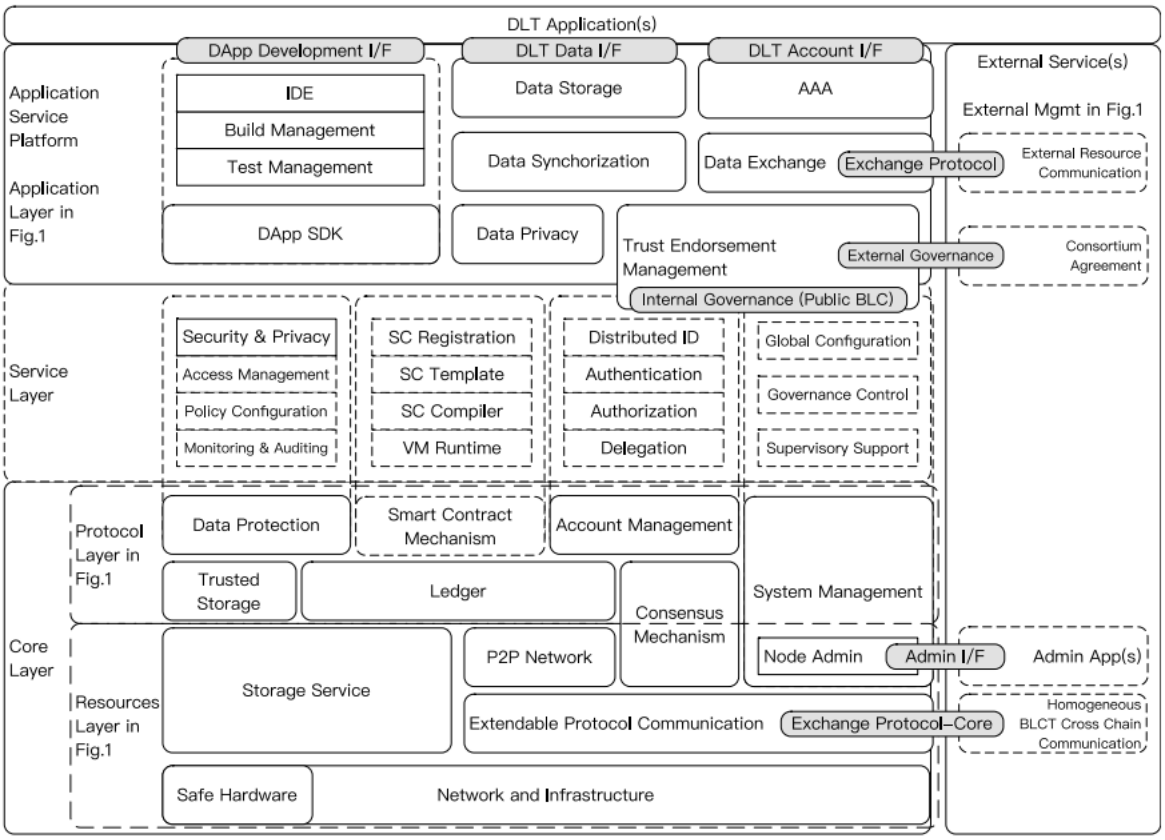

Figure 20 Architecture Diagram from ITU-T Focus Group [48]

ITU-T's trust ecosystem is developed within *Y.3050 series*, developed primarily under the auspices of Study Group 13 (SG13) Future Networks. This series establishes a complete taxonomy and functional architecture for "Trustworthy Networking" *Recommendation ITU-T Y.3051,* provides "The basic principles of trusted environment in information and communication technology infrastructure" [46]. This document establishes the philosophical and regulatory baseline. It states that a "Trusted Environment" is a multidimensional concept requiring a set of technical and regulatory conditions sufficient to establish trust between interacting entities.

ITU-T Y.3052 transforms trust into an operational mechanism known as *Trust Provisioning* [47]. This recommendation is pivotal because it disaggregates the abstract notion of "trustworthiness" into three quantifiable attributes derived from social psychology but applied to network nodes: *(i) Ability* refers to the competence of an entity to perform a required task. In a network context, ability is measured through metrics such as stability, reliability, scalability, robustness, and safety. If a router claims to handle 10 Gbps throughput but consistently fails under load, its "Ability" score degrades, *(ii) Integrity* attribute measures the honesty and consistency of the entity. It encompasses data accuracy, correctness, certainty, and recency. For example, an IoT sensor that transmits outdated timestamped data would suffer a reduction in its Integrity score, *(iii) Benevolence* measures the willingness of an entity to act in the interest of the trustor. In technical terms, this translates to availability, cooperation, and assurance that a service provider will not exploit the user's data for unauthorized purposes. Y.3052 and subsequent supplements (such as Supplement 84) also define a tripartite functional architecture. This architecture standardizes how trust data is harvested, processed, and sold/shared within the network: *(i) The Trust Agent (TA)* is

the sensory organ of the trust framework. Deployed on the device or network node (the trustee), the TA collects raw evidence regarding the entity's behaviour. This includes logs of transaction success rates, latency measurements, and security posture data. The TA acts as the interface between the physical/cyber resource and the broader trust management system, *(ii) The Trust Broker (TB)* serves as the intermediary and aggregator. In a large-scale network, it is inefficient for every device to query every other device directly. The TB collects trust information from multiple Trust Agents, effectively creating a "reputation system." When a Trustor (e.g., a user application) needs to select a service, it queries the Trust Broker for the Trust Index of available candidates. This mechanism is essential for scalability in environments like Social IoT, where billions of devices interact. (iii) *Trust Analysis and Management (TAM)* is the "brain" of the system. It receives raw data from Agents and Brokers and applies algorithms, potentially utility theory-based or AI-driven, to calculate the final Trust Index. The TAM is also responsible for the *Trust Information Lifecycle Management*. Trust is dynamic; it decays over time. The TAM manages the creation, updating (re-evaluation based on new feedback), and abolition of trust scores. If an entity is compromised or decommissioned, the TAM revokes its trust credentials.

2.4.5 IEEE Standardization

Wi-Fi networks typically need to support several security objectives. This is intended to be accomplished through a combination of security features built into the wireless networking standard. The most common security objectives for WLANs are as follows: Confidentiality, ensure that communication cannot be read by unauthorized parties, Integrity, detect any intentional or unintentional changes to data that occur in transit , Availability, ensure that devices and individuals can access a network and its resources whenever needed, Access Control, restrict the rights of devices or individuals to access a network or resources within a network.
The history of security and trust in IEEE for Wi-Fi has evolved significantly, due to the growing need for robust protection and trust. The first attempt goes back to the creation of WEP until the most recent WPA3.
The IEEE 802.11 standardization issues related to trust and security are typically addressed through specific task groups within the IEEE 802.11 Working groups. The IEEE 802.11i amendment allows for enhanced security features beyond WEP and the simple IEEE 802.11 shared key challenge-response authentication. The amendment introduces the concepts of Robust Security Networks (RSN) and Robust Security Network Associations (RSNA). The table below summarizes the main features introduced in the standard to enable trust and security:

Table 2 IEEE standard and security

Task Group	Improvement	Main Impact
802.11ax	Data handling	Indirect Improved QoS handling
802.11w	Protected Management Frames	Protecting management frames prevents spoofing and unauthorized access
802.11i	Wi-Fi Protected Access 2/3	More secure authentication process across all WPA3-supported devices

2.4.6 Summary

Table 3 summarises how several major Standards Development Organisations (SDOs) are converging on distributed (permissioned) ledger technologies from complementary angles. Within ETSI, ISG PDL leads the effort with specifications that range from a PoC framework and smart-contract guidance to a landscape of DLT standards/technologies, concrete application scenarios, and inter-ledger interoperability. ITU-T contributions are centred around the Focus Group on DLT, which has produced reference terms and definitions, an overall DLT ecosystem and architecture view, use cases, and a regulatory framework, complemented by Study Group 20 deliverables that anchor DLT in IoT and NGN environments. IEEE's work is organised through a set of working groups that address data management, IoT/CAV/energy applications, identity and access management, and access-control and interoperability standards. Together, these tracks show a maturing standards ecosystem in which ETSI, ITU-T, and IEEE cover everything from conceptual frameworks and architectures to protocol-level specifications, testing/PoCs, and security and governance requirements for DLT in telecom and 6G.

Table 3: Standards Development Organisations (SDOs), involved in standardising Distributed Electronic Ledgers

Standardization Body	Working Group	Focus Area	Reference / Standard
ETSI	ISG PDL	PoC Framework	GS PDL 005
	ISG PDL	Smart Contracts	GS PDL 011
	ISG PDL	Landscape of Standards and Technologies	GS PDL 001
	ISG PDL	Application Scenarios	GS PDL 003
	ISG PDL	Inter-Ledger Interoperability	GS PDL 006
ITU-T	FG DLT	DLT Terms and Definitions, DLT Overview, Concepts, Ecosystem, Standardization Landscape and Reference Architecture	FG DLT D1.1, D1.2, D1.3
	FG DLT	DLT Use Cases	FG DLT D2.1
	FG DLT	DLT Regulatory Framework	FG DLT D4.1
	FG DLT	Blockchain & DLT	REC-F.751.2
	SG20	IoT	Y.dec-IoT-arch, REC-Y.4476
	SG20	NGN	REC-Y.2342
IEEE	P2144	Data Management	Data Management Standards
	P2418	IoT, CAV, Energy	IoT and Energy Standards
	P2958	Identity & Access Management	Identity and Access Standards

3 UNITY-6G TRUST DOMAIN ARCHITECTURE

The analysis of existing research initiatives, standardization efforts, and architectural approaches in Section 2 highlights a clear trend: while trust is widely recognized as a cornerstone of future 6G systems, it is often addressed in a fragmented manner, either at the level of entities, services, execution environments, or individual domains. UNITY-6G addresses this gap by introducing a *dedicated trust domain architecture* that elevates trust from a collection of isolated mechanisms to *a high priority, system-wide capability* embedded into the 6G control and management plane. This section presents the architectural realization of that vision.

The UNITY-6G Trust Domain Architecture is designed to support *AI-native, intent-driven, and zero-touch operation* across highly distributed, multi-domain 6G environments. It provides the structural foundation needed to collect heterogeneous trust evidence, reason over trustworthiness in a coherent manner, and enforce trust-aware decisions throughout the full-service lifecycle. By integrating orchestration, analytics, policy management, and distributed ledger technologies within a unified architectural framework, UNITY-6G enables trust to be *measurable, explainable, verifiable, and enforceable* across the cloud-edge-IoT and TN-NTN continuum.

This section first introduces a *high-level reference trust architecture*, outlining the main design principles and global trust abstractions that guide UNITY-6G. It then details the functional blocks, interfaces, and communication mechanisms that implement the trust domain, including the role of DLT-based components and security considerations. Together, these architectural elements provide the necessary foundation for the trust model presented in Section 4, ensuring that UNITY-6G's approach to trust is not only conceptually sound, but also technically realizable and scalable in real-world 6G deployments.

3.1 HIGH-LEVEL REFERENCE TRUST DOMAIN ARCHITECTURE OVERVIEW

3.1.1 Design Principles of UNITY-6G Trust Architecture

Beyond detailing the functional components, the UNITY-6G trust architecture also emphasizes a set of foundational design principles that guide how these modules should be integrated and operated in practice.

- First, interoperability-by-design ensures that trust information such as identities, policies, and evidence can flow seamlessly across heterogeneous domains, independent of vendor-specific implementations. This is achieved through standardized interfaces in the adapters and the SBMA, enabling consistent enforcement of policies even when domains differ in their underlying infrastructure.
- Second, the architecture follows a least-privilege and zero-trust mindset, where every action, request, and cross-domain interaction must be explicitly authenticated, authorized, and continuously validated. The Access Manager and Credential Manager embody this philosophy by ensuring that orchestrators interact with the blockchain and other trust systems only through tightly controlled pathways.

- Third, UNITY-6G adopts evidence-driven automation, where operational and trust evidence collected through the Data Continuum directly influences orchestration decisions. SCs and blockchain oracles make this evidence immutable and verifiable, supporting robust auditing and dispute resolution.
- Finally, extensibility and modularity allow the trust layer to evolve with emerging cryptographic methods, identity frameworks, or DLT platforms, ensuring long-term adaptability for 6G networks.

3.1.2 The Global UNITY-6G Trust Architecture

The transition toward 6G networks introduces an unprecedented level of heterogeneity, autonomy, and scale, where services are dynamically instantiated across multiple technological layers (radio, transport, compute), administrative domains, and infrastructure ownership models. In such environments, trust can no longer be established or managed within isolated domains or through static, pre-configured security mechanisms. Instead, 6G requires a *global trust architecture* capable of coherently integrating diverse trust signals, enforcing consistent trust policies, and supporting end-to-end assurance across the entire cloud-edge-IoT and TN-NTN continuum. This necessity motivates the UNITY-6G Trust Architecture.

The UNITY-6G Trust Architecture provides a *system-wide abstraction of trust*, decoupling trust reasoning from individual network components while enabling coordinated, intent-driven trust enforcement across domains. It is designed to aggregate heterogeneous trust evidence, including behavioural analytics, compliance checks and results, AI model confidence, and operational context, into a unified trust representation that can be consumed by orchestration, control, and management functions. By doing so, UNITY-6G avoids fragmented or conflicting trust decisions that would otherwise arise from independently operating trust mechanisms, and instead enables *consistent, explainable, and policy-aligned trust behaviour* at the system level.

A key motivation for this global architecture is to support *AI-native and zero-touch operation* without sacrificing accountability or user control. As AI-driven orchestration increasingly takes responsibility for critical decisions such as service placement, scaling, migration, or isolation, trust must be evaluated holistically and continuously, rather than at discrete checkpoints. The UNITY-6G Trust Architecture therefore establishes a closed-loop trust framework in which trust assessment, decision-making, enforcement, and feedback are tightly integrated with the orchestration lifecycle. This enables UNITY-6G to dynamically adapt trust policies to changing conditions, user intents, and regulatory constraints, while maintaining transparency and auditability across domains. In this way, the global trust architecture serves as the conceptual backbone that unifies UNITY-6G's trust model, functional blocks, and DLT-based realization into a coherent and scalable trust management framework for future 6G systems.
The Figure 21 provides a unifying abstraction for how trust is established, enforced, and audited across heterogeneous network and cloud environments. As depicted in Figure 21, the architecture is organized in four logical layers: the Inter-Domain Management Orchestrator (IDMO) layer, the Cross-Domain Service-Based Management Bus (SBMA), the Infrastructure Layer, and the Trust Layer, which are interconnected by a transversal Data Continuum.

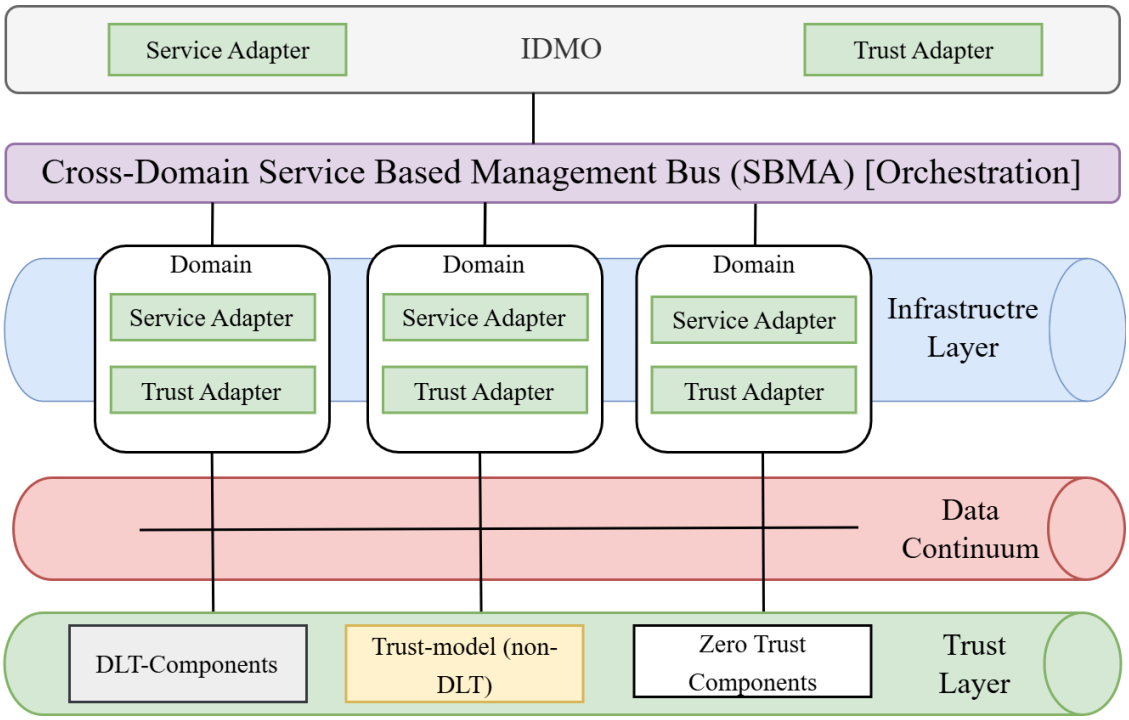

Figure 21 High-level reference trust domain architecture

- *At the top*, the IDMO layer exposes global Service Adapters and Trust Adapters. The Service Adapter translates high-level service objectives and intents into domain-agnostic management operations. The Trust Adapter performs an analogous role for trust, exposing functions such as identity management, attestation, reputation assessment, and policy retrieval in a technology-neutral manner. Both adapters allow the IDMO to coordinate service and trust decisions consistently across domains.
- *In the middle*, the SBMA acts as the common orchestration and management bus. Each administrative domain connects to the SBMA through its own pair of Service and Trust Adapters, which map generic SBMA operations to local interfaces while hiding domain-specific details.
- *Beneath these domains*, the Infrastructure Layer aggregates the underlying radio, transport, core, edge, and cloud resources that are configured and monitored via the adapters.
- *Above bottom*, the Data Continuum provides a shared, end-to-end data fabric where operational metrics, logs, and trust evidence are collected and made available to both management and trust functions.
- *At the bottom*, the *Trust Layer* hosts the concrete mechanisms used to realise UNITY-6G trust, including *distributed-ledger-based components*, *non-DLT trust models*, and *Zero-Trust functions*. Evidence flows from the domains to this layer through the Data Continuum, while trust decisions and policies are propagated back to the domains via their *Trust Adapters*.

3.1.3 DLT-based UNITY-6G Architecture

The increasing decentralization, autonomy, and multi-stakeholder nature of 6G networks fundamentally challenge traditional, centralized trust and security

mechanisms. In future 6G ecosystems, services will be dynamically composed across multiple administrative domains, cloud-edge-IoT infrastructures, and terrestrial-non-terrestrial network segments, often without pre-established trust relationships between involved actors. In such environments, trust can no longer rely on bilateral agreements, centralized authorities, or opaque decision logic. Instead, it must be *verifiable, auditable, tamper-resistant, and shareable across domains*, while preserving privacy and operational autonomy. This motivates the adoption of *Distributed Ledger Technology (DLT)* as a foundational component of the UNITY-6G trust architecture.

The DLT-based UNITY-6G architecture provides a *decentralized trust anchor* that complements AI-native orchestration and intent-based management by ensuring that trust-related evidence, policies, and decisions are recorded in an immutable and verifiable manner. DLT enables multiple stakeholders, such as infrastructure providers, service providers, operators, and verticals, to *independently verify trust claims* without relying on a single trusted third party. By anchoring identities, trust scores, policy commitments, and orchestration events on a shared ledger, UNITY-6G ensures non-repudiation, accountability, and cross-domain transparency, which are essential for trustworthy automation and zero-touch operation in 6G.

Beyond auditability, the use of DLT in UNITY-6G supports *programmable trust enforcement* through smart contracts, enabling trust policies and service-level agreements to be automatically evaluated and enforced as part of the orchestration workflow. Off-chain trust analytics and AI-driven trust assessments can be securely bound to on-chain commitments, allowing UNITY-6G to combine scalable computation with strong trust guarantees. In this way, the DLT-based architecture does not replace existing security or trust mechanisms, but rather *integrates and elevates them*, providing a unifying substrate for consistent, explainable, and enforceable trust management across heterogeneous 6G domains. This positions DLT as a key enabler for realizing UNITY-6G's vision of trustworthy, autonomous, and federated 6G network operation.

Figure 22 shows *a modular blockchain trust integration architecture* that connects both the Domain Management Orchestrator (DMO) and the IDMO to a blockchain network through a shared Service Bus. Rather than interacting with the blockchain directly, the orchestrators rely on dedicated functional blocks that manage all trust-critical operations: the Access Manager enforces authorization policies; the Credential Manager handles identities, keys, and transaction signing; the RPC Manager exposes a standardized interface for submitting transactions and querying on-chain state; and the Smart Contract Manager (SCM) deploys and executes programmable policies that govern cross-domain behaviour.

On the IDMO side, a Blockchain Oracle injects off-chain evidence, such as monitoring data, attestation outputs, or SLA metrics, into SCs, while a Blockchain Adapter translates orchestration requests into blockchain-compatible instructions. In general, these components form a trust layer that provides immutable auditing, verifiable policy enforcement, and secure credential handling, enabling trustworthy and interoperable 6G service management across domains.

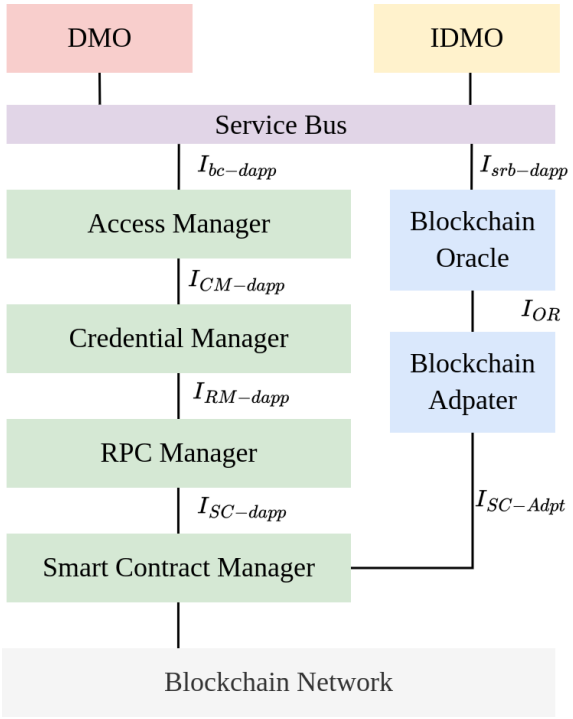

Figure 22 DLT components in the UNITY-6G trust layer

3.2 FUNCTIONAL BLOCKS, COMPONENTS AND FEATURES

3.2.1 Service Domain Orchestrator

3.2.1.1 Introduction

NearbyOne [35] will instantiate a DMO of the UNITY-6G trust layer (Figure 22). It is a powerful, cloud-native orchestration and automation platform designed to streamline and simplify the management of complex cloud, edge, and private network infrastructures. NearbyOne provides a unified operational layer, that enables businesses and service providers to automate deployments, manage infrastructure, and orchestrate services efficiently across multi-cloud, edge computing, and telecom environments.

3.2.1.2 Motivation

NearbyOne offers a multi-dashboard interface that provides users with a tailored experience, allowing them to manage their resources, services, and configurations with precision. Given the transversal capabilities offered across different operational domains, the security model is fundamental. All user actions are secure and follow Oauth2 [49] approval processes, identifying and authenticating the user and ensuring that the user has sufficient authorization level. These processes are subject to strict traceability and accountability rules for adhering to QoS policies.

3.2.1.3 Function of the Block

NearbyOne follows a Role-Based Access Control (RBAC) model, which allows to specify permissions for individuals and groups. The identity management system is based on three core entities: Organizations, Groups, and Users. These resources define how access and permissions are structured within the platform. Users gain access to system resources based on their group membership. The permissions system ensures secure and scalable identity management by mapping users to groups and groups to organizations.

1. An Organization is the top-level entity in NearbyOne. It serves as an independent unit that owns and manages resources, users, and permissions. Each organization functions as a separate entity with its own set of configurations and access control rules.
2. A Group is a collection of users within an organization. Groups are used to manage permissions efficiently, ensuring that users inherit roles and access rights based on their membership.
3. A User represents an individual with login credentials who interacts with the NearbyOne platform. Users are authenticated via LDAP/OIDC or managed locally.

This structure enables multi-tenancy, ensuring that each organization has isolated access control while maintaining centralized management.

3.2.2 Underlying Trust Block

3.2.2.1 Introduction

The Trust Layer is one of the four logical layers of UNITY-6G, on which the trust domain architecture is built, as shown in Figure 21. It provides the mechanisms that establish, enforce, and audit trust in heterogeneous domain environments. Positioned at the bottom of the architecture, represents the foundation of the system. The Trust Layer interacts with the Data Continuum to collect evidence from domains and propagate trust decisions and policies back through Trust Adapters. This layer is a set of trust-enabling technologies, including distributed-ledger-based components, non-DLT trust models, and Zero-Trust functions. In this context, IOTA is adopted as the DLT for its trust and identity capabilities.

3.2.2.2 Motivation

In multi-domain ecosystems, trust cannot be assumed. The Trust Layer is used to ensure that operations within the system occur in a trusted way. This enables control over both the devices and their behaviour as they interact with the system. Without a Trust Layer, vulnerabilities can arise from unverified identities or unauthorized access. Trust Layer addresses these challenges by introducing a decentralized trust foundation that guarantees authenticity, integrity, and privacy. Authentication prevents unauthorized devices from joining the network. Authorization defines the boundaries within a device is allowed to operate, it ensures that an authenticated device performs only the actions it is permitted to do and prevents it from executing unauthorized operations. Data integrity guarantees that information has not been altered. IOTA offers immutable proofs of data exchanges and credentials,

ensuring transparency and auditability. IOTA supports also secure identity management through decentralized identifiers (DIDs) and verifiable credentials, allowing entities to authenticate and authorize interactions without exposing sensitive information, preserving data integrity.

3.2.2.3 Function of the Block

The Trust Layer functions as the system's enforcement point for secure and trustworthy interactions across all domains. It integrates distributed ledger technology's identity management systems and trust frameworks to form a unified trust layer. Through IOTA, the layer ensures that all transactions are recorded immutably, providing transparency and auditability across a distributed and decentralized environment. IOTA trust framework enables entities to establish identities and verifiable credentials, supporting privacy-preserving authentication and authorization.
With these capabilities, the Trust Layer delivers a resilient and interoperable environment where trust decisions are based on cryptographic proofs, real-time evidence, and adaptive policies, ensuring secure collaboration across heterogeneous infrastructures.

3.2.3 Smart Contract Manager Block

3.2.3.1 Introduction

SCs are software programs executed in distributed and decentralized environments such as DLTs. The input of a SC call is recorded immutably on the DLT, leveraging its inherent properties. Execution occurs independently on every validator node, and a consensus on the results must be reached before storing the outcome on the ledger.

3.2.3.2 Motivation

Running SC is safer than executing traditional computer programs because tampering a SC is impossible without breaking the DLT consensus. However, SCs are not free from bugs, so their code must be carefully reviewed to prevent vulnerabilities that could be exploited. This gives the assurance that once a SC has been reviewed and deployed, its results are guaranteed to be correct, public and tamperproof.

3.2.3.3 Function of the Block

The SCM block is responsible to handle the interactions of the system with the DLT. The DLT expose the EVM that is a distributed virtual machine where the code is executed. This block ensures that contract calls are properly formatted and then it submits them to the ledger where the EVM will execute those transactions. Moreover the SCM block can monitor the SCs retrieving information about the SC's state.

3.2.4 Blockchain Oracle Block

3.2.4.1 Introduction

A blockchain oracle is an entity that provides external data to a DLT environment. Since a DLT operates in a closed and isolated context, it cannot directly access information

from the outside world without an oracle. Oracle acts as bridge between off-chain data sources and on-chain SCs, enabling decentralized applications to interact with real-world information.

3.2.4.2 Motivation

DLTs are designed as isolated environments to maintain security and consensus. However, SCs running on these networks often require external data to execute complex logic. Oracles offer a secure and reliable mechanism to deliver this data to SCs, allowing them to operate based on real-world inputs. To ensure accuracy and trustworthiness, oracles typically aggregate data from multiple sources, apply redundancy checks, and validate the information before transmitting it to the blockchain. This approach minimizes the risk of manipulation and guarantees that SCs can make decisions based on verified, tamper-resistant data.

3.2.4.3 Function of the Block

In the trust architecture, the *Blockchain Oracle Block* implemented acts as the trusted gateway between the cyber-range environment and the underlying DLT/PDL infrastructure. It exposes northbound interfaces toward emulated network functions, RIC/xApps, UE traffic generators, and monitoring systems, and southbound interfaces toward the SCs deployed on the ledger. Whenever a scenario requires on-chain logic to take decisions based on KPIs (e.g., QoS, energy use, FL accuracy, security alerts), the oracle is responsible for acquiring the relevant measurements from the cyber-range, transforming them into a canonical format, and issuing the corresponding transactions or function calls to the target SC. From a functional point of view, the block is organised around four core tasks:

1. Data acquisition and pre-processing: the oracle subscribes to telemetry streams or polling endpoints (e.g., REST/gRPC/Message Bus) and aggregates, filters, or normalises the raw metrics required by the scenario.
2. Validation and trust checks: it can apply sanity checks, threshold rules, or cross-source correlation to detect inconsistent or suspicious values before they are propagated on-chain, acting as a first line of defence against faulty probes or simple data poisoning.
3. Transaction preparation and submission: once validated, measurements are encoded into the appropriate smart-contract call (e.g., SLA update, QoS violation event, FL reputation update), signed with the oracle's key material, and submitted to the permissioned ledger through the PDL client libraries.
4. Feedback and event handling: the oracle listens for contract events (e.g., SLA breach flags, reputation changes, incentive triggers) and forwards them back into the cyber-range via its northbound APIs, so that other components (orchestrators, test controllers, UE/UE emulators) can adapt their behaviour accordingly.

3.2.5 Blockchain Adaptor Block

3.2.5.1 Introduction

The Adaptor Block acts as an intermediary between the Oracle and the SCM block. It collects data provided by the Oracle and makes it available to SCs in a structured and reliable way, ensuring compatibility with the DLT environment.

3.2.5.2 Motivation

This block is essential to establish a reliable and standardized communication channel between the Oracle and the SCM block. By introducing the Adaptor Block, the system gains a controlled interface that validates, normalizes, and securely transfers external data from the Oracle to the SCs. This approach ensures that SCs receive accurate and trusted information while maintaining the integrity and security of the DLT environment.

3.2.5.3 Function of the Block

The Blockchain Adaptor Block provides a controlled, bi-directional interface between the Oracle and the SCM block. On the ingress path, it receives data items and events from the Oracle, performs syntactic and semantic validation, and maps them to the data structures and types expected by the SCs. This includes normalisation (e.g., units, identifiers, timestamps), policy checks (e.g., admissible sources, freshness constraints), and optional enrichment with metadata such as domain identifiers or trust scores. On the egress path, the Adaptor assembles transaction payloads, selects the appropriate contract methods, and exposes them to the SCM through an API. It may additionally support batching, rate limiting, and retry strategies, thereby decoupling the timing and reliability of off-chain data delivery from on-chain transaction processing. Finally, the Adaptor exposes monitoring hooks and logging capabilities, enabling auditing and troubleshooting of the end-to-end data path between Oracle outputs and smart-contract state changes.

3.2.6 Communication Between Blocks

3.2.6.1 Introduction

The blocks introduced in the previous subsections (Access Manager, Credential Manager, RPC Manager, SCM, Blockchain Oracle, and Blockchain Adaptor) form a modular DLT service stack. The blocks introduced in the previous subsections (Access Manager, Credential Manager, RPC Manager, SCM, Blockchain Oracle, and Blockchain Adaptor) form a modular DLT service stack. The *Access Manager* acts as the policy enforcement point of the trust layer, evaluating authentication and authorization rules to control which entities may invoke ledger-backed services or trigger smart-contract actions. The *Credential Manager* provides the identity and credential backbone of the stack by managing the lifecycle of cryptographic identities and proofs. The *RPC Manager* abstracts and secures ledger connectivity by managing the remote procedure call interface to the permissioned DLT nodes ensuring reliable submission of transactions and consistent access to on-chain state. Together, these blocks decouple trust enforcement, identity management, and ledger communication from application logic, enabling the SCM, oracle, and adaptor to focus on executing governance rules and anchoring verifiable evidence. To operate coherently within the proposed framework, these blocks must interact through well-defined interfaces that are consistent with the overarching *Service Bus* and *trust architecture*. This subsection

outlines the general communication principles and patterns governing these interactions, with the goal of ensuring interoperability across implementations, simplifying integration with existing management systems, and facilitating future evolution of the DLT components without disrupting other functions.

The interfaces shown in Figure 22 define the logical interactions among the DLT components and the UNITY-6G service bus.

- ***I*$_{bc\text{-}dapp}$**: is the northbound interface between the Service Bus and the DMO-side DLT stack. It exposes high-level operations such as “deploy contract”, “submit transaction”, or “query ledger state” in a technology-agnostic way, allowing orchestration entities to invoke blockchain functionality without being aware of specific clients or networks.
- **I$_{srb\text{-}dapp}$**: connects the Service Bus with the IDMO-side Blockchain Oracle. Through this interface, management and assurance functions can request off-chain data retrieval, verification tasks, or evidence collection that will later be forwarded towards the blockchain via the Adaptor and SCM.
- **I$_{CM\text{-}dapp}$**: is the internal interface between the Access Manager and the Credential Manager. It is used to request, update, and revoke credentials keys required to interact with the blockchain network, and to obtain the associated authorisation policies.
- **I$_{RM\text{-}dapp}$**: links the Credential Manager with the RPC Manager. It transports validated credentials and configuration parameters so that the RPC Manager can establish authenticated and authorised connections to blockchain nodes and sign transactions.
- **I$_{SC\text{-}dapp}$**: is the interface between the RPC Manager and the SCM. It provides methods for invoking smart-contract functions, submitting state-changing transactions, and executing read-only calls. It also conveys status information such as transaction hashes, receipts, and emitted events back to the upper layers.
- **I$_{OR}$**: connects the Blockchain Oracle with the Blockchain Adaptor. The Oracle uses this interface to push normalised off-chain data items, events, or proofs to the Adaptor, which can acknowledge reception, request retransmission, or perform additional checks before the data is considered ready for on-chain use.
- **I$_{SC\text{-}Adpt}$**: is the interface between the Blockchain Adaptor and the SCM. Through this interface, the Adaptor hands over validated and formatted payloads, together with the target contract and method information, so that the SCM can trigger the corresponding on-chain transactions. This interface thus closes the loop from off-chain evidence to on-chain state updates.

3.2.7 Security Considerations

3.2.7.1 Introduction

Authentication, authorization, and encryption form the backbone of modern digital security. A secure system must implement these three properties, which together create the “triangle of trust.” Authentication ensures that only permitted clients can access the system, establishing a boundary between authorized and unauthorized entities. Once authenticated, authorization determines which operations a client is allowed to perform, preventing excessive privileges. Throughout these interactions,

encryption protects all messages and data exchanged within the system, ensuring that information can only be read by the sender and the intended recipient, or an authorized third party with the secret key. These three properties are interdependent; if one is missing, the overall security posture is compromised.

3.2.7.2 Addressing Cybersecurity Threats

Mitigating cybersecurity threats is complex, but implementing the triangle of trust provides multiple layers of defence. Authentication acts as the first barrier, preventing illegitimate users from accessing the system. Authorization serves as the second layer, restricting the actions of authenticated clients to only those necessary for their role, thereby minimizing the risk of privilege escalation or system compromise. This layered approach significantly reduces attack surfaces and ensures that even if one control is bypassed, others remain in place to protect the system.

3.2.7.3 Securing Proprietary Data

Proprietary data must be always encrypted, and the access controlled through authentication and authorization mechanisms. This ensures data confidentiality both in transit and at rest, as encrypted information can only be decrypted by authorized entities possessing the appropriate secret keys. By combining encryption with strict access controls, the system guarantees that sensitive data remains protected against unauthorized access or tampering.

3.2.8 Implementation Aspects

3.2.8.1 Underlying Trust Layer

The Trust Layer is implemented by using the IOTA network and its Tangle-based architecture (Figure 23). The IOTA trust framework can verify the digital identity of entities and enforce authorization policies. Cryptographic proof and signatures are generated and validated through the IOTA trust protocol ensuring that only trusted entities can interact within the system. The Tangle guarantees data integrity by making all operations tamper-proof and auditable. This implementation enables the system to evaluate trust dynamically and communicate securely between nodes, forming the backbone of the system's security and trust model. Furthermore, it leverages a decentralized trust fashion, eliminating the need for a central authority to rely on. It acts as a unifying layer of trust between nodes enabling secure interactions without intermediaries.

3.2.8.2 Smart Contract Block

SCs are implemented using Solidity. It provides a robust and widely adopted environment for defining decentralized logic and it is fully integrated with the EVM, so it allows an easy access to all the DLT utilities. Once the contract code has been fully tested and validated, it can be deployed to the target network, ensuring immutability and the decentralized and BFT execution. An approach, where SCs are thoroughly tested before deployment, allows that their execution aligns with the system's trust requirements since once deployed SCs cannot be modified anymore.

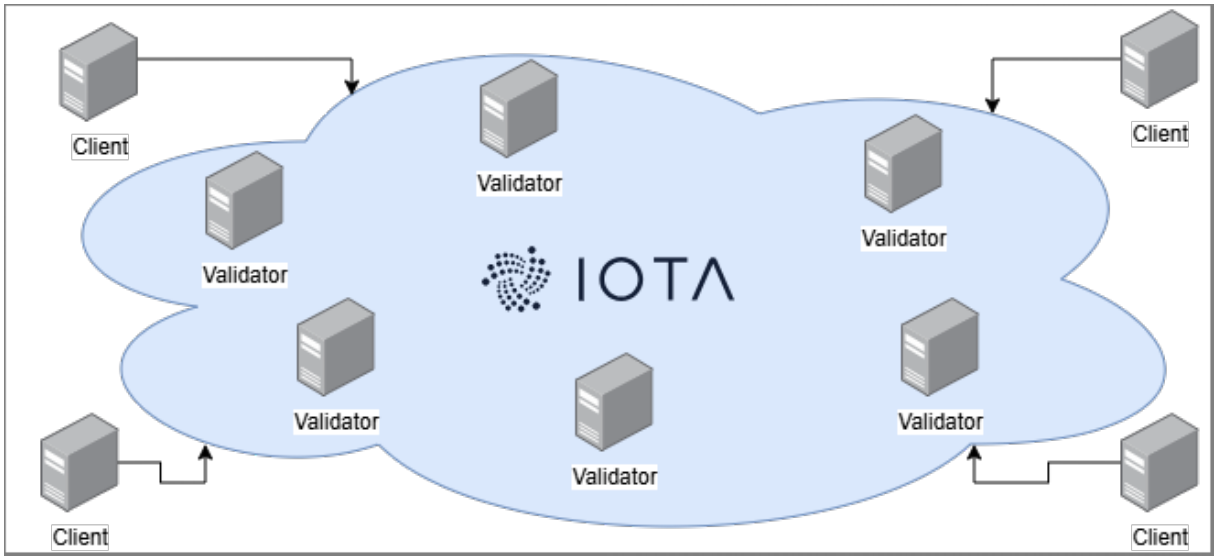

Figure 23 IOTA DLT network example

4 UNITY-6G TRUST MODEL

While Sections 2 and 3 established the state of the art and presented the architectural foundations of the UNITY-6G trust domain, this section focuses on the conceptual and operational trust model that underpins UNITY-6G. The objective of the UNITY-6G trust model is not to introduce a single, monolithic definition of trust, but to provide a flexible, multi-dimensional, and AI-assisted framework capable of unifying heterogeneous trust signals and translating them into actionable decisions across the full 6G service lifecycle. In contrast to many existing approaches that compute trust in isolation, at the level of entities, services, or domains, UNITY-6G positions trust as a system-level construct, dynamically derived from security, privacy, reliability, resilience, and semantic correctness, and continuously adapted through closed-loop orchestration.
The UNITY-6G trust model is explicitly designed to operate in highly distributed, multi-domain, and heterogeneous 6G environments, spanning the cloud-edge-IoT-TN/NTN continuum. It embraces the reality that trust in 6G cannot be static, binary, or solely credential-based; instead, it must be context-aware, intent-driven, explainable, and verifiable. To this end, UNITY-6G integrates user-centric requirements, AI-assisted decision making, and DLT-enabled evidence management, allowing trust to be measured, reasoned upon, and enforced in a transparent and auditable manner. Trust is therefore treated as a living property of the system, evolving in response to changes in context, behaviour, policies, and operational conditions.
This section first introduces the general approach to trust in UNITY-6G, clarifying the relationship between trust, trustworthiness, and Level of Trust (LoT), and highlighting the role of user intent and AI assistance in trust evaluation. It then demonstrates how the trust model can be customized for different scenarios, including IoT, O-RAN, federated learning, TrustNet-based networking, and WLAN environments, illustrating the adaptability of the model across diverse technologies and threat landscapes. Finally, the section presents the security validation framework that supports the UNITY-6G trust model, defining the principles, agreements, and validation mechanisms required to assess, benchmark, and continuously validate trust in practical deployments. Together, these elements establish UNITY-6G's trust model as a unifying, extensible, and operational foundation for trustworthy 6G systems.

4.1 GENERAL APPROACH TO TRUST

In 6G, trust surpasses conventional security-focused frameworks to emerge as a multidimensional, user-centric foundation for next-generation networks. The overall approach for trust within the EU SNS framework is comprehensive: trustworthiness is characterized as the unified guarantee of safety, security, privacy, resilience, and reliability, all continually tuned to address the requirements and intentions of users, tenants, and stakeholders.

This new model is achieved by user-centric designs that utilize distributed artificial intelligence. Cognitive management systems continuously analyse user needs and convert them into implementable configurations, policies, and resource distributions throughout all stages of the 6G lifecycle, from onboarding and deployment to operation

and disconnection. Integrating explainable AI and intent-driven orchestration enables measurable and adaptive trust in 6G, enabling customized security measures and detailed policy enforcement while maintaining usability, agility, and system openness. The EU SNS vision for trust in 6G networks calls for flexible, scalable, and interoperable trust frameworks. These frameworks integrate zero-trust concepts with comprehensive verification methods, allowing the network to swiftly address emerging risks, service requests, and ecosystem needs. Ultimately, trust serves as the cornerstone for new Key Value Indicators, such as Level of Trustworthiness (LoTw), and establishes the basis for user empowerment, safe and dependable service delivery, and public confidence in the 6G digital society.

4.1.1 Trustworthiness and Level of Trust Relation

The vision for 6G embraces an open, distributed, and user-driven evolution of today's Service-Based Architecture (SBA) core networks, presenting new complexities for trust and security. A key challenge arises with the heterogeneous, disaggregated cloud continuum, where multiple stakeholders operate across regions, leveraging private, public, and hybrid cloud systems. Combined with extensive softwarization and IT-centric infrastructure operations, this creates a landscape rich in flexibility but also exposed to unique risks and increased attack surfaces.

While 5G security architectures succeed in protecting largely centralized network designs with protocol-level trust relationships, 6G requires security and trust mechanisms that extend deeper. In future networks, trust connections must evolve to encompass not just technical protocols but the behaviour of devices, networks, and users themselves, reflecting a far broader set of assurances. The move to a user-centric model in 6G calls for making trustworthiness a native feature of the system, integrating robust guarantees for all entities in the ecosystem. Therefore, the most significant paradigm shift in 6G network design is expanding from a sole focus on security, toward a holistic approach to trustworthiness. This encompasses not only security, but also safety, privacy, resilience, and reliability, dimensions that must be designed, orchestrated, and measured natively in next-generation user-centric infrastructures [63].

IMT-2030 designates trustworthiness as a central feature of future 6G systems, driving industry-wide focus among standards groups, suppliers, and researchers to define, implement, and measure it [50]. Trust refers to user confidence, while trustworthiness is the network's actual capability to deliver secure, reliable, private, and resilient service. As standards bodies and technical projects work to address trustworthiness through user-centric architecture, AI-driven evaluation, and security-by-design, it's essential to clarify these concepts. A trustworthy 6G system not only boosts user trust but also streamlines continuous assessment, adaptive management, and robust assurance of safety, security, privacy, resilience, and reliability throughout the network. Consistent terminology and clear definitions will strengthen future research, deployment, and stakeholder engagement.

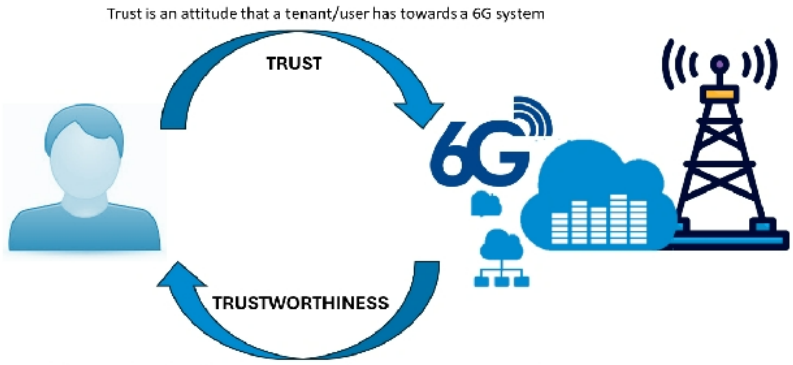

Figure 24 6G Level of Trustworthiness

4.1.2 User-centric and AI-assisted in 6G Systems Trustworthiness

A key paradigm shift in user-centric 6G systems is moving from security-only architectures to a comprehensive trustworthiness model, where safety, security, privacy, resilience, and reliability are equally integral dimensions. Addressing trustworthiness means balancing robust protection with usability, agility, and responsiveness by embedding security-by-design principles and leveraging cognitively coordinated elements such as AI-driven, intent-based trust management [51]. Given that each tenant or user may require different levels of assurance from a 6G network, dynamic adaptation is essential: the network must tailor trustworthiness to individual trust expectations and needs. Instead of static guarantees, 6G networks should provide flexible, user-centric services, enabling granular configuration of core network functions to match the real-time requirements of each stakeholder [52].

The general approach to trust described above is performed by the UNITY-6G trust domain architecture presented in Section 3. In that architecture, the multidimensional notion of trustworthiness (safety, security, privacy, resilience, reliability) is realised through dedicated functional blocks, such as the underlying Trust Block, Smart Contract Manager, Blockchain Oracle and Blockchain Adaptor, which provide verifiable evidence, policy enforcement and closed-loop orchestration across domains. The trust model introduced in this Section (Section 4), therefore defines *what* is evaluated and how Level of Trust (LoT/LoTw) is computed, while the trust architecture in Section 3 defines *where* these computations and enforcement actions are executed within the UNITY-6G system.

In the context of the UNITY6G concept of trust, the terms introduced above such as trustworthiness, Level of Trust, directly support the implementation of the project's trust model and trust layer as follows:

5. The general adoption of trust and trustworthiness in this section provides a common conceptual basis for defining quantitative Levels of Trust that can be integrated as control inputs to the UNITY-6G trust layer and orchestrator, instead of being handled as informal or ad-hoc indicators, thus directly supporting the architectural goal of trust-aware, closed-loop control.
6. The user-centric, AI-assisted view of trustworthiness mentioned above, supports the UNITY6G objective of translating heterogeneous user and tenant

requirements into machine processable trust policies that span terrestrial and non-terrestrial networks (NTNs), O-RAN, core, edge, and IoT domains.

7. By framing trust within continuous assessment and explainability principles, the proposed approach directly underpins the UNITY6G design choice of combining AI-driven analytics with distributed ledger-based evidence anchoring and smart contract based enforcement in the Trust Domain architecture.
8. The abstract LoT concepts introduced in this section enable a uniform vocabulary and metric space that can be specialised in the scenario specific models of Section 4.2, ensuring that UNITY6G maintains consistent trust semantics across IoT, ORAN, Wi-Fi, and AI security use cases.
9. The resulting conceptual framework positions trust as a first-class, measurable property that can be monitored, audited, and adapted over time, and it explicitly targets the UNITY-6G architectural goals of enabling trust evidence collection, evaluation, and enforcement through the Trust Domain's functional blocks and DLT-based mechanisms across heterogeneous 6G environments.

4.2 CUSTOMIZATIONS PER SCENARIO

4.2.1 DLT-enabled trustworthy and FL for IoT

The 3GPP has driven the evolution of fifth generation (5G) and beyond systems. Release 16 introduced the Network Data Analytics Function (NWDAF) for data-driven intelligence, followed by Release 17's broader analytical capabilities. Release 18, known as 5G Advanced, integrates AI and ML to support more flexible and distributed network operations. These advances align with the move toward edge computing in the IoT, where applications like smart cities and healthcare demand low latency and real-time analytics. FL enables privacy-preserving intelligence by allowing devices to train locally and share model updates instead of raw data ideal for heterogeneous IoT environments [55][53].

However, FL faces challenges in IoT settings. Large model transfers increase overhead, gradient sharing risks privacy, and the central aggregator creates a single point of failure. Organizations such as ETSI emphasize the need for decentralized, trustworthy ML with mechanisms to verify model integrity without centralizing data.

Blockchain integration can address these issues by replacing the central aggregator with a tamper-evident, transparent infrastructure. It enables secure recording and verification of model updates while reducing single points of failure. Recent 3GPP and ETSI efforts explore permissioned distributed ledgers (PDLs) [54] to enhance trust and accountability in FL systems. Still, IoT constraints demand lightweight blockchain integration. Only essential data should be anchored on-chain, with gateways filtering updates. On-chain reputation and verification mechanisms are needed to ensure model integrity under resource limitations.

Existing studies partially tackle these challenges some remove central servers, others add reputation systems or optimize FL for IoT, but none combine all these features effectively. Many retain centralized elements or ignore energy constraints. This work addresses these gaps by combining feeless IOTA transactions, a lightweight on-chain reputation manager, and gateway-assisted offloading, achieving decentralized aggregation, verifiable trust, and energy-efficient operation for FL in IoT environments.

4.2.1.1 Considered Problem

The problem under consideration lies at the intersection between FL, the IoT constraints, and the need for trustworthy, standards-aligned network intelligence. As 5G evolves toward 5G-Advanced, 3GPP and O-RAN-based architectures push AI/ML capabilities closer to the edge, including support for FL in RAN and IoT scenarios. However, most standardized FL architectures still assume a logically centralized aggregation server that collects model updates from devices and computes a global model. This central entity becomes a single point of failure and a concentration of trust: if compromised or misconfigured, it can degrade model quality, leak sensitive information, or act as a bottleneck for latency-sensitive IoT services [56].
In IoT deployments, these limitations are exacerbated by stringent resource constraints. Battery-powered or low-power devices must repeatedly upload large model parameters over unreliable links, which increases energy consumption and occupies scarce radio resources. Even when raw data remain local, gradients and parameters can still leak private information, making it difficult to reconcile FL with strict regulatory and privacy requirements in domains such as smart cities or healthcare. At the same time, public and even many permissioned blockchains are not directly suitable as drop-in replacements for the aggregator: transaction fees, heavy consensus mechanisms, and the requirement to maintain full ledger replicas conflict with the capabilities of typical IoT nodes.
Recent research on blockchain-enabled FL introduces decentralization and transparency but often addresses only isolated parts of the problem. Some schemes remove the central server but ignore resource-constrained IoT devices; others introduce on-chain reputation or auditing yet rely on expensive, fee-based ledgers; still others examine IOTA- or DAG-based architectures without integrating standard-compliant trust and management functions. In short, there is no integrated solution that (i) eliminates the single FL aggregator, (ii) supports verifiable, reputation-aware model updates, and (iii) respects the tight energy and bandwidth budgets of IoT devices while remaining compatible with ongoing 3GPP and ETSI efforts. The core problem addressed in this work is thus to design a blockchain-based FL architecture that offers tamper-evident model aggregation and trust management, without pushing consensus or ledger maintenance burdens onto constrained IoT devices [56].

4.2.1.2 Motivation

The motivation for our approach is twofold: (i) to close the gap between research on blockchain-enabled FL and the practical constraints of IoT deployments, and (ii) to align this solution with emerging standardization trends. On the research side, FL has been widely investigated to embed privacy-preserving intelligence into 5G and beyond network management, spanning use cases from RAN optimization to core-network analytics. Parallel work on blockchain-enabled FL shows how distributed ledgers can provide tamper-evident logging, decentralized coordination, and reputation mechanisms for FL participants. Yet, many of these proposals either assume relatively powerful edge nodes, accept non-negligible transaction fees, or lack a detailed treatment of energy consumption and communication overhead in realistic IoT environments.
From a standards perspective, several organizations are converging on the need for decentralized and trustworthy AI/ML. 3GPP study items on FL and AI/ML in 5G-Advanced highlight the importance of member selection assistance, data analytics,

and exposure of network information to FL applications, but deliberately leave trust and ledger mechanisms outside their scope. ETSI's ZSM framework introduces AI enablers for zero-touch network and service management, stressing governance and trustworthiness of AI/ML functions across domains. In parallel, ETSI's PDL work item (e.g., GR PDL 032) explicitly explores AI for permissioned distributed ledgers and identifies FL as a key privacy-preserving application area but does not define a concrete IoT-tailored FL-over-DLT architecture [54].

These developments motivate an architecture that can serve as a concrete, standards-friendly blueprint. The goal is to show how a permissioned DLT, in our case, a private IOTA Tangle can act as a lightweight yet trustworthy aggregation substrate, while IoT devices interact only through simple, low-overhead messaging (e.g., MQTT). By anchoring only essential metadata and cryptographic hashes on-chain, and by delegating heavy operations such as consensus and proof-of-work to dedicated validator nodes, we aim to reconcile the need for verifiable, auditable FL with the practical realities of IoT hardware. In addition, embedding on-chain reputation and verification mechanisms addresses open concerns about malicious or low-quality model updates, further justifying the proposed framework as a natural evolution of current standardization efforts [42].

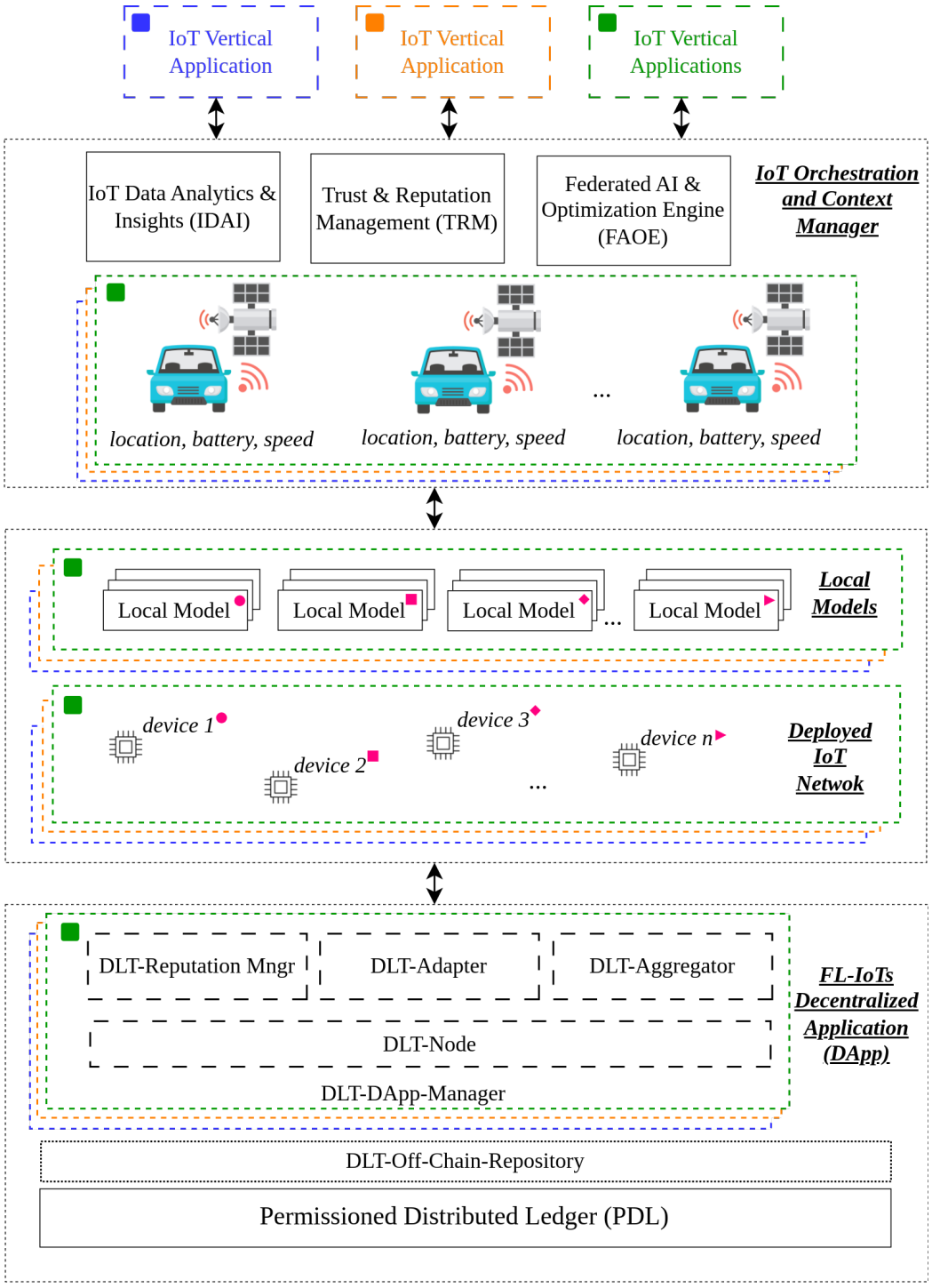

Figure 25 High-level view of system architecture for DLT-enabled trustworthy and FL for IoT.

While previously presented high-level architecture presents the generic cross-domain orchestration and trust-plane architecture, below we present a proposed architecture that provides its instantiation for an FL-driven IoT vertical. The SBMA's Service/Trust Adapters and IDMO functions are reflected in the orchestration and trust/reputation components and the trust layer is concretely realized via the DLT-backed FL-DApp that anchors evidence and reputation state on a permissioned ledger.

4.2.1.3 Proposed Approach

To address the above challenges, we propose a blockchain-based FL framework in Figure 25 that leverages a permissioned IOTA Tangle as a decentralized aggregation and trust layer, while keeping IoT devices as light as possible. The overall architecture is structured into three main components: (i) an IoT Orchestration and Context Manager, responsible for device analytics and FL coordination; (ii) local models trained on heterogeneous IoT devices (or network functions) using their own data; and (iii) an FL-IoT DApp that manages interactions with the permissioned DLT. Within the FL-IoT dApp, a DLT-dApp Manager orchestrates three core modules: the DLT-Adapter, DLT-Verifier, and DLT-Aggregator. IoT devices never interact directly with the Tangle. Instead, they send local model updates over a lightweight protocol such as MQTT to the DLT-Adapter, which authenticates devices based on information from the Trust & Reputation Management (TRM) module, discards invalid or redundant updates, and forwards valid updates to the DLT-Verifier and DLT-Aggregator. The full model parameters are kept off-chain in an external repository; only cryptographic hashes and minimal metadata are anchored on the Tangle using zero-value, feeless transactions, with remote Proof-of-Work performed by permissioned IOTA nodes. This design protects model confidentiality, minimizes on-chain storage, and shifts computational effort away from constrained IoT devices.
The DLT-Verifier evaluates incoming model updates against validation data or quality metrics and maintains a ledger-based reputation score for each device. Updates from unreliable nodes are penalized or discarded, whereas reputable contributors gain influence in the aggregation step. The DLT-Aggregator performs a weighted aggregation (e.g., reputation-aware Federated Averaging (FedAvg)) of verified updates to produce a new global model. A hash of this global model, together with references to contributing updates and reputation adjustments, is recorded on the Tangle, enabling auditable traceability of FL rounds. Experimental evaluation on a private Hornet-based IOTA network with multiple FL rounds and simulated IoT clients demonstrates that this approach sustains stable transactions per second and bounded block processing times under increasing FL workloads, supporting its suitability for trustworthy, energy-aware FL in IoT settings.

4.2.2 DLT-enabled trustworthy and FL for O-RAN

4.2.2.1 Considered Problem

In O-RAN, intelligence is pushed into disaggregated, multi-vendor RAN components and the SMO framework, with ML training and inference distributed between the Non-RT and Near-RT RICs [68]. This openness creates powerful opportunities for collaborative optimization (e.g., cross-operator traffic prediction or resource allocation) but also amplifies trust and accountability challenges. Service providers are expected to contribute model updates to shared FL processes, yet there is no built-in mechanism to verify the quality of these updates, nor to prevent a single actor from unduly influencing the global model.
Traditional FL still assumes a logically centralized aggregator that coordinates model updates and acts as a de-facto trust anchor. In a multi-vendor O-RAN deployment, that assumption is problematic: an aggregator operated by one stakeholder may not be fully trusted by others, and its misbehaviour or compromise could degrade models used across the RAN. At the same time, naïve blockchain integration recording every

partial update on-chain, would be too costly and complex to operate at O-RAN scale, especially when dozens of clients participate in each FL round. Many existing blockchain-for-FL proposals use private ledgers or simulations and do not expose realistic gas usage, latency, or concurrency effects on a live network [57].
Furthermore, while prior O-RAN-blockchain works have targeted zero-trust security, resource sharing, or identity management, they generally stop short of offering a concrete reputation mechanism tailored to FL clients and integrated into the O-RAN AI/ML workflow. The result is a gap between the need for verifiable, reputation-aware FL in standardized O-RAN architecture and the lack of a practical, evaluated solution that can run over real-world DLT infrastructure [58].

4.2.2.2 Motivation

DLT-enabled trustworthy and FL for O-RAN is motivated by the need to move from implicit to explicit trust in O-RAN FL deployments. O-RAN specifications emphasize embedded intelligence and flexible placement of ML training and inference across Non-RT and Near-RT RICs, including support for FL in their AI/ML workflow descriptions. However, these specifications do not prescribe how multiple providers should build verifiable trust in each other's FL contributions, especially when models affect shared RAN behaviour. A blockchain-backed reputation layer offers a natural way to make this trust explicit and auditable without centralizing raw data [58].

From a research and standardization perspective, the aim is to provide a concrete reference implementation that complements ongoing work in O-RAN, ETSI PAS adoption of O-RAN specs, and broader industry efforts exploring blockchain for telecom trust. The prototype seeks to show that a realistic O-RAN FL scenario with tens of clients and multi-round training can be mapped onto an EVM-compatible Layer-2 network and instrumented to expose gas consumption, block size, transaction counts, and latency under both sequential and concurrent loads. This fills an evidence gap left by prior work based on private chains or purely analytical models [57].

There is also a strong "lightweight-by-design" motivation. In O-RAN, control-plane and management-plane functions operate at large scale and must remain cost- and energy-efficient. The proposed framework explicitly targets minimal on-chain footprint (e.g., batching reputation updates into a single transaction per FL round and keeping heavy computations off-chain), while still delivering transparent, tamper-evident records of client participation and reputation. By demonstrating this on the Polygon Amoy testnet, DLT-enabled trustworthy and FL for O-RAN illustrates how FL trust mechanisms can benefit from current and future L2 scalability improvements without requiring O-RAN operators to run bespoke blockchains.

4.2.2.3 Proposed Approach

The proposed solution instantiates a blockchain-enabled reputation framework tightly aligned with the O-RAN architecture [68]. ML training for the considered use case (CPU resource prediction from MonB5G) is hosted in the SMO (FL Training Host), while inference runs in the Non-RT RIC (FL Inference Host), corresponding to an O-RAN deployment option where the SMO orchestrates training across multiple Client Service Providers (CLSPs) and a single Aggregator Service Provider (AGSP). Each CLSP

trains local models on its own data and interacts with a blockchain DApp that provides registration, performance reporting, and reputation management.
The DApp (BC-DApp) is deployed on the Polygon Amoy Layer-2 testnet and is composed of several SCs: an access-control and *registrationClient* contract for onboarding CLSPs and the AGSP; a *performanceSubmission* contract to record performance metrics (Normalized Mean Squared Error, NMSE) per FL round; and a *reputationCalculation* contract that maintains client reputation scores and selects the top-performing 90% of clients for the next round. A Chainlink-based oracle path (fetchOracle) bridges off-chain FL infrastructure and on-chain contracts, allowing NMSE values computed by FL clients to be securely injected into the ledger without exposing model weights. The reputation $R_i$of client i is computed as:

$$R_i = \frac{\alpha}{(\ NMSE_i + \ \epsilon)}$$

where $\text{NMSE}_i$is the NMSE of client i, $\alpha = 10^{18}$is a scaling factor used for fixed-point arithmetic in the SC, and $\varepsilon > 0$ is a small constant (e.g., $\varepsilon = 10^{-3}$) that stabilizes the calculation and avoids division by zero hence lower NMSE therefore yields a higher reputation score.
The workflow is as follows: the AGSP deploys the dApp and opens a registration phase; CLSPs register and receive blockchain credentials. In each FL round, the AGSP distributes the current global model; CLSPs train locally, compute NMSE, and submit their metrics via the oracle to the *performanceSubmission* contract. The *reputationCalculation* contract then updates each client's reputation according to the above inverse-NMSE formula and selects the top-reputation clients to participate in aggregation. The AGSP performs a weighted aggregation of model updates, favouring higher-reputation clients, and distributes the updated global model. All key events registrations, metric submissions, and reputation updates are immutably logged on-chain.
To keep the blockchain integration lightweight, reputation scores are computed largely off-chain and committed as a single batched transaction per round, reducing the number of on-chain operations from O(number of clients) to O(1) per round. The implementation with 50 clients and one aggregator is evaluated in terms of gas usage, block size, transaction count, and latency (both sequential and concurrent), demonstrating that the approach can operate within practical cost and performance bounds while providing an auditable trust layer for O-RAN FL [59].

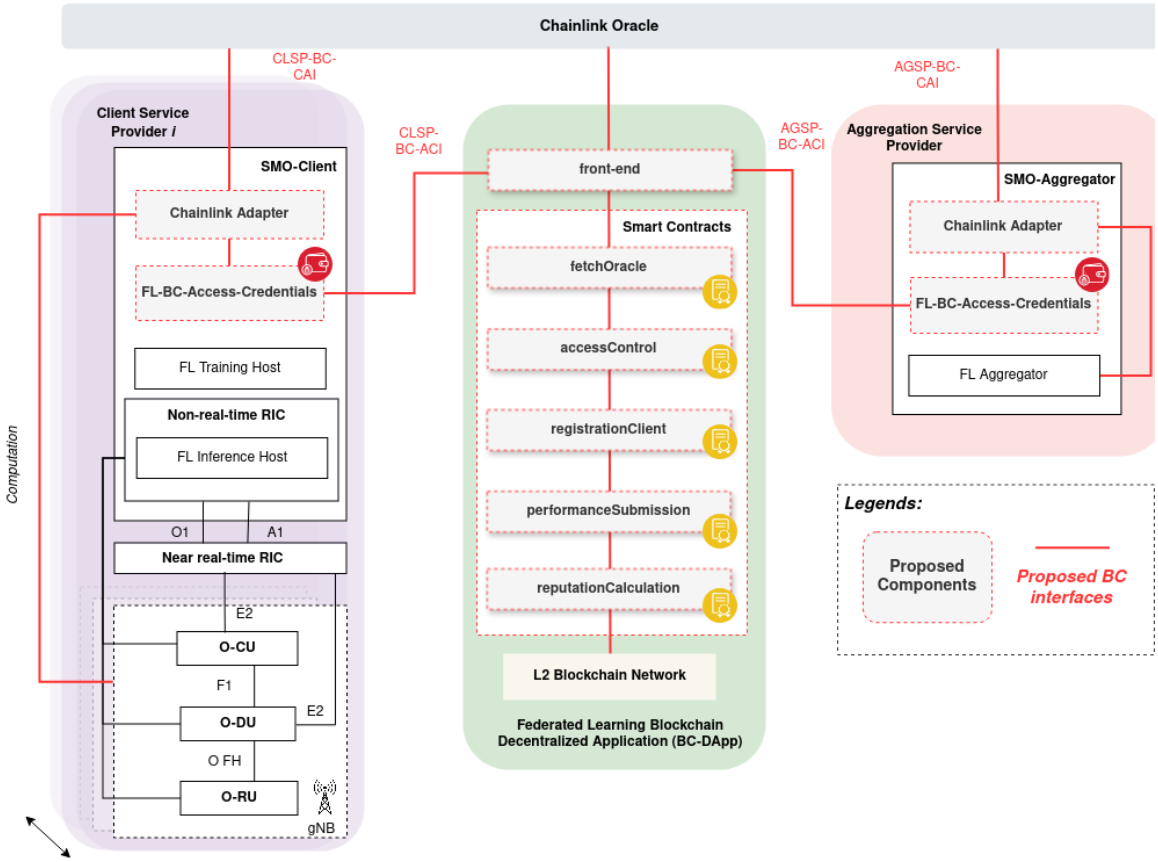

Figure 26 Proposed Framework: Functional blockchain-enabled O-RAN architecture for trustworthy FL using smart contracts

Implementation Aspects:

From an implementation standpoint, the FL-dApp is deployed on an Ethereum-compatible Layer-2 network (Polygon Amoy) and relies on a lightweight tool chain to keep the setup reproducible. A dedicated Smart Contract Orchestrator (SCO) acts as the SCM, encapsulating deployment, configuration, and interaction with the core contracts. Off-chain components run in a Linux/Node.js environment using Hardhat, Ethers.js, and dotenv to manage RPC connectivity (via, e.g., Alchemy), private keys, and chain IDs. Minimal *.env* file stores the RPC URL, private key, and network parameters, while scripts handle deployment (*deploy.ts*), binding of the oracle job, round management, and score updates. SMO clients and the SMO aggregator connect through dApp and Oracle adapters: the dApp Adapter issues registration and policy queries, managing gas estimation, nonces, and retries; the Oracle Adapter signs NMSE metrics and pushes them to the blockchain oracle endpoint, which in turn forwards validated payloads to the SCO. All training and aggregation logic remains off-chain and interacts with the FL-DApp only through these adapters, ensuring that blockchain integration is modular and non-intrusive.

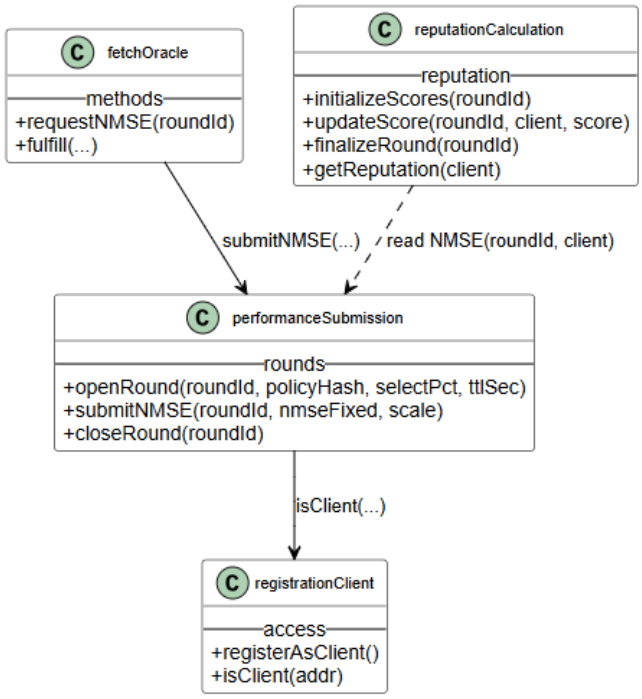

Figure 27 Class Diagram of Smart Contracts

Within this setup, the SCO coordinates a set of Solidity SCs that implement registration, metric submission, and reputation management. The *registrationClient* contract centralizes identity and role management. It maintains mappings to track which addresses are registered as clients or as the aggregator and exposes helper functions such as *registerAsClient()* and *isClient(addr)*. These functions are used by the SCO as read-only guards so that only onboarded CLSPs can submit performance evidence and only the AGSP can exercise round-control operations like opening, closing, or finalizing rounds. This ensures that state transitions are driven exclusively by authorized actors while keeping the access-control logic compact (see Figure 27).

The *fetchOracle* and *performanceSubmission* contracts jointly implement the oracle boundary and round-scoped metric storage. *fetchOracle* registers the trusted oracle address, binds feed identifiers, and enforces freshness via a time-to-live (TTL) parameter. When the Oracle Adapter pushes a signed payload, *fetchOracle* validates signature, membership (client role), and timestamp before forwarding the fixed-point NMSE value to *performanceSubmission*. The latter maintains per-round structures that store a single effective *nmseFixed* value per ⟨client, roundId⟩, along with round flags (e.g., Open, Closed). Submissions are accepted only while the round is open and are quantized as $nmse_{fixed} = |NMSE \cdot 10^6|$ to reduce storage cost and avoid floating-point arithmetic on-chain. Events such as *PerformanceSubmitted* are emitted to allow the off-chain Reputation Manager to subscribe to metric updates.

Finally, the *reputationCalculation* contract implements the on-chain reputation layer. It maintains a round-scoped mapping from client addresses to fixed-point reputation scores, along with configuration parameters such as the selection percentage (e.g., top 90%). Reputation scores are computed off-chain by the AGSP using an inverse-NMSE rule and then committed on-chain in batches via *updateScore(roundId, clients[], scores[])*. Guard checks ensure non-empty, length-matching arrays, score bounds

(e.g., ≤ 1e18), and that the round has not yet been finalized. Once all scores for a round are written, *finalizeRound(roundId)* seals the snapshot, preventing further mutations and marking the reputation state as ready for consumption by downstream selection logic. This batched design minimizes the number of on-chain writes from O(number of clients) to effectively one transaction per round, reducing gas while preserving a tamper-evident, auditable history of client performance and reputation.

4.2.3 TrustNet: Trust-Based Networking Architecture

4.2.3.1 Considered Problem

Figure 28 shows a multi-domain network architecture in which several Autonomous Systems (AS-1, AS-2, AS-3) are connected to routers that use AI-based routing decisions. The figure shows three possible paths (Path-1, Path-2, and Path-3) between Router-1 and a data center in AS-3. Each path is evaluated by the AI at the edge nodes, which assign probabilities reflecting the trustworthiness of the paths based on various factors such as latency, congestion, or past performance. However, there are several challenges to this scenario. First, the variability of path probabilities between different routers can lead to inconsistent decisions. For example, Router-1 assigns the highest trust to Path-1 (93%), while Router-2 prioritizes Path-3 (95%) based on its local AI analysis. This discrepancy in probability assignments can lead to inconsistencies in routing decisions, resulting in *suboptimal or conflicting data flows*. Secondly, a lack of global coordination between the AI systems of different routers can lead to inefficiencies. Since each router operates independently, *the lack of a unified trust assessment* across the network leads to fragmented decision making. This fragmentation can lead to bottlenecks or increased latency if the routers prioritize different paths.

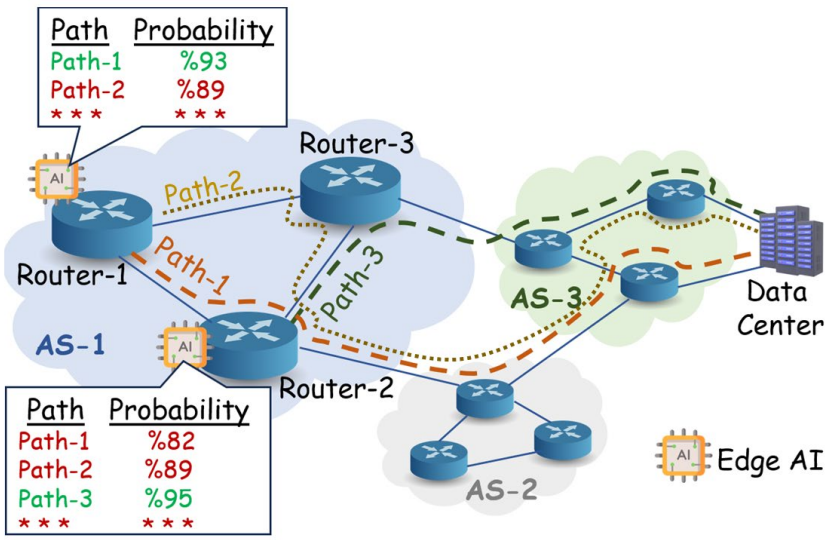

Figure 28 Routers with AI capabilities in the network can

The AI models in the edge routers do provide initial probability calculations or path selection, but their ability to adapt to rapid changes in the network is limited without real-time updates from other nodes or a centralized coordination mechanism. The

multi-domain nature of the modern networks brings challenges in interoperability and trust assessment. Path-2 in Figure 28, for example, spans three autonomous systems (AS-1, AS-2, AS-3), which requires trust assessments in different administrative domains with different policies, metrics, and standards. This lack of standardization can undermine the accuracy of trust-based routing decisions and increase the complexity of ensuring secure and reliable communication. Although AI-assisted routing can offer significant potential to improve trust in path selection strategy, issues such as conflicting probability assessments, lack of coordination, limited adaptability to real-time conditions, and interoperability between multiple domains need to be addressed to fully realize the benefits. A mechanism for global trust coordination, dynamic trust updates, and the use of standardized trust metrics to effectively address these challenges can be used to solve the problem stated above.

4.2.3.2 Motivation

Trust is essential to AI-native network management, enabling secure, context-aware operations and resilience against cascading failures from compromised devices. In highly dynamic environments such as edge and multi- domain networks, trust must be continuously evaluated and integrated into automated decision processes.

4.2.3.3 Proposed Approach

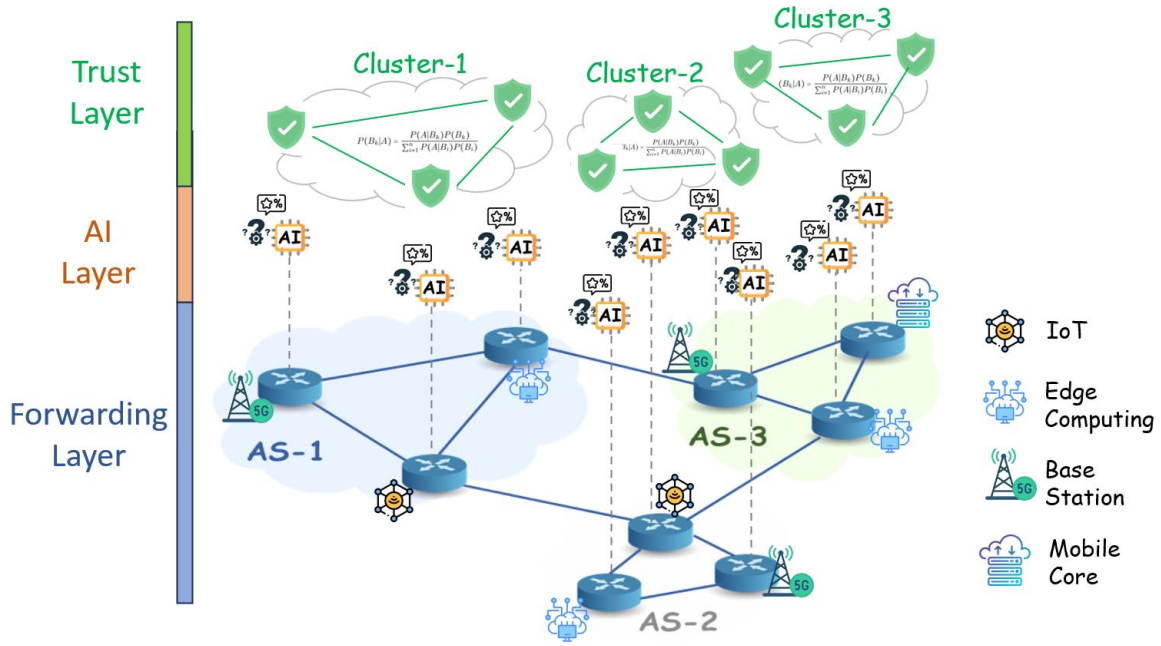

Figure 29 The proposed trust layer combines the individual AI results within the cluster and computes a single trusted

Figure 29 illustrates the high-level structure of the TrustNet architecture, which augments conventional network stacks, comprising data, control, and management planes, with a dedicated trust layer. Traditional infrastructures often lack mechanisms to dynamically evaluate node trustworthiness. Trust layer is located above AI layer and is used to combine the individual AI results of routers path selections within a predefined cluster and computes a single trusted path for a given cluster.

4.2.4 Trust modelling and Security Customization for 802.11 networks

In IEEE 802.11 networks, trust and trustworthiness are emerging as critical factors in network management and optimization.

These factors are particularly important in scenarios where a centralized controller manages a network of access points (APs), grouping them via clustering techniques to optimize the network. In this scenario, the controller clusters APs based on activity patterns (e.g. high/low traffic, time of day) by collecting data from different access points in the network. Subsequently, the controller may use clustered FL approaches to optimize the network (see Figure 30).

Therefore, incorporating the concept of trust into clustering activities improves the network's capacity to provide secure, dependable and robust services, thus maintaining user confidence.

In the context of Unity-6G, we first focus on detecting data poisoning and corruption, and second on identifying anomalies when collecting data from different access points for a clustering algorithm and a clustered FL solution.

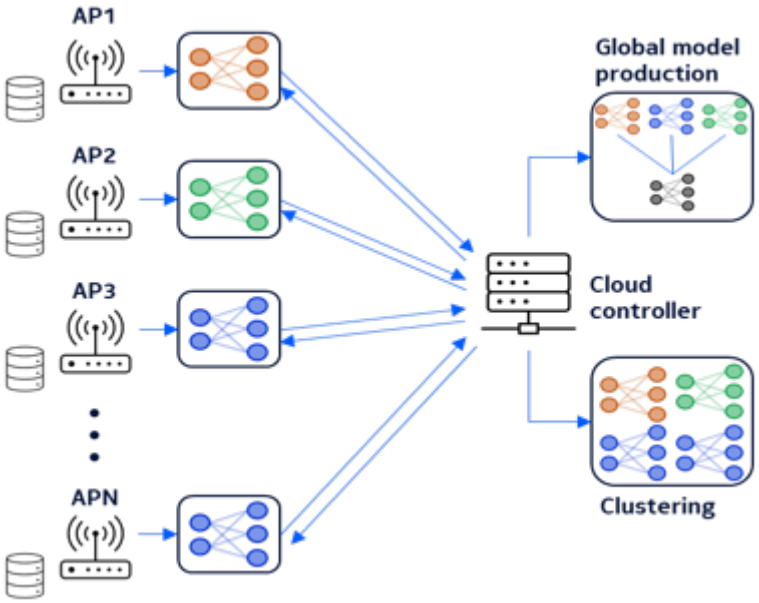

Figure 30 Cloud aggregation and clustering of AP-trained local models

4.2.4.1 Proposed anomaly detection activity for trust in 802.11 networks

In IEEE 802.11 networks, a centralized controller can evaluate the trustworthiness of connected access points (APs) based on the exchanged data. Specifically, we propose a refined trust modelling framework for IEEE 802.11 networks, where trust is defined in terms of anomaly detection. Trust, for example, can be established in this model by identifying access points that transmit corrupted (or not complete) data or fail to comply with Wi-Fi security and encryption requirements. In this model, trust is characterised by an AP's ability to transmit data in the expected format and comply with mandatory Wi-Fi security standards such as WPA3 and IEEE 802.11w. Anomalies are defined as deviations in data format, encryption or security compliance, which may indicate

tampering, transmission errors or malicious activity. On the other hand, trustworthiness can be determined through objective metrics such as security compliance (e.g. adherence to WPA3 and IEEE 802.11w encryption standards), reliability (measured by uptime and failure rates), privacy (to prevent data leaks) and resilience (in recovering from attacks or failures).

To implement this approach, a centralized controller monitors incoming data from APs, analysing each packet for anomalies in format, encryption, and compliance. We consider to develop an outlier/anomaly detection algorithm, trained on normal data patterns can be employed to detect deviations indicative of corruption or non-compliance. Data packets are evaluated in real-time to promptly identify and isolate problematic APs. As new data arrives and anomalies are identified, the online trust evaluation model assigns dynamic trust scores to APs based on the frequency and severity of the detected anomalies. Those with high trust scores are grouped into clusters that handle sensitive traffic, while those with low trust scores are isolated or assigned to less critical clusters. Those that consistently transmit corrupted or non-compliant data are excluded from the network or subjected to stricter monitoring. This protects sensitive traffic from compromised APs, optimises network performance and resource allocation, and maintains secure communication channels.

Additionally, we consider to implement anomaly detection for a clustered FL solution. In this approach, we propose the same centralized controller to monitor the quality and the quantity of the clusters formed to contribute to the global model of the FL solution. In this case, clusters that do not satisfy a certain quality threshold or do not contain a high enough number of Aps will be identified as an anomaly.

These approaches have significant benefits: they prevent untrusted access points (APs) from compromising sensitive traffic; ensures that clusters for advanced machine learning techniques (e.g. clustered federated learning for network optimisation) are formed with reliable APs; and builds user confidence by demonstrating a commitment to secure and dependable service delivery. By integrating these concepts, IEEE 802.11 networks can achieve greater operational efficiency and resilience while maintaining high security and reliability standards.

4.2.5 Security of AI algorithms in O-RAN

4.2.5.1 Motivation

The shift toward open, disaggregated O-RAN architectures significantly increases both transparency and complexity of the RAN. AI/ML algorithms, deployed in the Non-RT RIC, Near-RT RIC, and edge modules, support critical functions such as resource optimization, interference mitigation, mobility management, and anomaly detection. Their performance directly influences network reliability.

At the same time, AI becomes a high-value target. Open interfaces broaden the attack surface, while the presence of AI/ML introduces entirely new classes of attacks specific to machine learning systems. Securing AI in O-RAN is therefore essential to maintain

trustworthy automation, protect sensitive data, and ensure resilience against adversarial manipulation.

4.2.5.2 Security Opportunities and Challenges in O-RAN

Opportunities:

- Programmable RIC platforms enable specialized security xApps and rApps for continuous monitoring, threat detection, and behavioural analysis.
- Distributed AI allows attacks to be detected closer to their origin with reduced latency.
- Openness increases visibility and testability compared to proprietary RAN systems.

Challenges:

- Expanded exposure of interfaces and components increases vulnerability.
- AI/ML models are susceptible to adversarial manipulation during training and inference.
- ML-as-a-Service and API-based access create new risks of model theft and information leakage.
- Distributed learning (e.g., federated learning) introduces risks of poisoning and manipulation of global model parameters.

4.2.5.3 Threats Against AI in O-RAN

Attacks on AI/ML systems in O-RAN (Fig. 31) can be divided into three main groups, each affecting different stages of the model lifecycle:

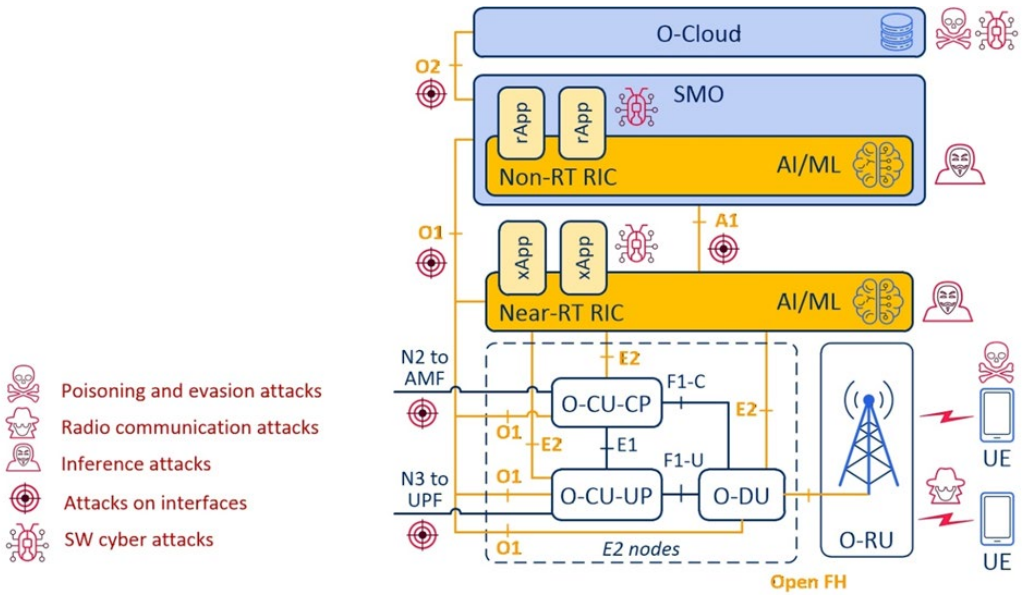

Figure 31 View of the attacks on the O-RAN infrastructure

4.2.5.3.1 Poisoning Attacks

These attacks target the training phase and aim to corrupt the model before deployment. They are achieved by injecting malicious samples, manipulating training data, or tampering with the learning logic (e.g., in federated learning). Their goal is to force the AI model to learn distorted patterns, which later degrade prediction accuracy or introduce backdoors.

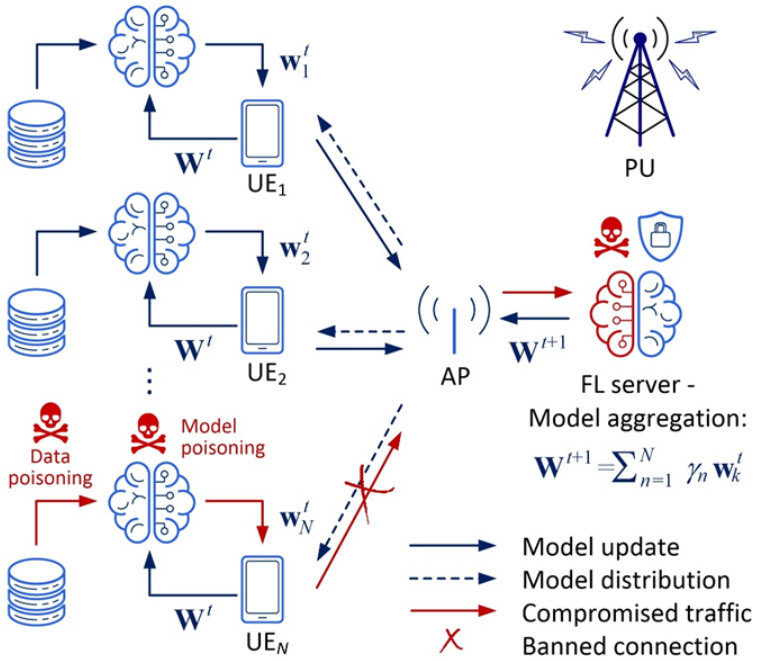

Figure 32 View of attack

4.2.5.3.2 Evasion Attacks

Evasion attacks occur during inference. An adversary introduces small perturbations into input signals or traffic patterns to make the AI model misclassify malicious behaviour as legitimate. In O-RAN, this may include crafting waveforms that mimic legal user signals or altering radio features to fool anomaly detection systems.

4.2.5.3.3 Model Theft and Privacy Attacks

With AI models increasingly exposed through APIs, attackers may attempt to:

- extract the model (model extraction),
- reconstruct training data (model inversion),
- verify whether specific samples were used during training (membership inference).

These attacks threaten confidentiality, intellectual property, and user privacy.

4.2.5.4.1 Securing Training and Data Pipelines

Securing the training pipeline requires strict validation of data integrity and authentication of all training sources to prevent the introduction of corrupted or manipulated inputs. In federated learning scenarios, secure aggregation mechanisms

must be applied to protect local model updates from tampering. Continuous monitoring of gradients and model behaviour supports early detection of poisoning attempts or unauthorized interference in the learning process.

4.2.5.4.2 Defending Against Adversarial Inputs

Strengthening inference-time robustness involves techniques such as adversarial training and input sanitization, which reduce model sensitivity to manipulated inputs. Anomaly detection helps identify suspicious patterns designed to evade classification, while ensemble-based approaches add further resilience by mitigating the impact of single misclassifications.

4.2.5.4.3 Protecting Models and APIs

AI models exposed through APIs require strict access control and authentication to prevent unauthorized use. Privacy-preserving techniques, including differential privacy, can mitigate the risk of inversion and membership inference attacks. Monitoring and limiting API queries helps prevent model extraction attempts, and encrypting model parameters in transit and at rest ensures confidentiality even if system components are compromised.

4.2.5.4.4 Continuous Monitoring and Verification

Long-term protection of AI systems requires continuous oversight. Security-oriented xApps and rApps can assess model behaviour and integrity in real time, while behavioural and anomaly analysis across RAN nodes enables early detection of deviations that may indicate attacks or model degradation.

4.2.5.5 Summary

AI algorithms are central to O-RAN's automation and optimization capabilities, but their integration introduces new risks that must be systematically addressed. O-RAN's openness offers strong opportunities for advanced security monitoring, yet also exposes AI/ML models to poisoning, evasion, and model-theft attacks. Ensuring robust AI security, through hardened training pipelines, adversarial defences, model confidentiality protections, and continuous monitoring, is essential for maintaining resilient and trustworthy O-RAN deployments.
UNITY-6G will harden AI in O-RAN by securing training/FL pipelines, adding adversarial robust inference with XAI logging, protecting models/APIs, and anchoring model/reputation evidence in the Trust Layer (smart contracts + oracles). Integrated through SBMA/IDMO, measured LoT/LoTw and on-chain events drive zero-touch actions. Validation in the cyber-range will report attack-success reduction, detection TPR/FPR, time-to-revoke, LoT deltas, and on-chain overheads, delivering xApp/rApp security tooling, FL-reputation contracts, oracle adapters, and orchestration playbooks.

4.3 SECURITY VALIDATION FRAMEWORK FOR UNITY-6G

Due to the inherent complexity and heterogeneity of cellular networks, ensuring robust cross-layer security is essential for maintaining a trustworthy and resilient communication environment. The trust-domain architecture proposed within UNITY-6G introduces new paradigms for safeguarding network infrastructure. Consequently, it requires innovative validation approaches capable of verifying not only whether the architecture is deployed correctly, but also whether it behaves securely in practice.
To meet this challenge, the UNITY-6G security validation strategy leverages a cross-layer instrumentation methodology grounded in network visibility. This approach enables comprehensive monitoring across multiple protocol layers, collecting and correlating data from various network elements and operational domains. By linking events and behaviours observed at different layers, the system becomes capable of detecting abnormal patterns, inconsistencies, or anomalies that may serve as indicators of emerging cybersecurity threats.
To systematically identify risks within the proposed architecture, a refined threat model is proposed to be developed using customizable parametrizations that adapt both to the horizontal deployment framework and to specific vertical use cases. This iterative modelling process will inform the selection of concrete security requirements, ensuring that they are context-aware and aligned with UNITY-6G's operational objectives. Building upon these foundations, the validation framework will assess the effectiveness of the implemented security controls, verifying that they provide adequate protection for the system and uphold the principles of the envisioned trust-domain architecture.

4.3.1 Foundations and Agreements for Validation on UNITY 6G

The validation framework for a Zero Trust 6G architecture leveraging Distributed Ledger Technologies (DLT) and built upon the principles of 3GPP TS 28.533 [65] must address both functional security enforcement and systemic behavioural assurance. Unlike traditional perimeter-based security paradigms, where trust is implicitly granted to internal components, a Zero Trust architecture assumes a hostile environment where all users, devices, network functions, slices, and services must continually prove their legitimacy. This foundational philosophy alone demands a validation methodology that is dynamic, context-aware, and able to detect deviations not at a single layer of the cellular stack, but across all operational layers simultaneously. [63] provides a robust foundation for assurance, focusing on lifecycle validation, closed-loop automation, data-driven decisioning, and observability. However, Zero Trust introduces an added dimension: validation must extend beyond state-of-the-art monitoring toward the active evaluation of identity correctness, policy enforcement integrity, and cryptographically verifiable transaction evidence through blockchain technology. In other words, the framework must continually ensure that trust is neither assumed nor static, it must be earned, validated, and auditable.
Within this context, [63] introduces assurance mechanisms that form the operational skeleton for validation. The standard emphasizes that assurance cannot be a passive, retrospective process; it must operate in continuous loops where monitoring data, analytics feedback, and automated policies coexist. The closed-loop automation models described in [63] inherently support Zero Trust by enabling a cycle of detection, analysis, and enforcement without reliance on long-lived assumptions of good behaviour. In practice, validation aims to verify that these loops respond appropriately

to real-world conditions such as fluctuating device integrity, changing application contexts, or deviations in traffic patterns and in network elements. A critical aspect of compliance with [63] lies in the degree of observability embedded in the architecture. To validate a Zero Trust 6G deployment, observability must extend beyond key performance indicators or fault counters into the semantic content of identity transactions, attestation claims, device health proofs, and execution context metadata. This results in a deeper operational footprint, extending assurance into behavioural analytics, identity confidence scoring, movement-based risk evaluation, attack trajectory mapping, and anomaly correlation across layers.

The Introduction of a blockchain or "DLT layer" fundamentally reshapes validation expectations because it decentralizes trust, enforces immutability, and distributes accountability. Traditional validation frameworks often analyse trust as a function of system configuration, for example, whether encryption protocols are applied, or whether authentication policies are consistent. In contrast, DLT validation requires demonstration that the network cannot violate trust even if multiple parties behave maliciously. Thus, the validation framework must examine whether the ledger consensus model remains stable under realistic 6G conditions and use cases, such as fluctuating edge connectivity, partial network partitioning, or large-scale device mobility. The validation must prove that ledger synchronization remains resilient, that blocks cannot be rewritten, and that the distributed membership model does not create performance bottlenecks during identity verification or microtransaction authorization. Most importantly, the validation must demonstrate that the immutable record of access decisions, attestation events, trust revocations, and cross-domain agreements provides cryptographic evidence that persists even if components are compromised. Cryptographic auditability replaces implicit trust in operational logs, and therefore the validation must demonstrate that blockchain anchoring is not merely passive logging, but a living enforcement mechanism against policy violation. Moreover, it is well known that blockchain-based mechanisms are weak during the initial phase of usage, making this phase critical in terms of security.

The framework must also validate that DLT-based policies do not become a static global lock. A common concern in blockchain-mediated architectures is the risk of policy rigidity, the idea that once a rule is committed to the ledger, it cannot be revised in time to address evolving threats. A properly designed validation methodology must therefore examine the flexibility of the smart-contract layer or equivalent DLT logic. It must confirm that policies can be dynamically updated, revoked, or overridden without compromising ledger integrity or introducing exploit pathways. For instance, the validation must ensure that time-bound credentials expire reliably, that SCs include multi-party revocation signals, and that role-based access indicators are context-sensitive rather than permanently asserted. This is especially significant in multi-domain federated deployments where multiple operators, vertical providers, or tenant entities maintain independent trust relationships. In such cases, validation must demonstrate that local enforcement decisions remain sovereign yet still produce global cryptographic evidence that other domains can rely upon.

A comprehensive validation framework must incorporate threat modelling not as a one-time document but as a continuously evolving analytical layer. The threat model must include conventional cyberattacks, insider misuse scenarios, orchestration poisoning, invalid ledger forks, manipulation of SC logic, and adversarial ML attacks against risk scoring modules. The validation effort must measure how the system detects deviations that do not manifest as raw network patterns but instead as subtle

manipulations of trust flows or metadata. An attacker may not flood a network with traffic; they may instead conduct privilege escalation by injecting misleading context into federated models or by exploiting timing gaps in ledger confirmation. The validation framework must therefore include adversarial simulation environments, stress tests against ledger participation rates, forced partitioning events, and evaluations where compromised identities are allowed to traverse the network until they are detected and cryptographically revoked. This holistic approach transforms validation from a compliance exercise into an empirical demonstration that the Zero Trust paradigm, reinforced by distributed ledger evidence and governed by [65] operational principles, can withstand real-world adversarial conditions.

4.3.2 Validation Framework for UNITY-6G

The validation framework for the UNITY-6G trust architecture (Figure 33) must consider the interaction between the Service Management and Orchestration layer (as defined in [65]) and the trust-enabling technologies that operate above and across it. [63] already establishes a comprehensive foundation for lifecycle assurance, closed-loop operational models, and semantic observability, but it remains agnostic to intrinsic trust primitives. The proposed architecture fills this gap by introducing the Trust Layer and its supporting blocks: NearbyOne, Smart Contract Manager, Blockchain Oracle, Blockchain Adaptor, and associated credential and RPC managers, which collectively convert trust from a static policy artifact into a continuous and verifiable property of operational environments. Validation must therefore no longer check whether a system is configured securely at deployment. Instead, the framework must demonstrate that security and trust mechanisms are enacted, enforced, recorded, and auditable over the entire execution lifecycle of services, devices, and actors.

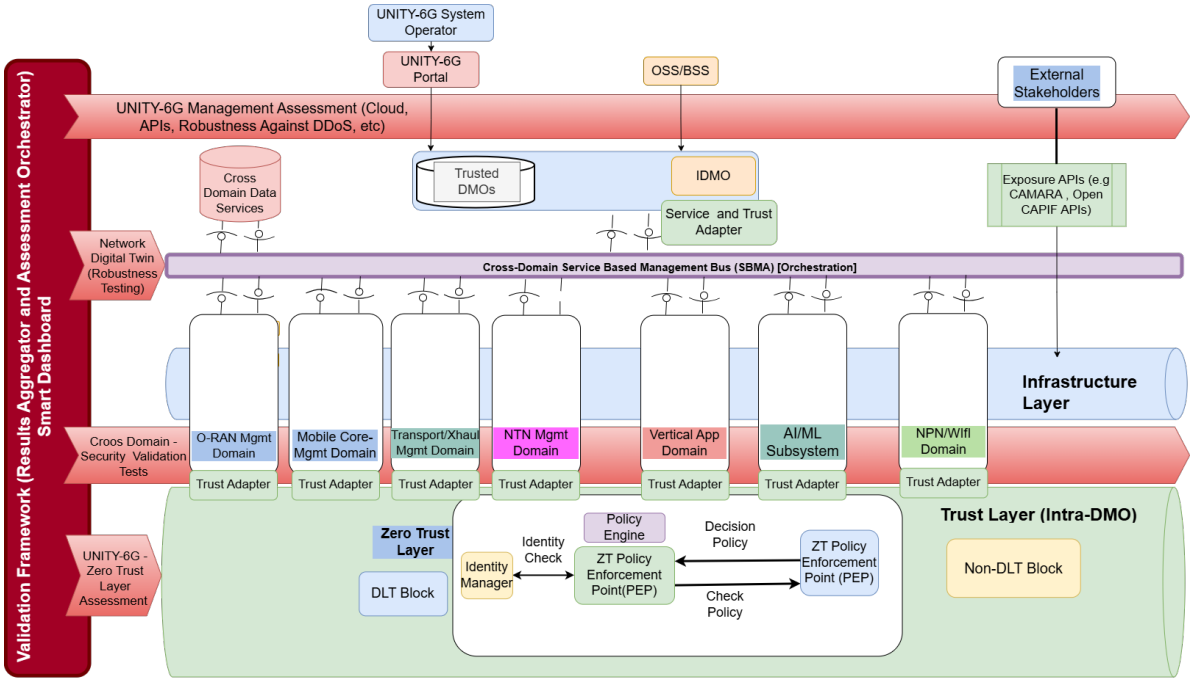

Figure 33 ***Validation framework on top of the UNITY-6G Architecture***

4.3.3 Validation of the orchestration entities

The validation process begins with the NearbyOne Service Orchestration entity. NearbyOne is a cloud-native orchestration layer that forms the operational entry point into UNITY-6G. All other blocks, including the Trust Layer, exist downstream from its control processes. Because NearbyOne is a multi-domain and multi-tenancy orchestrator that exposes a unified interface to users, tenants, or vertical owners, validation must first ensure that its authentication and authorization controls are airtight and correctly mediated. Its Role-Based Access Control (RBAC) model, combined with LDAP/OpenID Connect (OICD) authentication and Oauth2 authorization, cannot be treated as mere User Interface (UI) enforcement. Instead, it must be validated as a security boundary that prevents unauthorized tenants from influencing other organizations' resource operations or altering trust-critical configurations.
Roles, groups, and organizations must be validated not only at the application layer but also against their effects on orchestration actions, service instantiation, resource scaling, and cross-domain traffic steering. For example, a threat model must examine whether a privileged user could deploy a malicious containerized service, force an unauthorized SC invocation, or attempt to manipulate a blockchain credential via automation workflows. Proper validation demands evidence that all actions are traceable through the orchestration workflow, cryptographically bound to the user identity, and contextualized so that downstream trust blocks can make inference decisions when abnormal actions occur.

4.3.4 Validation of the Trust Layer

Once orchestration flows propagate into the Unity-6G architecture, the Trust Layer becomes the enforcement backbone. Validation of this layer requires a dual perspective: functional correctness and adversarial resilience. IOTA's Tangle-based DLT provides immutable event anchoring, decentralized identity management, and support for verifiable credentials. Validation activities must demonstrate that devices, users, agents, or services become participants only when authenticated using cryptographic credentials anchored on-chain or via verifiable DIDs. Likewise, authorization must not rely on static permission rules; rather, the system must evaluate the operational context of each interaction. For instance, even when a user or device has valid credentials, validation must show that the Trust Layer denies operations that violate contextual constraints (e.g., access to a sensitive function at a time window, cross-domain evidence request, or abnormal resource allocation pattern). The validation framework therefore tests not only the correctness of IOTA credential issuance, but also the consistency of revocation propagation, the recovery behaviour when ledger access temporarily degrades, the ability to re-establish trust after isolation, and the resilience of authorization under partial resource outages. A trusted system is not simply one where identities are correct; it is one in which identities cannot undermine the system even when compromised
SCs represent the formalization of trust rules and governance. In contrast to classical rule engines or policy databases, SCs embed execution logic in a tamper-resistant environment. Validation must establish that contract logic faithfully represents intended trust semantics and is not exposed to unexpected execution paths, race conditions, or unbounded state behaviours. The SCM block is therefore not merely an interface to the EVM; it is the governing element that ensures every on-chain invocation is justified, properly formatted, and cryptographically attributable. A robust validation framework

must simulate both valid and adversarial contract invocation patterns. It should confirm that contract roles cannot be escalated, that contract storage cannot be manipulated via re-entrancy, and that malformed inputs, intentional or accidental, are not silently dropped but instead emitted as auditable events. Because SCs execute deterministically across all validator nodes, validation must prove that multi-tenant environments do not cause information leakage through contract state or execution time patterns. The goal of validation is not that contracts "work," but that contracts enforce the Zero Trust principle: no actor, however legitimate, is ever beyond cryptographic scrutiny.

The Blockchain Oracle introduces the most delicate aspect of DLT validation: the interaction between the deterministic nature of blockchains and the chaotic nature of real-world systems. SCs cannot directly access hardware telemetry, ML analytics, or external KPIs; therefore, validation must establish that the Oracle's external inputs cannot introduce new attack vectors. An Oracle must demonstrate its ability to authenticate its own data sources, validate their correctness, and prevent single-source influence. Validation activities must simulate erroneous, stale, or manipulated off-chain inputs and confirm that these conditions trigger rejection rather than silent acceptance. The validation framework must also verify that trust decisions made based on Oracle inputs, such as revoking an identity, freezing a node's interactions, or activating compensating controls, are recorded immutably and propagated through the SC layer. Any mismatch between Oracle evidence and blockchain anchoring constitutes a systemic trust failure.

The Blockchain Adaptor block completes the trust-to-execution pipeline by translating off-chain events into on-chain transactions. From a validation standpoint, this is where semantic correctness is at the highest risk of failure. Data from the Oracle must be validated syntactically and semantically before being mapped to smart-contract structures; otherwise, the authenticity of an external measurement could corrupt smart-contract logic despite being internally consistent. The validation framework must establish that the Adaptor enforces input constraints (format, timestamp, trust score, provenance) and that edge cases such as partial data delivery, packet reordering, delayed telemetry, or domain ambiguity do not lead to silent ledger writes. For example, validation must ensure that a temperature reading from a manufacturing IoT device cannot be mistakenly interpreted as a traffic metric from a telecom service, or conversely, that economic values cannot be written into the ledger in a context designed purely for technical policy enforcement. In distributed architecture, integrity is not just cryptographic; it is representational.

4.3.5 System-level validation

At the system level, the validation framework must simulate failure modes across interfaces. Communication interfaces between the orchestration system, DLT components, and the Service Bus must be proven robust to identity spoofing, replay attacks, unauthorized data injections, and trust downgrade attempts. Since [64] emphasizes cross-layer observability, validation must ensure that every trust decision produces correlatable evidence: audit trails in NearbyOne, identity event logs in the Trust Layer, and immutable ledger entries in the SC Layer. This traceability must be demonstrable without requiring privileged access or post-hoc forensic reconstruction. The validation process must, in effect, prove that the architecture behaves like a distributed truth machine.

As is well known, a Zero Trust validation framework must demonstrate operational resilience. A malicious tenant might attempt to exploit automation workflows, inject misleading metrics through the Oracle, or leverage privilege boundaries to compromise other service domains. Validation must show that no such action can propagate unchecked. Even if NearbyOne correctly approves user actions, the Trust Layer should still reject them if context is insufficient or if ledger evidence contradicts the claimed privileges. Even if the Trust Layer approves a subject, the SCM should block action if identity has not been re-verified or if previous behaviour suggests risk. Even if on-chain logic validates an input, the Oracle and Adaptor should reject events that violate external trust conditions. Only when all layers continuously validate each other through cryptographic evidence (not assumptions) can the architecture be said to embody end-to-end Zero Trust.

5 ROAD AHEAD FOR TRUST MODELING

A prospective 6G architecture is envisioned in which DLT, AI/FL, and IoT services are designed together rather than bolted on afterward. At the top, a Function Layer bundles the main network capabilities: unified data management, policy control, AI functions (including FL-based analytics), sensing, computing, mobility management, and communication management. Many of these functions are conceived as DLT Trust Functions (DLT-TF), meaning that standard 6G functions can be "DLT-aware," exchanging model-update hashes and trust evidence with the lower layers. IoT vertical applications (e.g., smart city, Industry 4.0, healthcare) sit above this layer and consume these capabilities through the usual control, user, and data planes.

Beneath this lies a Trust Layer that hosts the concrete DLT infrastructure. It groups permissioned DLT nodes together with a DLT-dApp-Manager and an off-chain repository. This layer is responsible for validating transactions, maintaining tamper-evident logs of model updates and trust scores, and exposing trustworthy information back to the Function Layer through well-defined interfaces. Multiple DLT "islands" can coexist to serve different verticals or administrative domains, but all follow the same architectural pattern. As part of UNITY-6G, we will actively develop and mature these 6G trust mechanisms (DLT-TF, interfaces, and deployment options) through implementation and evaluation across representative verticals.

At the bottom, a PDL Layer represents the underlying permissioned distributed ledger fabric that ties everything together, integrating with control, user, and data planes. Overall, this architectural view conveys a 6G vision where communications, sensing, and computing are tightly coupled with a native trust substrate based on DLT. While this aligns with emerging views on trustworthy 6G networks, important gaps remain around scalability, interoperability, and common standards for DLT-TF components, especially for energy-constrained IoT and FL-driven services.

In this document, private and permissioned DLT solutions have been presented sometimes. However, the next steps of the project will explore the feasibility of using a public and permissionless DLT environment as well to have a secure and transparent trust layer for real deployments.

6 CONCLUSIONS

This deliverable has provided an extensive analysis of trust as a fundamental enabler for secure, reliable, and resilient 6G networks. Starting from the motivation for trust in highly dynamic and heterogeneous environments, the document reviewed the SotA in trust-related research, relevant EU projects, and ongoing standardization efforts, highlighting the evolution from traditional security models to comprehensive trust frameworks that integrate Zero-Trust principles, AI-driven mechanisms, and distributed ledger technologies.

Building on this foundation, the deliverable introduced conceptual and functional architectures for the UNITY-6G trust domain. These architectures outline design principles and functional blocks, including trust adapters, orchestration components, and DLT-based elements like SCs and blockchain oracles. Rather than prescribing a single definitive solution, the document presented multiple architectural options and design guidelines that will inform subsequent development phases.

The deliverable also proposed a general trust model, emphasizing user-centric and AI-assisted trustworthiness, and introduced the concept of LoT as a measurable indicator for dynamic trust evaluation. Scenario-specific customizations were discussed for IoT, O-RAN, and Wi-Fi environments, demonstrating how DLT-enabled federated learning and reputation mechanisms can enhance trust in distributed AI workflows and multi-vendor ecosystems.

This document consolidates the activities and preliminary results aimed at defining trust models and management approaches for 6G. By providing architectural options, trust modelling strategies, and scenario-driven adaptations, the deliverable lays the groundwork for future implementation and validation efforts within UNITY-6G. These contributions will serve as a reference for designing a unified trust layer capable of supporting secure, explainable, and adaptive trust management across terrestrial and non-terrestrial networks, defining the way for trustworthy 6G ecosystems.

REFERENCES

[1] Groen, Joshua, et al. "Implementing and evaluating security in O-RAN: Interfaces, intelligence, and platforms." *IEEE Network* (2024)

[2] S. Poretsky, "Open RAN is secure and ready for deployment," *Ericsson Blog*, Aug. 11, 2025. [Online]. Available: https://www.ericsson.com/en/blog/north-america/2025/open-ran-is-secure-and-ready-for-deployment

[3] U.S. Dept. of Commerce, National Telecommunications and Information Administration, *Open RAN Security Report*, May 2023. [Online]. Available: https://www.ntia.gov/sites/default/files/publications/open_ran_security_report_full_report_0.pdf

[4] D. Je, S. Kim, and D. Kim, "Zero Trust Architecture for 6G," Samsung Research Blog, Sep. 8, 2025. [Online]. Available: https://research.samsung.com/blog/Zero-Trust-Architecture-for-6G

[5] Fujitsu, *"A Brief Look at O-RAN Security"*, White Paper, 2022. [Online]. Available: https://www.fujitsu.com/global/documents/products/network/Whitepaper-A-Brief-Look-at-O-RAN-Security.pdf

[6] Brik, Bouziane, et al. "Explainable ai in 6g o-ran: A tutorial and survey on architecture, use cases, challenges, and future research." *IEEE Communications Surveys & Tutorials* (2024)

[7] Ma, Zhe, et al. "Blockchain-based zero-trust supply chain security integrated with deep reinforcement learning for inventory optimization." *Future Internet* 16.5 (2024): 163

[8] Garzon, Sandro Rodriguez, et al. "Beyond certificates: 6G-ready access control for the service-based architecture with decentralized identifiers and verifiable credentials." *2024 Joint European Conference on Networks and Communications & 6G Summit (EuCNC/6G Summit)*. IEEE, 2024

[9] El-Hajj, Mohammed. "Secure and Trustworthy Open Radio Access Network (O-RAN) Optimization: A Zero-Trust and Federated Learning Framework for 6G Networks." *Future Internet* 17.6 (2025): 233

[10] ITU-T Recommendation X.1400, "Terms and definitions for distributed ledger technology," Oct. 29, 2020. [Online]. Available: https://handle.itu.int/11.1002/1000/14449

[11] ETSI Industry Specification Group on Permissioned Distributed Ledger (ISG PDL), "Permissioned Distributed Ledgers (PDL)," ETSI. [Online]. Available: https://www.etsi.org/technologies/permissioned-distributed-ledgers

[12] Dimou, Stavros, Kostas Choumas, and Thanasis Korakis. "On using Blockchain in beyond 5G: Roaming Improvements." *2024 6th Conference on Blockchain Research & Applications for Innovative Networks and Services (BRAINS)*. IEEE, 2024

[13] iTrust6G - Intelligent Trust and Security Orchestration for 6G Distributed Cloud Environments

[14] Katsis, Charalampos, Imtiaz Karim, and Elisa Bertino. "Zero-Trust Strategies for O-RAN Cellular Networks: Principles, Challenges and Research Directions." *arXiv preprint arXiv:2511.18568* (2025)

[15] Gambo, Muhammad Liman, and Ahmad Almulhem. "Zero Trust Architecture: A systematic literature review." *Journal of Network and Systems Management* 34.1 (2026): 25

[16] Javed, Farhana, et al. "Trustworthy Reputation for Federated Learning in O-RAN Using Blockchain and Smart Contracts." IEEE Open Journal of the Communications Society 6 (2025): 1343-1362.

[17] Javed, Farhana, et al. "Blockchain for Federated Learning in the Internet of Things: Trustworthy Adaptation, Standards, and the Road Ahead." IEEE Communications Standards Magazine (2025).

[18] Zeydan, Engin, et al. "Transforming Open Radio Access Networks with Multi-Layered Architecture: A Decentralized Approach." *Distributed Ledger Technologies: Research and Practice* (2025).

[19] Chiejina, Azuka, et al. "System-level analysis of adversarial attacks and defenses on intelligence in O-RAN based cellular networks." *Proceedings of the 17th ACM Conference on Security and Privacy in Wireless and Mobile Networks*. 2024.

[20] M. Wasilewska, H. Bogucka and H. V. Poor, "Secure Federated Learning for Cognitive Radio Sensing," in IEEE Communications Magazine, vol. 61, no. 3, pp. 68-73, March 2023, doi: 10.1109/MCOM.001.2200465. keywords: {Federated learning;Cognitive radio;Spread spectrum management},

[21] Advanced Methods and Techniques for Detecting and Mitigating Attacks on 5G Access Infrastructure and Applications - 5gSTAR report B

[22] iTrust6G, Intelligent Trust and Security Orchestration for 6G Distributed Cloud Environments, Online: https://www.sns-itrust6g.com/, Available: November 2025D

[23] Robust6G, SMART, AUTOMATED AND RELIABLE SECURITY SERVICE PLATFORM FOR 6G, Online; https://robust-6g.eu/, Available: November 2025

[24] NATWORK, Net-Zero self-adaptive activation of distributed self-resilient augmented services, Online: https://natwork-project.eu/, Available: November 2025

[25] RIGOROUS, secuRe desIGn and deplOyment of trUsthwoRthy cOntinUum computing 6G Services, Online: https://rigourous.eu/, Available: November 2025

[26] MARE Project, Online: https://mare6g.eu/project-overview/, Available: November 2025

[27] 6G-IA Security WG, *Innovative Approaches for 6G Security*, Nov. 2025. [Online]. Available: https://6g-ia.eu/wp-content/uploads/2025/11/6g-ia_security-wg_white-paper_nov25_final.pdf

[28] RIGOROUS Consortium “D2.3: Final Rigorous Reference Architcture.” Zenodo, DOI: 10.5281/zenodo.15903015. [Online]. Available: https://zenodo.org/records/15903015

[29] SAFE-6G Project, “A Smart and Adaptive Framework for Enhancing Trust in 6G Networks,” SAFE-6G, 2024. [Online]. Available: https://safe-6g.eu. [Accessed: Jan-2026].

[30] ELASTIC Project, “Efficient, portabLe And Secure orchesTration for reliable servICes,” ELASTIC, 2024. [Online]. Available: https://elasticproject.eu/, [Accessed: Jan-2026].

[31] HORSE Project, “Holistic, omnipresent, resilient services for future 6G wireless and computing ecosystems,” HORSE Project, 2026. [Online]. Available: https://horse-6g.eu/ , [Accessed: Jan-2026].

[32] 6G-Cloud Project, “Service-oriented 6G Network Architecture for Distributed, Intelligent, and Sustainable Cloud-native Communication Systems,” 6G-Cloud, 2026. [Online]. Available: https://www.6g-cloud.eu/, [Accessed: Jan-2026].

[33] SUSTAIN-6G Project, “SUSTainability Advanced and Innovative Networking with 6G,” SUSTAIN-6G. [Online]. Available: https://sustain-6g.eu/, [Accessed: Jan-2026].

[34] PRIVATEER Project, “Privacy-first Security Enablers for 6G Networks,” PRIVATEER. [Online]. Available: https://www.privateer-project.eu/, [Accessed: Jan-2026].

[35] Nearby Computing, “NearbyOne: Orchestration platform for Telcos & Enterprises,” Nearby Computing. [Online]. Available: https://www.nearbycomputing.com/nearbyone/, [Accessed: Jan-2026].

[36] Zero Trust Maturity Model (ZTMM), version 2.0, US DHS CISA, April 2023.

[37] Zero Trust Architecture (ZTA), NIST SP 800-207, US DoC NIST, September 2020

[38] The O-RAN ALLIANCE Security Work Group Continues Defining O-RAN Security Solutions, https://www.o-ran.org/blog/the-o-ran-alliance-security-work-group-continues-defining-o-ran-security-solutions

[39] O-RAN ALLIANCE WG11, Zero Trust Architecture for Secure O-RAN, v1.0, White Paper, May 2024. [Online]. Available: https://mediastorage.o-ran.org/white-papers/O-RAN.WG11.ZTA%20for%20Secure%20O-RAN%20White%20Paper-2024-05.pdf

[40] ETSI TR 104 106 V3.0.0, "Publicly Available Specification (PAS); O-RAN Security Threat Modelling and Risk Assessment (O-RAN.WG11.Threat-Modeling.O-R003-v03.00)", European Telecommunications Standards Institute, June 2025. [Online]. Available: https://www.etsi.org/deliver/etsi_tr/104100_104199/104106/03.00.00_60/tr_104106v030000p.pdf

[41] O-RAN ALLIANCE WG11, “The O-RAN ALLIANCE Security Working Group Continues to Advance O-RAN Security,” O-RAN ALLIANCE e.V., White Paper, Feb. 9, 2024. [Online]. Available : https://mediastorage.o-ran.org/announcement/O-RAN.WG11.O-RAN%20ALLIANCE%20Continues%20to%20Advance%20O-RAN%20Security.pdf

[42] ETSI Permissioned Distributed Ledger (PDL); Trust in Telecom System, https://portal.etsi.org/webapp/WorkProgram/Report_WorkItem.asp?WKI_ID=70024, [Accessed: Jan-2026].

[43] ETSI, Permissioned Distributed Ledgers (PDL); Smart Contracts; System Architecture and Functional Specification, ETSI GR PDL 004 V1.1.1, Feb. 2021. [Online]. Available: https://www.etsi.org/deliver/etsi_gr/PDL/001_099/004/01.01.01_60/gr_pdl004v010101p.pdf, [Accessed: Jan-2026].

[44] ETSI, *Permissioned Distributed Ledger (PDL); Trust in Telecom System*, ETSI GR PDL 030 V1.1.1, May 2025. [Online]. Available: https://www.etsi.org/deliver/etsi_gr/PDL/001_099/030/01.01.01_60/gr_pdl030v010101p.pdf, [Accessed: Jan-2026].

[45] ITU-T Rec. Y.3053, “Framework of trustworthy networking with trust-centric network domains,” Jan. 2018. [Online]. Available: https://www.itu.int/rec/dologin_pub.asp?lang=e&id=T-REC-Y.3053-201801-I!!PDF-E&type=items

[46] ITU-T Recommendation Y.3051, “The basic principles of trusted environment in information and communication technology infrastructure,” Mar. 2017. [Online]. Available: https://www.itu.int/rec/dologin_pub.asp?lang=e&id=T-REC-Y.3051-201703-I!!PDF-E&type=items

[47] ITU-T Recommendation Y.3052 (03/2017), “Overview of trust provisioning in information and communication technology infrastructures and services,” Mar. 2017. [Online]. Available: https://www.itu.int/rec/dologin_pub.asp?lang=e&id=T-REC-Y.3052-201703-I!!PDF-E&type=items

[48] ITU-T Focus Group on Application of Distributed Ledger Technology, *Technical*

Specification FG DLT D3.1: Distributed Ledger Technology Reference Architecture, Aug. 2019. [Online]. Available: https://www.itu.int/en/ITU-T/focusgroups/dlt/Documents/d31.pdf

[49] OAuth-Working-Group, "OAuth 2.0," oauth.net, 2025. [Online]. Available: https://oauth.net/2/. [Accessed: Jan-2026]

[50] I. C. Secretary, «Systems and software engineering - Systems and software quality requirements and evaluation (SQuaRE) - Measurement of quality in use,» International Organization for Standardization, vol. Standard ISO/IEC 25022, 2016.

[51] X. Yan, X. An, W. Ye, M. Zhao, Y. Xi and J. Wu, "User-Centric Network Architecture Design for 6G Mobile Communication Systems," 2023 Joint European Conference on Networks and Communications & 6G Summit (EuCNC/6G Summit), Gothenburg, Sweden, 2023, pp. 305-310, doi: 10.1109/EuCNC/6GSummit58263.2023.10188283

[52] Innovative Approaches for 6G Security: Challenges, Solutions, and Impact, GA-IA WG Security, v0.1, 2024, 6g-ia.eu/wp-content/uploads/2025/01/wg_sec_position_paper-23.pdf

[53] O-RAN Working Group 2, AI/ML Workflow Description and Requirements (O-RAN.WG2.AIML-v01.02), https://www.o-ran.org/resources

[54] An Introduction of Permissioned Distributed Ledger (PDL) (ETSI White Paper No. 48), https://www.etsi.org/images/files/ETSIWhitePapers/ETSI-WP48-PDL.pdf

[55] 6G Trustworthiness Considerations, https://www.ngmn.org/publications/6g-trustworthiness-considerations.html

[56] J. Lee, F. Solat, T. Y. Kim, H. V. Poor, *Federated Learning-Empowered Mobile Network Management for 5G and Beyond Networks: From Access to Core*, IEEE Communications Surveys & Tutorials, 2024, https://doi.org/10.1109/COMST.2024.3352910

[57] ETSI TS 103 982 V8.0.0 (2024-01), "O-RAN Architecture Description (O-RAN.WG1.OAD-R003-v08.00)", https://www.etsi.org/deliver/etsi_ts/103900_103999/103982/08.00.00_60/ts_103982v080000p.pdf

[58] O-RAN ALLIANCE, "The O-RAN ALLIANCE Security Work Group Continues Defining O-RAN Security Solutions", https://www.o-ran.org/blog/the-o-ran-alliance-security-work-group-continues-defining-o-ran-security-solutions

[59] O-RAN Working Group 2, "AI/ML Workflow Description and Requirements (O-RAN.WG2.AIML)", https://www.o-ran.org/specification-access

[60] Xavier Titi, Carlos Ballester Lafuente, Jean-Marc Seigneur "Trust Management for Selecting Trustworthy Access Points", IJCSI International Journal of Computer Science Issues, Vol. 8, Issue 2, March 2011

[61] Seigneur, Jean-Marc. "Wi-trust: Improving Wi-Fi hotspots trustworthiness with computational trust management." *2015 ITU kaleidoscope: Trust in the information society (K-2015)*. IEEE, 2015.

[62] Khan, Tayyab, et al. "A novel and comprehensive trust estimation clustering based approach for large scale wireless sensor networks." IEEE Access 7 (2019): 58221-58240

[63] Karoliny, Julian, et al. "Network support layers trustworthiness computation for wireless networks." IEEE Transactions on Communications (2024)

[64] Alexandropoulos, I.; Koumaras, H.; Rentoula, V.; Papanikolaou-Ntais, G.; Georgoulas, S.; Makropoulos, G. Adversarial Robustness in Cognitive Systems: A Trustworthiness Assessment Perspective for 6G Networks. *Electronics* 2025, *14*, 2285.

https://doi.org/10.3390/electronics14112285

[65] 3GPP TS 28.533, Management and orchestration; Architecture framework, v19.3.0, 2025

[66] H. Bogucka, M. Hoffmann, P. Kryszkiewicz and Ł. Kułacz, "An Open-RAN Testbed for Detecting and Mitigating Radio-Access Anomalies," in IEEE Communications Magazine, vol. 63, no. 11, pp. 122-127, November 2025, doi: 10.1109/MCOM.003.2400513.

[67] IOTA Foundation, “IOTA,” IOTA Foundation. [Online]. Available: https://www.iota.org/, [Accessed: Jan-2026]

[68] O-RAN Software Community, O-RAN Architecture Overview — latest release, O-RAN SC. [Online]. Available: https://docs.o-ran-sc.org/en/latest/architecture/architecture.html, [Accessed: Jan-2026]