DOI. No. 10.13052/2794-7254.031

Towards a Unified 6G Network Architecture: Principles, Layers, and Key Enabling Technologies

Sudhir Dixit

Abstract: Sixth-generation (6G) wireless networks – formally designated IMT-2030 by the ITU – will require a fundamental rearchitecting of how wireless systems are conceived, standardised, and operated. Motivated by application classes that 5G cannot serve, this paper proposes a four-layer 6G network architecture comprising an Intelligent RAN (I-RAN), an AI-Native Core (ANC), a Semantic Communication Plane (SCP), and a Non-Terrestrial Network (NTN) Stratum, unified by a cross-cutting AI/ML orchestration plane and a zero-trust security framework. The paper draws on the Hexa-X use case taxonomy (HEX-D13) [1], the eight architecture principles of HEX-D51 [2], and the 5G PPP functional reference model [3].

Keywords: 6G network architecture, Hexa-X, IMT-2030, I-RAN, AI-Native Core, NTN, Semantic communications..

1. Introduction

The evolution from 5G to 6G is not an incremental upgrade – it will require a generational rearchitecting. The ITU IMT-2030 framework (Recommendation M.2160, 2023) [4] sets a capability envelope qualitatively and quantitatively different from any previous generation: peak data rates exceeding 1 Tbps, air interface latency below 100 μs, a hundredfold improvement in energy efficiency per bit, and native AI integration across every network layer. Five structural gaps in 5G motivate this rearchitecting: no native AI, insufficient programmability, incomplete cloud-nativity, no continuum orchestration, and sustainability treated as an afterthought rather than a foundational constraint.

This paper addresses each gap through a coherent four-layer architecture illustrated throughout with figures. Sections 2–5 establish the motivating context: use cases, why 6G is needed, structural gaps, and design principles. Sections 6–7 present the three reference architecture views from Hexa-X and the 5G PPP Working Group. Section 8 proposes the four-layer architecture with deep dives into each layer. Section 9 compares it with 5G. Section 10 analyses four open research challenges. Section 11 concludes.

Basic Internet Foundation, Oslo, Norway & University of Oulu, Oulu, Finland
E-mail: sudhir.dixit@ieee.org
Manuscript received 25 June 2026, accepted 07 July 2026, and ready for publication 22 July 2026.

2. 6G Use Cases and Application Drivers

Figure 1 reproduces the Hexa-X HEX-D13 use case taxonomy [1], which organises 6G applications into six families around the convergence of the physical, digital, and human worlds. These families provide the inputs to the requirements that ITU has defined for the 6G architecture that must be satisfied.

Robots to Cobots

Consumer robots, AI partners, interacting and cooperative mobile robots, flexible manufacturing, and situation-aware device reconfiguration. A cobot (collaborative robot) is a robotic system designed to operate safely in close proximity to human workers, allowing them to share the same workspace. The defining requirements are ultra-reliable low-latency communication (URLLC) at density levels beyond 5G capability, deterministic latency guarantees, and distributed sensing to support real-time collaborative autonomy.

Telepresence

Fully merged cyber-physical worlds, mixed reality co-design, immersive sport events, and merged reality games and work environments. These applications require multi-Gbps bidirectional video streams with sub-millisecond synchronisation between endpoints – a throughput and latency combination that exceeds 5G's capability by more than an order of magnitude, and which requires semantic compression at the network layer to be bandwidth-feasible.

Massive Twinning

Digital twins for manufacturing, immersive smart cities, and the Internet of Tags (IoTa) designed for passive tracking using technologies of RFID and NFC. Digital twins are persistent, real-time computational mirrors of physical entities and environments. Sustaining them at scale – a city, a supply chain, a national infrastructure – requires continuous, low-latency telemetry from millions of sensors, edge computing resources co-located with physical assets, and a network fabric capable of supporting planetary-scale data flows with deterministic quality of service.

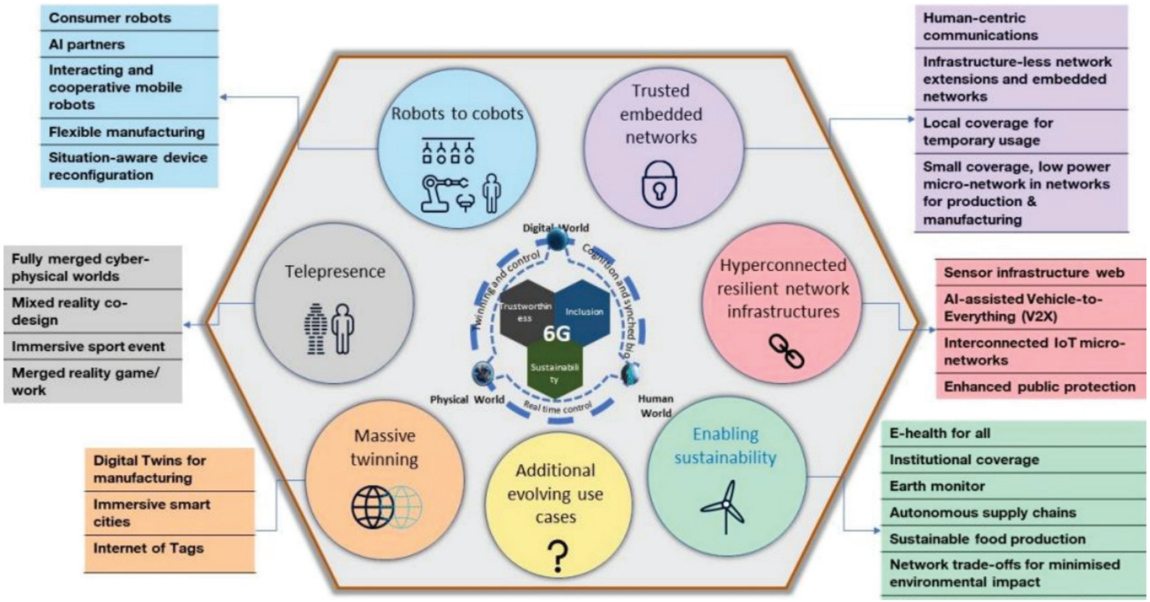

Figure 1.
Families of 6G use cases [1].

Hyperconnected Resilient Infrastructures

Sensor infrastructure webs, AI-assisted Vehicle-to-Everything (V2X), interconnected IoT micro-networks, and enhanced public protection systems [6]. These require the 6G network to act as a sensing and inference platform, not merely a connectivity conduit – integrating radar-like environmental sensing directly into the radio access network through integrated sensing and communications (ISAC).

Trusted Embedded Networks

Infrastructure-less network extensions, local coverage for temporary usage, and small, low-power micro-networks for production and manufacturing. These require 6G to function without fixed infrastructure – in disaster recovery, remote industrial, and tactical military contexts – using non-terrestrial networks (NTN) and device-to-device relay architectures as primary means to connect with the data network.

Enabling Sustainability

E-health for all, institutional coverage, earth monitoring, autonomous supply chains, sustainable food production, and network functionality for crisis resilience. This family collectively establishes that 6G must serve not just the richest markets and most profitable verticals, but must be affordable, energy-efficient, and globally deployable – making sustainability not an add-on requirement but a core architectural constraint.

3GPP, "TR 22.870" has also studied the use cases and requirements for IMT-2030 [5]. From a North American perspective, the ATIS Next G Alliance (NGA) has played a central role in articulating a coherent 6G vision for the region. Established in 2020 by the Alliance for Telecommunications Industry Solutions (ATIS), the NGA brings together more than 80 member organisations spanning operators, equipment vendors, chipmakers, government agencies, and academia to develop a coordinated roadmap for 6G leadership. The Alliance aims to establish North American preeminence in 5G evolution and 6G development. Its work focuses on creating a comprehensive national roadmap, influencing government policy, and driving rapid commercialization of beyond 5G, including 6G, technologies. More specifically, its foundational "Roadmap to 6G" [7], published in 2022 and updated since, identifies six audacious goals addressing North America's priorities in standardisation, R&D, spectrum policy, sustainability, and supply chain resilience. The NGA's National 6G Roadmap Working Group has subsequently published North American 6G Roadmap Priorities, identifying nine representative use cases – spanning immersive communications, connected intelligence, and resilient infrastructure – that directly inform the use case taxonomy discussed in Section 2 of this paper [7]. The NGA has also produced targeted technical reports on ISAC channel modelling in the 7–24 GHz mid-band range, spectrum needs assessments, and component technology white papers covering antennas, RF circuits, and transceivers. Taken together, this body of work establishes that the architectural shifts proposed in this paper – AI-native design, ISAC integration, Open RAN, and post-quantum security – are fully consistent with North American industry priorities and the NGA's roadmap for achieving global 6G leadership.

3. Why 6G? The Case for Rearchitecting

Early work on 6G began in the Finnish 6G Flagship project in year 2018 [8]. Around the same time, other interesting papers were published articulating the 6G vision and requirements [9–11],

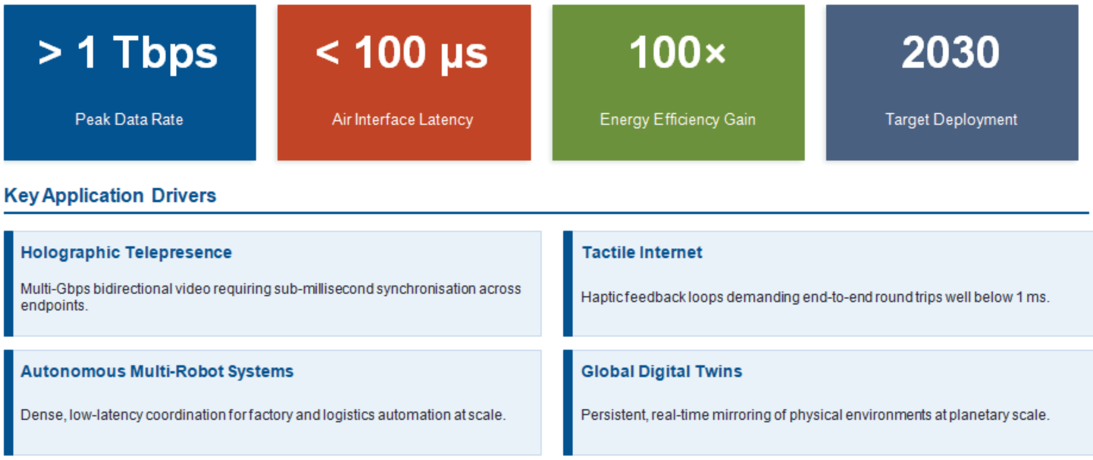

Figure 2.
Why 6G? Four IMT-2030 headline targets – >1 Tbps peak data rate, <100 μs air interface latency, 100× energy efficiency gain, 2030 commercial deployment – and four key application driver families: Holographic Telepresence, Tactile Internet, Autonomous Multi-Robot Systems, and Global Digital Twins. Adapted from [4].

including discussions on whether 6G is needed beyond 5G [9]. ITU continued the momentum on 6G, by now called IMT-2030, and summarized the four headline targets and the four principal application driver families as depicted in Figure 2.

4. Structural Gaps From 5G to 6G

There are five structural gaps in 5G that 6G must close – each representing a dimension in which 5G is fundamentally, not merely quantitatively, insufficient.

- **Enabling AI and Agentic AI** – 5G has no native AI. 6G must embed autonomous, goal-directed AI inference in every network function from the outset. This would enable more automation, self-learning, cost reduction, and increased efficiency.
- **Programmability** – 5G's Open RAN and network slicing do not provide closed-loop policy generation. 6G must be fully programmable from radio to application layer.
- **Cloud-native softwarisation and service-based architecture** – 5G softwarisation is partial, particularly in the RAN. 6G requires complete softwarisation to the radio edge to enable open source implementation, programmability, faster configuration/changeover, faster upgrades and operational efficiency.
- **Continuum Orchestration** – 5G cannot seamlessly orchestrate across device, edge, cloud, and NTN. 6G must treat these as a unified compute-communication fabric.
- **Cross-cutting concerns** – sustainability, trustworthiness, cybersecurity, and regulatory compliance must be first-principles architectural constraints in 6G, not post-deployment overlays.

5. Architectural Design Principles

The Hexa-X deliverable HEX-D51 establishes eight guiding principles for 6G architecture design [2]. These principles function as an evaluation lens: every architectural decision must be assessed against them, and any decision that violates more than one principle without compelling justification should be reconsidered. Table 1 summarises the eight principles.

Two principles deserve particular emphasis in the context of the proposed architecture. Principle 2 – AI for full automation – implies that the network intelligence is not an xApp or an external optimiser but an intrinsic property of every network function: session management, handover control, interference coordination, and security enforcement are all AI-driven from initial design, not retrofitted. Principle 8 – network simplification – is a counterweight to the complexity that the preceding principles tend to add: the architecture must be measurably simpler to operate than 5G, or it will remain a research artefact rather than a commercially deployable system.

6. High Level View of the 6G System Architecture

The Hexa-X high-level architecture view (HEX-D13) [1] organises the 6G system into three horizontal strata and two vertical cross-cutting dimensions, as illustrated conceptually in Figure 3. The three strata are:

- Infrastructure Layer – the physical and virtualised compute, transport, and radio infrastructure spanning extreme/far edge, edge cloud, and central cloud, together with non-terrestrial assets (LEO satellites, HAPS, UAVs).
- Network Service Layer – the set of network functions (control plane, user plane, management plane) exposed through an Exposure Framework and Integration Fabric. This stratum is explicitly multi-tier: functions may run at any edge or cloud location, and the Exposure Framework provides a unified northbound interface regardless of function placement.
- Application Layer – vertical service providers access the network through the Exposure Framework, running application instances that span the use case families of Section 2, from immersive smart cities to AI-assisted V2X to fully merged cyber-physical worlds.

Table 1.

6G architecture design principles [HEX-D51] [2]	
Principle	**Description**
1 – Exposure of capabilities	Well-defined northbound APIs enable vertical service providers to programmatically consume network capabilities, decoupling application logic from network implementation.
2 – AI for full automation	The network manages itself autonomously through embedded AI/ML inference engines; human operators declare intent rather than configure parameters.
3 – Flexibility to different topologies	The architecture operates identically in dense urban, rural, aerial, maritime, and satellite environments without topology-specific customisation.
4 – Scalability	From a handful of devices to trillions of connected objects, the architecture scales horizontally without redesign.
5 – Resilience and availability	Self-healing network functions, geographic redundancy, and automatic failover ensure continuity under hardware failures, cyber-attacks, or natural disasters.
6 – Servicebased exposed interfaces	All inter-function communication uses standardised, loosely coupled service-based interfaces – enabling independent evolution and multi-vendor deployments.
7 – Separation of concerns	Management, control, and user plane functions are cleanly separated, allowing independent scaling, updating, and policy enforcement at each plane.
8 – Network simplification	6G must be measurably simpler to operate than 5G – fewer parameters, more automation, and reduced total cost of ownership – enabling global reach including low-income markets.

The two vertical cross-cutting dimensions are: Management and Orchestration with AI/ML (left axis), which operates continuously across all three strata with CI/CD (Continuous Integration/Continuous Delivery/Deployment) pipelines enabling live network updates; and Security (right axis), which implements a multi-layer security stack including physical layer security, privacy-enhancing technologies, quantum security, AI/ML security, distributed ledger technologies, and trust foundations. The critical architectural insight from this view is that both AI/ML orchestration and security are pervasive system properties – not components at any single layer – and the architecture must be designed accordingly.

7. Proposed 6G Reference Architecture -- Functional View

The 5G PPP Architecture Working Group's functional reference model (from Working Party documentation) further decomposes the Network Service Layer into a Security Stratum, Network Intelligence Stratum, Sensing Stratum, and Network Stratum, each running vertically through the Management Plane, Control Plane Network Functions, User Plane Network Functions, and Infrastructure [3]. This functional view makes the separation of concerns explicit (Principle 7) and shows how the Sensing Stratum – representing integrated sensing and communications – is a first-class architectural element in 6G, not an application-layer overlay as in 5G. Figure 4 shows the 5G PPP Architecture Working Group's functional reference model [3], which decomposes the Network Service Layer into four vertical strata running through all horizontal planes.

The Network Intelligence Stratum – the AI/ML engine – continuously optimises every function. The Security Stratum enforces security at every plane, not only at the perimeter. The Sensing Stratum makes ISAC a native architectural element: the same physical-layer waveform simultaneously carries user data and performs environmental sensing, localisation, and radar-like ranging. The Sensing Stratum outputs are consumed both by applications (autonomous vehicles, digital twins) and by the network's own resource management functions.

8. Proposed Four-Layer Architecture

Building on the principles and high-level views, we propose organising the 6G architecture into four functional layers, each with well-defined responsibilities, northbound APIs to the layer above, and southbound service consumption from the layer below. A cross-cutting AI/ML orchestration plane and a unified data fabric span all four layers. Figure 5 shows this stack with the layers numbered 1 (bottom) to 4 (top), with the AI/ML orchestration plane as a vertical strip and the Unified Data Fabric plus Zero-Trust Security Framework as a horizontal base. Figure 6 provides detailed deep dives into each layer's key functions.

The sub-100-μs latency target enables haptic feedback loops in the tactile internet – human reflexes operate at approximately 10 ms, and realistic haptic response requires network round trips well below this threshold. The 100× energy efficiency target addresses the unsustainable energy trajectory of globally deployed wireless networks. The 2030 commercial target is driven by the competitive strategies of Korea (targeting 2028), Japan, China, and others, as detailed in the WWRF 6G Global Pathways position paper (under development).

Layer 1 -- Intelligent RAN (I-RAN)

The Intelligent RAN extends the 5G Open RAN paradigm [17] along three dimensions. First, the functional split between the Distributed Unit (DU) and Central Unit (CU) is made dynamic and context-aware: latency-sensitive processing migrates toward the edge in real time based on traffic conditions, interference state, and application requirements, without manual reconfiguration. This dynamic split is enabled by a federated learning-capable Near-Real-Time RAN Intelligent Controller (Near-RT RIC), which hosts xApps that train interference coordination and handover prediction models across base stations without centralising raw channel data – preserving user privacy while maintaining system optimality.

Second, Reconfigurable Intelligent Surfaces (RIS) are managed as first-class network elements under the RAN controller, not as external accessories. RIS panels – large arrays of passive reflective elements – enable passive beamforming with

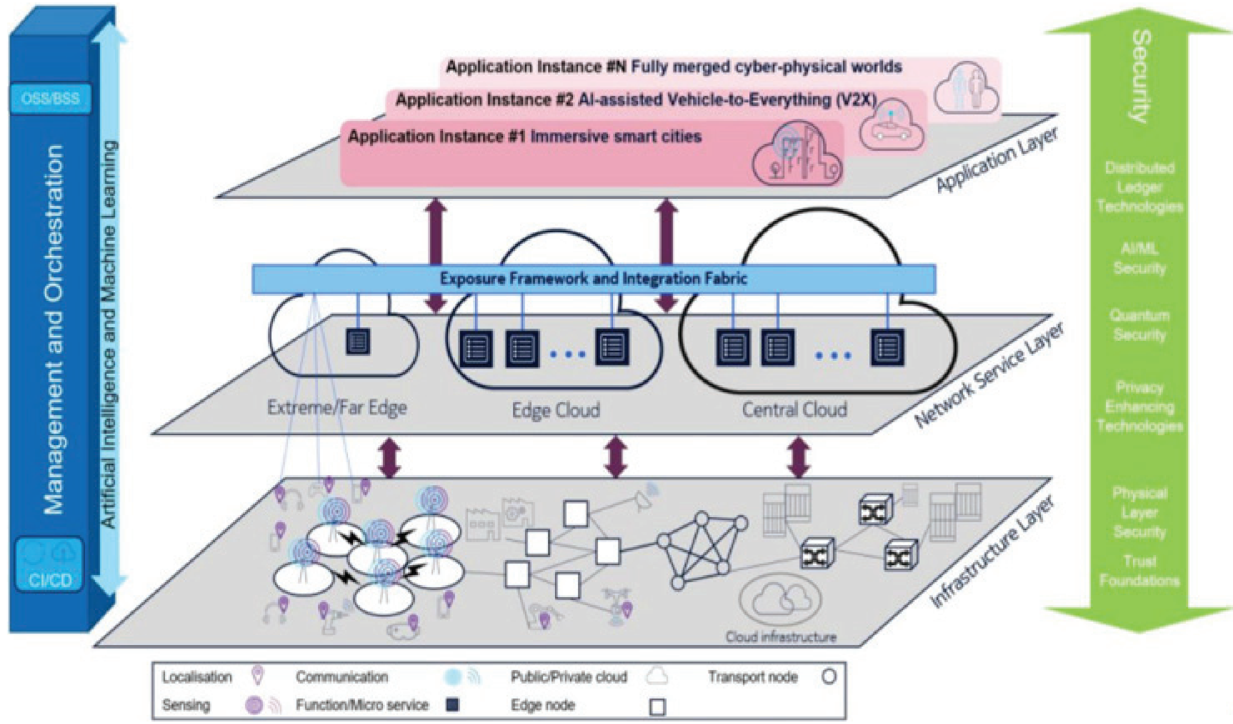

Figure 3.
High-level view of the 6G architecture [1].

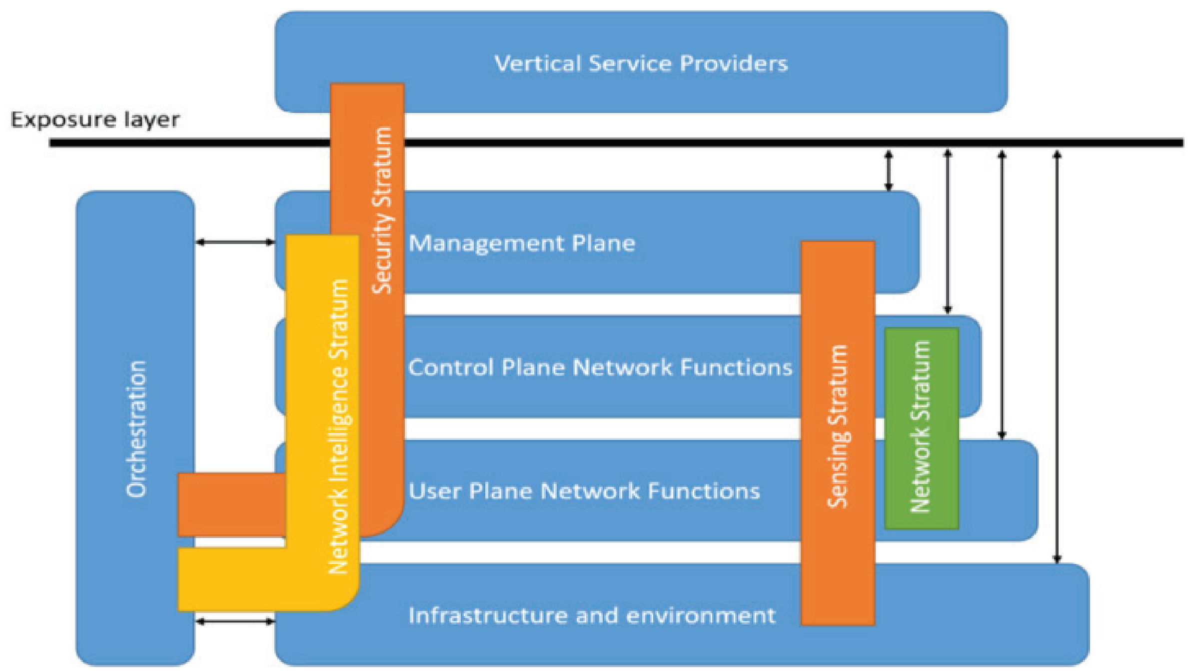

Figure 4.
Proposed 6G reference architecture – functional view [3]. Four vertical strata – Network Intelligence Stratum, Security Stratum, Sensing Stratum, and Network Stratum – run through the Management Plane, Control Plane Network Functions, User Plane Network Functions, and Infrastructure. The Exposure Layer provides northbound access to Vertical Service Providers. Orchestration with CI/CD runs on the left.

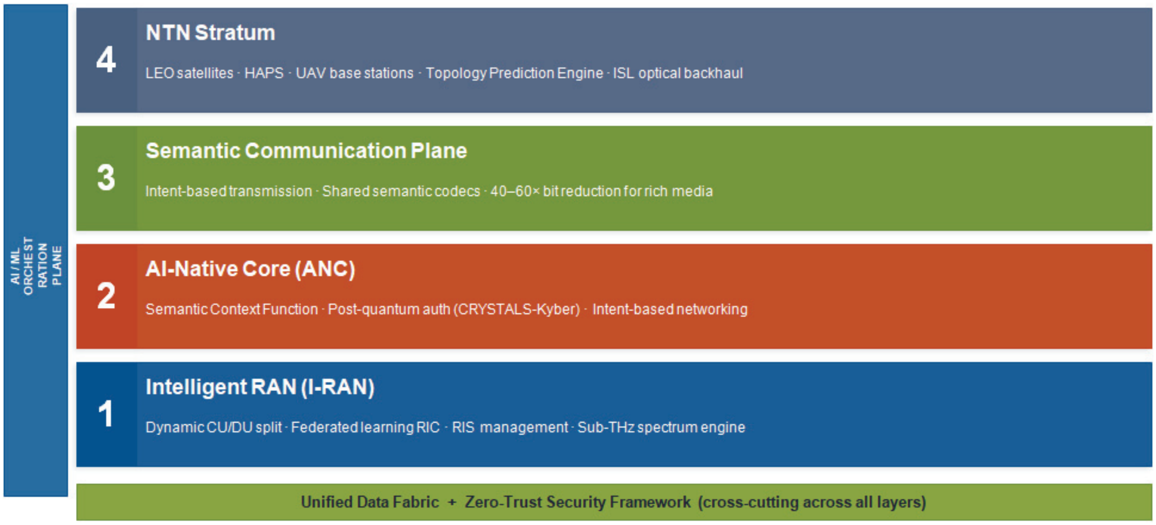

Figure 5.

Proposed 6G four-layer architecture stack. Layer 1 (navy): Intelligent RAN (I-RAN) – dynamic CU/DU split, federated learning RIC, RIS, sub-THz spectrum engine. Layer 2 (orange): AI-Native Core (ANC) – Semantic Context Function, CRYSTALS-Kyber auth, intent-based networking. Layer 3 (olive): Semantic Communication Plane – 40–60× bit reduction, semantic codecs. Layer 4 (slate): NTN Stratum – LEO/HAPS/UAV, Topology Prediction Engine, ISL backhaul. Cross-cutting: AI/ML Orchestration Plane (vertical) and Unified Data Fabric + Zero-Trust Security Framework (horizontal base).

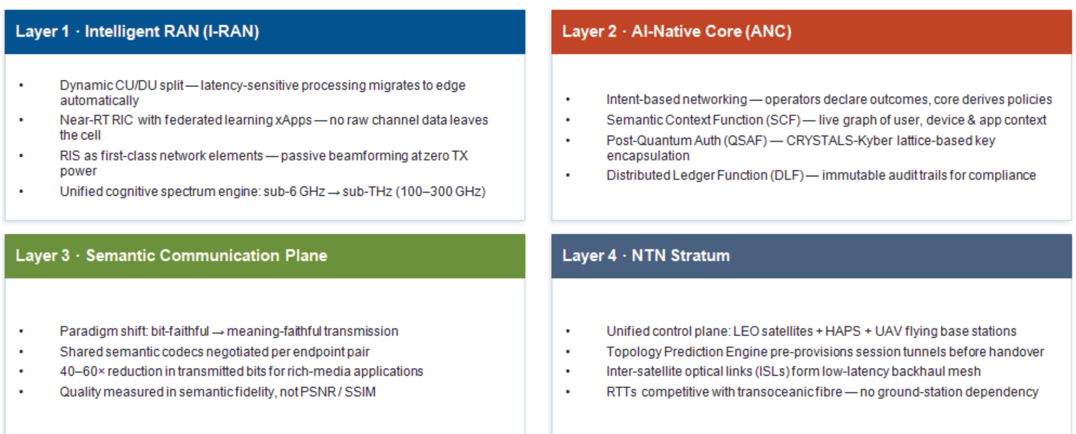

Figure 6.

Layer deep dives. I-RAN: dynamic CU/DU split, federated learning Near-RT RIC xApps (no raw channel data leaves the cell), RIS as first-class elements (passive beamforming at zero TX power), cognitive spectrum engine sub-6GHz through sub-THz (100–300 GHz). ANC: intent-based networking, Semantic Context Function (SCF), Post-Quantum Auth via CRYSTALS-Kyber (QSAF) [12, 13], Distributed Ledger Function (DLF). SCP: bit-faithful to meaning-faithful paradigm shift, shared semantic codecs, 40–60× bit reduction [14, 15], semantic fidelity metrics. NTN Stratum: unified LEO+HAPS+UAV control plane, Topology Prediction Engine [16], intersatellite optical links, RTTs competitive with transoceanic fibre.

zero transmission power, extending coverage into blind spots and improving spectral efficiency without adding interference. Treating them as native network functions allows the AI/ML orchestration plane to include them in system-wide resource allocation.

Third, the I-RAN's spectrum management engine spans the full 6G spectrum range: sub-6 GHz for coverage, millimetre-wave (24–100 GHz) for capacity, and sub-terahertz (100–300 GHz) for peak-rate hotspot scenarios. It should be noted that the most essential spectrum is in lower part of FR3 (Frequency Range 3), like 7–8 GHz as well as around 6 GHz. Sub-THz and mm-wave are not central at least not in the beginning of 6G, due to the "terahertz wall" of severe physical and technical hurdles, such as extreme path loss, short range, water absorption, health/biological effects. A unified cognitive engine arbitrates access based on real-time propagation sensing enabled by ISAC – the same waveform serves both communication and environmental sensing simultaneously, providing the network with a continuous, high-resolution map of the radio environment.

It is to be noted that the efforts required for the optimization of radio layer functions is going to add complexity in the higher layers of the four-layer architecture. Since the spectrum is a critical resource, intra-layer (radio-layer) design and optimization is extremely important. In addition, because the same waveform can be used for sensing and communication, etc, this results in the need to develop MAC layer enhancements caused by the massive sensing. Such an enhanced MAC layer would have to be capable of managing heavy load of sensing packets (passive and active sensing) coming from the "surrounding" targets in high density environments. The timely availability of this information would be essential for tuning AI mechanisms.

Layer 2 -- AI-Native Core (ANC)

The AI-Native Core departs from the 5G Service-Based Architecture by embedding AI inference engines directly into network functions, rather than post-hoc optimisation through external controllers. The defining paradigm shift is intent-based networking: operators and vertical applications declare outcomes (e.g., 'ensure 99.999% reliability for this surgical robotics slice') and the core autonomously derives and enforces the necessary policies across session management, UPF placement, and mobility without human configuration.

Three new network functions are central to the ANC. The Semantic Context Function (SCF) maintains a live knowledge graph of user, device, and application context – a structured, queryable representation of who is on the network, what they are doing, and what the network knows about their environment. This graph is the substrate for AI-driven policy generation. The Quantum-Secured Authentication Function (QSAF) implements lattice-based key encapsulation mechanisms standardised by NIST – specifically the CRYSTALS-Kyber algorithm (FIPS 203, ML-KEM) [12, 13] – providing authentication that is secure against both classical and quantum adversaries [12]. The Distributed Ledger Function (DLF) maintains an immutable, tamper-evident audit trail of network events and policy decisions, supporting regulatory compliance requirements across jurisdictions without requiring a centralised trusted authority.

Layer 3 -- Semantic Communication Plane (SCP)

The Semantic Communication Plane is the most architecturally novel layer of the proposed stack. It represents a fundamental shift in what the network transmits: from bit-faithful reproduction of source signals to meaning-faithful transmission of communicative intent. In the SCP paradigm, a video conferencing application need not transmit full pixel streams; instead, the SCP negotiates a shared semantic codec with the remote endpoint – a learned model that both parties agree encodes the relevant features of the communication – and transmits only the semantic content, with the receiver reconstructing a photorealistic representation locally.

The SCP interposes between applications and the transport network as a managed service layer, allowing the network operator to compress, prioritise, and reconstruct information at the semantic level. Quality of service is measured not in PSNR or SSIM – signal-fidelity metrics inherited from information theory – but in semantic fidelity: did the receiver understand the intent of the communication? Initial projections, consistent with results reported in [14, 15], suggest 40–60× reduction in transmitted bits for rich-media applications with no perceptible quality degradation. At 6G scale – billions of simultaneous high-bandwidth sessions – this reduction is not merely commercially attractive but is arguably a necessary condition for meeting the energy efficiency targets of IMT-2030.

Layer 4 -- NTN Stratum

Global 6G coverage – including the remote, maritime, and aerial environments required by the use case families of Section 2 – cannot be achieved by terrestrial base stations alone. The NTN Stratum integrates Low Earth Orbit (LEO) satellite constellations, High-Altitude Platform Systems (HAPS), and Unmanned Aerial Vehicles (UAV) acting as flying base stations into a single, unified control plane. Integration is not merely an interconnection of separate systems: NTN nodes are managed as first-class network elements by the same control plane that manages terrestrial base stations, with seamless handover between segments.

One of the critical enabling innovations for the NTN Stratum is the Topology Prediction Engine (TPE), which anticipates satellite visibility windows and handover events – using orbital mechanics, atmospheric models, and real-time traffic state – and pre-provisions session continuity tunnels before they are needed. This eliminates the handover latency spikes that are the principal quality-of-service problem in current LEO systems [16]. Inter-satellite optical links (ISLs) form a low-latency backhaul mesh between satellites, reducing reliance on ground stations and enabling round-trip times competitive with transoceanic fibre for data paths that cross oceans or poles.

Cross-Cutting Elements

Three mechanisms operate across all four layers. The AI/ML Orchestration Plane provides a hierarchical intelligence controller: at the RAN level, Near-RT RIC xApps handle sub-second resource allocation; at the core level, a Non-RT RIC provides network-wide policy; and at the system level, a global orchestrator balances load across terrestrial and non-terrestrial segments using reinforcement learning for long-horizon optimisation and transformerbased anomaly detection for real-time fault isolation. The Unified Data Fabric exposes standardised APIs for telemetry collection, model training, and inference – with differential privacy mechanisms ensuring that sensitive data never leaves its origin domain. The Zero-Trust Security Framework mandates continuous device authentication, microsegmentation of network slices, and hardware-rooted trust anchors (e.g., TPM 3.0) at every layer, eliminating the implicit trust relationships that are the principal vulnerability in current mobile network designs. Microsegmentation of network slices is a Zero Trust security technique that further segments broad, virtualized network slices into smaller, strictly isolated security zones down to the individual workload or service level [18].

Table 2.

6G vs. 5G: eight-dimension architectural comparison		
Parameter	**5G**	**Proposed 6G**
Peak Data Rate	20 Gbps	>1 Tbps
Air Interface Latency	1 ms	<100 μs
AI Integration	Post-hoc, external	Native & embedded
Coverage	Terrestrial + limited NTN	Unified terrestrial + NTN
Security	EAP-AKA / 5G-AKA	Post-quantum + zero-trust
Communication Model	Bit-faithful	Semantic/ goal-oriented
Energy Efficiency	Baseline	100× improvement per bit
Spectrum Bands	Sub-6GHz + mmWave	Sub-6/mmWave/ sub-THz

9. Architectural Comparison with 5G

Table 2 compares the proposed 6G architecture with 5G NR / 5G Core across eight dimensions. The qualitative shifts are architecturally more significant than the quantitative ones.

The quantitative dimensions – peak data rate, air interface latency – are well-established in the IMT-2030 framework. The qualitative shifts are architecturally more significant. The change in AI integration from post-hoc to native fundamentally changes the operational model: 6G networks are not configured, they are trained. The change in communication model from bit-faithful to semantic changes what the network's service boundary is: in 5G the network delivers bits; in 6G the network delivers meaning. The change in security from EAP-AKA to post-quantum plus zero-trust reflects not an incremental hardening but a change in threat model – one that accounts for the cryptographically relevant quantum computers that are expected to be operational before 6G reaches end of life [13, 19]. These qualitative shifts interact: semantic communication requires AI to generate and decode semantic codecs; AI requires zero-trust security to prevent adversarial manipulation of training data; zero-trust security requires the Distributed Ledger Function (DLF) to maintain tamper-evident audit trails. The four layers are not independent – they are architecturally coupled.

10. Open Research Challenges

Figure 7 summarises the four open research challenges alongside the architecture's key takeaways and some near-term validation opportunities to build confidence in the paradigm shift from 5G to 6G.

10.1 Semantic Codec Standardisation

The Semantic Communication Plane requires cross-industry agreement on a shared semantic codec interface – the protocol by which two endpoints negotiate a joint learned model for encoding and decoding communicative intent. This requires consensus on ontologies: the structured vocabulary of concepts and relationships that the codec uses to represent meaning. No such standard exists. The problem is deeper than a naming convention: different application domains (healthcare, automotive, manufacturing) have different semantic structures, and a codec that performs well in one domain may fail catastrophically in another. Research challenges include domain-adaptive semantic codecs, adversarial robustness of learned encoders, and objective quality metrics that capture semantic fidelity independently of signal-level measures such as PSNR or SSIM [14, 15]. This work is planned to be done by ETSI ISG.

10.2 NTN Topology Prediction at Orbital Scale

The Topology Prediction Engine must operate with high fidelity across constellations of thousands of LEO satellites – each moving at approximately 7.5 km/s relative to the ground, with inter-satellite geometry changing continuously. At this scale, the prediction problem is not tractable with classical orbital mechanics alone: atmospheric drag variations, satellite manoeuvres for collision avoidance, and traffic-dependent routing decisions all introduce stochastic perturbations that must be modelled probabilistically. Research challenges include large-scale graph neural networks for orbital topology prediction, handover decision policies under uncertainty, and the design of session continuity protocols that tolerate prediction errors without service interruption [16]. While each NTN will have its own TPE, it is unclear whether there is a need for an overarching TPE as well.

10.3 Post-Quantum Cryptography Compute Overhead

CRYSTALS-Kyber (ML-KEM, FIPS 203) introduces key encapsulation operations that are computationally more expensive than the elliptic-curve Diffie-Hellman operations used in 5G-AKA. For 6G authentication at scale – where a single base station may simultaneously authenticate thousands of devices following a power-up event or NTN handover – this overhead must be absorbed without degrading authentication latency below the 100 μs air interface target. Research challenges include hardware acceleration architectures for lattice-based cryptography at base station scale, lightweight variants of ML-KEM for constrained IoT devices, and hybrid classical/post-quantum authentication protocols that allow graceful migration from legacy 5G credentials [13, 19].

10.4 Federated Learning Convergence in Heterogeneous RAN

The I-RAN's Near-RT RIC uses federated learning to train xApps across base stations without sharing raw channel data. In practice, the data distributions across cell sites are highly heterogeneous and non-independent-and-identically-distributed (non-IID): urban macro-cells see different traffic patterns, channel conditions, and interference topologies than suburban micro-cells or industrial private networks. Standard federated averaging algorithms have poorly understood convergence properties under non-IID conditions, particularly when the number of participating

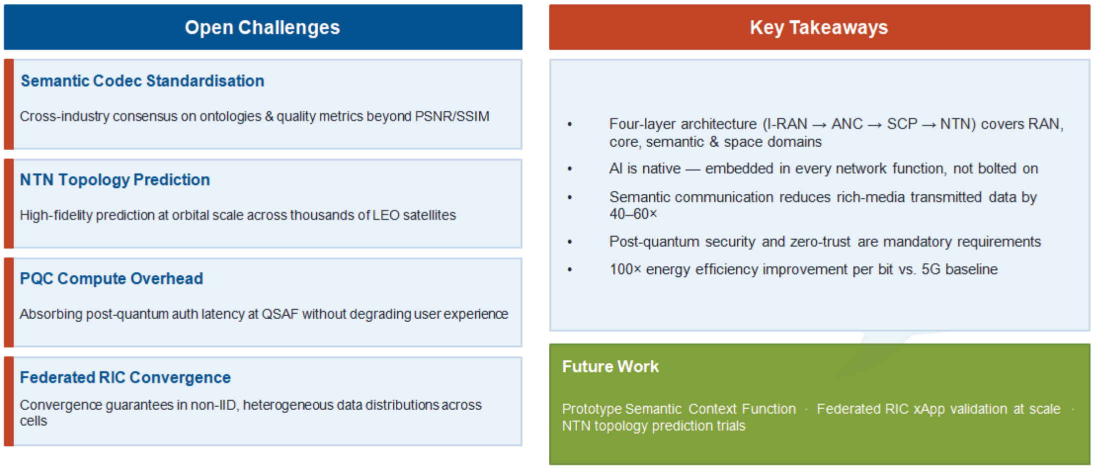

Figure 7.
Challenges and conclusions.

nodes is large and communication rounds are rate-limited by the RIC's near-real-time constraint. Research challenges include federated optimisation algorithms with provable convergence guarantees under non-IID data, privacy-preserving gradient aggregation mechanisms, and adaptive client selection policies that balance convergence speed with communication cost [20, 21].

11. Conclusion

This paper has proposed and analysed a unified four-layer 6G network architecture – Intelligent RAN, AI-Native Core, Semantic Communication Plane, and NTN Stratum – designed to satisfy the six Hexa-X use case families and the eight HEX-D51 design principles while closing the five structural gaps left by 5G, all in conformance with the requirements set out by the ITU. The architecture is distinguished by four qualitative shifts from 5G: post-hoc to native AI; bit-faithful to semantic communication; unified TN-NTN coverage; and perimeter-based to zero-trust security. These shifts are mutually reinforcing and architecturally coupled, forming a coherent system rather than a collection of independent improvements. This reinforcing coupling is the result of: semantic communication requires native AI; native AI requires zero-trust security; zero-trust security requires the distributed ledger infrastructure of the AI-Native Core.

The architecture is grounded in three established reference frameworks: the Hexa-X high-level layered view (HEX-D13), the functional decomposition of the 5G PPP Architecture Working Group, and the eight design principles of HEX-D51. These have in turn guided the ITU to define the IMT-2030 requirements. Its four open research challenges – semantic codec standardisation, NTN topology prediction, post-quantum cryptography compute overhead, and federated RIC convergence – define the engineering agenda that must be resolved before 2030 commercial deployment. Progress on each will require close coordination between 3GPP standardisation, ITU-R spectrum work, and academic research communities should benefit from the kind of cross-regional collaboration that the WWRF and its member organisations are uniquely positioned to facilitate.

Future work will focus on prototype implementation of the Semantic Context Function within an open-source 5G core, experimental validation of federated Near-RT RIC xApps under realistic non-IID traffic distributions, and the design of a hardware-accelerated QSAF compliant with the FIPS 203 ML-KEM standard.

References

[1] Hexa-X Consortium, "Deliverable D1.3 – Hexa-X Use Cases and Requirements (HEX-D13)," EU Horizon 2020, 2022.
[2] Hexa-X Consortium, "Deliverable D5.1 – 6G Architecture Principles and Design (HEX-D51)," EU Horizon 2020, 2023.
[3] 5G PPP Architecture Working Group, "View on 6G Architecture, v2.0," 5G Infrastructure Association, 2023.
[4] ITU-R, "Framework and overall objectives of the future development of IMT for 2030 and beyond," Recommendation ITU-R M.2160, ITU, Geneva, 2023.
[5] 3GPP, "TR 22.870 – Study on use cases and requirements for IMT-2030," Release 20 Study Item, 2025.
[6] Samsung Research, "The Next Hyper-Connected Experience for All: A White Paper on 6G," 2020 (updated 2022).
[7] ATIS Next G Alliance, "Roadmap to 6G: Building the Foundation for North American Leadership in 6G and Beyond," ATIS, Washington, DC, 2022 (updated 2024–2025). [Online]. Available: https://nextgalliance.org.
[8] University of Oulu 6G Flagship, "White Paper on 6G Networking," 6G Research Visions No. 6, 2020.
[9] David K. and Berndt H.., "6G vision and requirements: Is there any need for beyond 5G?", IEEE Vehicular Technology Magazine, vol. 13, no. 3, pp. 72–80, September 2018.
[10] Andrews, J. G. et al., "What will 5G be?" IEEE JSAC, vol. 32, no. 6, pp. 1065–1082, 2014.
[11] IMT-2030 (6G) Promotion Group, "6G Vision and Candidate Technologies White Paper," MIIT, China, 2021 (updated 2024).
[12] NIST, "Module-Lattice-Based Key-Encapsulation Mechanism Standard (ML-KEM)," FIPS 203, 2024.
[13] Bos, J. et al., "CRYSTALS-Kyber: a CCA-secure module-lattice-based KEM," IEEE EuroS&P, pp. 353–367, 2018.
[14] Qin, Z. et al., "Semantic communications: principles and challenges," IEEE Communications Magazine, vol. 61, no. 3, pp. 26–32, 2022.

[15] Shi, J. et al., "Semantic communication systems for speech transmission," IEEE JSAC, vol. 39, no. 8, pp. 2434–2444, 2021.
[16] Del Portillo, I. et al., "A technical comparison of three LEO satellite constellation systems," Acta Astronautica, vol. 159, pp. 216–225, 2019.
[17] O-RAN Alliance, "O-RAN Architecture Description v07.00," Technical Specification, 2023.
[18] Olli Mammel, et al, "Towards Micro-Segmentation in 5G Network Security," European Conference on Networks and. Communications (EuCNC), Athens, 2016.
[19] Bernstein, D. J. and Lange, T., "Post-quantum cryptography," Nature, vol. 549, pp. 188–194, 2017.
[20] McMahan, B. et al., "Communication-efficient learning of deep networks from decentralised data," AISTATS, vol. 54, pp. 1273–1282, 2017.
[21] Li, T. et al., "Federated learning: challenges, methods, and future directions," IEEE Signal Processing Magazine, vol. 37, no. 3, pp. 50–60, 2020.

Biography

Sudhir Dixit is Co-Founder and COO of the Basic Internet Foundation in Oslo, Norway, where he also heads US operations. He is also a Docent at the University of Oulu, Finland. Previously, he co-founded Skydoot, Inc. in Davis, California (2015–2017) and was its CEO, and spent five years at Hewlett-Packard Enterprise Services as Distinguished Chief Technologist, CTO for Communications and Media Services in the Americas, and Director of HP Laboratories India. Earlier in his career, he held engineering and leadership roles at BlackBerry, Nokia, NSN, and NYNEX (now Verizon Communications).

A prolific figure in the IEEE community, Dixit serves as Co-Chair of the Connecting the Unconnected working group in the IEEE Future Networks Initiative; WG 6G Chair, and Vice Chair for Americas at the WWRF, where he is also the Editor in Chief of the WWRF magazine, Wireless World Research and Trends. He has served as a technical editor of IEEE Communications Magazine and was on the editorial board of IEEE Spectrum. Dixit has published eight books and holds 21 U.S. patents.

A Life Fellow of the IEEE, IET and IETE, Dixit holds a Ph.D. from the University of Strathclyde, Glasgow, an M.B.A. from the Florida Institute of Technology, an M.E. in Electronics from BITS Pilani, and a B.E. in Electrical Engineering from MANIT, Bhopal.