Review

Computational Architectures for 6G Networks: Integrating Distributed Computing and Edge Artificial Intelligence

Evelio Astaiza Hoyos , Héctor Fabio Bermudez-Orozco * , Nasly Cristina Rodriguez-Idrobo

Posted Date: 12 February 2026

doi: 10.20944/preprints202602.1019.v1

Keywords: distributed computing; edge artificial intelligence; 6G network architectures; computational frameworks; AI-native networks; intelligent orchestration

Review

Computational Architectures for 6G Networks: Integrating Distributed Computing and Edge Artificial Intelligence

Evelio Astaiza Hoyos, Héctor Fabio Bermúdez-Orozco * and Nasly Cristina Rodriguez-Idrobo

University of Quindio

* Correspondence: hfbermudez@uniquindio.edu.co; Tel.: +057 3206671107

Abstract

The sixth generation of mobile networks (6G) is envisioned as an AI-native and computation-driven infrastructure capable of supporting ultra-low latency, massive connectivity and intelligent services across highly heterogeneous environments. Achieving these objectives challenges traditional centralised architectures and motivates a shift towards distributed computing and intelligence at the network edge. This study presents a structured computational analysis of architectural approaches that integrate distributed computing paradigms and Edge Artificial Intelligence (Edge AI) as core enablers of 6G networks. The methodology follows PRISMA guidelines for systematic reviews and is based on a comprehensive analysis of peer-reviewed literature, architectural proposals and standardisation documents retrieved from major scientific databases, including IEEE Xplore, Scopus, Web of Science, MDPI and arXiv, as well as reports from ITU-R, 3GPP and ETSI. The analysis examines the evolution from cloud-centric to edge-centric computing, key Edge AI techniques—such as Federated Learning, Split Learning and edge-adapted large AI models—and their role in enabling intelligent orchestration, resource optimisation and context-aware services. The results indicate that the tight integration of distributed computing and Edge AI enhances network responsiveness, scalability and adaptability, while also revealing persistent challenges related to orchestration complexity, resource constraints, security and interoperability. The study concludes that holistic computational architectures and AI-native design principles are essential for the effective realisation of 6G networks and for guiding future research and standardisation efforts.

Keywords: distributed computing; edge artificial intelligence; 6G network architectures; computational frameworks; AI-native networks; intelligent orchestration

1. Introduction

The evolution of mobile communication systems is increasingly characterised by the convergence of networking, distributed computing and artificial intelligence, transforming communication infrastructures into large-scale computational systems. As network complexity, data volume and service heterogeneity continue to grow, the efficient distribution and orchestration of computational and intelligent functions across the network have become central design challenges. In this context, future mobile networks are no longer conceived merely as platforms for data transmission, but as computation-driven and intelligence-native infrastructures in which communication, computing and data processing are tightly integrated. This computational perspective provides the necessary foundation for understanding the architectural transformations required to support the next generation of mobile systems and frames the discussion of sixth-generation (6G) networks presented in the following sections.

1.1 The 6G Vision: Beyond Connectivity

 doi:10.20944/preprints202602.1019.v1

Mobile networks have evolved dramatically from the first generation (1G), which introduced mobile telephony, to the fifth generation (5G), which began incorporating intelligence into mobile communications [1]. The sixth generation (6G), expected to reach commercial deployment around 2030, promises an even deeper transformation, marking the transition from "mobile intelligence" to an "Ubiquitous Intelligent Mobile Society" [2]. The 6G vision transcends incremental improvements in connectivity and aspires instead to achieve a seamless fusion of the physical, digital and human worlds [3], where artificial intelligence (AI) is not an overlaying application but an intrinsic capability of the network itself [4].

The International Telecommunication Union (ITU), through its Radiocommunication Sector (ITU-R), has established the foundational framework for 6G under the designation "IMT-2030" [5]. This framework defines several evolved use scenarios compared with 5G, such as immersive communication (enabling interactive experiences like XR and holographic telepresence), massive communication (supporting the Internet of Everything – IoE), and high-reliability, low-latency communication (HRLLC) for mission-critical applications [5]. It also introduces new scenarios enabled by emerging capabilities, including integrated sensing and communication (ISAC), native AI integration and ubiquitous connectivity spanning terrestrial, aerial, space and underwater environments [6].

To support these scenarios, 6G sets key performance indicators (KPIs) that are significantly more ambitious than those of 5G. These include peak data rates on the order of 1 Tbps [7], user-experienced data rates of up to 10 Gbps [2], end-to-end (E2E) latency below 1 ms—and even in the range of 0.1 ms to 100 μs for specific cases [6]—connection densities potentially reaching 10^8 devices per km² [7], extremely high reliability (success probability of 0.99999 to 0.9999999) [8], mobility support for speeds up to 1000 km/h [2], and substantial improvements in energy and spectral efficiency [6]. New capabilities such as centimetre-level or even sub-centimetre-level positioning accuracy [6] and integrated sensing [9] also form an integral part of the vision.

Realising this vision and meeting such demanding KPIs critically depend on the development and integration of radically new enabling technologies and network architectures [10]. Technologies such as terahertz (THz) and visible light communication (VLC), reconfigurable intelligent surfaces (RIS), ultra-massive MIMO (UM-MIMO), non-terrestrial networks (NTN) and AI are considered fundamental pillars [1].

1.2 *The Critical Role of Distributed Computing and Edge AI*

Among the most crucial enabling technologies for 6G are distributed computing—particularly in the form of Edge Computing and Multi-access Edge Computing (MEC)—and Edge Artificial Intelligence (Edge AI) [10]. Edge Computing consists of bringing computational and storage capabilities closer to end users or data sources by situating them at the edge of the access network [3]. This contrasts with the traditional centralised cloud model and is essential for mitigating latency and reducing the load on transport networks [11].

Edge AI, in turn, refers to the execution of AI algorithms—both for training and inference—directly on edge nodes or even on end devices [12]. The convergence of distributed edge computing with AI is essential for realising the 6G vision of "connected intelligence" [13]. This synergy not only enables intelligent and autonomous optimisation of the 6G network itself but also supports a new generation of services requiring ultra-low latency, extensive local data processing and customised, context-aware AI capabilities. The deep integration of AI into the network architecture, often termed "AI-native" [4], fundamentally depends on distributed computing infrastructure at the edge.

This article aims to provide an expert and comprehensive analysis of existing proposals for integrating distributed computing and Edge AI as key elements for implementing 6G networks. It seeks to identify the synergies between these two technological domains, examine the inherent challenges of their joint deployment within the 6G context and explore future perspectives and directions for research and standardisation.

 doi:10.20944/preprints202602.1019.v1

Section 2 examines the distributed computing paradigms relevant to 6G, tracing their evolution from cloud to edge and highlighting the benefits and challenges of Edge Computing. Section 3 delves into the concept of Edge AI, describing key techniques (such as FL, SL and Edge LAMs), their applications for optimising the 6G network and enabling intelligent services, and the associated challenges. Section 4 analyses various architectural and orchestration proposals that aim to effectively integrate distributed computing and AI into 6G networks, including interactions with technologies such as Digital Twins and ISAC. Section 5 reviews the current state of standardisation within major organisations (ITU-R, 3GPP, ETSI) and discusses open challenges and future research directions. Section 6 presents the main conclusions of the analysis, synthesising the key findings and reinforcing the transformative role of distributed computing and Edge AI for the future of mobile communications.

1. Distributed Computing Paradigms in the 6G Context

The evolution towards sixth-generation (6G) mobile networks involves not only advances in radio technologies, but also a fundamental transformation in how computation, data processing and intelligence are distributed across the network. From a computational perspective, 6G can be understood as a large-scale, heterogeneous distributed computing system in which communication, computing and storage resources must be jointly orchestrated to meet extreme performance requirements. In this context, distributed computing—particularly edge-oriented paradigms—emerges as a core architectural shift required to support ultra-low latency, massive scalability and intelligent network operation. These paradigms redefine where computational tasks are executed, how resources are allocated and how system-level efficiency, responsiveness and reliability are achieved in next-generation mobile networks.

2.1 Historical Evolution: From Centralised Cloud to the Distributed Edge

For the past several decades, the dominant paradigm has been Cloud Computing. According to the NIST definition, it is a model that enables ubiquitous, convenient and on-demand access to a shared pool of configurable computing resources (networks, servers, storage, applications and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction [13]. This centralised model has been fundamental for large-scale data processing and the deployment of massive applications [14].

However, the growing proliferation of connected devices (Internet of Things – IoT) and the emergence of applications with strict real-time requirements—such as augmented/virtual reality (AR/VR), autonomous driving and industrial automation—have highlighted inherent limitations of the centralised cloud model [3]. The physical distance between end devices and cloud data centres introduces significant latency, incompatible with the millisecond- or even microsecond-level requirements of many 6G applications [14]. Furthermore, transmitting the massive volumes of data generated at the edge—estimated to represent 75% of enterprise data by 2025 [15]—to the cloud consumes substantial bandwidth and raises serious concerns regarding the privacy and security of sensitive information [3].

Fog Computing emerged in response to these limitations. This paradigm introduces an intermediate layer of distributed computing infrastructure positioned between end devices and the centralised cloud [11]. Fog nodes, although more resource-constrained than the cloud [16], are geographically closer to users, enabling the local processing of latency-sensitive tasks or smaller data volumes, thereby reducing dependence on the cloud [17]. The typical architecture of Fog Computing is often described as comprising three layers: the terminal/IoT layer, the Fog layer and the Cloud layer [17].

Edge Computing represents the logical extension of this decentralisation trend, bringing computational and storage capabilities even closer to the end user, directly to the edge of the access

 doi:10.20944/preprints202602.1019.v1

network [3]. Within this broad concept, Multi-access Edge Computing (MEC) is a key initiative, particularly relevant for mobile networks.

This progression from centralised cloud to distributed edge is not merely another technological option but a necessary and inevitable response to the fundamental requirements of the 6G vision. Applications that will define 6G—such as immersive XR experiences [10], cooperative autonomous driving [18] and the Tactile Internet [1]—impose ultra-low latency requirements (<1 ms) [7] and generate massive locally produced data streams [19]. The physics of signal propagation and the bandwidth limitations of transport networks make it infeasible to satisfy these requirements from distant centralised cloud infrastructures [14]. The need to process critical information in real time and to preserve the privacy of sensitive data (such as biometric data in XR or vehicular data) [14] further reinforces the imperative of bringing computation closer to the source. Therefore, the adoption of paradigms such as Fog Computing and, especially, Edge Computing becomes a sine qua non condition for enabling many of the most transformative 6G use cases.

2.2 Multi-access Edge Computing (MEC) as a Key Enabler in 6G

The European Telecommunications Standards Institute (ETSI) defines MEC as a technology that provides application developers and content providers with cloud-computing capabilities and an IT service environment at the edge of the multi-access network (including mobile networks such as the RAN, as well as Wi-Fi and others) [20]. This environment is characterised by ultra-low latency, high bandwidth and real-time access to radio network information, which can be exploited by applications [20].

MEC, already a significant component in the evolution of 5G [1], is considered even more critical for 6G [7]. The MEC architecture is expected to undergo significant transformation in 6G, becoming more highly distributed, extending further towards the "Far Edge" (computing nodes located closer to the user, potentially within radio units or cellular sites) [21], and adopting principles of openness and disaggregation [22].

The deep integration of MEC into the 6G architecture is essential for supporting the most demanding use cases [23]. Examples include:

2.2.1. Connected and Autonomous Vehicles (V2X)

MEC provides the low latency required for safety-critical V2V and V2I communications (collision avoidance, coordinated driving) and enables local processing of sensor data for environmental perception [3].

2.2.2. Extended Reality (XR) and the Metaverse

Immersive experiences require intensive graphical rendering and extremely low-latency responses to user actions. MEC enables the offloading of part of this computation to the edge, improving Quality of Experience (QoE) and reducing the processing burden on end devices [11].

2.2.3. Industrial Automation and Robotics

Real-time control of robots and cyber-physical systems (CPS) in Industry 4.0 demands microsecond-level latency and high reliability, both of which MEC can support [7].

2.2.4. Drones and UAVs

Fleet management, autonomous navigation and onboard sensor-data processing can be enhanced through computational support provided by terrestrial or aerial MEC nodes [9].

2.2.5. Telemedicine

Applications such as remote surgery or AI-assisted diagnosis require low latency and secure processing of sensitive medical data, making MEC an ideal enabler [7].

Moreover, the evolution towards Open-Source MEC (OS-MEC) is actively being explored [22]. Traditional MEC architectures may be rigid, as they depend on specialised hardware integrated with proprietary software. OS-MEC proposes decoupling MEC functions (software) from underlying

 doi:10.20944/preprints202602.1019.v1

resources (hardware) using technologies such as Network Function Virtualisation (NFV) and Software-Defined Networking (SDN). This would allow dynamic reconfiguration of functions and resources to create customised MEC services tailored to specific 6G scenarios, promoting innovation and flexibility [22].

2.3 Fundamental Benefits of Edge Computing for 6G

The widespread adoption of Edge Computing in 6G networks offers several essential advantages:

2.2.1. Drastic Reduction in Latency

By processing data and executing applications closer to the user or the data source, the round-trip time to a centralised cloud is eliminated. This is absolutely critical for meeting the ultra-low latency requirements (<1 ms, and even 0.1 ms) of 6G scenarios such as HRLLC, the Tactile Internet, immersive XR, cloud gaming, remote surgery and autonomous driving [3].

2.2.2. Optimised Bandwidth Usage

Local processing significantly reduces the amount of data that must be transmitted across backhaul networks and the network core [3]. This is vital in 6G, where an explosion in data volume is expected from billions of IoT devices, sensors and multimedia applications [19]. Alleviating backhaul congestion improves overall network performance.

2.2.3. Enhanced Privacy and Security

Keeping sensitive data (personal, medical, industrial, vehicular) at the edge, without the need to transmit it to a potentially less trustworthy or more exposed central cloud, intrinsically enhances privacy [4]. Local processing reduces the attack surface associated with long-distance data transmission.

2.2.4. Enablement of Context Awareness and Localisation

Physical proximity allows edge applications and services to access and react to local information in real time, such as radio channel conditions, precise device location or events in the physical environment [3]. This is crucial for personalised and adaptive services.

2.2.5. Greater Scalability and Reliability

Distributed Edge Computing architectures can be inherently more scalable, allowing capacity to be added incrementally where needed. Moreover, they avoid the single points of failure typical of centralised systems, thereby improving overall system resilience [11].

2.4 Inherent Challenges of Distributed Edge Computing

Despite its advantages, the implementation of distributed edge computing for 6G presents significant challenges that must be addressed:

2.3.1. Resource Constraints

Edge devices and nodes (ranging from sensors and smartphones to MEC servers) possess considerably lower computational (CPU, GPU, NPU), storage and energy capabilities than centralised cloud data centres [3]. These limitations hinder the execution of computationally intensive tasks, particularly complex AI algorithms such as large language models (LLMs) [11].

2.3.2. Complex Management and Orchestration

Managing an ecosystem of heterogeneous, geographically distributed and dynamic computing resources is inherently complex [10]. It requires sophisticated solutions for resource discovery, task allocation (offloading), load balancing, application lifecycle management and Quality of Service (QoS) assurance [3].

2.3.3. Security and Privacy

Although the edge enhances privacy by localising data, the distributed and often open nature of edge infrastructures increases the attack surface [24]. Challenges include physical security of edge

 doi:10.20944/preprints202602.1019.v1

nodes, heterogeneity of operating systems and protocols, limited user interfaces on IoT devices, weak computational capabilities at peripheral devices for robust defences, container security in MEC, vulnerabilities in SDN (centralised controller, switch saturation), complex access management in mobile and multi-domain environments, risks associated with open interfaces (such as in O-RAN) and multi-vendor interoperability issues [24]. Robust mechanisms for authentication, authorisation, encryption, data protection, intrusion detection and trust management are essential [11].

2.3.4. Mobility and Intermittent Connectivity

Ensuring service continuity and maintaining application state for mobile users and devices as they move across different access points or edge nodes is a major challenge [3]. This is particularly critical in V2X and UAV scenarios, where network topology changes rapidly [18]. Intermittent connectivity can also disrupt distributed tasks such as AI model training [25].

2.3.5. Interoperability and Standardisation

The diversity of hardware, software and protocols in the edge ecosystem can lead to interoperability issues unless clear and open standards are established [24]. The lack of standardisation hinders the creation of a unified market and limits application portability [26].

The transition towards more open and disaggregated architectures—such as Open-Source MEC (OS-MEC) [22] or Open RAN (O-RAN)-based frameworks [24]—while promising in terms of flexibility, innovation and reduced vendor lock-in, intensifies several of these challenges. Open interfaces, by definition, increase the surface exposed to potential attacks [24]. Multi-vendor component management introduces additional complexity in ensuring security and interoperability [24]. Functional decoupling requires more sophisticated orchestration mechanisms to compose and manage end-to-end services [22]. This dynamic highlight a fundamental tension: the pursuit of openness and flexibility at the 6G edge creates an even more urgent need for advanced and automated solutions for managing security, trust and orchestration—possibly through AI-driven approaches or distributed ledger technologies such as blockchain [11]. Table 1 summarises the key characteristics of the computing paradigms discussed.

Table 1. Comparison of Computing Paradigms: Cloud, Fog and Edge

Characteristic	Cloud Computing	Fog Computing	Edge Computing (MEC)
Location of Compute/Storage	Centralised, remote data centres	Distributed nodes between Cloud and Edge	At/near the network edge (e.g., RAN, cellular sites)
Typical Latency	High (tens to hundreds of ms) [46]	Medium/Low (ms to tens of ms) [53]	Ultra-low (sub-ms to a few ms) [4]
Bandwidth (Backhaul)	High consumption [41]	Reduced consumption compared with Cloud [53]	Minimal/optimised consumption [47]
Management/Orchestration	Centralised, mature	Distributed, more complex	Highly distributed, complex, requires automation [33]
Scalability	Very high (virtually unlimited resources)	Moderate	High (distributed), but limited by local resources
Main Limitations	Latency, bandwidth, privacy [41]	Per-node resource constraints, complexity [53]	Very limited per-node resources, mobility, security [33]
Typical 6G Use Cases	Large-scale offline processing, storage	Industrial IoT, local video analytics [42]	XR, V2X, robotics, gaming, real-time AI [4]

 doi:10.20944/preprints202602.1019.v1

2. Edge AI: Artificial Intelligence Integration at the 6G Network Edge

The convergence of artificial intelligence with edge computing infrastructure is one of the most significant conceptual and technological pillars of 6G. Edge AI is not merely an application running on the edge network; it is an intrinsic capability that redefines both network operation and the services it can deliver as is illustrated in Figure 1.\

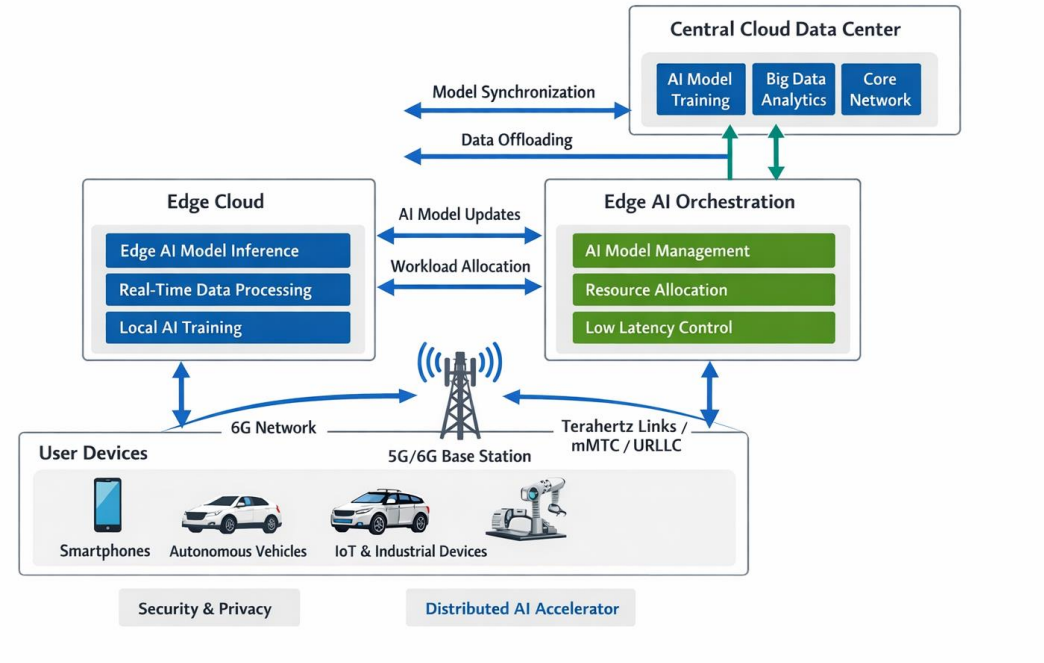

Figure 1. Possible Edge AI Architecture in 6G

2.1. *Concept and Relevance of Edge AI in 6G*

Edge AI refers to the deployment and execution of artificial intelligence algorithms—including both training and inference phases—directly on edge devices (such as smartphones, sensors or vehicles) or on infrastructure nodes located close to the data source (such as MEC servers) [27]. This approach represents a paradigm shift compared with traditional cloud-centred AI.

Its relevance to 6G is fundamental. The 6G vision of "connected intelligence" [13] or "AI-native networks" [4] implies that AI must permeate all layers and domains of the network [28]. Edge AI is the disruptive technology that enables this deep integration, fostering a synergy among communication, computing, sensing and intelligence [10]. Pervasive AI is considered a key candidate for managing the complexity of 6G networks, enabling dynamic resource allocation, adaptive traffic flow management and advanced signal processing in interference-rich environments [29].

The key advantages of Edge AI over cloud-centred AI are particularly significant in the 6G context [4]:

- Lower Latency: Local execution eliminates communication delays with the cloud, which is crucial for real-time 6G applications.
- Reduced Bandwidth Consumption: Large volumes of raw data do not need to be transmitted to the cloud, as they are processed locally.
- Enhanced Privacy and Security: Sensitive data remain on the local device or edge node, reducing exposure and attack surface.
- Context Awareness: AI models can exploit real-time, locally available information to make more adaptive and context-specific decisions.
- Offline Operation: Intelligent services can function even without a continuous connection to the cloud.

 doi:10.20944/preprints202602.1019.v1

2.2. *Key Models and Techniques for Edge AI in 6G*

Given the resource constraints at the edge, not all AI models and techniques are directly applicable. Specific approaches have therefore been developed and adapted for Edge AI, with the most relevant for 6G being the following:

2.2.1. Federated Learning (FL)

Federated Learning is a distributed learning framework in which multiple clients (edge devices) collaboratively train a global model coordinated by a central server, without sharing their private local data [30]. Only model updates (e.g., gradients or parameters) are exchanged [31]. *Benefits:* Its main advantage is the preservation of local data privacy [11]. It also reduces the need to transmit large volumes of data to a central server and enables model training with distributed datasets that cannot be centralised for legal or privacy-related reasons [31]. It is considered ideal for the distributed and privacy-sensitive environment of 6G [32]. *Challenges:* FL faces several challenges, such as statistical heterogeneity (non-IID data across clients), system heterogeneity (differences in computing capacity, connectivity and availability), communication overhead (transmitting updates of large models can be costly), scalability to large numbers of devices, and security issues (susceptibility to poisoning attacks or inference over shared updates). Efficient and secure aggregation is also a key difficulty [12].

The development of FL is not coincidental but a direct consequence of operating at the edge. The impossibility of centralising massive and private datasets, together with bandwidth limitations [14], made traditional centralised approaches impractical. FL emerged as an architectural and algorithmic solution specifically designed to overcome these constraints imposed by the distributed and privacy-restricted edge environment [31].

2.2.2. Split Learning (SL)

In Split Learning, a neural network model is divided into two (or more) parts [30]. An initial portion (typically smaller) runs on the client device, processing raw data and generating an intermediate representation (activations or "smash data"). This representation is transmitted to an edge server, which runs the remaining (typically larger and computationally intensive) part of the model to complete the forward inference and/or backward training pass [33]. *Benefits:* The main advantage of SL is the significant reduction in computational load on the client device compared with FL, since most computation is performed on the server [32]. This makes it more suitable for devices with very limited resources. SL also offers privacy benefits, as raw data never leave the client device, and intermediate activations are generally less informative than full gradients or raw data [33]. It aligns well with the 6G vision of leveraging dispersed resources [32]. *Challenges:* SL introduces additional latency due to the bidirectional communication required per sample or batch during training (activations to the server, gradients back to the client) [33]. Deciding where to split the model optimally is a complex problem. Moreover, although SL enhances privacy compared with full centralisation, exchanged activations and gradients may still leak information, requiring additional protection mechanisms [33].

As with FL, SL is a response to edge limitations. When FL becomes computationally infeasible for extremely resource-constrained devices, SL offers an alternative by offloading most computation to the server, addressing the client-side resource bottleneck directly [32].

2.2.3. Edge Large AI Models (Edge LAMs)

Edge LAMs refer to the adaptation, deployment and execution of large pre-trained AI models (PFMs), such as GPT-type large language models or multimodal/vision models, within the 6G edge infrastructure [28]. *Potential:* These models offer unprecedented capabilities in generalisation, knowledge transfer and handling complex, diverse tasks (cross-modal reasoning, few-/zero-shot learning) [28]. They have enormous potential to revolutionise 6G network management (intelligent orchestration, semantic air-interface optimisation) and enable highly personalised and interactive services (conversational agents, advanced virtual assistants) [28]. *"Giant" Challenges:* The main obstacle is the stark mismatch between the enormous requirements of LAMs (billions of parameters,

 doi:10.20944/preprints202602.1019.v1

terabytes of training data, massive computation) and the severely limited resources available at the edge (compute, memory, storage, energy, bandwidth) [28]. Traditional Edge AI approaches such as FL or SL may be insufficient or infeasible for full LAMs [12].

This tension is driving research into extreme optimisation techniques, such as:

- Decomposition and Distributed Deployment: Dividing the LAM into smaller modules and distributing them across devices, edge nodes and the cloud [12].
- Efficient Fine-Tuning: Techniques such as Parameter-Efficient Fine-Tuning (PEFT), which update only a small fraction of a LAM's parameters [12], as well as distributed adaptations such as FedFT or Split FedFT [30].
- Compression and Quantisation: Reducing model size and computational precision to lower memory and compute requirements [33].
- Collaborative and Efficient Inference: Architectures in which multiple nodes collaborate to perform inference by sharing parameters or intermediate outputs [12], while also avoiding redundant computation in microservice-based deployments [12].

The emergence of Edge LAMs marks a critical inflection point. The ambition to harness the transformative power of large AI models directly conflicts with the practical constraints of edge environments. Addressing this tension requires fundamental innovations not only in AI model efficiency but also in the design of edge-computing and network architectures—potentially redefining both fields.

Other Relevant Techniques Include:

- Tiny Machine Learning (TinyML): Approaches designed to execute machine-learning models on devices with extremely limited resources, such as microcontrollers [11].
- Over-the-Air Computation (AirComp): A technique that exploits the superposition properties of the wireless channel to perform aggregations (such as those required in FL) directly over the air, thereby reducing latency and spectrum usage [30].
- Split Inference: Similar to Split Learning (SL) but applied solely to the inference phase rather than training [34]. It is useful for accelerating the inference of complex models on resource-constrained devices.

2.2.4. AI Applications for 6G Network Optimisation at the Edge

Edge AI not only enables new end-user services but also serves as a powerful tool for optimising the 6G network itself in an intelligent and adaptive manner:

- Intelligent Resource Management: AI can address the complex joint resource-optimisation problems in 6G (spectrum, power and channel allocation, user association, beamforming, network slicing), many of which are NP-hard [35]. ML/DL/RL algorithms can learn from network data and make real-time decisions to maximise efficiency (spectral and energy), capacity, fairness and QoS, while adapting to dynamic environmental conditions [3].
- Air-Interface Optimisation: AI can significantly enhance physical-layer performance. Examples include accurate channel state information (CSI) prediction, intelligent beam management in massive MIMO systems, and the development of adaptive modulation and coding schemes [9]. An emerging area is semantic communication, where AI extracts and transmits only the relevant (semantic) information rather than raw bits, thereby improving efficiency and robustness [36]. Edge LAMs are also proposed for intelligent air-interface design and optimisation [12].
- Intelligent Mobility and Handover Management: AI algorithms can predict user mobility patterns and optimise handover decisions between cells or edge nodes, minimising service interruptions and ensuring seamless QoE [37].

 doi:10.20944/preprints202602.1019.v1

- Efficient Operation and Maintenance (O&M): AI enables advanced Self-Organising Network (SON) capabilities, including proactive fault detection, predictive maintenance based on data analytics, self-configuration and self-optimisation of network parameters, and autonomous fault recovery [37]. This reduces the need for human intervention and lowers operational costs.
- AI-Enhanced Security: Edge AI can strengthen 6G network security through intelligent real-time intrusion and anomaly detection, behavioural analysis for threat identification and automated response mechanisms [38]. It can also be used to reinforce physical-layer security (PLS) approaches [34].

2.2.5. Benefits of Edge AI for 6G Services

The integration of AI at the edge enables a new generation of services and significantly enhances the user experience:

- Enablement of Real-Time Intelligent Services: Edge AI is essential for applications requiring low latency and on-site intelligent decision-making. This includes autonomous driving (environment perception, manoeuvre planning) [11], collaborative industrial robotics [7], truly interactive and personalised XR and metaverse experiences [11], advanced virtual assistants and context-aware personalised services [12].
- Improved Efficiency (Energy, Spectrum): Intelligent optimisation of network resources (radio and computing) and reduced data traffic towards the cloud contribute to greater energy and spectral efficiency, both of which are critical for the sustainability of 6G [6].
- Enhanced Quality of Experience (QoE): By reducing latency, increasing reliability and enabling personalised services, Edge AI directly improves the quality of experience for end users [4].
- New Network Capabilities: 6G may provide Artificial Intelligence as a Service (AIaaS) directly from the network [4]. Furthermore, synergy with integrated sensing opens the door to Integrated Sensing Edge Intelligence (ISEA), where the network not only communicates and computes but also intelligently perceives the environment [30].

2.2.6. Challenges of Edge AI in the 6G Environment

Despite its enormous potential, the effective deployment of Edge AI in 6G faces several major challenges:

- Computational Requirements vs. Limited Resources: This remains the primary challenge. AI models—particularly advanced ones such as LAMs—require substantial computational resources (measured in FLOPs), memory and energy, often exceeding the capabilities of edge devices and servers [28]. A holistic optimisation strategy is needed across data (cleaning, compression, augmentation), models (compression, quantisation, pruning, efficient architectures such as NAS) and systems (specialised hardware such as NPUs/TPUs, efficient resource allocation) [12].
- Data and Model Privacy and Security: Although techniques such as FL and SL aim to preserve privacy, they are not immune to sophisticated attacks. There is a risk of inferring sensitive information from model updates (gradients) or intermediate activations [32]. Models may also be vulnerable to poisoning attacks (manipulated training data) or adversarial attacks (inputs crafted to deceive the model during inference) [34]. Ensuring privacy and security in distributed AI environments requires integrating advanced cryptographic techniques such as Differential Privacy (DP), Homomorphic Encryption (HE) or Secure Multiparty Computation (SMC) [39], together with robust access control and auditing mechanisms [11].
- Communication Overhead: Distributed training (FL/SL) and collaborative inference involve frequent exchanges of information (parameters, gradients, activations) between nodes, which can impose significant overhead on the wireless network, consuming bandwidth and energy [12]. Efficient model/update-compression techniques and optimised communication strategies (e.g., AirComp) are required [30].

 doi:10.20944/preprints202602.1019.v1

- Robustness and Generalisation: AI models must operate reliably in the 6G wireless environment, which is inherently dynamic, noisy and error-prone [12]. They must be robust to channel variations, mobility and interference. Moreover, especially for LAMs, they must generalise well across different tasks, scenarios and domains with minimal re-adaptation [37].
- Data Collection and Quality: AI critically depends on the availability of large volumes of high-quality training data ("Garbage in, Garbage out" [15]). Collecting, labelling and managing such data in a distributed and heterogeneous environment like the 6G edge poses both logistical and cost challenges [28]. Incomplete, noisy or biased data can lead to unreliable or unfair models [40]. The use of synthetic data generated by Digital Twins is a promising avenue to mitigate this issue [28].
- Ethics, Transparency and Explainability: As AI becomes responsible for increasingly critical decisions in 6G networks and services, ethical concerns intensify. These include potential algorithmic biases, lack of transparency in decision-making (the "black-box" problem) and the need for explainability, particularly in sensitive applications [41].

Managing AI in 6G will likely require a hierarchy and coexistence of different types of models. While large foundational models (PFMs/LLMs) may reside in more capable nodes (edge servers or regional clouds) to perform complex tasks such as orchestration, semantic understanding or planning [28], smaller, specialised and efficient models (task-oriented AI, TinyML) will be deployed on end devices or highly resource-constrained edge nodes for specific, low-latency tasks [28]. The intelligent orchestration of this hierarchical and heterogeneous collaboration across the Cloud–Edge–Device continuum will be a key challenge, but also an opportunity to leverage the strengths of each approach [28]. Table 2 summarises and compares the key Edge AI techniques discussed.

Table 2. Summary of Key Edge AI Techniques for 6G

Technique	Brief Description	Main Advantage for 6G Edge	Main Challenge for 6G Edge
Federated Learning (FL)	**Collaborative training on local devices without sharing raw data; only updates are exchanged [31].**	**Preservation of local data privacy; use of distributed data [31].**	**Heterogeneity (data/systems); communication overhead; security of updates [36].**
Split Learning (SL)	**Neural model split between client and server; client processes early layers, server processes the rest [35].**	**Reduces computational load on the client compared with FL; privacy of raw data [31].**	**Latency due to bidirectional communication; difficulty in deciding split point; privacy of intermediate activations [35].**
Edge Large AI Models (LAMs)	**Adaptation and deployment of large models (LLMs, PFMs) at the edge [8].**	**Generalised capabilities; multimodal processing; few-/zero-shot learning [8].**	**Massive compute/memory/data demands vs. limited edge resources [36].**

 doi:10.20944/preprints202602.1019.v1

Technique	Brief Description	Main Advantage for 6G Edge	Main Challenge for 6G Edge
TinyML	Execution of ML on devices with extremely limited resources (microcontrollers) [33].	Enables intelligence on very small/low-power devices.	Very limited model capacity; potentially reduced accuracy.
AirComp	Aggregation of signals (e.g., FL gradients) leveraging the wireless channel [10].	Reduces latency and spectrum use for distributed aggregation.	Sensitivity to channel noise; requires precise synchronisation.

3. Architectural and Orchestration Proposals for 6G with Distributed AI

Realising the 6G vision—with its emphasis on ubiquitous intelligence and distributed computing—requires a significant evolution, and in some cases a reinvention, of the network architecture. Current proposals explore how to structure and manage these complex networks in order to integrate edge computing and artificial intelligence as native capabilities, proposed intelligent distributed architecture and orchestration is shown in Figure 2.

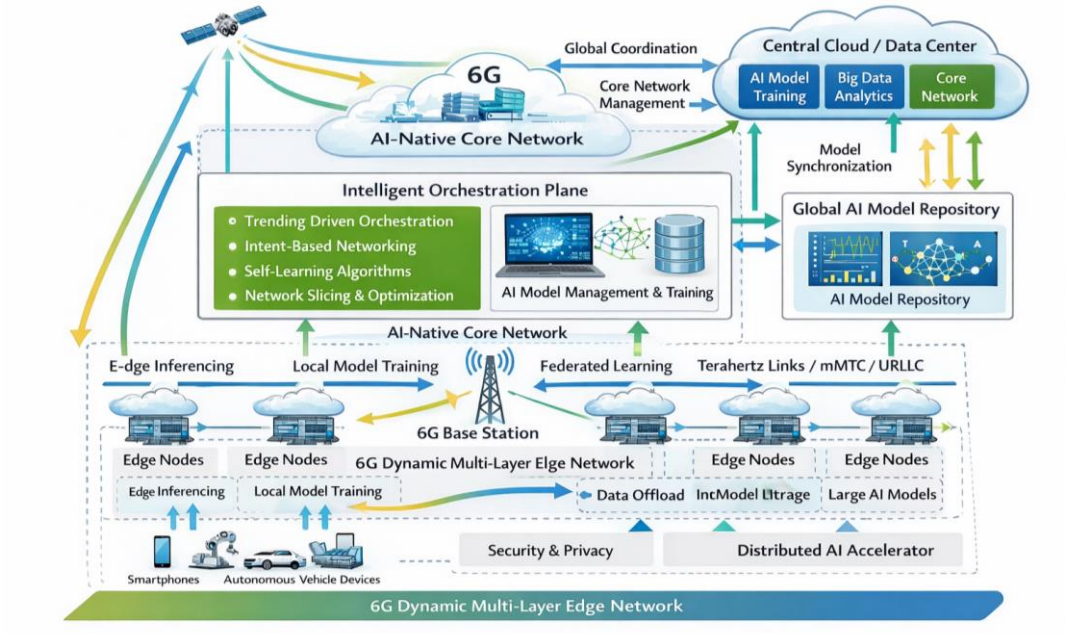

Figure 2. Proposed intelligent distributed architecture and orchestration

3.1. *Evolution of Network Architecture Towards 6G*

The 6G architecture will not simply be an extrapolation of 5G; rather, it will incorporate new design principles and foundational capabilities:

3.1.1. Design Principles

The aim is to develop an architecture that is modular, supports the agile introduction of new functionalities, is simple and sustainable in the long term, intrinsically trustworthy, cloud-native, facilitates a seamless migration from 5G and, above all, enables innovation [42].

3.1.2. Multidimensional Integration

 doi:10.20944/preprints202602.1019.v1

A defining characteristic of 6G will be the seamless integration of different communication domains: terrestrial networks; non-terrestrial networks (NTN)—including satellites (LEO, MEO, GEO), high-altitude platforms (HAPS) and unmanned aerial vehicles (UAVs); and even aerial and underwater communication systems [7]. The objective is to achieve truly global and ubiquitous coverage.

3.1.3. Horizontalisation and Disaggregation

The trend towards separating software from hardware, initiated in 5G, will deepen in 6G. Increased adoption of virtualisation (NFV), software-defined networking (SDN), service-based architectures (SBA), cloud-native principles and open interfaces (such as those promoted by O-RAN) is expected across all network domains (RAN, Core, Transport) [1]. This aims to enhance flexibility, efficiency, automation and the overall capacity for innovation.

3.1.4. AI-Native Design and Distributed Computing

The 6G architecture must be designed from the outset to natively support artificial intelligence and distributed computing [9]. This not only involves enabling AI to optimise the network, but also allowing the network itself to operate as a platform for executing and orchestrating distributed AI functions and edge-computing services. Integration with sensing (ISAC) is also an emerging architectural requirement [10].

3.2. Proposed Reference Architectures

Given the early stage of 6G development, several architectural proposals have been put forward, each reflecting different approaches and priorities:

3.2.1. Evolutionary Vision

This proposal advocates a single-step migration from 5G Standalone (SA) to 6G SA, evolving the 5G Core (5GC) based on the Service-Based Architecture (SBA) rather than creating an entirely new core [42]. In the RAN domain, it supports native lower-layer split (LLS) functionality to optimise performance and enable intent-based automation through Service Management and Orchestration (SMO) functions [42]. This approach prioritises the reuse of existing investments and a smoother transition path.

3.2.2. Task-Oriented Native AI Architecture (TONA)

This is a more revolutionary proposal, suggesting a fundamental shift from the traditional communication "session" management paradigm to an AI "task"-centric paradigm [4]. TONA explicitly manages multidimensional resources—communication, computing, data and AI models—and employs a task-based control plane to orchestrate multi-node collaborations and guarantee customised AI Quality of Service (QoAIS). It promises significant advantages in latency, efficiency and privacy for AI services integrated within the network [4].

3.2.3. Integrated Satellite–Terrestrial Architectures (SAGIN):

Reference architectures have been proposed for integrating 6G terrestrial and non-terrestrial networks (NTN), evolving from non-virtualised schemes to fully virtualised architectures with a unified management and orchestration (MANO) system across both segments [43]. These architectures are crucial for enabling global coverage and supporting use cases in remote or high-mobility environments (maritime, aerial).

3.2.4. O-RAN-Based Architectures with AI

These architectures leverage the open and disaggregated design of O-RAN, particularly the RAN Intelligent Controller (RIC) in its Near-Real-Time and Non-Real-Time variants, to introduce data-driven intelligence and optimisation within the access network [44]. Recent proposals explore the use of intelligent agents, including LLM-based agents, implemented as xApps or rApps on top of the RIC to perform automated and intuitive orchestration of Edge AI services and network adaptation [44].

3.2.5. Other Specific Proposals

 doi:10.20944/preprints202602.1019.v1

Additional approaches focus on particular dimensions, such as data-oriented architectures (e.g., Data Plane for Data as a Service – DaaS) to facilitate data management and processing for distributed AI [40]; architectures designed for the seamless integration of sensing and Edge AI (ISEA) [45]; or architectures optimised to support the unique demands of the Metaverse at the edge [46].

This diversity of architectural proposals reflects an active exploration phase across both industry and academia. While some approaches prioritise continuity and pragmatic evolution from 5G, others advocate more radical shifts to integrate intelligence and distribution natively from the outset. This early lack of consensus, although fostering innovation, also introduces the risk of fragmentation or the recurrence of complexities observed in 5G, such as multiple competing architectural options (NSA vs. SA) [42]. It is likely that the standardisation process will seek convergence, although the final direction has yet to be defined.

3.3. Intelligent Orchestration of Services and Resources

Regardless of the specific architecture adopted, the inherent complexity of 6G networks—with their distributed, heterogeneous, virtualised and highly dynamic nature—makes intelligent orchestration an absolutely indispensable component [47]. Manual or static rule-based management is infeasible [41]. Orchestration acts as the "central nervous system" of the network, enabling coherent, adaptive and efficient operation. Key aspects of intelligent orchestration in 6G include:

3.3.1. Joint Communication, Computing and Storage Management (JCC)

It is essential to optimise communication resources (bandwidth, spectrum), computing resources (CPU, GPU, NPU at edge nodes) and storage along the entire network continuum—from devices to the cloud, via the edge [1]. This is particularly critical in integrated networks such as SAGIN [48]. Since these joint-optimisation problems are often complex (NP-hard), advanced heuristics or, increasingly, AI-based approaches (RL, DL) are required to obtain efficient real-time solutions [35].

3.3.2. Intent-Based Networking (IBN)

IBN represents a higher level of abstraction in network management [49]. Instead of configuring low-level parameters, users or applications declare high-level goals or requirements as "intentions" (e.g., "ensure < 5 ms latency for VR traffic", "deploy object-detection service at the edge near camera X") [44]. The intelligent orchestration system (NMS/Orchestrator) is responsible for interpreting the intention, translating it into specific network and resource configurations, activating those configurations and continuously ensuring intention fulfilment [42]. IBN architectures for 6G are being actively explored [50], including the use of LLMs to interpret natural-language intents and automate the entire intent lifecycle [44].

3.3.3. Orchestration of AI Functions at the Edge

Beyond network-resource orchestration, 6G requires the specific orchestration of AI functions and models deployed at the edge [28]. Tasks include selecting the appropriate AI model for a given purpose, deciding where to place it (which edge node or device), deploying it (e.g., as a container), allocating the necessary computing resources, monitoring its performance and managing its lifecycle (updates, retraining). Intelligent-agent-based frameworks (potentially LLM-driven) are being proposed to automate these complex decisions, considering factors such as resource availability, QoS requirements and latency constraints [28].

3.3.4. Multi-domain and Multi-agent Federation and Collaboration

Orchestration in 6G must operate across multiple administrative domains (e.g., different operators in roaming or network-sharing scenarios) and technological domains (e.g., coordination between edge and cloud, or between RAN and Core) [20]. This requires federation mechanisms that enable controlled and secure sharing of resources and information. Moreover, in architectures with multiple distributed AI agents, orchestration must also facilitate effective collaboration among them [28].

 doi:10.20944/preprints202602.1019.v1

3.4. *Integration with Key Enabling Technologies*

The architecture and orchestration of 6G do not exist in isolation; rather, they must be synergistically integrated with other key technologies to achieve their full potential:

3.4.1. Digital Twins (DT)

A Digital Twin is a dynamic and real-time virtual representation of a physical object, process or system [23]. In 6G, DTs may be created for the network itself, its components or the environment in which it operates.

- Roles: DTs enable scenario simulations ("what-if" analyses), testing of configurations or optimisations before deploying them in the real network, real-time monitoring and behavioural prediction of the network [10], and the generation of realistic synthetic data for training AI models—overcoming limitations of real data [28]. They can also support industrial, logistics or smart-city applications [23], and enhance security via threat modelling [38]. Architectures combining DT with blockchain and FL have been proposed to achieve secure and efficient edge networks [51].
- Implementation: DTs require distributed IoT–Edge–Cloud platforms to collect data from the physical world and maintain synchronisation with the virtual replica [23].

3.4.2. Integrated Sensing and Communication (ISAC)

ISAC leverages existing communication signals and infrastructure to perform environmental sensing tasks such as object detection, distance/velocity/angle estimation, localisation and mapping [1].

- Synergy with Edge AI: ISAC generates substantial sensing data that can be processed by Edge AI algorithms to extract useful information and support intelligent decision-making. In turn, Edge AI can optimise ISAC processes themselves. This synergy leads to the paradigm of Integrated Sensing and Edge AI (ISEA) [10], in which communication, edge computing, sensing and AI are jointly designed and optimised for a given task.
- Challenges: The main difficulty lies in achieving true integration at the hardware, algorithmic and signal-design levels, rather than mere coexistence [10].

3.4.3. Blockchain

Distributed ledger technology can play a crucial role in 6G by addressing security, privacy and trust issues in distributed and multi-stakeholder environments [52].

- Applications: Blockchain is proposed for secure and decentralised identity and access management, secure sharing of spectrum and resources, data and transaction traceability and auditing, and establishing trust in distributed AI systems such as FL or in interactions with DTs [41].

The integration of these technologies creates synergistic feedback loops that enhance the overall intelligence and efficiency of the 6G system. ISAC functions as the network's "sensory system", capturing data from the physical world [10]. Edge AI acts as the distributed "brain", analysing these and other data to understand the environment, optimise the network and make decisions [45]. The Digital Twin provides a "virtual space for testing and prediction", fuelled by ISAC data and used by Edge AI to simulate scenarios, validate strategies and anticipate behaviours before acting in the real world [23]. This continuous cycle of *Sense (ISAC)* → *Analyse/Decide (Edge AI)* → *Simulate/Validate (DT)* → *Act (Network/Control)* enables far more robust, adaptive and intelligent cyber-physical systems, in which each component reinforces the capabilities of the others. Table 3 compares several of the key architectural proposals discussed.

Table 3. Comparative Analysis of Architectural and Orchestration Proposals for 6G with Distributed AI

 doi:10.20944/preprints202602.1019.v1

Proposal / Origin	Main Approach	Role of Distributed Computing / Edge AI	Associated Key Technologies	Potential Strengths / Weaknesses
Ericsson 5GC Evolved [5]	Pragmatic evolution of the 5G Core; RAN with native LLS	Edge AI for automation (SMO), MEC supported by the evolved Core	5GC SBA, LLS, IBN, SMO	(+) Smooth migration, investment reuse. (–) Potentially less optimised for AI-native designs.
TONA [7]	AI task-oriented network management	Central: AI is the task to be managed. Edge AI/MEC treated as native resources orchestrated by the network	Task-Control, QoAIS, Multi-Resource Management	(+) AI-native optimisation, fine granularity. (–) Breaks current paradigms; high complexity.
O-RAN AI-enabled [45]	Intelligence and openness in the RAN	Edge AI implemented as xApps/rApps in the RIC for RAN optimisation and edge-service orchestration	O-RAN, RIC, xApps/rApps, LLM Agents	(+) RAN flexibility, open innovation. (–) RAN-centric scope; O-RAN complexity; interface security.
Integrated SAGIN [28]	Terrestrial–Non-Terrestrial Integration (NTN)	MEC/Edge Computing deployed on NTN platforms (satellites, HAPS) for global intelligent-service coverage	NTN, Virtualisation (NFV), Integrated MANO	(+) Ubiquitous coverage, new use cases. (–) NTN latency; MANO integration complexity.
Data Plane for DaaS [40]	Data-centric architecture for AI	Facilitates collection, transmission, processing and storage of data for distributed AI	Data Plane, DaaS APIs	(+) Optimised for AI data pipelines. (–) Less focus on compute/communication optimisation.
ISEA Framework [17]	Deep integration of Sensing and Edge AI	Co-design of communication, edge computing, sensing and AI for specific tasks	ISAC, Edge AI, Joint Optimisation	(+) Optimal performance for intelligent sensing tasks. (–) Task specificity; high complexity.

4. Standardisation and Future Perspectives

The successful development and deployment of 6G—with its deep integration of distributed computing and Edge AI—depend heavily on global standardisation efforts and the resolution of substantial research challenges.

4.1. Status of Standardisation

Several international organisations are actively working on defining 6G:

4.1.1. ITU-R (IMT-2030)

The ITU establishes the global vision and requirements for future generations of IMT.

- Process: Following the established processes for IMT-2000, IMT-Advanced and IMT-2020, work on IMT-2030 (6G) began with the definition of the vision and overarching framework, culminating in Recommendation ITU-R M.2160 ("IMT-2030 Framework"), approved in November 2023 [6]. The next phase (2024–2027) will focus on defining detailed technical requirements and evaluation criteria for candidate Radio Interface Technologies (RITs). RIT submissions will be accepted between 2027 and early 2029, with final specifications expected around 2030 [6].

 doi:10.20944/preprints202602.1019.v1

- Vision and Capabilities: The IMT-2030 framework identifies key use scenarios, including enhanced versions of those in 5G (immersive communication, massive communication, HRLLC) and new scenarios such as ISAC, AI integration and ubiquitous connectivity [5]. It defines 15 target capabilities, with significant improvements over 5G in data rate, latency, density and more, while introducing new capabilities such as centimetre-level positioning, integrated sensing and, crucially, AI-related capabilities [5].
  These AI-related capabilities explicitly include support for distributed data processing, distributed learning (such as FL), execution of AI models and inference within the network [6]. Sustainability is also identified as a key pillar [6].
- Spectrum: ITU-R recognises the need for additional spectrum across a wide range of bands, from sub-1 GHz (for broad coverage) to millimetre-wave bands and, potentially, frequencies above 92 GHz or even into the THz range (to support extreme data rates and applications such as ISAC) [2]. Spectrum harmonisation is crucial [53].

4.1.2. 3GPP (Releases 19, 20 and 21)

3GPP is the main organisation responsible for developing technical specifications for mobile networks and will play a central role in defining the technical foundations of 6G.

- Timeline: Formal work on 6G within 3GPP began in Release 19 (main work through ~Q3 2025) with studies on use cases and service requirements (led by SA1, resulting in TR 22.870) [54]. Release 20 (expected Q3 2025 to ~Q1 2027) will host the major technical studies in the RAN working groups, investigating candidate technologies and defining detailed technical requirements for the 6G radio interface [10]. Normative work (specification development) for the first version of 6G will take place in Release 21, with the goal of completion by late 2028 or early 2029 to align with the IMT-2030 submission schedule [10]. It is important to note that Releases 19 and 20 will also continue advancing 5G-Advanced in parallel with 6G studies [54].
- Focus on AI/ML: 3GPP has been working on AI/ML since earlier releases (e.g., Release 18, which studied AI/ML for the air interface) [55]. In Release 19, this work continues through the specification of use cases such as CSI prediction, beam management and positioning [55]. The general 3GPP approach is not to standardise AI algorithms or models themselves, given their rapid rate of evolution [56]. Instead, the goal is to standardise the infrastructure and interfaces required to support and manage AI/ML in the network. This includes mechanisms for operator-controlled data collection, model transfer and lifecycle management (activation, deactivation, performance monitoring), as well as the necessary air-interface extensions [38]. AI/ML is expected to be an integral part of 6G specifications from the outset (Release 21) [56], with potential to support online learning and greater flexibility in implementing specific functionalities [56]. Functions such as NWDAF (Network Data Analytics Function), introduced in 5G, are seen as foundational components for AI-driven analytics and automation in 6G [38].

4.1.3. ETSI MEC

The ETSI ISG MEC continues to develop standards for multi-access edge computing.

- Evolution: Recent phases (Phase 3 and 4) focus on improving security, enabling federation across different MEC platforms (critical for roaming and multi-operator scenarios), supporting network slicing at both MEC and application levels, and aligning with the needs of vertical industries (e.g., V2X through 5GAA) [20].
  There is ongoing alignment with 3GPP, particularly with SA6's work on edge-application architectures (EDGEAPP) [20].
- Relevance for 6G: The work of ETSI MEC is fundamental in providing a standardised environment that enables the deployment of edge-native applications in 6G [20]. However, some perspectives highlight that the current MEC architecture may need to evolve towards more open and flexible approaches (such as OS-MEC) to fully meet the dynamism and personalisation requirements of 6G [22].

 doi:10.20944/preprints202602.1019.v1

There is an inherent tension in the standardisation process. On the one hand, early definition of a framework and requirements (as IMT-2030 does) is essential to guide research and investment [57]. On the other hand, there is a risk of standardising specific technologies or architectures before they are fully mature or before their implications are fully understood, which could limit future innovation or lead to suboptimal solutions [42]. The experience with multiple initial 5G deployment options (NSA/SA) suggests caution [42]. The phased approach adopted by ITU-R and 3GPP (*Vision → Requirements → Technical Study → Normative Specification*) [6] seeks to balance this tension by allowing research to inform standardisation, although the challenge of making key decisions at the right moment persists.

A notable aspect is the strategic decision by standardisation bodies—particularly 3GPP—to focus on enabling the infrastructure for AI (data collection, model management, interfaces) rather than standardising AI models themselves [38]. This recognises the extremely dynamic nature of AI research and aims to create a standardised yet open ecosystem in which different vendors can innovate and compete with their own AI solutions on top of a common, interoperable foundation. In other words, standardisation defines how to integrate and manage AI, not which AI should be used.

4.2. *Open Challenges and Future Research Directions*

Despite progress in defining the 6G vision and initial standardisation, numerous areas require intensive research to make 6G with distributed AI a practical reality:

4.2.1. Efficient and Scalable Edge Resource Optimisation

Developing algorithms capable of jointly managing heterogeneous resources (communication, computing, storage, energy) in large-scale, dynamic and resource-constrained edge environments in real time [35]. This includes optimisation specifically tailored for the deployment and execution of Edge LAMs [12].

4.2.2. Trustworthy Distributed AI

Enhancing the robustness, efficiency, fairness and—critically—the privacy and security of distributed learning paradigms such as FL and SL [32]. Research is needed in areas such as asynchronous FL, handling non-IID data, label-privacy preservation in SL and effective integration of DP, HE, SMC and ZKP in 6G edge environments [32]. Developing frameworks resilient to poisoning and adversarial attacks, specifically tailored to distributed AI, is also essential [34].

4.2.3. AI-Native Architectures

Defining and validating 6G network architectures that integrate AI as a foundational component rather than an overlay [4]. This includes exploring task-oriented architectures (TONA), intelligent-agent-based designs (PFM/LLM) and architectures supporting intent-based end-to-end orchestration [47].

4.2.4. Holistic Integration and Co-design

Advancing the synergistic integration of communication, computing, sensing and AI [45]. Co-design approaches are needed to jointly optimise, for example, data/model compression, task partitioning and communication protocols in scenarios such as Semantic Edge Computing (SEC) or Semantic Communications (SemCom) [33].

4.2.5. Holistic Security and Privacy

Addressing end-to-end security across all layers: physical layer security (PLS [34]), network-level security (distributed infrastructure, open interfaces [24]), application-layer protection and—increasingly—the security of distributed AI systems themselves (data privacy, model integrity) [30].

4.2.6. Sustainability and Energy Efficiency

 doi:10.20944/preprints202602.1019.v1

Investigating and developing techniques to minimise the energy consumption of both the 6G network infrastructure and distributed AI processes (training and inference), which can be highly demanding [38].

4.2.7. Standardisation and Ecosystem Development

Achieving global consensus on key technical standards for 6G is essential to avoid market fragmentation [57]. Encouraging open software- and hardware-based ecosystems can accelerate innovation and adoption [38].

It is essential to recognise that these open challenges are not isolated problems. There is deep interdependence among them. For example, resource optimisation [35] depends on trustworthy AI algorithms to make decisions [37], yet trustworthy AI (such as FL/SL) itself requires efficient resource management (communication and computing) to operate effectively [32]. Both, in turn, rely on AI-native architectures that can support them adequately [4]. Security must protect both the network infrastructure and the components of distributed AI [24]. Sustainability is a cross-cutting constraint influencing architectural choices, AI algorithm design and resource-management strategies [6]. Thus, advancing towards 6G requires a systemic and multidisciplinary approach that addresses these challenges in a coordinated and holistic manner. Table 4 summarises the current state of standardisation across key organisations.

Table 4. Summary of Ongoing 6G Standardisation Efforts Across Key Organisations

Organisation	Relevant Group / Initiative	Specific Focus	Current Milestone / Status	Key Upcoming Steps
ITU-R	WP 5D / IMT-2030	Global vision, use scenarios, capabilities (incl. AI, ISAC), spectrum [14]	Rec. M.2160 (Framework) published (Nov 2023) [14]	Definition of technical requirements and evaluation criteria (2024–2027) [14]
3GPP	TSG SA (SA1), TSG RAN (RAN1/2/3/4)	6G requirements (Rel-19), 6G technical studies (Rel-20), 6G specs (Rel-21) [55]	Rel-19: Initial 6G studies in progress [56]	Rel-20: Main technical studies (starting Q3 2025) [55]; Rel-21: Specs (starting ~2027) [55]
3GPP	Various WGs	Infrastructure for AI/ML (data collection, model management) [57]	Ongoing work in Rel-19/20 [57]	Native integration in Rel-21 specifications [57]
ETSI	ISG MEC	MEC architecture, APIs, federation, slicing, vertical-industry support [48]	Phase 3 completed, Phase 4 ongoing [48]	Continued evolution to support 6G; alignment with 3GPP SA6 [48]

5. Conclusions

The transition towards 6G networks marks a defining inflection point in the evolution of mobile communications, driven by the vision of a hyperconnected and intelligently integrated world. This article has examined how distributed computing and edge artificial intelligence (Edge AI) emerge as fundamental and mutually interdependent technologies for realising this vision.

 doi:10.20944/preprints202602.1019.v1

5.1. *Recapitulation of Key Proposals*

The analysis has highlighted the unavoidable need to migrate from centralised computing paradigms to distributed architectures, with Multi-access Edge Computing (MEC) acting as a key enabler for bringing computing and storage capabilities closer to the end user. This proximity is essential for meeting the stringent latency, bandwidth and privacy requirements demanded by the most transformative 6G applications.

In parallel, Edge AI consolidates itself as the intelligence engine of these distributed networks. Techniques such as Federated Learning (FL) and Split Learning (SL) have emerged as direct responses to the resource constraints and privacy concerns at the edge, enabling collaborative training of AI models.

The rise of Large AI Models (LAMs) at the edge represents a significant challenge but also an unprecedented opportunity, driving research on extreme optimisation and distributed inference architectures. Edge AI not only enables intelligent services for end users but is also crucial for autonomous optimisation and efficient management of the inherent complexity of 6G networks.

The diversity of architectural proposals—evolutionary approaches such as Ericsson's, revolutionary ones such as TONA, open architectures such as O-RAN-based designs, and integrated approaches such as SAGIN and ISEA—reflects the intense exploratory phase in which the field currently finds itself. Nevertheless, all converge on the need for intelligent and automated orchestration, with approaches such as Intent-Based Networking (IBN) emerging as promising solutions to abstract and manage the complex joint optimisation of communication, computing and storage resources in this distributed environment.

5.2. *Transformative Potential*

The synergy between distributed computing and Edge AI has the potential to unlock truly revolutionary capabilities. By integrating intelligence natively into the network infrastructure, 6G can move beyond mere data transmission to become a platform for ubiquitous intelligence. This will enable genuinely immersive experiences (XR, the metaverse, holographic communication), advanced automation (Industry 4.0, cooperative autonomous driving), personalised and context-aware services and new capabilities such as integrated sensing and the provision of Artificial Intelligence as a Service (AIaaS) directly from the network. The expected benefits include dramatic improvements in performance (latency, capacity), efficiency (energy and spectrum) and user experience quality.

5.3. *Need for Continued Research and Standardisation*

Despite the enthusiasm and significant progress achieved, the full realisation of the 6G vision with distributed AI faces formidable challenges. Resource limitations at the edge, ensuring security and privacy in complex distributed systems, managing communication overhead, guaranteeing robustness and reliability of AI models, securing data quality and availability and addressing ethical and sustainability considerations all demand ongoing research and innovative solutions.

The path to 6G requires a concerted and collaborative global effort. Standardisation within organisations such as ITU-R and 3GPP is essential to ensure interoperability and to create a unified global market; however, it must balance the need for early direction with the flexibility to incorporate emerging research advances. Fostering open ecosystems and multidisciplinary research will be key to overcoming the remaining challenges and harnessing the full transformative potential offered by distributed computing and Edge AI for the next generation of mobile networks.

Author Contributions: Conceptualization, E.A.H. and H.F.B.-O.; methodology, N.C.R.-I.; software, H.F.B.-O.; validation, E.A.H., H.F.B.-O. and N.C.R.-I.; formal analysis, E.A.H.; investigation, E.A.H. and H.F.B.-O.; resources, N.C.R.-I.; data curation, N.C.R.-I.; writing—original draft preparation, E.A.H.; writing—review and editing, H.F.B.-O. and N.C.R.-I.; visualization, H.F.B.-O.; supervision, E.A.H. and N.C.R.-I; project

 doi:10.20944/preprints202602.1019.v1

administration, E.A.H.; funding acquisition, H.F.B.-O and N.C.R.-I. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding, and the APC was funded by University of Quindío [100016837].

Data Availability Statement: Not applicable. This study is based on publicly available and cited literature; no new data or supplementary materials were generated or deposited.

Acknowledgments: The authors would like to acknowledge the support of the Telecommunications Research Group (GITUQ) for its contribution to the development and technical discussion of this study.

Conflicts of Interest: The authors declare no conflict of interest

References

1. H. Pennanen, T. Hänninen, O. Tervo, A. Tölli, and M. Latva-aho. 6G: The Intelligent Network of Everything. IEEE Access, 2024. https://doi.org/10.1109/ACCESS.2024.3521579
2. S. Chen, Y.-C. Liang, S. Sun, S. Kang, W. Cheng, and M. Peng. Vision, Requirements, and Technology Trend of 6G: How to Tackle the Challenges of System Coverage, Capacity, User Data-Rate and Movement Speed. arXiv, 2020. http://arxiv.org/abs/2002.04929
3. A. Al-Ansi, A. M. Al-Ansi, A. Muthanna, I. A. Elgendy, and A. Koucheryavy. Survey on Intelligence Edge Computing in 6G: Characteristics, Challenges, Potential Use Cases, and Market Drivers. Future Internet, 2021, 13(5): 118. https://doi.org/10.3390/fi13050118
4. Y. Yang, X. Wang, H. Huang, J. Zhang, Z. Zhang, L. Li, and H. Ji. Task-Oriented 6G Native-AI Network Architecture. IEEE Network, 2024, 38(1): 219–227. https://doi.org/10.1109/MNET.2023.3321464
5. R. Liu, L. Zhang, R. Y.-N. Li, and M. Di Renzo. The ITU Vision and Framework for 6G: Scenarios, Capabilities and Enablers. arXiv, 2023. http://arxiv.org/abs/2305.13887
6. ITU-R. IMT Towards 2030 and Beyond (IMT-2030), 2025. https://www.itu.int/en/itu-r/study-groups/rsg5/rwp5d/imt-2030/pages/default.aspx
7. P. Li, J. Fan, and J. Wu. Exploring the key technologies and applications of 6G wireless communication network. iScience, 2025, Article 112281. https://doi.org/10.1016/j.isci.2025.112281
8. A. A. Shamsabadi, A. Yadav, Y. Gadallah, and H. Yanikomeroglu. Exploring the 6G potentials: Immersive, hyper reliable, and low-latency communication. arXiv, 2024. http://arxiv.org/abs/2407.11051
9. H. Pennanen, T. Hänninen, O. Tervo, A. Tölli, and M. Latva-aho. 6G: The intelligent network of everything – A comprehensive vision, survey, and tutorial. IEEE Access, 2024. https://doi.org/10.1109/ACCESS.2024.3521579
10. R. Liu, L. Zhang, R. Y-N. Li, and M. Di Renzo. The ITU vision and framework for 6G: Scenarios, capabilities and enablers. arXiv, 2023. http://arxiv.org/abs/2305.13887
11. IEEE WCNC. WS-11: Distributed and intelligent edge computing for 6G communications, 2023. https://wcnc2023.ieee-wcnc.org/workshop/ws-11-distributed-and-intelligent-edge-computing-6g-communications
12. Z. Wang, Y. Shi, Y. Zhou, J. Zhu, and K. B. Letaief. Edge large AI models: Revolutionizing 6G networks. arXiv, 2025. http://arxiv.org/abs/2505.00321
13. Y. Wang and J. Zhao. Mobile edge computing, metaverse, 6G wireless communications, artificial intelligence, and blockchain: Survey and their convergence. arXiv, 2022. http://arxiv.org/abs/2209.14147
14. N. A. Angel, D. Ravindran, P. M. D. R. Vincent, K. Srinivasan, and Y. C. Hu. Recent advances in evolving computing paradigms: Cloud, edge, and fog technologies. Sensors, 2022, 22(1): 196. https://doi.org/10.3390/s22010196
15. X. Wang and W. Jia. Optimizing edge AI: A comprehensive survey on data, model, and system strategies. arXiv, 2025. http://arxiv.org/abs/2501.03265v1
16. I. A. Alimi, K. M. S. Huq, F. Aurzada, L. A. DaSilva, X. Han, and M. R. K. Aziz. Trends in cloud computing paradigms: Fundamental issues, recent advances, and research directions toward 6G fog networks. In:

Moving Broadband Mobile Communications Forward – Intelligent Technologies for 5G and Beyond. IntechOpen, London, 2021. https://doi.org/10.5772/intechopen.98315

17. S. S. Gill. A manifesto for modern fog and edge computing: Vision, new paradigms, opportunities, and future directions. In: EAI/Springer Innovations in Communication and Computing. Springer, Cham, 2022: 237–253. https://doi.org/10.1007/978-3-030-74402-1_13
18. A. Biswas and H. C. Wang. Autonomous vehicles enabled by the integration of IoT, edge intelligence, 5G, and blockchain. Sensors, 2023, 23(4): 1963. https://doi.org/10.3390/s23041963
19. C.-X. Wang, J. Huang, H. Wang, X. Gao, X. You, and C. Huang. On the road to 6G: Visions, requirements, key technologies and testbeds. arXiv, 2023. http://arxiv.org/abs/2302.14536
20. ETSI. Multi-access edge computing – Standards for MEC, 2025. https://www.etsi.org/technologies/multi-access-edge-computing
21. M. Ishtiaq, N. Saeed, and M. A. Khan. Edge computing in IoT: A 6G perspective. IEEE Internet of Things Magazine, 2024. https://doi.org/10.1109/MITP.2024.3366778
22. L. Zhao, G. Zhou, G. Zheng, I. Chih-Lin, X. You, and L. Hanzo. Open-source multi-access edge computing for 6G: Opportunities and challenges. IEEE Access, 2021, 9: 158426–158439. https://doi.org/10.1109/ACCESS.2021.3130418
23. M. Crespo-Aguado, R. Lozano, F. Hernandez-Gobertti, N. Molner, and D. Gomez-Barquero. Flexible hyper-distributed IoT–edge–cloud platform for real-time digital twin applications on 6G-intended testbeds for logistics and industry. Future Internet, 2024, 16(11): 431. https://doi.org/10.3390/fi16110431
24. H. Rifa-Pous, V. Garcia-Font, C. Nunez-Gomez, and J. Salas. Security, trust and privacy challenges in AI-driven 6G networks. Computer Science and Information Technology, 2024, 14(14): 95–116. https://doi.org/10.5121/csit.2024.141408
25. L. Lovén, T. Leppänen, E. Peltonen, and J. Partala. EdgeAI: A vision for distributed, edge-native artificial intelligence in future 6G networks. 6G Wireless Summit, March 24-26, 2023. Levi, Finland. http://www.6gsummit.com/wp-content/uploads/2021/04/Loven_EdgeAI-Vision.pdf
26. P. K. Gkonis, A. Giannopoulos, N. Nomikos, P. Trakadas, L. Sarakis, and X. Masip-Bruin. A survey on architectural approaches for 6G networks: Implementation challenges, current trends, and future directions. Preprints, 2025. https://doi.org/10.20944/preprints202502.2175.v1
27. K. B. Letaief, Y. Shi, J. Lu, and J. Lu. Edge artificial intelligence for 6G: Vision, enabling technologies, and applications. IEEE Journal on Selected Areas in Communications, 2022, 40(1): 5–36. https://doi.org/10.1109/JSAC.2021.3126076
28. X. Chen, Y. Huang, C. Li, J. Wang, H. Zhang, A. Liu, Y. Zhang, and X. You. Toward 6G native-AI network: Foundation model based cloud-edge-end collaboration framework. arXiv, 2023. http://arxiv.org/abs/2310.17471
29. C. Watson, K. Woods, and D. J. Shyy. 6G and AI/ML. MITRE Technical Report, 2021. https://www.mitre.org/sites/default/files/2021-11/pr-21-0214-6g-and-artificial-intelligence-and-machine-learning.pdf
30. F. Zhu, W. Shi, S. Han, T. Q. S. Quek, Y. Shi, and K. B. Letaief. Wireless large AI model: Shaping the AI-native future of 6G and beyond. https://doi.org/10.48550/arXiv.2504.14653
31. H. G. Abreha, M. Hayajneh, and M. A. Serhani. Federated learning in edge computing: A systematic survey. Sensors, 2022, 22(2): 450. https://doi.org/10.3390/s22020450
32. Z. Lin, G. Qu, X. Chen, and K. Huang. Split learning in 6G edge networks. arXiv, 2023. https://doi.org/10.48550/arXiv.2306.12194
33. M. Zhang, M. Abdi, V. R. Dasari, and F. Restuccia. Semantic edge computing and semantic communications in 6G networks: A unifying survey and research challenges. arXiv, 2024. https://doi.org/10.48550/arXiv.2411.18199
34. H. Chou, J. Zhao, X. Ma, Y. Liu, S. Wang, X. Zhang, and J. Li. Edge AI empowered physical layer security for 6G NTN: Potential threats and future opportunities. arXiv, 2023. https://doi.org/10.48550/arXiv.2401.01005
35. H. F. Alhashimi, M. A. Al-Rubaye, M. A. Al-Juboori, M. S. Aljumaili, M. J. Alabboodi, and J. Balasubramaniam. Survey on AI-enabled resource management for 6G heterogeneous networks: Recent

 doi:10.20944/preprints202602.1019.v1

research, challenges, and future trends. Computers, Materials & Continua, 2025, 83(3): 3585–3622. https://doi.org/10.32604/CMC.2025.062867

36. Q. Cui, F. Zheng, F. Fang, L. Liu, Y. Huang, Z. Han, G. Y. Li, and H. V. Poor. Overview of AI and communication for 6G network: Fundamentals, challenges, and future research opportunities. Science China Information Sciences, 2024. https://doi.org/10.1007/s11432-024-4337-1
37. H. Yang, A. Alphones, Z. Xiong, D. Niyato, J. Zhao, and K. Wu. Artificial-intelligence-enabled intelligent 6G networks. IEEE Network, 2020, 34(6): 272–280. https://doi.org/10.1109/MNET.011.2000195
38. Nokia. Unlocking the full potential of AI-native 6G through standards, 2025. https://www.nokia.com/6g/unlocking-the-full-potential-of-ai-native-6g-through-standards/
39. C. Sandeepa, E. Zeydan, T. Samarasinghe, and M. Liyanage. Federated learning for 6G networks: Navigating privacy benefits and challenges. IEEE Open Journal of the Communications Society, 2024. https://doi.org/10.1109/OJCOMS.2024.3513832
40. X. Yan, J. Wang, X. Zhang, and Y. Zhang. Data plane design for AI-native 6G networks. Huawei Technologies, 2025. https://www.huawei.com/en/huaweitech/future-technologies/data-plane-design-ai-native-6g-networks
41. Y. Sanjalawe, S. Fraihat, S. Al-E'mari, M. Abualhaj, S. Makhadmeh, and E. Alzubi. A review of 6G and AI convergence: Enhancing communication networks with artificial intelligence. IEEE Open Journal of the Communications Society, 2025. https://doi.org/10.1109/OJCOMS.2025.3553302
42. Ericsson. 6G network architecture: A proposal for early alignment, 2025. https://www.ericsson.com/en/reports-and-papers/ericsson-technology-review/articles/6g-network-architecture
43. D. Dalai, S. Babu, and M. B. S. Satellite–6G network integration roadmap on reference architectures. TechRxiv, 2023. https://doi.org/10.36227/TECHRXIV.20624685.V1
44. Y. Tang, U. C. Srinivasan, B. J. Scott, O. Umealor, D. Kevogo, and W. Guo. End-to-end edge AI service provisioning framework in 6G ORAN. arXiv, 2025. https://doi.org/10.48550/arXiv.2503.11933
45. Z. Liu, X. Chen, H. Wu, Z. Wang, X. Chen, D. Niyato, K. Huang. Integrated sensing and edge AI: Realizing intelligent perception in 6G. arXiv, 2025. https://doi.org/10.48550/arXiv.2501.06726
46. L. Chang, Z. Zhang, P. Li, S. Xi, W. Guo, Y. Shen. 6G-enabled edge AI for metaverse: Challenges, methods, and future research directions. Journal of Communications and Information Networks, 2022, 7(2): 107–124. https://doi.org/10.23919/JCIN.2022.9815195
47. IEEE WCNC. WS-01: AI-enabled network orchestration – Design challenges and opportunities for 6G networks, 2023. https://wcnc2023.ieee-wcnc.org/workshop/ws-01-ai-enabled-network-orchestration-design-challenges-and-opportunities-6g-networks
48. Q. Chen, Z. Guo, W. Meng, S. Han, C. Li, and T. Q. S. Quek. A survey on resource management in joint communication and computing-embedded SAGIN. IEEE Communications Surveys & Tutorials, 2024. https://doi.org/10.1109/COMST.2024.3421523
49. A. Mekrache, A. Ksentini, and C. Verikoukis. Intent-based management of next-generation networks: An LLM-centric approach. IEEE Network, 2024, 38(5): 29-36. https://doi.org/10.1109/MNET.2024.3420120
50. A. Boutouchent, A. Mekrache, A. Ksentini, G. Adhane, J. da Fonseca, J. McNamara. 6G-INTENSE: Intent-driven native artificial intelligence architecture supporting network–compute abstraction and sensing at the deep edge. IEEE Vehicular Technology Magazine, 2025. 20(1): 44-55. https://doi.org/10.1109/MVT.2024.3525001
51. Y. Lu, X. Huang, K. Zhang, S. Maharjan, and Y. Zhang. Low-latency federated learning and blockchain for edge association in digital twin empowered 6G networks. IEEE Transactions on Industrial Informatics, 2021, 17(7): 5098–5107. https://doi.org/10.1109/TII.2020.3017668
52. S. Dang, O. Amin, B. Shihada, and M. S. Alouini. What should 6G be? Nature Electronics, 2020, 3(1): 20–29. https://doi.org/10.1038/s41928-019-0355-6
53. U. Löwenstein. WRC-23 and IMT-2030 (6G), 2023. https://summit2025.one6g.org/wp-content/uploads/Session-1_Uwe-Loewenstein_ITU-R-Status-update-on-WRC-and-IMT-2030.pdf

 doi:10.20944/preprints202602.1019.v1

54. D. C. Larsson, A. Grövlen, S. Parkvall, and O. Liberg. 6G standardization timeline and principles. Ericsson Blog, 2024. https://www.ericsson.com/en/blog/2024/3/6g-standardization-timeline-and-technology-principles
55. W. Chen. RAN Rel-19 status and a look beyond. 3GPP, 2025. https://www.3gpp.org/technologies/ran-rel-19
56. 3GPP. Overview of AI/ML-related work in 3GPP, 2025. https://www.3gpp.org/news-events/3gpp-news/ai-ml-2025
57. R. Daws. IMT-2030 vision: Industry experts outline the path to 6G. Telecoms Tech News, 2025. https://www.telecomstechnews.com/news/imt-2030-vision-industry-experts-outline-path-to-6g/