Beyond digital twins: The rise of cognitive twins in 6G and beyond wireless systems

***Ian F. Akyildiz*[1], *Tuğçe Bilen*[2]**

[1] International Telecommunication Union, Geneva, Switzerland (e-mail: ian.akyildiz@itu.int), and the Technology Innovation Institute, Abu Dhabi, United Arab Emiratess (e-mail: ian.akyildiz@tii.ae), [2] Department of Artificial Intelligence and Data Engineering, Istanbul Technical University, Istanbul, Turkey

Corresponding author: Tuğçe Bilen, bilent@itu.edu.tr

Digital twins have emerged as a key enabler for monitoring, simulation, and optimization in complex systems. However, existing approaches remain fundamentally centered on state replication and prediction, lacking the ability to systematically reuse experience, interpret system behavior, and support interactive decision-making under dynamic conditions. In this article, we introduce the *cognitive twin* as a new system paradigm that extends state synchronization with structured cognitive capabilities, including perception, memory, causal reasoning, and interaction. Unlike traditional digital twins and their learning-based extensions, a cognitive twin maintains an evolving cognitive state that supports experience-aware reasoning, causal interpretation, and context-aware interaction, enabling a transition from data-driven monitoring toward understanding-driven operation. We develop a four-layer reference architecture that captures how cognitive processes are distributed across device, edge, and cloud domains and realized through the continuous exchange of structured state, contextual knowledge, and inferred decisions. A central insight is that this paradigm fundamentally reshapes the role of communication. Rather than solely supporting data exchange, communication directly influences the formation, synchronization, and evolution of cognitive state under heterogeneous constraints on latency, bandwidth, reliability, and synchronization. A key hypothesis explored in this article is that communication impairments affect cognitive fidelity asymmetrically with respect to semantic importance, such that degradation of semantically-critical information may produce disproportionately large reasoning errors relative to equivalent losses in redundant or low impact data streams. This perspective motivates wireless-native system designs in which communication, computation, memory, and cognition are jointly optimized rather than treated as independent subsystems. We show how emerging 6G capabilities, including ultra-low-latency communication, integrated sensing and communication, semantic communication, and the edge–cloud continuum, provide the foundation for distributed cognitive operation at scale. Finally, we identify key open research problems for the wireless research community and present a structured roadmap toward the practical realization of cognitive twin systems.

Keywords: 6G networks, causal reasoning, cognitive twin, digital twin, distributed cognition, edge–cloud continuum, semantic communication

1. INTRODUCTION

Over the past two decades, the representation and control of physical systems have undergone a fundamental transformation. Early engineering workflows relied primarily on offline simulations to evaluate system behavior prior to deployment. The emergence of the Internet of Things (IoT) subsequently enabled continuous observation of physical entities through large-scale sensing and data collection [1]. More recently, digital twins have unified these developments into a continuously synchronized virtual representation that supports monitoring, prediction, and control [2, 3, 4]. While this progression has significantly reduced the gap between physical and digital systems, it has simultaneously exposed deeper limitations in how current systems interpret, contextualize, and act upon observed behavior.

Existing digital twins are highly effective at maintaining and forecasting system state, yet their intelligence remains largely centered on correlation-driven prediction and optimization [5]. Although modern frameworks increasingly incorporate learning-based analytics and adaptive control, most systems continue to focus on state estimation, anomaly detection, and predictive monitoring rather than on explicit reasoning over experience, context, and causality. As a result, recurring events are rarely operationalized as reusable experiences, explanations remain limited, and decision-making often continues to depend on human interpretation layered atop otherwise advanced digital infrastructure.

These limitations become increasingly significant in emerging 6G-scale environments characterized by ultra-dense deployments, strong non-stationarity, high mobility, and stringent reliability and latency requirements [6, 7, 8]. Applications such as autonomous mobility, industrial automation, immersive reality, and remote healthcare demand not only real-time awareness but also context-aware and experience-driven operation. In such environments, reacting solely to instantaneous observations is insufficient. Systems must continuously interpret evolving conditions, relate them to prior operational context, and support decisions under uncertainty.

To address this gap, this paper introduces the concept of the *cognitive twin*. A cognitive twin is a wireless-native and communication-aware virtual entity that extends state synchronization with structured cognitive capabilities, including perceptual grounding, episodic and semantic memory, causal reasoning, and interactive decision-making. Perceptual grounding refers to the construction of a temporally-consistent representation of system state from distributed multi-modal observations, while cognitive state denotes the internal representation combining current observations, retrieved memory, contextual knowledge, and inferred relationships used during reasoning and decision-making. Unlike digital twins augmented with standalone machine learning models or LLM-based agents connected to external memory, a cognitive twin explicitly integrates perception, memory, reasoning, and communication into a continuously operating closed-loop system executed under latency, synchronization, reliability, and resource constraints.

Rather than acting as a passive mirror of the physical environment, the cognitive twin operates as an active cognitive entity that continuously interprets observations, relates them to prior experience, constructs causal explanations, and supports decisions through interaction [9]. As illustrated in Fig. 1, this progression reflects a broader transition from synchronizing the system state to synchronizing the system understanding. In this context, understanding refers to the ability of the system to contextualize observations using memory and causal relationships in order to support interpretable and experience-driven decision-making. Consequently, while a conventional digital twin primarily answers *what* is happening, a cognitive twin is designed to reason about *why* it is happening, whether similar situations have occurred previously, and what actions should be taken next.

A central thesis of this paper is that future communication infrastructures must evolve beyond reliable bit transport toward supporting distributed cognitive execution. In cognitive twin systems, communication directly influences the consistency and fidelity of perception, memory retrieval, reasoning, and interaction across distributed execution domains. Here, cognitive fidelity refers to the consistency and reliability with which distributed cognitive processes preserve semantically-meaningful state, contextual relationships, and reasoning outcomes under communication and system constraints. Latency affects temporal alignment during multimodal perception, synchronization errors distort causal consistency, and packet loss may degrade memory retrieval or reasoning accuracy, depending on the semantic importance of the missing information.

Importantly, the motivation for beyond 5G and 6G capabilities is not that cognitive twins become entirely non-functional under modest communication impairments. Rather, different cognitive operations exhibit highly asymmetric sensitivity to latency, synchronization, and semantic information loss. For example, a few milliseconds of additional delay may have a negligible impact on archival memory synchronization, yet significantly affect temporally-aligned multimodal perception or distributed edge-assisted control operating over rapidly evolving physical processes. Similarly, moderate packet loss may be tolerable for redundant monitoring streams, while loss of semantically-critical observations can disproportionately distort reasoning outcomes and decision consistency. This asymmetry motivates communication-aware cognitive execution strategies that dynamically adapt computation, communication, and reasoning according to system conditions and semantic relevance.

Realizing such capabilities fundamentally reshapes the role of system infrastructure. cognitive twins operate across distributed, resource-constrained, and latency-sensitive environments in which perception, memory, reasoning, and interaction are partitioned across device, edge, and cloud domains. Lightweight perception and inference must often execute under strict real-time constraints at the edge, whereas computationally-intensive reasoning and long-horizon causal analysis are carried out through edge–cloud collaboration [10]. This introduces a fundamental trade-off between latency, energy consumption, communication overhead, and reasoning fidelity. Accordingly, the cognitive twin should be understood not as a monolithic architecture but as an adaptive, distributed cognitive system whose execution continuously evolves in response to communication conditions, resource availability, and application requirements. While conventional digital twins primarily address a synchronization problem, cognitive twins introduce a distributed cognition problem

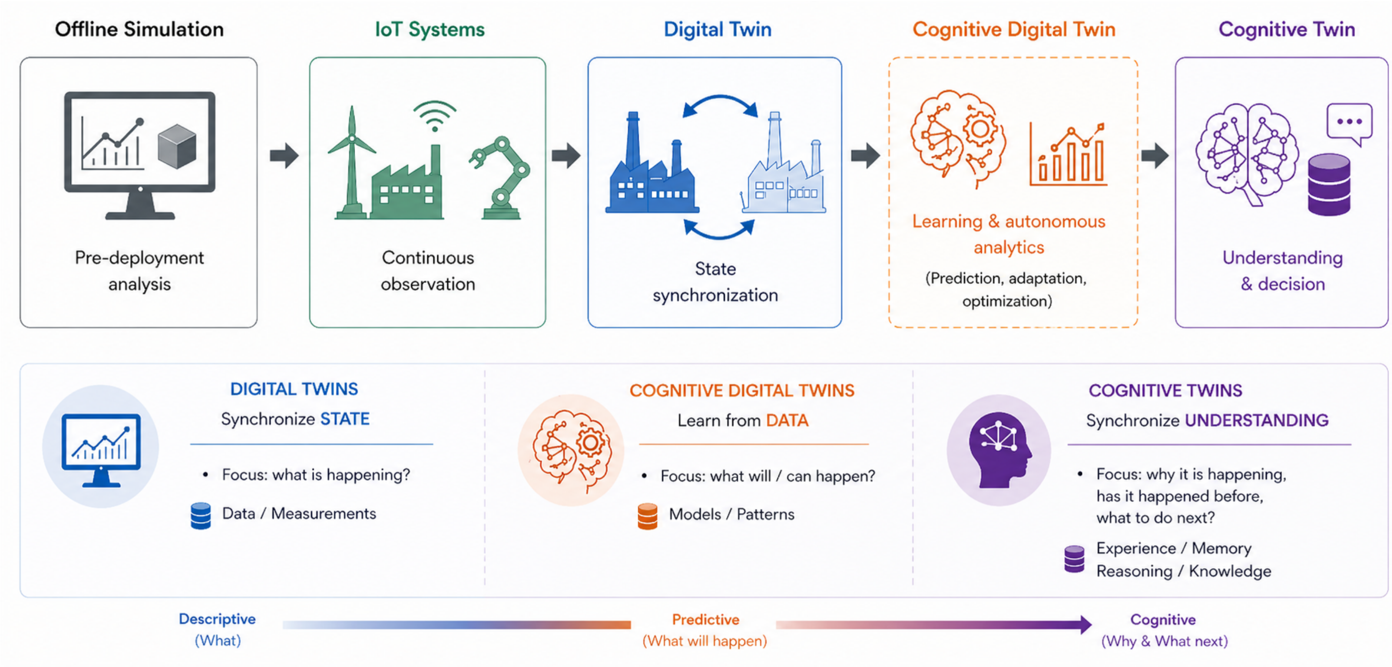

Figure 1 – Evolution of system paradigms toward cognitive twins. While digital twins synchronize system state, cognitive twins shift toward synchronizing understanding by integrating memory, experience, and reasoning for decision-making under dynamic conditions.

in which communication, computation, memory, and reasoning become tightly coupled system components. The main contributions of this paper are as follows:

- We formalize the notion of the *cognitive twin* as a communication aware cognitive system with explicit perceptual, memory, reasoning, and interaction capabilities, clearly distinguishing it from traditional digital twins, cognitive digital twins, and LLM-based agentic systems.
- We introduce a unified capability and function model that provides a structured operational interpretation of cognition in engineered systems by explicitly linking perception, memory, reasoning, and interaction within a closed-loop process.
- We propose a four-layer reference architecture organized into perceptual, memory, reasoning, and expression layers, and systematically map these layers onto device, edge, and cloud domains under realistic communication and execution constraints.
- We analyze the communication-aware realization of cognitive twins in future wireless systems, demonstrating how cognitive operations impose heterogeneous requirements on latency, synchronization, bandwidth, reliability, and semantic communication efficiency.
- We identify key open research problems and outline a structured roadmap spanning communication systems, distributed AI, semantic communication, causal reasoning, and cognitive system orchestration.

This work is motivated by the convergence of three developments that have only recently become practically realizable: the maturation of edge inference hardware enabling distributed cognitive execution at scale, advances in resource-efficient causal reasoning that make structured inference increasingly tractable in real-world environments, and the emergence of semantic communication as a design primitive for exchanging contextual and inferred information rather than raw data. Together, these developments shift the realization of cognitive twins from a conceptual direction toward an actionable systems' research agenda.

Accordingly, this article is positioned as a roadmap rather than a validation study. The quantitative values presented throughout should be interpreted as representative operating regimes grounded in existing wireless standards, edge computing studies, and analytical models, rather than strict deterministic guarantees. The objective is not to make definitive performance claims but to establish a structured, testable research agenda for cognitive twin systems. This perspective positions cognitive twin as a new interdisciplinary research direction at the intersection of communication systems, distributed intelligence, and cognitive system design.

The remainder of this paper is organized as follows. Section 2 reviews related work and positions the cognitive twin within the literature. Section 3 introduces the capability and function model and highlights differences between digital twins. Section 4 details the reference architecture for cognitive twins. Section 5 presents the wireless-native realization of cognitive twins in 6G. Section 6 discusses a representative application scenario. Section 7 outlines open research challenges, Section 8 presents the research roadmap, and Section 9 concludes the paper.

2. RELATED WORK

The concept of the cognitive twin does not emerge in isolation but from the gradual evolution of how physical systems are modeled, observed, and controlled. As illustrated in Fig. 1, this evolution spans from offline simulation and IoT-based monitoring through digital twins and their learning-augmented extensions. In this section, we trace this progression, identify the architectural limitations that persist in existing approaches, and motivate the need for a communication-aware distributed cognitive system abstraction.

2.1 Toward networked digital twins

The digital twin paradigm was initially introduced to create high-fidelity virtual representations of physical assets supporting monitoring, simulation, and lifecycle management [11, 12]. Subsequent studies extended this vision toward continuously synchronized, data-driven system representations [13, 14], generally characterized by bidirectional data exchange and persistent integration across design, operation, and maintenance phases. In parallel, digital twins have been widely studied in the context of industrial and cyber-physical systems. In industrial IoT environments, they have been applied to predictive maintenance, fault diagnosis, and process optimization [15, 16, 17]. More recently, the notion of digital twin Networks has emerged, where multiple interconnected twins collectively represent large-scale environments such as smart cities, transportation systems, and communication infrastructures [18, 19]. This represents an important transition from isolated, asset-centric models to distributed, system-level representations. Despite this progress, most digital twin frameworks remain centered on state synchronization, monitoring, and predictive analytics. Distribution improves scalability and coordination, yet the dominant operational abstraction continues to rely on maintaining a synchronized physical-system representation and performing prediction or optimization based on observed data. The nature of intelligence in these systems remains state-centric and reactive.

2.2 Toward cognitive digital twins

To address the limitations of purely descriptive digital twins, recent studies have incorporated learning-based and autonomous capabilities into twin architectures [20, 21, 22]. In particular, Cognitive digital twins have been proposed to combine learning, reasoning, and knowledge-driven optimization [23]. Additional developments include graph-based system representations [24, 25] and knowledge-driven frameworks that model relationships between operational states and system components [26, 27].

However, most existing Cognitive digital twin approaches remain primarily focused on predictive optimization and decision support rather than continuously evolving distributed cognition. Memory is typically incorporated implicitly via learned model parameters or historical datasets, rather than explicitly via episodic and semantic structures designed for real-time retrieval and contextual reasoning. Similarly, reasoning is often embedded within learned inference pipelines, limiting interpretability and restricting explicit causal or counterfactual analysis [28, 26]. Although these systems advance predictive intelligence, they only partially address the broader challenge of experience-driven and communication-aware cognitive operation.

2.3 Relation to LLM-based agents and AI memory systems

Recent advances in Large Language Models (LLMs), Retrieval-Augmented Generation (RAG), and agentic AI architectures have enabled systems capable of combining reasoning, external memory, and tool use [29]. These systems exhibit several cognitive-like behaviors, including contextual dialogue, retrieval-based interaction, and adaptive task execution.

Nevertheless, these architectures are primarily designed for textual and symbolic domains rather than for continuously evolving physical systems grounded in real-time sensory observations. Memory in such systems is typically based on embedding retrieval or external databases, where context is reconstructed at query time rather than maintained as a continuously synchronized operational state. Moreover, existing agentic AI systems do not explicitly model how communication latency, synchronization errors, bandwidth limitations, or distributed execution constraints affect reasoning consistency and system behavior. As a result, while LLM-based agents can generate plausible responses and support high-level interaction, they do not maintain a temporally-grounded and communication-aware cognitive representation of a physical system operating under real-world constraints.

2.4 Architectural distinction from agentic AI systems

The distinction between cognitive twins and existing agentic AI systems is architectural rather than terminological. While LLM-based agents combine memory, reasoning, and interaction, cognitive twins additionally require continuous perceptual grounding, distributed execution, and communication-aware cognitive consistency.

Specifically, a cognitive twin maintains an internal cognitive state continuously constructed from temporally-aligned multimodal observations rather than reconstructed on demand from retrieved textual representations. It explicitly accounts for the effect of communication impairments, latency, packet loss, synchronization errors, and distributed execution constraints on perception coherence, memory consistency, and reasoning fidelity. Cognitive processes are dynamically partitioned across the device, edge, and cloud domains based on communication conditions, resource availability, and application requirements.

Consequently, a cognitive twin operates as a continuously evolving, distributed, dynamic system whose memory, contextual knowledge, and causal representations jointly evolve with the physical environment. This introduces three defining properties: *grounding*, in which cognitive representations remain tied to temporally-consistent physical observations; *consistency*, in which memory and reasoning remain coherent under partial or delayed information; and *communication-aware execution*, in which cognitive processes adapt dynamically to network and system conditions. To the best of our knowledge, existing agentic AI frameworks do not model these properties as first-class system constraints, which establishes the cognitive twin as a fundamentally distinct system abstraction rather than an incremental extension.

2.5 Limitations of existing approaches

Taken together, the existing literature reveals a clear progression toward increasingly intelligent and distributed system representations, yet several structural limitations persist. First, many current digital twin and Cognitive digital twin frameworks rely on historical data retention or model-based learning rather than explicit operational memory structures designed for real-time retrieval and contextual reuse [28]. While past observations influence model training and optimization, prior operational experiences are often not maintained as queryable episodic memory supporting experience-driven reasoning at runtime. Second, most existing approaches remain dominated by correlation-based prediction and optimization. Although some recent studies incorporate knowledge graphs or reasoning mechanisms, explicit causal reasoning and intervention-aware decision-making remain limited [30]. Many systems can predict system behavior but provide only restricted support for interpretable causal analysis or counterfactual reasoning. Third, existing digital twin systems generally treat communication as a supporting infrastructure for data exchange rather than as an active component of distributed cognition. As systems scale across device, edge, and cloud environments, the synchronization of contextual knowledge, memory consistency, and reasoning outputs becomes increasingly sensitive to communication conditions. Latency, synchronization errors, and packet loss may therefore affect not only data delivery but also the consistency and fidelity of distributed cognitive processes. Importantly, this limitation does not imply that existing digital twins are incapable of operating over current wireless or hybrid infrastructures. Rather, future large-scale cognitive twin systems will increasingly require tight integration among communication, computation, memory, and reasoning as cognitive operations become more distributed, latency-sensitive, and semantically-coupled.

2.6 Toward cognitive twins

The observations above motivate the need for a new abstraction that moves beyond state-centric synchronization toward communication-aware distributed cognition. The proposed cognitive twin represents this transition. Rather than extending the digital twin paradigm solely through improved prediction accuracy or larger-scale synchronization, the cognitive twin introduces a distributed cognitive abstraction in which perception, memory, reasoning, interaction, and communication participate jointly in system operation.

Unlike traditional digital twins, which primarily maintain system state, or Cognitive digital twins, which enhance predictive analytics and decision support, a cognitive twin maintains structured representations of experience, supports contextual and causal reasoning, and continuously adapts its execution according to communication and system conditions. Crucially, communication is no longer limited to transmitting sensor measurements or model updates; it supports the synchronization and evolution of cognitive state, including contextual knowledge, memory consistency, inferred relationships, and reasoning outputs, across distributed execution domains. In this sense, the transition from digital twins to cognitive twins can be understood as a transition from synchronizing system state to synchronizing system understanding, defining a new research space at the intersection of communication systems, distributed intelligence, semantic communication, and cognitive system design.

From this perspective, cognitive twins should not be viewed as a replacement for existing digital twin technologies, but rather as a next-stage architectural evolution motivated by the increasing coupling among communication, computation, memory, and reasoning in future large-scale intelligent systems. As systems become increasingly distributed, dynamic, and autonomous, maintaining an accurate representation of physical state alone is no longer sufficient. Future infrastructures must also preserve contextual knowledge, support experience-driven adaptation, and coordinate reasoning processes across heterogeneous execution domains. Cognitive twins provide a conceptual foundation for this transition by treating cognition itself as a distributed systems function, thereby extending the role of communication from information delivery to the maintenance of shared understanding.

Table 1 – Comparison of digital twin, cognitive digital twin, and cognitive twin paradigms across representative system dimensions.

Dimension	digital twin	Cognitive digital twin	cognitive twin (Proposed)
System paradigm	State synchronization	Data-driven inference	Distributed cognition
Core objective	State fidelity	Context-aware recommendation	Experience-driven reasoning
Memory model	Time-series storage	Limited contextual memory	Episodic + semantic memory with active retrieval
Reasoning	Simulation / rule-based	Learned inference	Multi-level causal reasoning
Communication role	Data transport	Event-driven updates	Cognitive state synchronization
Information exchanged	Raw sensor data	Features / embeddings	Context, intent, inferred knowledge
Failure behavior	State staleness	Missing context	Graceful degradation under bounded uncertainty
Interaction model	Passive monitoring	Assisted recommendation	Context-aware interactive dialogue
Execution model	Centralized / static	Learning-assisted	Adaptive edge-cloud cognition
Explainability	Threshold-based	Feature attribution	Causal + counterfactual reasoning
Data modality	Structured sensor data	Semi-structured data	Multi-modal cognitive state

3. FROM DIGITAL TWINS TO COGNITIVE TWINS: A CAPABILITY AND FUNCTION PERSPECTIVE

The evolution from digital twins to cognitive twins can be understood through two dimensions: how system knowledge is represented and how decisions are produced under operational constraints. Traditional digital twins primarily focus on maintaining synchronized representations of physical systems to support monitoring, visualization, and predictive analysis. Cognitive digital twins extend this paradigm by incorporating learning-based inference and adaptive analytics. The proposed cognitive twin introduces a further transition toward distributed cognition, where perception, memory, reasoning, interaction, and communication jointly participate in system operation.

Importantly, these paradigms correspond to different operational abstractions rather than incremental improvements. Digital twins remain centered on state synchronization and predictive monitoring. cognitive twins, in contrast, maintain an evolving cognitive state composed of contextual memory, inferred relationships, uncertainty representations, and reasoning outputs distributed across device, edge, and cloud domains. One of the key distinctions lies in the treatment of memory and reasoning. In conventional digital twins, memory mainly serves as historical storage for synchronization and offline analytics. Cognitive digital twins improve prediction quality through learning-based inference, yet memory and reasoning are often implicitly embedded in trained models. cognitive twins instead treat memory as an active computational component that directly participates in contextual retrieval, interpretation, and decision-making. Reasoning also extends beyond predictive inference to causal and intervention-aware analysis that supports explanations, counterfactual evaluation, and experience-driven adaptation.

Another important distinction concerns communication and execution. Existing digital twin systems generally treat communication as supporting infrastructure for synchronization and data exchange. cognitive twins, however, rely on distributed cognitive execution across device, edge, and cloud domains. In this setting, latency, synchronization, reliability, and semantic consistency directly influence perception coherence, memory accessibility, and reasoning fidelity. Importantly, this does not imply that current wireless or hybrid infrastructures are incapable of supporting cognitive twin functionality. Rather, future large-scale cognitive twin systems will increasingly require tighter integration among communication, computation, memory, and reasoning as cognitive processes become more distributed, latency-sensitive, and semantically-coupled.

The transition from state synchronization to distributed cognition is summarized in Table 1. The table highlights that the cognitive twin is not simply a more accurate or larger-scale digital twin. Instead, it represents a different system abstraction in which cognition itself becomes a distributed systems process shaped by communication and execution constraints.

To make the notion of distributed cognition more concrete and operational, we characterize the cognitive twin through two complementary abstractions: a capability layer and a function layer. The capability layer captures what the system is fundamentally capable of representing and processing. The function layer captures how these capabilities manifest at runtime. The capabilities define the structural building blocks of cognition, whereas the functions describe their observable behavioral manifestation during system operation.

3.1 Core capabilities of a cognitive twin

The cognitive twin is defined by seven core capabilities, summarized in Table 2. Together, these capabilities operationalize the distributed cognition model introduced in Table 1.

- *Perceptual grounding:* Constructing a temporally-consistent, uncertainty-aware representation of system state from distributed multimodal observations. Rather than simply aggregating sensor measurements, the objective is to maintain a coherent state representation suitable for

Table 2 – Core capabilities of a cognitive twin.

Capability	What It Means	Concrete Example
Perceptual grounding	Real-time fusion of multimodal observations into a temporally consistent state representation	Video, vibration, thermal, and audio streams fused into a unified system view
Episodic memory	Retrieval of past events through contextual similarity	"Recall the turbine failure last March with the rapid 12°C temperature rise"
Semantic memory	Structured representation of causal and relational knowledge	"Lubrication loss → temperature rise → bearing spalling"
Causal reasoning	Intervention-aware and counterfactual reasoning	"Would earlier replacement have prevented this failure?"
Natural language interaction	Context-aware interaction with human operators	"Why did the temperature spike?" with a grounded explanation
Meta-cognition	Awareness of uncertainty and incomplete information	"87% confidence due to inconsistent sensor readings"
Value alignment	Trade-off reasoning across competing objectives	"Option A reduces downtime; Option B is safer but more costly"

downstream reasoning and decision-making. Because this representation is constructed across device and edge domains, its consistency depends directly on latency, synchronization, and communication reliability.

- *Episodic memory:* Structured storage and retrieval of prior operational experiences inspired by episodic memory models in cognitive science [31]. The cognitive twin maintains experience-level representations that can be queried based on contextual relevance and similarity, enabling systematic reuse of past experience during decision-making.
- *Semantic memory:* Structured representation of domain knowledge in the form of graphs, relational dependencies, or causal models [32]. This layer captures causal mechanisms and system-level dependencies, enabling interpretation and generalization beyond individual observations.
- *Causal reasoning:* Intervention-aware and counterfactual analysis that extends beyond correlation-based prediction. Grounded in structural causal models and causal graphs, this capability supports explanation, planning, and evaluation of alternative operational outcomes. The fidelity of causal reasoning depends on the availability and temporal consistency of underlying observations.
- *Natural language interaction:* Context-aware dialogue grounded in operational system state, retrieved experience, and reasoning outputs rather than unconstrained generative interaction. This capability is essential for transparent and interpretable human-system collaboration.
- *Meta-cognition:* Awareness of uncertainty, incomplete information, and confidence limitations. By quantifying confidence and detecting inconsistencies, the cognitive twin avoids overconfident decisions when observations are partial or noisy.
- *Value alignment:* Reasoning over competing objectives such as safety, latency, energy, performance, and operational cost under changing system conditions. This capability is particularly critical in wireless environments where resources are limited and conditions evolve rapidly.

These capabilities do not operate independently. In practice, perception constructs system state, memory provides contextual grounding, reasoning produces decisions, and interaction communicates outcomes to external entities. Together, these components form a closed-loop process continuously shaped by system constraints. Consequently, cognition becomes a distributed systems process whose fidelity depends jointly on computation, memory consistency, and communication conditions.

3.2 Core functions of a cognitive twin

While the capabilities above define what a cognitive twin contains, its operational distinction becomes clearer through what it does. The seven capabilities jointly produce four observable cognitive functions, summarized in Table 3. These functions represent the externally-observable behavioral manifestation of the underlying cognitive process. This relationship establishes a direct mapping between the structural properties of the cognitive twin and its runtime behavior.

- *Explains:* The cognitive twin generates interpretable causal explanations grounded in retrieved operational context rather than isolated anomaly alerts. When full causal inference is infeasible under real-time constraints, the system relies on partial or approximate explanations while explicitly representing the associated uncertainty.
- *Remembers:* The system retrieves structurally similar prior experiences to contextualize current observations and support experience-driven reasoning. Retrieval is performed through approximate similarity search over episodic memory, enabling low-latency access at the edge while maintaining consistency with long-term memory stored in the cloud.
- *Recommends:* Decisions are produced together with confidence estimates, trade-offs, and contextual reasoning outputs rather than opaque predictions. Real-time decisions rely on lightweight approximations and locally-available knowledge, whereas more complex evaluations may involve deferred or cloud-assisted processing.
- *Interacts:* The cognitive twin supports iterative, context-aware interaction while preserving consistency with the underlying cognitive state. Responses are grounded in current perception, retrieved memory, and previously computed reasoning outputs, ensuring that interaction remains coherent with actual system behavior.

As summarized in Table 4, each function emerges from

Table 3 – Core functions of a cognitive twin.

Function	Description	Representative Output
Explains	Generates causal interpretations rather than isolated alerts	Causal hypothesis with confidence estimate
Remembers	Retrieves relevant past experiences without explicit manual queries	Retrieved episode with temporal progression
Recommends	Produces actions together with associated trade-offs and uncertainty	Ranked action set with risk/cost estimates
Interacts	Supports context-aware dialogue grounded in operational state	Context-aware response linked to system conditions

Table 4 – Relationship between core functions and supporting capabilities.

Function	Primary supporting capabilities
Explains	Causal reasoning, semantic memory, perceptual grounding
Remembers	Episodic memory, semantic memory
Recommends	Causal reasoning, episodic memory, value alignment
Interacts	Natural language interaction, meta-cognition

the coordinated interaction of multiple capabilities, forming a many-to-many dependency structure. For example, the *Explains* function draws jointly on causal reasoning, semantic memory, and perceptual grounding. Similarly, *Recommends* additionally relies on episodic memory and value alignment to produce trade-off-aware decisions. This dependency structure implies that degradation in any single capability, whether caused by insufficient memory retrieval, impaired perception, or limited reasoning fidelity, can propagate across multiple functions simultaneously.

Taken together, the capability and function models establish the cognitive twin as a continuously evolving distributed cognitive system rather than a passive synchronized representation. The effectiveness of each function depends not only on local computation but also on the consistency and timeliness of information exchanged across distributed execution domains. This dependence motivates a concrete execution architecture that jointly coordinates perception, memory, reasoning, interaction, and communication under realistic system constraints, which is developed in the following section.

4. REFERENCE ARCHITECTURE OF A COGNITIVE TWIN

The capability and function model introduced in the previous section defines what a cognitive twin represents and how cognitive behavior emerges during operation. Realizing these abstractions in practice requires a concrete system-level organization in which perception, memory, reasoning, and interaction are distributed across heterogeneous execution domains operating under different latency, energy, and communication constraints.

To capture this structure, we propose the reference architecture illustrated in Fig. 2. The architecture is organized into four tightly coupled layers: the perceptual layer, memory layer, reasoning layer, and expression layer. These layers operate across device, edge, and cloud domains, forming a hierarchical yet continuously interacting cognitive pipeline.

A defining property of this architecture is its directional information flow. Multimodal observations are first transformed into a structured system representation. This representation is then enriched through memory retrieval and contextualization. The reasoning layer generates decisions and explanations from the enriched context, and the resulting outputs are communicated through the expression layer and translated into system actions. These actions subsequently influence the physical environment and generate new observations. As a result, the architecture forms a continuously evolving closed-loop cognitive process rather than a static monitoring pipeline.

The layers are not executed uniformly across the infrastructure. Different cognitive processes impose heterogeneous timing, synchronization, computational, and communication requirements. Lightweight perception and low-latency interaction typically require execution close to the edge. In contrast, large-scale reasoning, long-term memory maintenance, and complex causal analysis are better suited to cloud-assisted execution. Accordingly, the cognitive twin architecture should be interpreted as an adaptive execution framework in which cognition is dynamically partitioned across device, edge, and cloud domains according to system conditions and operational constraints.

4.1 Perceptual layer

The perceptual layer transforms raw multimodal observations into a structured and temporally-consistent system representation, as illustrated in Fig. 2. This layer spans both device and edge domains, separating lightweight sensing operations from computationally-intensive fusion processes.

At the device level, raw sensory signals are preprocessed through lightweight filtering, compression, and feature extraction. These observations may originate from heterogeneous sources, including video streams, vibration sensors, thermal measurements, communication-level indicators, and environmental sensing systems. The processed signals are transmitted to the edge, where multimodal fusion constructs a unified representation of system state by integrating heterogeneous inputs into a coherent state space.

The objective of this layer extends beyond simple data aggregation. Instead, the goal is to construct a temporally-aligned and uncertainty-aware representation suitable for

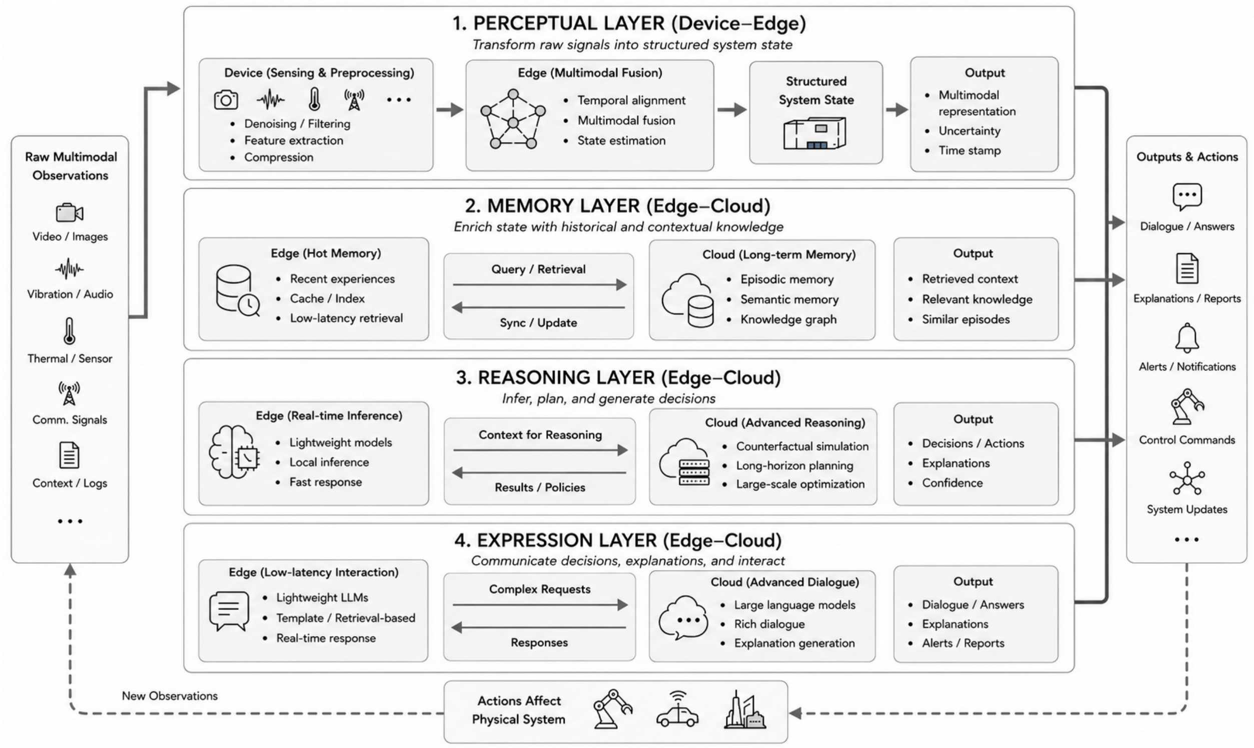

Figure 2 – Reference architecture of a cognitive twin, organized into perceptual, memory, reasoning, and expression layers. The system operates as a closed-loop pipeline in which multimodal observations are transformed into a structured state, enriched through memory retrieval, processed via distributed causal reasoning, and translated into decisions and explanations. Edge and cloud components jointly support real-time inference and large-scale optimization across heterogeneous execution domains.

downstream reasoning. Because observations arrive from distributed sensing modalities operating under heterogeneous delays and communication conditions, synchronization consistency becomes a critical requirement. Even small temporal inconsistencies may propagate upward as distorted state estimation and degraded reasoning fidelity. The output of the perceptual layer is a structured multimodal cognitive state composed of fused observations, temporal relationships, confidence estimates, and contextual metadata. Together, these components provide the foundation for memory retrieval and reasoning.

4.2 Memory layer

The memory layer augments the perceived system state with contextual and historical knowledge [33]. It maintains both episodic memory, which captures prior operational experiences, and semantic memory, which encodes structured relationships, causal dependencies, and domain-level abstractions. This layer operates in a distributed manner. The edge maintains a hot memory of recently accessed experiences, supporting low-latency reasoning. The cloud, on the other hand, maintains long-term structures including archival experiences, knowledge graphs, and large-scale relational representations. This organization balances scalability and responsiveness while enabling continuous accumulation of operational experience. Memory operations are inherently bidirectional. The edge continuously retrieves contextually-relevant experiences to support real-time inference, while updates are periodically synchronized with cloud-based long-term memory to preserve global consistency.

Unlike conventional digital twin storage architectures, memory in the proposed cognitive twin is not treated as passive historical retention. Instead, memory actively participates in contextual interpretation and reasoning through retrieval-driven cognitive execution. This distinction directly corresponds to the transition from passive synchronization to distributed cognition, summarized in Table 1. The output of this layer is contextualized knowledge obtained by combining the current system state with retrieved episodic and semantic information. This enriched representation is subsequently passed to the reasoning layer.

4.3 Reasoning layer

The reasoning layer transforms contextualized knowledge into decisions, explanations, and action policies. It operates across both edge and cloud domains with a functional separation between low-latency inference and computationally-intensive reasoning. At the edge, lightweight models support rapid local inference under strict timing constraints.

These models operate on the enriched contextual representation produced by the memory layer. More computationally-intensive operations are executed in the cloud, where larger models and broader contextual information support long-horizon planning, counterfactual evaluation, and deeper causal analysis. Fast edge-level reasoning prioritizes responsiveness and operational continuity, whereas cloud-assisted reasoning prioritizes fidelity, causal depth, and global contextual consistency. The interaction between edge and cloud remains bidirectional. The edge provides contextual observations and intermediate reasoning states, while the cloud returns refined outputs, distilled policies, and reusable reasoning artifacts.

Reasoning fidelity is inherently constrained by communication conditions and memory consistency. Incomplete observations, stale contextual retrieval, or synchronization errors may degrade causal consistency and introduce uncertainty into downstream decisions. Consequently, the quality of reasoning depends not only on model capability but also on the integrity of the distributed cognitive pipeline as a whole. This creates a strong coupling between communication quality and cognitive correctness, particularly in distributed deployments where perception, memory retrieval, and reasoning depend on timely and semantically-consistent information exchange.

4.4 Expression layer

The expression layer communicates the cognitive twin's outputs to external entities, including human operators, autonomous agents, and other system components. It transforms internal cognitive state into interpretable and actionable outputs such as decisions, explanations, recommendations, and interactive dialogue. At the edge, lightweight interaction models support low-latency responses for operationally-critical tasks. More sophisticated dialogue generation and explanation mechanisms are delegated to cloud-assisted processing when richer contextualization or deeper reasoning becomes necessary. Crucially, interaction in the cognitive twin is grounded in operational state, retrieved memory, and reasoning outputs rather than unconstrained generative behavior. This ensures consistency between communicated outputs and the underlying cognitive state, preserving interpretability and operational reliability under real-world system constraints. The expression layer, therefore, acts not simply as a user interface, but as the external manifestation of the distributed cognitive process itself.

4.5 Closed-loop cognitive execution

The four layers form a tightly coupled closed-loop cognitive architecture in which information flows upward from perception through memory and reasoning to expression, while control flows downward as decisions influence the physical environment and generate new observations. The perceptual layer constructs a structured system state. The memory layer contextualizes this state using prior experience. The reasoning layer produces decisions and explanations, and the expression layer communicates these outputs to external entities.

A key implication of this architecture is that cognition emerges not from any individual layer in isolation, but from the coordinated interaction between perception, memory, reasoning, interaction, and communication across distributed execution domains. This distinguishes the proposed cognitive twin from conventional pipeline-based digital twin systems and motivates the need for communication-aware execution strategies. Importantly, different applications may place different emphasis on individual layers, resulting in diverse realizations of the cognitive twin architecture while preserving the same underlying cognitive principles. Understanding how this pipeline behaves under realistic wireless conditions, and how communication constraints shape the fidelity and placement of each cognitive operation, requires a more explicit operational treatment, which is developed in the following section.

5. WIRELESS-NATIVE REALIZATION OF COGNITIVE TWINS IN 6G

The adaptive execution framework developed in the previous section establishes how cognitive processes are dynamically partitioned across device, edge, and cloud domains under system constraints. A natural follow-on question is what specific communication capabilities are required to sustain this execution in practice, and how emerging 6G technologies address these requirements. Unlike conventional digital twin systems, where communication primarily serves state synchronization, cognitive twins require communication to sustain their cognitive processes. Perception depends on temporally-aligned observations; memory relies on timely and consistent retrieval; reasoning requires reliable contextual information; and interaction depends on preserving semantic integrity across distributed execution domains. Consequently, latency, synchronization, reliability, and semantic consistency become fundamental determinants of cognitive fidelity rather than secondary implementation concerns.

Fig. 3 illustrates how 6G capabilities provide the substrate required for this distributed cognitive execution. In the perceptual layer, real-time multimodal fusion requires ultra-low latency and tight synchronization to keep observations temporally-aligned; any degradation at this stage propagates upward as an inconsistency in the perceived system state. In the memory layer, edge-local retrieval must be fast enough to support real-time reasoning, whereas cloud-based archival access tolerates higher delays but requires

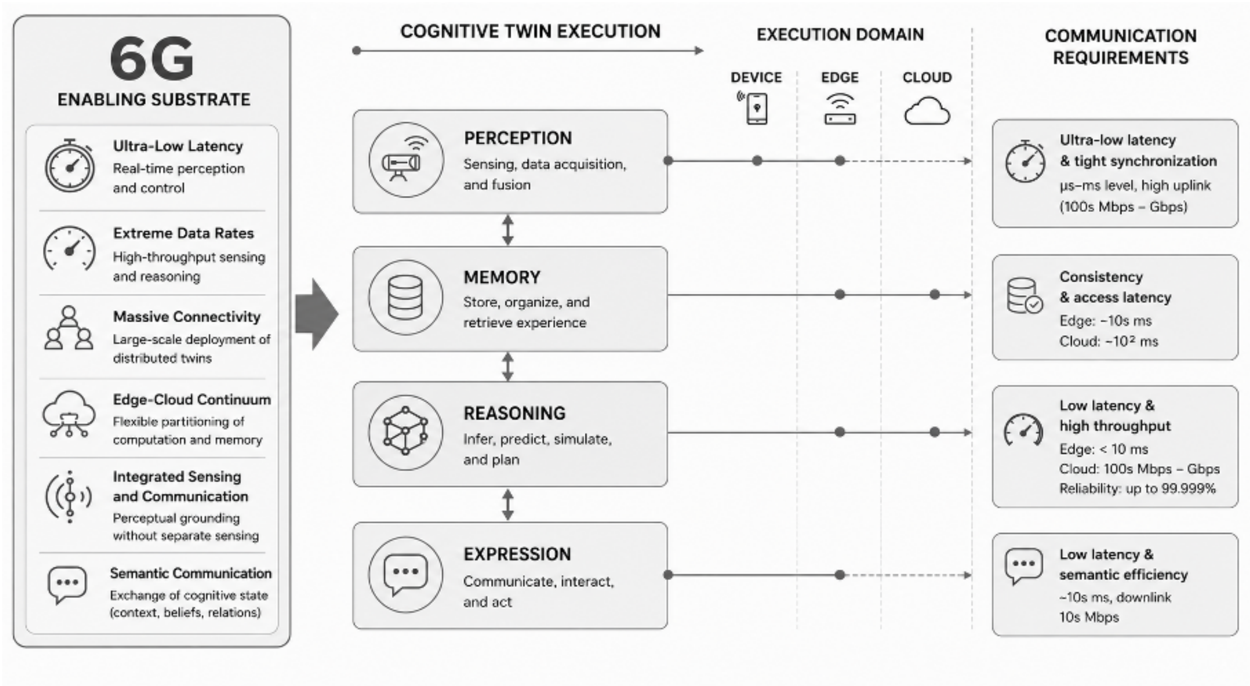

Figure 3 – 6G as an enabling substrate for cognitive twin execution. Cognitive operations across perception, memory, reasoning, and expression layers impose heterogeneous requirements on latency, synchronization, reliability, throughput, and semantic consistency.

Table 5 – Capability-driven cognitive operations and their mapping to execution layers, communication requirements, and 6G enablers. Values are indicative and reflect typical operating regimes rather than strict guarantees.

Capability	Representative Operation	Execution Layer	Wireless Requirement (Indicative)	6G Enabler
Perceptual grounding	Real-time multimodal fusion	Device → Edge	Latency < 10 ms, uplink 10–100 Mbps, synchronization < 100 μs	URLLC, ISAC
Episodic memory	Hot memory retrieval	Edge	Latency < 20 ms, low–moderate throughput	Edge intelligence
	Cold memory retrieval	Cloud	Latency < 200 ms, sustained throughput 10–100 Mbps	High data rate
Semantic memory	Causal graph query	Edge ↔ Cloud	Consistency-aware access, latency 10–100 ms, periodic synchronization	Semantic communication
Causal reasoning	Simulation and planning	Edge ↔ Cloud	Latency 10–100 ms (edge), higher for cloud processing; high throughput and reliability	URLLC, extreme data rate
Interaction	Routine response	Edge	Latency < 30 ms, downlink 10–50 Mbps	Edge intelligence
	Complex dialogue / reasoning	Cloud-assisted	Latency 50–300 ms, semantic-aware communication efficiency	Semantic communication
Meta-cognition	Policy adaptation / meta-learning	Edge ↔ Cloud	Periodic updates (seconds–minutes), moderate throughput	Edge–cloud continuum
Value alignment	Trade-off reasoning	Edge / Cloud	Latency 10–100 ms, high reliability	URLLC

sustained, reliable synchronization. In the reasoning layer, edge inference must satisfy strict latency constraints for immediate decisions, whereas cloud-assisted reasoning requires bursty high-throughput communication and high reliability for complex analysis. In the expression layer, interaction shifts the objective from raw data delivery to low-latency semantic integrity. These layer-specific requirements confirm that the same cognitive function may impose fundamentally different communication constraints depending on its execution domain, necessitating adaptive strategies that dynamically align network resources with cognitive objectives.

Table 5 maps each cognitive capability to its representative operation, execution domain, communication requirement, and enabling 6G technology. The quantitative values are indicative operating regimes grounded in wireless standards and edge computing studies, including URLLC latency targets, industrial uplink throughput requirements, and ISAC synchronization constraints. In practice, achievable performance depends on deployment-specific factors such as channel conditions, interference, system scale, and hardware constraints. The table should therefore be read as a set of reference design points for analyzing how communication conditions shape cognitive execution.

The primary implication of Table 5 is that communication constraints directly determine which cognitive functions can be executed, where they can be placed, and at what fidelity they can operate. Packet loss, latency, and synchronization errors propagate through the cognitive pipeline, degrading perceptual coherence, memory retrieval accuracy, and reasoning outcomes. Different cognitive operations exhibit asymmetric sensitivity to such impairments: a few milliseconds of additional delay may have a negligible impact on archival memory synchronization, yet significantly affect temporally-aligned multimodal perception or latency-sensitive edge-assisted reasoning. This asymmetry motivates a central hypothesis of this work: communication impairments affect cognitive fidelity in ways that vary with semantic importance. Degradation of semantically-critical observations may produce disproportionately large reasoning errors relative to equivalent losses in redundant or low-impact information streams, motivating communication strategies that prioritize semantic relevance rather than treating all transmitted data uniformly.

5.1 Causal consistency under imperfect wireless observations

Within the distributed cognitive pipeline, causal reasoning plays a central role because it determines how observations are interpreted and how decisions are made. Unlike purely data-driven inference, causal models rely on temporally-consistent and semantically-complete observations to correctly capture dependencies and interventions, a requirement that wireless impairments directly challenge.

Synchronization errors across distributed sensing streams lead to temporally-misaligned observations, resulting in incorrect event ordering and distorted causal dependencies. Packet loss in feedback loops removes semantically-important state transitions, producing incomplete reasoning about system dynamics. Bandwidth and resource constraints at the edge further limit the complexity of causal representations that can be maintained under strict latency constraints.

These interactions reveal a key systems-level implication: causal consistency is not solely determined by model design, but emerges from the joint behavior of communication conditions and cognitive execution. The same causal model may produce reliable or unreliable outcomes depending on latency, packet loss, synchronization accuracy, and memory consistency across distributed execution domains. Moreover, graceful degradation is not guaranteed under all operating conditions. Severe synchronization inconsistencies, persistent loss of semantically-critical observations, or stale contextual retrieval may propagate through the pipeline, leading to unstable or systematically biased reasoning outcomes. cognitive twins, therefore, require uncertainty-aware execution strategies that can dynamically reduce reasoning fidelity, simplify causal representations, or revert to lower-complexity operational modes under adverse conditions.

5.2 Adaptive execution and control of cognitive processes

The causal consistency requirements identified above reinforce the need for cognitive execution strategies that continuously adapt to communication conditions rather than assuming stable wireless environments. We formalize this by representing the cognitive twin as a structured dynamic system and modeling cognitive execution as an online control problem.

5.2.1 Cognitive state representation

At time t, the cognitive twin is defined as:

$$C_t = \left(s_t,\ m_t^{(e)},\ m_t^{(s)},\ \mathcal{G}_t,\ \pi_t\right), \tag{1}$$

where s_t denotes the perceived system state, $m_t^{(e)}$ episodic memory, $m_t^{(s)}$ semantic memory, $\mathcal{G}_t$ the causal model, and π_t the execution policy. The system evolves through a closed-loop process:

$$s_t = f_{\text{perc}}(o_t), \quad c_t = f_{\text{mem}}\left(s_t, m_t^{(e)}, m_t^{(s)}\right), \quad a_t = f_{\text{reason}}(c_t, \mathcal{G}_t), \tag{2}$$

with memory and causal structures updated as:

$$m_{t+1}^{(e)} = \mathcal{U}_e\left(m_t^{(e)}, s_t, a_t\right), \quad \mathcal{G}_{t+1} = \mathcal{U}_c(\mathcal{G}_t, s_t). \tag{3}$$

This formulation captures cognition as a distributed closed-loop process in which the fidelity of each step is directly constrained by communication dynamics, influencing temporal consistency, memory synchronization, and reasoning reliability across execution domains.

5.2.2 Execution as a control problem

Let $\mathcal{F} = \{f_1, \dots, f_n\}$ denote cognitive functions and $\mathcal{E} = \{\text{device, edge, cloud}\}$ denote execution domains. Each function f_i executed at domain e incurs a cost:

$$C_i(e) = [L_i(e),\ E_i(e),\ B_i(e)], \quad e \in \mathcal{E}, \tag{4}$$

where L_i, E_i, and B_i denote latency, energy, and communication cost, respectively. The execution policy maps cognitive functions to execution domains:

$$\pi_t : f_i \mapsto e \in \mathcal{E}, \tag{5}$$

and is obtained by solving:

$$\min_{\pi_t} \sum_i \left(w_L\, L_i(\pi_t(f_i)) + w_E\, E_i(\pi_t(f_i)) + w_B\, B_i(\pi_t(f_i))\right), \tag{6}$$

subject to:

$$L_i(\pi_t(f_i)) \le L_i^{\max}, \qquad E_i(\pi_t(f_i)) \le E_i^{\max}. \tag{7}$$

Edge execution minimizes latency and supports real-time response, whereas cloud execution enables higher-fidelity reasoning at the cost of additional communication delay. The optimal policy π_t therefore evolves continuously with both system dynamics and network conditions.

5.2.3 Multi-fidelity cognitive execution

Cognitive functions do not require uniform execution fidelity. Reasoning operates across three regimes reflecting different trade-offs between latency, complexity, and accuracy. L1 corresponds to lightweight real-time inference at the edge under strict latency constraints. L2 captures localized reasoning over partial system representations with moderate latency. L3 involves high-fidelity causal reasoning and long-horizon planning executed through cloud-assisted processing. The execution policy π_t dynamically selects among these regimes based on task criticality, resource availability, and communication conditions. Latency-critical tasks are preferentially mapped to L1, while non-critical or exploratory tasks leverage L2 or L3 reasoning. Cognitive execution thus follows a hierarchical structure in which fast, approximate decisions are continuously complemented and refined by slower, higher-fidelity processes.

5.2.4 Communication as a control variable

Network conditions directly constrain the feasible policy π_t. Increased latency, synchronization degradation, or bandwidth limitations may force transitions from L3 toward L1 execution, trading reasoning fidelity for responsiveness, while packet loss degrades memory retrieval and contextualization. This dependence is expressed as:

$$\pi_t = \arg\min_{\pi} \mathcal{J}(\pi, \mathcal{N}_t), \tag{8}$$

where $\mathcal{N}_t$ denotes the network state at time t. Cognition is therefore not statically deployed but continuously reconfigured in response to communication dynamics. The network does not merely support cognition; it actively shapes its structure, fidelity, and temporal behavior, motivating a unified design perspective in which communication, computation, and cognition are jointly optimized.

From a standardization perspective, this coupling aligns naturally with emerging O-RAN and 3GPP architectures. Near-real-time adaptive cognitive control aligns with edge-native orchestration frameworks such as O-RAN near-RT RIC xApps, whereas longer-horizon reasoning, policy adaptation, and memory coordination align with non-RT RIC rApps and cloud-assisted orchestration layers. Semantic state exchange and distributed cognitive coordination further motivate extensions to semantic communication primitives and distributed AI management frameworks envisioned within ongoing 3GPP, O-RAN, and ITU-T standardization activities.

6. WHAT COGNITIVE TWINS ENABLE: A CITY-SCALE EMERGENCY COORDINATION SCENARIO

The cognitive twin architecture enables a class of applications that extend beyond monitoring, prediction, recommendation, or autonomous control in isolation. Its distinguishing capability emerges from the continuous interaction of perception, memory, reasoning, communication, and adaptation under real-world operational constraints. To make this concrete, we examine a representative city-scale emergency coordination scenario that illustrates how these capabilities operate jointly within a large-scale 6G environment.

The purpose of this scenario is not to demonstrate application feasibility in isolation, but to illustrate operational conditions under which communication quality directly influences cognitive execution. In such systems, latency, synchronization accuracy, and the availability of semantically-critical information affect not only data delivery but also memory consistency, causal reasoning quality, and collective decision-making. This provides a concrete context for understanding why future cognitive twin systems increasingly require communication infrastructures capable of maintaining cognitive state in addition to transporting data.

Consider a large metropolitan area equipped with a 6G-native intelligent transportation and public safety infrastructure. A major traffic accident occurs at a busy intersection during peak hours, triggering a cascade of secondary effects, including road congestion, ambulance dispatch, adaptive traffic control, pedestrian safety management, and temporary spectrum congestion caused by increased communication activity. The affected environment comprises thousands of interacting entities, including vehicles, roadside units, cameras, drones, emergency responders, edge servers, and cloud-based management systems.

Under conventional digital twin architectures, the system can maintain synchronized representations of the physical environment and provide predictive analytics, yet decision-making remains largely dependent on centralized processing and current observations. LLM-based agents may generate recommendations based on retrieved information, but they do not maintain a continuously evolving cognitive state that is synchronized with distributed physical observations under strict latency constraints. A cognitive twin operates differently across each layer of the cognitive pipeline, as summarized in Table 6.

At the perceptual layer, multimodal observations from cameras, connected vehicles, infrastructure sensors, and aerial drones are fused into a temporally-consistent representation of the evolving situation. This fusion must occur under strict synchronization constraints because de-

Table 6 – Role of cognitive twin layers in the emergency coordination scenario.

Layer	Role in Scenario
Perception	Fusion of multimodal observations from vehicles, sensors, cameras, and drones into a temporally consistent system representation
Memory	Retrieval of relevant experiences and operational knowledge from prior incidents to contextualize current conditions
Reasoning	Evaluation of interventions and causal impact on future system evolution under dynamic system conditions
Interaction	Communication of decisions, explanations, confidence estimates, and coordination actions to human operators and distributed entities

layed or misaligned observations can produce conflicting interpretations of the environment, directly affecting the quality of downstream reasoning. At the memory layer, the system continuously retrieves relevant experiences from prior incidents. Rather than searching raw historical logs, structured episodes describing previous accidents with similar traffic density, environmental conditions, response times, and coordination outcomes are recalled and used to contextualize the current situation. Simultaneously, semantic memory provides domain knowledge regarding traffic management policies, emergency response procedures, communication constraints, and safety regulations. At the reasoning layer, current observations are combined with retrieved experience to construct causal explanations and evaluate potential interventions. Rather than predicting future congestion levels in isolation, the system estimates how alternative actions such as rerouting traffic, modifying signal schedules, reallocating communication resources, or dispatching additional emergency units influence future system behavior. Decisions, therefore, emerge from both current observations and accumulated operational experience.

A defining property of this scenario is that cognition continues even when communication quality deteriorates. Temporary network congestion, partial infrastructure failures, or localized coverage disruptions may prevent complete system synchronization. Rather than halting operation, local cognitive twins maintain reduced but functional cognitive states using cached memory, localized reasoning, and semantic state representations. Upon reconnection, compact cognitive updates are exchanged to restore global consistency without requiring full retransmission of raw observations.

Communication, therefore, becomes part of the cognitive process itself. The system does not simply exchange sensor measurements; it exchanges compressed representations of beliefs, intentions, confidence estimates, inferred relationships, and coordination decisions. This significantly reduces communication overhead while preserving the information most relevant to collective reasoning. At a broader scale, experience acquired during one incident becomes reusable across future events through structured cognitive updates rather than centralized collection of massive datasets, enabling the system to develop a distributed form of collective intelligence over time.

This scenario highlights the defining characteristics of cognitive twins as maintaining perception under communication uncertainty, retrieving and reusing structured experience, performing causal and intervention-aware reasoning, adapting execution across device, edge, and cloud domains, and exchanging cognitive state rather than raw data. Taken together, these properties establish communication quality as a direct determinant of cognitive fidelity, making the joint design of communication, computation, memory, and reasoning a fundamental requirement for future 6G-enabled cognitive twin systems.

7. OPEN RESEARCH PROBLEMS FOR THE WIRELESS COMMUNITY

The cognitive twin paradigm introduces a fundamentally new operating regime for wireless systems in which communication, computation, memory, and reasoning become tightly coupled. The application scenario examined in the previous section illustrates concretely how communication quality shapes cognitive fidelity: degraded synchronization distorts causal inference, packet loss disrupts memory retrieval, and bandwidth constraints limit the complexity of maintainable cognitive representations. These dependencies transform classical design objectives such as throughput, latency, and reliability into cognitive systems problems. The key question is no longer how to deliver data efficiently, but how to preserve, synchronize, and evolve *cognitive state* under dynamic wireless conditions. This shift expands the scope of wireless system design beyond information delivery toward maintaining the consistency, integrity, and evolution of distributed cognitive processes. Table 7 summarizes the resulting research challenges and their implications for wireless system design. A defining property of cognitive twin systems is that communication impairments not only degrade data delivery but also directly affect cognitive correctness. Latency, packet loss, and synchronization errors propagate through perception, memory, and reasoning, leading to inconsistencies in the inferred system state. This fundamentally changes how reliability should be interpreted: packet-level success does not guarantee correct reasoning. Systems must instead ensure *semantic reliability*, where the objective is to preserve the information required for accurate and consistent cognitive inference.

An important open question concerns the limits of graceful degradation. While cognitive twins are designed to operate under partial observations and communication impairments, severe synchronization errors, persistent loss of semantically-critical information, or cascading memory

Table 7 – Open research problems in cognitive twin systems and their implications for wireless communication design.

Research problem	Cognitive impact	Wireless challenge	Research direction
Real-time causal inference under imperfect observations	Degraded, inconsistent, or misleading reasoning under missing data	Packet loss, latency jitter, synchronization errors affecting temporal consistency	Causality-aware communication, semantic prioritization, cross-layer co-design
Privacy-preserving cognitive state synchronization	Limited sharing of episodic memory and contextual knowledge	Latency–privacy–overhead trade-off under distributed execution	Over-the-air aggregation, lightweight privacy mechanisms, distributed secure learning
Energy-constrained edge cognition	Limited local execution of perception, memory, and reasoning	Joint computation–communication energy trade-offs at the edge	Energy-aware offloading, wireless power transfer, hardware-aware model design
Semantic compression of cognitive state	Loss of reasoning fidelity under aggressive compression	Bandwidth constraints for high-dimensional cognitive representations	Semantic communication, joint source–channel coding, learned compression for structured state
Mobility-aware cognitive state management	Loss of cognitive continuity during handoff and mobility events	Low-latency state migration under dynamic network topology	Predictive state migration, cognition-aware handoff, context-aware mobility management
Verification and validation of cognitive behavior	Difficulty in defining correctness, reliability, and consistency of reasoning	Lack of evaluation frameworks under realistic wireless impairments	Over-the-air testing, adversarial evaluation, cognitive benchmarks, and stress testing
Standardization and interoperability	Lack of interoperable cognitive-state representations and coordination mechanisms	Integration with heterogeneous wireless, edge, cloud, and AI-native infrastructures	3GPP/O-RAN/ITU-T alignment, cognitive state formats, interoperability frameworks

inconsistencies may eventually exceed the system's ability to maintain coherent reasoning. Characterizing these breakdown thresholds and designing mechanisms that detect and mitigate impending cognitive instability remain largely unexplored research directions.

7.1 Real-time causal inference under imperfect observations

Causal reasoning in cognitive twins assumes temporally-aligned and semantically-complete observations, an assumption that is frequently violated in practical wireless environments due to missing, delayed, or inconsistent data. As a result, inferred causal relationships may become unreliable or misleading. Designing inference mechanisms that remain robust under such conditions is a key challenge. Rather than producing brittle outputs, cognitive twins must generate uncertainty-aware reasoning results that explicitly reflect missing or inconsistent information. The impact of communication impairments on causal inference is inherently asymmetric: loss of semantically-critical inputs can disproportionately degrade reasoning outcomes compared to equivalent losses in redundant data streams. Communication strategies must therefore become causality-aware, incorporating priority-driven scheduling, adaptive retransmission, and feedback mechanisms that reflect semantic completeness rather than packet success. A fundamental challenge is that causal completeness cannot be directly observed at the communication layer. Determining which observations are causally critical may itself require contextual reasoning, creating a tightly coupled dependency between communication and cognition.

7.2 Privacy-preserving distributed cognitive state synchronization

cognitive twins maintain rich internal representations, including episodic memory and contextual knowledge, which may contain sensitive operational information. Enabling distributed cognition while preserving privacy introduces a fundamental tension between information sharing and protection. Conventional techniques such as differential privacy and secure computation incur additional communication and computation overhead that may conflict with real-time constraints. Privacy must therefore be integrated into the communication design rather than treated as an external constraint. Emerging directions include over-the-air aggregation, channel-aware privacy mechanisms, and distributed learning schemes that align privacy preservation with communication efficiency.

Unlike conventional digital twins, cognitive twins introduce additional attack surfaces because episodic memory, semantic knowledge, and reasoning outputs may reveal behavioral patterns, operational intent, or historical decision traces. Potential threats include cognitive state inference attacks, memory poisoning, semantic manipulation, and unauthorized reconstruction of contextual knowledge from exchanged cognitive updates. Understanding how such attacks propagate through distributed cognitive systems, and how communication protocols can mitigate them, remains an important open challenge.

7.3 Energy-constrained cognitive processing at the edge

Real-time cognitive operation requires executing perception, memory, and reasoning functions at the edge, where energy resources are inherently limited. Even lightweight models can impose significant energy demands under continuous execution, introducing a fundamental trade-off between latency, energy consumption, and reasoning fidelity. Edge execution reduces latency but increases energy cost, whereas cloud execution improves reasoning capability at the cost of responsiveness. Addressing this challenge requires adaptive execution strategies that jointly optimize computation and communication under energy constraints. Promising directions include energy-aware offloading, wireless power transfer, and hardware-aware model design that explicitly accounts for cognitive workload characteristics.

7.4 Semantic compression and communication of cognitive state

The internal state of a cognitive twin is inherently high-dimensional, making direct transmission infeasible at scale. Systems must instead exchange compact representations that preserve the information required for consistent reasoning. This introduces a fundamental trade-off between communication efficiency and cognitive fidelity: semantic compression reduces bandwidth usage but may degrade reasoning accuracy if critical information is lost. A key research challenge is to characterize this trade-off formally and design communication schemes that prioritize semantic relevance over bit-level fidelity. Relevant directions include joint source–channel coding and learning-based compression techniques tailored to structured cognitive representations.

7.5 Mobility-aware cognitive state management

Mobility requires cognitive twins to maintain continuity as devices move across network regions. Unlike traditional handoff mechanisms, this involves transferring cognitive state, including memory, context, and inferred relationships, under strict latency constraints. A central challenge is determining the minimal subset of states required to preserve cognitive continuity while minimizing communication overhead. Predictive approaches such as trajectory-aware state migration offer potential solutions but introduce additional uncertainty and complexity. This problem calls for mobility management mechanisms that tightly integrate communication and cognition, extending traditional handoff protocols toward context-aware and predictive designs.

7.6 Verification and validation of cognitive behavior

Verification and validation are among the most critical challenges for cognitive twin systems. Unlike conventional systems evaluated using numerical accuracy metrics, cognitive twins produce probabilistic, context-dependent outputs often expressed in structured or natural language forms. This necessitates new evaluation paradigms based on measurable proxies such as reasoning consistency, robustness to perturbations, and alignment with system objectives. Validation must explicitly account for the impact of wireless impairments on cognitive behavior. A promising direction is hypothesis-driven validation, where system behavior is evaluated under controlled scenarios with explicit performance criteria. Over-the-air testing, adversarial evaluation, and benchmark design for cognitive workloads are essential to establishing trust and deployment readiness.

7.7 Standardization and ecosystem challenges

Realizing cognitive twins at scale will ultimately require support from future standardization frameworks. Several cognitive twin functions naturally align with ongoing activities within 3GPP, O-RAN, and ITU-T. Semantic information exchange may build upon emerging semantic communication frameworks, while distributed cognitive orchestration can leverage AI-native management functions developed for beyond 5G and 6G systems. Cognitive state management, memory synchronization, and reasoning coordination may further require standardized representation formats, interoperability interfaces, and lifecycle management procedures. An important research challenge is therefore to identify which cognitive twin abstractions should be exposed to standardization bodies and how they can be integrated into existing communication and network management architectures without sacrificing interoperability.

These challenges collectively define a fundamental shift in wireless system design. Future communication infrastructures must not only deliver data but actively support the formation, synchronization, protection, and evolution of distributed cognitive state. Advancing this agenda will require coordinated progress across communication theory, distributed AI, edge computing, privacy engineering, and standardization, establishing a new interdisciplinary research space at the intersection of communication systems and cognitive system design.

8. TOWARD REALIZING COGNITIVE TWINS: A RESEARCH ROADMAP

While the cognitive twin paradigm introduces a new class of systems, its practical realization requires both a structured development pathway and a systematic validation methodology. Unlike conventional systems evaluated through well-defined metrics such as accuracy or throughput, cognitive twins must be assessed in terms of cognitive behavior, including reasoning consistency, robustness under uncertainty, and alignment with system objectives. This shift implies that evaluation must move beyond data-centric metrics toward cognition-centric criteria, as illustrated in Fig. 4. The central question is no longer whether a system predicts accurately, but whether it maintains a consistent, reliable, and actionable cognitive state under real-world constraints.

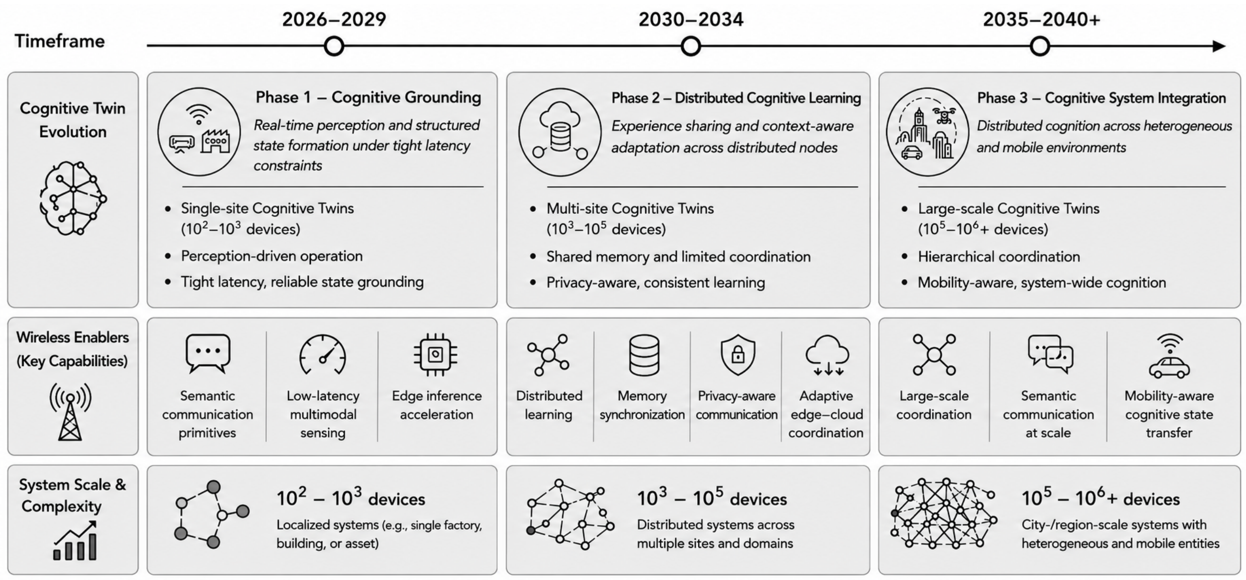

Figure 4 – Progressive realization of cognitive twin systems. The roadmap illustrates the co-evolution of wireless communication capabilities, cognitive functionalities, and system scale, progressing from perception-centric cognitive twins toward large-scale distributed cognition.

Table 8 – Indicative research roadmap for cognitive twin systems and their wireless enablers. Timeframes are approximate and reflect the progressive development of capabilities rather than guaranteed deployment milestones.

Timeframe	Wireless Research Priorities	cognitive twin Milestones	Phase Description
2026–2029	Semantic communication primitives; ultra-low-latency multimodal sensing; edge inference acceleration	Single-site cognitive twins (10^2–10^3 devices) with perception-centric operation	Phase 1 – Cognitive Grounding: Establishing reliable perception and temporally consistent state formation under strict latency and synchronization constraints
2030–2034	Distributed learning; memory synchronization; privacy-aware communication; adaptive edge–cloud orchestration	Multi-site cognitive twins (10^3–10^5 devices) with shared memory and coordinated adaptation	Phase 2 – Distributed Cognitive Learning: Enabling experience sharing and context-aware adaptation across distributed nodes under communication constraints
2035–2040	Large-scale coordination; semantic communication at scale; mobility-aware cognitive state transfer	Large-scale cognitive twin systems (10^5–10^6 devices) with hierarchical and mobile coordination	Phase 3 – Cognitive System Integration: Achieving distributed cognition across heterogeneous, large-scale, and dynamic environments

8.1 Toward validation of cognitive twin systems

A central challenge in advancing cognitive twins is the absence of established validation frameworks. Cognitive outputs are inherently context-dependent, probabilistic, and often involve structured or natural language explanations, so correctness cannot be defined solely by numerical error metrics. To address this, we adopt a hypothesis-driven validation perspective. Rather than treating system properties as established results, each claim is formulated as a testable hypothesis with a corresponding evaluation methodology. For example, the effectiveness of semantic communication can be evaluated by comparing cognitive state transmission with raw data transmission in terms of bandwidth efficiency and downstream reasoning accuracy. Similarly, the impact of communication impairments can be quantified by measuring degradation in memory retrieval consistency or decision stability under controlled conditions of latency and packet loss. This approach shifts validation from static benchmarking to controlled experimentation in which communication conditions, system scale, and reasoning complexity are systematically varied, requiring integrated simulation environments that jointly model wireless networks and cognitive processing pipelines.

A particularly important challenge is reproducibility. Future cognitive twin evaluations should report not only communication parameters such as latency, packet loss, synchronization accuracy, and bandwidth availability, but also cognitive metrics including memory retrieval consistency, reasoning stability, uncertainty calibration, and cognitive state divergence across distributed nodes. Establishing such evaluation practices will be essential for enabling fair comparison across future cognitive twin implementations.

While existing digital twin platforms and open datasets provide a useful starting point, they are insufficient to capture the interaction between cognition and communication. New benchmarks are therefore required to evaluate the impact of communication constraints on perception coherence, memory consistency, and reasoning fidelity. Establishing standardized evaluation methodologies is a critical step toward transitioning cognitive twins from conceptual architectures to deployable systems.

8.2 A progressive roadmap for cognitive twin realization

Realizing cognitive twins requires a gradual co-evolution of wireless communication capabilities and cognitive system design. This evolution should not be interpreted as a fixed deployment timeline but as a progression of capabilities constrained by practical system limitations. Large-scale cognitive twin systems involving hundreds of thousands or millions of nodes, as well as high-fidelity causal reasoning at scale, remain long-term objectives whose realization depends on advances that are currently open research problems, including scalable semantic communication, distributed memory consistency, and energy-efficient edge intelligence. Table 8 summarizes an indicative roadmap capturing this progression across three phases.

The first phase, *cognitive grounding*, focuses on establishing reliable perception and structured state representation under strict wireless constraints. Advances in ultra-low-latency communication and integrated sensing enable systems to construct temporally-consistent multimodal representations at the edge. At this stage, cognitive twins remain localized and primarily support perception-driven applications. The central challenge is ensuring that sensed data can be fused and synchronized in real time without introducing inconsistencies that propagate to higher layers. The objective is therefore not full cognition, but stable and reliable grounding of the system state.

The second phase, *distributed cognitive learning*, introduces memory sharing and collaborative adaptation across distributed nodes. Cognitive twins begin exchanging experience through federated, privacy-preserving mechanisms, enabling coordinated learning across multiple sites. The central challenge becomes maintaining consistency of cognitive state under communication constraints: synchronizing memory, context, and learned representations across nodes must be achieved without excessive communication overhead or privacy leakage. This phase marks the transition from isolated intelligence to distributed cognition. From a standardization perspective, this phase is likely to coincide with the emergence of interoperable cognitive state representations and semantic information exchange mechanisms. Existing efforts within 3GPP, O-RAN, and ITU-T provide a potential foundation, particularly in semantic communication, AI-native network management, and edge–cloud orchestration, though substantial work remains in defining interoperable formats for memory synchronization and distributed reasoning coordination.

The final phase, *cognitive system integration*, targets large-scale deployment under dynamic conditions. cognitive twins operate in heterogeneous, mobile environments, requiring robust synchronization of memory, context, and reasoning among distributed entities. Mobility-aware protocols, semantic communication, and hierarchical coordination mechanisms become essential. Even at this stage, high-fidelity causal reasoning is not assumed to operate in real time at the system-wide scale. Full counterfactual analysis is performed asynchronously or over localized subsets of the system, while real-time decisions rely on approximate or learned models. This explicitly reflects the computational and communication limits of large-scale cognitive systems.

Three fundamental observations emerge from this roadmap. First, realizing cognitive twins requires coordinated advances across communication, computation, and cognitive system design rather than isolated improvements in any single domain. Second, different cognitive capabilities become feasible at different stages, depending on the maturity of underlying wireless technologies and system infrastructure. Third, the transition from state synchronization to distributed cognition is inherently gradual and constrained by system-level limitations in both communication and computation.

From a wireless perspective, this progression reinforces a central conclusion: future communication systems must evolve beyond data transport to support cognition-aware operation. The ability to exchange not only data but also context, intent, and inferred knowledge will be a defining characteristic of next-generation networks. The realization of cognitive twins is therefore not only an application of 6G systems, but a driver for fundamentally rethinking their design. In this sense, cognitive twins represent not merely a new application domain but a potential organizing principle for future communication systems in which communication, computation, memory, and reasoning evolve as a unified system architecture.

9. CONCLUSION

This article introduced the cognitive twin as a new system paradigm that extends beyond traditional digital twins and their learning-based variants. By integrating perception, structured memory, reasoning, and interaction into a unified closed-loop architecture, cognitive twins enable a transition from data-driven monitoring to understanding-driven operation. This paradigm represents a qualitative shift from state replication to distributed cognition: unlike existing approaches, cognitive twins explicitly incorporate episodic and semantic memory, support multi-level causal reasoning, and enable interaction grounded in system state and experience, giving rise to operational capabilities that cannot be realized within conventional digital twin frameworks.

A central insight of this work is that enabling such capabilities requires rethinking system design rather than simply improving models. In cognitive twin architectures, communication is elevated from a transport mechanism to a core component of cognition. Maintaining consistent context, synchronizing distributed memory, and supporting real-time decision-making impose tightly coupled requirements on latency, bandwidth, reliability, and synchronization. As a result, communication, computation, and cognition must be jointly designed rather than treated as independent subsystems. From this perspective, cognitive twins are inherently aligned with the vision of 6G systems. Capabilities such as ultra-low latency communication, integrated sensing and communication, semantic communication, and the edge–cloud continuum provide the necessary substrate for distributed cognitive operation. At the same time, the requirements imposed by cognitive twins offer a new system-level lens for the design, optimization, and evaluation of next-generation wireless networks.

This work is a roadmap rather than a validation study. The quantitative claims presented throughout should be interpreted as testable hypotheses that motivate future simulation, prototyping, and system-level evaluation. The open research problems and structured roadmap outlined here define a concrete path from conceptual abstraction to engineering realization.

Ultimately, cognitive twins should not be viewed as an incremental extension of existing paradigms or a direct application of current AI systems. They represent a new class of cyber-physical systems in which distributed cognition emerges from the interaction of perception, memory, reasoning, and communication under real-world constraints. The transition from digital twins to cognitive twins is therefore not merely a technological evolution, but a shift toward cognition-centric system design. Realizing this vision will require rethinking wireless networks not only as communication infrastructures but as programmable substrates that actively shape the fidelity, consistency, and scalability of distributed intelligence.

REFERENCES

[1] Yuanhao Cui, Fan Liu, Xiaojun Jing, and Junsheng Mu. "Integrating Sensing and Communications for Ubiquitous IoT: Applications, Trends, and Challenges". In: *IEEE Network* 35.5 (2021), pp. 158–167. DOI: 10.1109/MNET.010.2100152.

[2] Latif U. Khan, Walid Saad, Dusit Niyato, Zhu Han, and Choong Seon Hong. "Digital-Twin-Enabled 6G: Vision, Architectural Trends, and Future Directions". In: *IEEE Communications Magazine* 60.1 (2022), pp. 74–80. DOI: 10.1109/MCOM.001.21143.

[3] Huan X. Nguyen, Ramona Trestian, Duc To, and Mallik Tatipamula. "Digital Twin for 5G and Beyond". In: *IEEE Communications Magazine* 59.2 (2021), pp. 10–15. DOI: 10.1109/MCOM.001.2000343.

[4] Elif Bozkaya, Müge Erel-Özçevik, Tuğçe Bilen, and Yusuf Özçevik. "Proof of Evaluation-based energy and delay aware computation offloading for Digital Twin Edge Network". In: *Ad Hoc Networks* 149 (2023), p. 103254. ISSN: 1570-8705. DOI: https://doi.org/10.1016/j.adhoc.2023.103254. URL: https://www.sciencedirect.com/science/article/pii/S1570870523001749.

[5] Müge Erel-Özçevik, Yusuf Özçevik, Elif Bozkaya-Aras, and Tuğçe Bilen. "6G-enabled digital twin-based smart warehouse for supply chain of things". In: *Computing* 108.6 (2026), p. 79. ISSN: 1436-5057. DOI: 10.1007/s00607-026-01668-3. URL: https://doi.org/10.1007/s00607-026-01668-3.

[6] Ian F. Akyildiz, Ahan Kak, and Shuai Nie. "6G and Beyond: The Future of Wireless Communications Systems". In: *IEEE Access* 8 (2020), pp. 133995–134030. DOI: 10.1109/ACCESS.2020.3010896.

[7] Guangyi Liu, Yuhong Huang, Na Li, Jing Dong, Jing Jin, Qixing Wang, and Nan Li. "Vision, requirements and network architecture of 6G mobile network beyond 2030". In: *China Communications* 17.9 (2020), pp. 92–104. DOI: 10.23919/JCC.2020.09.008.

[8] Tuğçe Bilen and Berk Canberk. "Q-Learning Driven Routing for Aeronautical Ad-Hoc Networks". In: *Pervasive and Mobile Computing* 87 (2022), p. 101724. ISSN: 1574-1192. DOI: https://doi.org/10.1016/j.pmcj.2022.101724. URL: https://www.sciencedirect.com/science/article/pii/S1574119222001377.

[9] Ian F. Akyildiz and Tuğçe Bilen. "The Network That Thinks: Kraken and the Dawn of Cognitive 6G". In: *IEEE Network* (2026), pp. 1–13. DOI: 10.1109/MNET.2026.3699341.

[10] Weisong Shi, Jie Cao, Quan Zhang, Youhuizi Li, and Lanyu Xu. "Edge Computing: Vision and Challenges". In: *IEEE Internet of Things Journal* 3.5 (2016), pp. 637–646. DOI: 10.1109/JIOT.2016.2579198.

[11] Mahmoud Abdelrahman, Edgardo Macatulad, Binyu Lei, Matias Quintana, Clayton Miller, and Filip Biljecki. "What is a Digital Twin anyway? Deriving the definition for the built environment from over 15,000 scientific publications". In: *Building and Environment* 274 (2025), p. 112748. ISSN: 0360-1323. DOI: https://doi.org/10.1016/j.buildenv.2025.112748.

[12] Tuğçe Bilen, Elif Ak, Bahadır Bal, and Berk Canberk. "A Proof of Concept on Digital Twin-Controlled WiFi Core Network Selection for In-Flight Connectivity". In: *IEEE Communications Standards Magazine* 6.3 (2022), pp. 60–68. DOI: 10.1109/MCOMSTD.0001.2100103.

[13] Sekione Reward Jeremiah, Abir El Azzaoui, Neal N. Xiong, and Jong Hyuk Park. "A comprehensive survey of digital twins: Applications, technologies and security challenges". In: *Journal of Systems Architecture* 151 (2024), p. 103120. ISSN: 1383-7621. DOI: https://doi.org/10.1016/j.sysarc.2024.103120. URL: https://www.sciencedirect.com/science/article/pii/S1383762124000572.

[14] Yuntao Wang, Zhou Su, Shaolong Guo, Minghui Dai, Tom H. Luan, and Yiliang Liu. "A Survey on Digital Twins: Architecture, Enabling Technologies, Security and Privacy, and Future Prospects". In: *IEEE Internet of Things Journal* 10.17 (2023), pp. 14965–14987. DOI: 10.1109/JIOT.2023.3263909.

[15] Hansong Xu, Jun Wu, Qianqian Pan, Xinping Guan, and Mohsen Guizani. "A Survey on Digital Twin for Industrial Internet of Things: Applications, Technologies and Tools". In: *Commun. Surveys Tuts.* 25.4 (Oct. 2023), pp. 2569–2598. ISSN: 1553-877X. DOI: 10.1109/COMST.2023.3297395.

[16] Marko Vuković, Daniele Mazzei, Stefano Chessa, and Gualtiero Fantoni. "Digital Twins in Industrial IoT: a survey of the state of the art and of relevant standards". In: *IEEE International Conference on Communications Workshops (ICC Workshops)*. 2021, pp. 1–6. DOI: 10.1109/ICCWorkshops50388.2021.9473889.

[17] Camilo E. Rojas-Sanchez, Juan Antonio Paredes, Pau Ferrer-Cid, Jose M. Barcelo-Ordinas, and Jorge Garcia-Vidal. "IoT-Based Digital Twin Model for Industrial Applications". In: *International Conference on Modeling, Analysis and Simulation of Wireless and Mobile Systems (MSWiM)*. 2025, pp. 771–774. DOI: 10.1109/MSWiM67937.2025.11309159.

[18] Yiwen Wu, Ke Zhang, and Yan Zhang. "Digital Twin Networks: A Survey". In: *IEEE Internet of Things Journal* 8.18 (2021), pp. 13789–13804. DOI: 10.1109/JIOT.2021.3079510.

[19] Raymon van Dinter, Bedir Tekinerdogan, and Cagatay Catal. "Predictive maintenance using digital twins: A systematic literature review". In: *Information and Software Technology* 151 (2022), p. 107008. ISSN: 0950-5849. DOI: https://doi.org/10.1016/j.infsof.2022.107008.

[20] Tim Kreuzer, Panagiotis Papapetrou, and Jelena Zdravkovic. "Artificial intelligence in digital twins—A systematic literature review". In: *Data Knowledge Engineering* 151 (2024), p. 102304. ISSN: 0169-023X. DOI: https://doi.org/10.1016/j.datak.2024.102304.

[21] Mike Reichardt, Michael Gundall, Daniel Buch, Philip Stricker, and Hans D. Schotten. "Artificial Intelligence Meets Digital Twin: The AI Asset Administration Shell". In: *2024 IEEE 3rd Industrial Electronics Society Annual On-Line Conference (ONCON)*. 2024, pp. 1–6. DOI: 10.1109/ONCON62778.2024.10931277.

[22] Tuğçe Bilen, Berk Canberk, and Trung Q. Duong. "Digital Twin Evolution for Hard-to-Follow Aeronautical Ad-Hoc Networks in Beyond 5G". In: *IEEE Communications Standards Magazine* 7.1 (2023), pp. 4–12. DOI: 10.1109/MCOMSTD.0001.2200040.

[23] Muhammad Intizar Ali, Pankesh Patel, John Breslin, Ramy Harik, and Amit Sheth. "Cognitive Digital Twins for Smart Manufacturing". In: *IEEE Intelligent Systems* 36 (Mar. 2021), pp. 96–100. DOI: 10.1109/MIS.2021.3062437.

[24] Trier Mortlock, Deepan Muthirayan, Shih-Yuan Yu, Pramod P. Khargonekar, and Mohammad Abdullah Al Faruque. " Graph Learning for Cognitive Digital Twins in Manufacturing Systems ". In: *IEEE Transactions on Emerging Topics in Computing* 10.01 (Jan. 2022), pp. 34–45. ISSN: 2168-6750. DOI: 10.1109/TETC.2021.3132251.

[25] Luwen Huang, Michael Kapteyn, and Karen E. Willcox. "Digital Twin: Graph Formulations for Managing Complexity and Uncertainty". In: *IEEE Smart World Congress (SWC)*. 2024, pp. 2141–2148. DOI: 10.1109/SWC62898.2024.00327.

[26] Georgia Stavropoulou, Konstantinos Tsitseklis, Lydia Mavraidi, Kuo-I Chang, Anastasios Zafeiropoulos, Vasileios Karyotis, and Symeon Papavassiliou. "Digital Twin Meets Knowledge Graph for Intelligent Manufacturing Processes". In: *Sensors (Basel, Switzerland)* 24 (2024).

[27] Franz Listl, Daniel Dittler, Gary Hildebrandt, Valentin Stegmaier, Nasser Jazdi, and Michael Weyrich. *Knowledge Graphs in the Digital Twin: A Systematic Literature Review About the Combination of Semantic Technologies and Simulation in Industrial Automation*. June 2024. DOI: 10.48550/arXiv.2406.09042.

[28] Mohammad Abdullah Al Faruque, Deepan Muthirayan, Shih-Yuan Yu, and Pramod P. Khargonekar. "Cognitive Digital Twin for Manufacturing Systems". In: *Design, Automation Test in Europe Conference Exhibition (DATE)*. 2021, pp. 440–445. DOI: 10.23919/DATE51398.2021.9474166.

[29] Patrick Lewis, Ethan Perez, Aleksandra Piktus, Fabio Petroni, Vladimir Karpukhin, Naman Goyal, Heinrich Küttler, Mike Lewis, Wen-tau Yih, Tim Rocktäschel, Sebastian Riedel, and Douwe Kiela. "Retrieval-augmented generation for knowledge-intensive NLP tasks". In: *Proceedings of the 34th International Conference on Neural Information Processing Systems*. NIPS '20. Vancouver, BC, Canada: Curran Associates Inc., 2020. ISBN: 9781713829546.

[30] Kevin Hoover. "Causality: Models, Reasoning, and Inference." In: *The Economic Journal* 113 (June 2003). DOI: 10.1111/1468-0297.13919.

[31] Endel Tulving. "1O Episodic and Semantic". In: *Organization of Memory* (1972), p. 381.

[32] Aidan Hogan, Eva Blomqvist, Michael Cochez, Claudia D'amato, Gerard De Melo, Claudio Gutierrez, Sabrina Kirrane, José Emilio Labra Gayo, Roberto Navigli, Sebastian Neumaier, Axel-Cyrille Ngonga Ngomo, Axel Polleres, Sabbir M. Rashid, Anisa Rula, Lukas Schmelzeisen, Juan Sequeda, Steffen Staab, and Antoine Zimmermann. "Knowledge Graphs". In: *ACM Comput. Surv.* 54.4 (July 2021). ISSN: 0360-0300. DOI: 10.1145/3447772. URL: https://doi.org/10.1145/3447772.

[33] Tuğçe Bilen. "KDN-Driven zero-shot learning for intelligent self-healing in 6G small cell networks". In: *Ad Hoc Networks* 178 (2025), p. 103984. ISSN: 1570-8705. DOI: https://doi.org/10.1016/j.adhoc.2025.103984. URL: https://www.sciencedirect.com/science/article/pii/S1570870525002320X.

AUTHORS

IAN F. AKYILDIZ received his B.S., M.S., and Ph.D. degrees in electrical and computer engineering from the University of Erlangen–Nürnberg, Germany, in 1978, 1981, and 1984, respectively. From 1985 to 2020, he held the Ken Byers Chair Professorship at Georgia Tech, where he directed the Broadband Wireless Networking Laboratory. A visionary entrepreneur, he is the President of Truva Inc. and a key advisor to global institutions like TII (Abu Dhabi) and Odine Labs (Istanbul). Since 2020, he has served as the founding Editor-in-Chief of the ITU Journal on Future and Evolving Technologies. His pioneering research spans 6G/7G systems, molecular communication, terahertz technology, and underwater networking. As of March 2026, he holds an H-index of 146 with over 155 000 citations. Dr. Akyildiz is an ACM Fellow and recipient of prestigious honors, including the Humboldt (Germany) and TÜBİTAK (Türkiye) Awards.

TUĞÇE BILEN received her B.Sc., M.Sc., and Ph.D. degrees in computer engineering from Istanbul Technical University (ITU) in 2015, 2017, and 2022, respectively. She is currently an assistant professor in the Department of Artificial Intelligence and Data Engineering at ITU, where she previously served as a research and teaching assistant. Her doctoral research earned several prestigious honors, including the 2025 IEEE Turkey Section Ph.D. Thesis Award, the 2023 Turkish Academy of Sciences (TÜBA) First Prize in Science and Technology, the 2023 Serhat Özyar Young Scientist Honorary Award, and the 2022 ITU Best Ph.D. Thesis Award. Her research focuses on 6G networks, Knowledge-Defined Networking (KDN), AI-driven network management, and digital twins, with a particular emphasis on integrating intelligent systems into future architectures. Dr. Bilen also serves as a reviewer for leading international journals.