从数字孪生到世界模型：移动边缘通用智能的机遇、挑战与应用

郑杰[1]，Dusit Niyato[2]，赵常源[2]，康嘉文[3]，王嘉诚[2]

（1.西北大学计算机学院，陕西 西安 710127；2.新加坡南洋理工大学计算与数据科学学院，海外 其他市 639798；3.广东工业大学自动化学院，广东省 广州市 510006）

摘　要：未来6G通信系统的快速演进正加速数字孪生与世界模型在网络边缘的融合。传统数字孪生为物理系统提供高保真表征，并支持监控、分析与离线优化，但在高度动态的边缘环境中，其在自主性、适应性和可扩展性方面存在不足。本文系统综述了从数字孪生向世界模型的演进及其对边缘通用智能的支撑作用。首先，对比两者的概念差异：由基于物理、集中式、以系统为中心的数字孪生，转向数据驱动、去中心化、以智能体为中心的内部模型，从而实现更具适应性、自主性与资源效率的边缘智能。其次，介绍世界模型的设计原则、体系结构及关键组件，包括感知、潜在状态表征、动力学学习、基于想象的规划与记忆机制。进一步讨论世界模型与数字孪生在无线边缘通用智能系统中的融合，并分析其在通感一体化、语义通信、空地一体网络及低空无线网络中的应用。最后，梳理面向无线与边缘计算环境的设计路线与开源项目，并总结面向边缘原生智能体AI的可扩展、可靠与可互操作世界模型的关键挑战与未来方向。

关键词：移动边缘通用智能；数字孪生；世界模型
中图分类号：TP309
文献标志码：A
DOI：10.11959/j.issn.1000

From Digital Twins to World Models: Opportunities, Challenges, and Applications for Mobile Edge General Intelligence

Abstract: The rapid evolution of future 6G communication systems is accelerating the integration of digital twins and world models at the network edge. Traditional digital twins provide high-fidelity representations of physical systems and support monitoring, analysis, and offline optimization. However, in highly dynamic edge environments, they face limitations in autonomy, adaptability, and scalability. This paper presents a systematic survey of the transition from digital twins to world models and their role in enabling edge general intelligence. First, we compare their conceptual differences. Digital twins are typically physics-based, centralized, and system-centric, while world models are data-driven, decentralized, and agent-centric internal models. This shift enables more adaptive, autonomous, and resource-efficient edge intelligence. Next, we introduce the design principles, architecture, and key components of world models, including perception, latent state representation, dynamics learning, imagination-based planning, and memory mechanisms. We then discuss the integration of world models and digital twins in wireless edge general intelligence systems, and analyze their applications in integrated sensing and communication, semantic communication, air-ground integrated networks, and low-altitude wireless networks. Finally, we outline design guidelines and open-source efforts for wireless and edge com-

收稿日期：XXXX–XX–XX；修回日期：XXXX–XX–XX
通信作者：郑杰，jzheng@nwu.edu.cn
基金项目：国家重点研发计划资助项目(No.2023YFC3304605)；国家自然科学基金资助项目（No. 62572132）；
Foundation Items：The National Key R&D Program of China (No.2023YFC3304605),The National Natural Science Foundation of China (No. 62572132)

puting environments, and summarize key challenges and future directions for building scalable, reliable, and interoperable world models for edge-native intelligent agents.

Keywords: mobile edge general intelligence, digital twin, world model

0 INTRODUCTION

A. Background

Edge computing is shifting from task-specific edge artificial intelligence (Edge AI) to edge general intelligence (EGI)[1]. EGI refers to intelligent systems deployed near the physical world that support long-term autonomous operation across multiple tasks and environments. It enables low or ultra-low latency inference while reducing energy use and cost, which is suitable for resource-limited devices[2]. Traditional Edge AI focuses on running predefined algorithms near data sources to meet low-latency and high-efficiency needs, such as autonomous driving[3]. In contrast, EGI agents can perceive, reason, and act in dynamic and partially observable environments.

By running models on edge devices, edge intelligence supports local data processing and reduces dependence on the cloud. This improves latency, privacy, and security since data stay local[4]. Typical applications include UAV networks[5], intelligent transportation systems[6], and smart industry systems[7]. In these cases, EGI works closely with physical systems under real-time and resource constraints. This forms a closed loop of perception, decision, and action, showing clear embodied features. Embodied intelligence relies on interaction with the physical world and uses simulators and world models for training and prediction[8].

High autonomy requires more than simple reactive inference. Methods based only on direct perception-to-action mapping are often unstable under environmental changes, partial observability, and delays. Therefore, EGI systems need world models, which represent the external environment[9]. World models help estimate current states and predict future changes. They support state transition modeling, reasoning under uncertainty, and action evaluation for long-term planning. Without them, edge intelligence often relies on short-term decisions, which limits reliability and generalization in dynamic environments[10].

B. Motivation and Contribution

From early ideas to recent developments, digital twins have become a key paradigm for modeling physical objects in cyber – physical systems[11]. They are implemented as high-fidelity virtual models of physical systems, supporting context modeling, real-time state updates, and modular interactions through application programming interfaces (APIs). This enables industrial analysis and decision-making[12]. Built on physical laws, domain equations, and expert rules, digital twins are especially useful for offline analysis and system design.

In recent years, digital twins have been widely used in robotics, healthcare monitoring, and wireless communications. For example, RoboTwin combines 3D generative models with large language models (LLMs) to generate data for robotic tasks, improving performance[13]. In smart healthcare, digital twins support monitoring, health prediction, and control[14]. In wireless and edge systems, they are used for network planning, testing, and what-if analysis, and enable joint optimization of communication and computing under latency and energy constraints[15].

However, digital twin technology still faces key challenges, including complex data integration, limited training data, high cost, and the lack of unified frameworks[16]. Traditional digital twins are designed for offline use and do not meet the needs of online and autonomous EGI, especially in heterogeneous environments with missing standards[11]. They also rely on predefined rules, which limits generalization. High-fidelity models are hard to run in real time on resource-limited devices. In addition, system-focused design ignores the agent’s perception, decision, and action. As a result, many digital twins remain static and lack dynamic and intelligent capabilities[17].

To complement digital twins, world models have emerged for similar tasks[18]. They shift the focus from high-fidelity replication to modeling environment changes that support decision-making. World models use representation learning, such as variational autoencoders, to map high-dimensional inputs into low-dimensional latent states[9]. They learn action-conditioned transitions to predict future dynamics, enabling efficient planning and integration with reinforcement learning. Recent work studies world models in areas such as multi-scale modeling, controllable prediction, structured reasoning, and dynamics modeling[10]. For example, DriveDreamer-2 supports diverse predictions in autonomous driving[19]. GLAM models global and local state changes for better learning[20]. Drive-OccWorld uses a 4D vision-based model for planning[21]. SWAP enables structured reasoning, and MoSim supports long-term physical prediction[22][23][24]. These works show that focusing on task-related dynamics allows world models to support long-term autonomy in resource-limited edge systems.

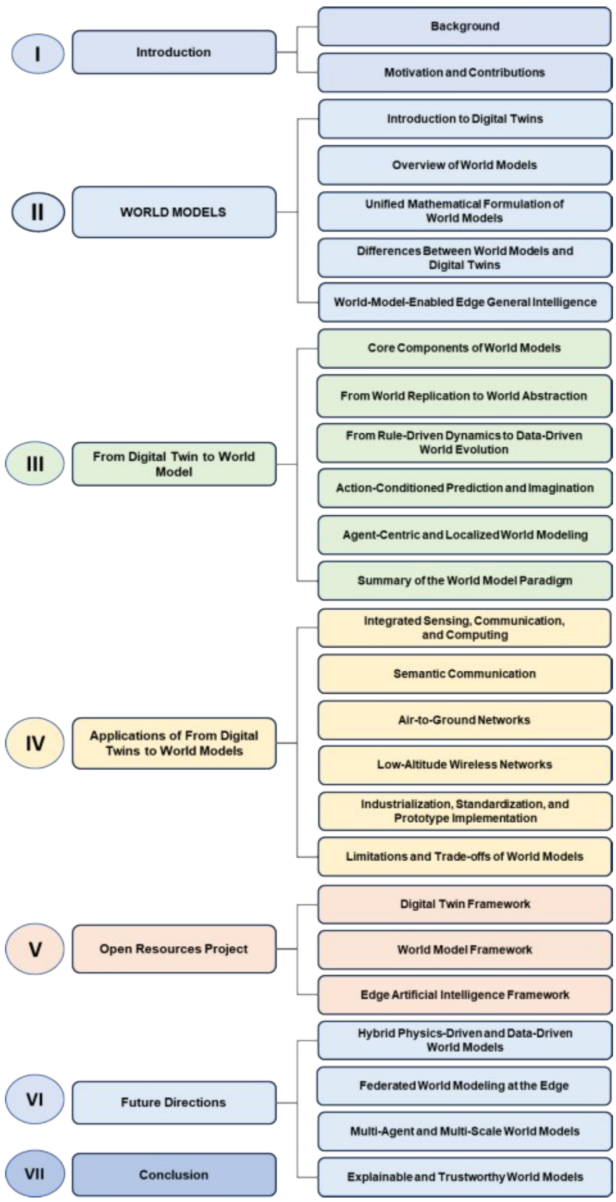

Fig. 1 Conceptual architecture from digital twins to world models for edge general intelligence.

This survey reviews the evolution from digital twins to world models through the lens of EGI. Figure 1 outlines the structure of this paper, covering motivation, world models, comparison with digital twins, technical evolution, applications, open resources, and future directions. Building on the background and motivation discussed above, the contributions of this work are as follows:

● We provide a comprehensive review of world models for EGI in wireless and edge systems, and compare them with traditional digital twin approaches. We present a unified view of these two paradigms as complementary but fundamentally different ways to model the physical world. Particularly, we highlight shifts to decision-oriented abstraction, data-driven dynamics, and agent-centric modeling. These clarify the role of world models in EGI and guide autonomous operation at the edge.

● We establish a taxonomy of world models tailored to wireless and edge scenarios, and decompose them into core components, including representation learning, dynamics modeling, observation and action interfaces, and imagination-based planning. We review representative methods from machine learning, robotics, and control, and reinterpret them in terms of edge deployment constraints, communication awareness, and integration with existing digital twin infrastructures.

● We demonstrate the potential of world models in edge systems by mapping classical digital twin applications in integrated sensing, communication, and computing (ISCC), semantic communication, air-to-ground network, and low-altitude platforms, and industrial edge infrastructures to their corresponding world model-based counterparts. We identify key open challenges and outline future research directions, including hybrid physics – data-driven world models, federated world modeling at the edge, multi-agent and multi-scale modeling, and explainable and trustworthy world models for safety-critical EGI.

offline planning to online action.

By organizing existing knowledge and open questions along these dimensions, this survey offers both a conceptual foundation and a practical roadmap for researchers and practitioners moving from digital-twin-centric design toward world-model-centric EGI.

1 WORLD MODELS

This section introduces the basic concepts of digital twins and world models, and then discusses their main differences. We further explain how world models can serve as a core enabler for EGI. The definition of EGI refers to the integration of edge computing and artificial intelligence to enable human-like cognition on edge devices. Its core capabilities include multi-modal perception, reasoning, and autonomous decision-making in dynamic, multi-task settings, with continuous learning from limited data. The boundaries of EGI lie in its support for local optimization and coordination in scenarios such as vehicular and IoT networks, while being limited by resource constraints for large-scale training and global reasoning. The evaluation criteria of EGI focus on prediction and generalization, adaptability, and decision performance under resource constraints. The detailed framework is shown in Figure 2.

1.1 Introduction to Digital Twins

Digital twins are high-fidelity virtual replicas of physical systems, governed by physical laws, domain equations, and expert rules, supporting simulation, evaluation, and optimization[11].

In wireless and edge systems, they are widely used for

network planning, configuration validation, and what-if analysis under known conditions[14]. For example, operators use them to evaluate deployment strategies and test policies before live deployment, demonstrating the value of physics- and rule-based

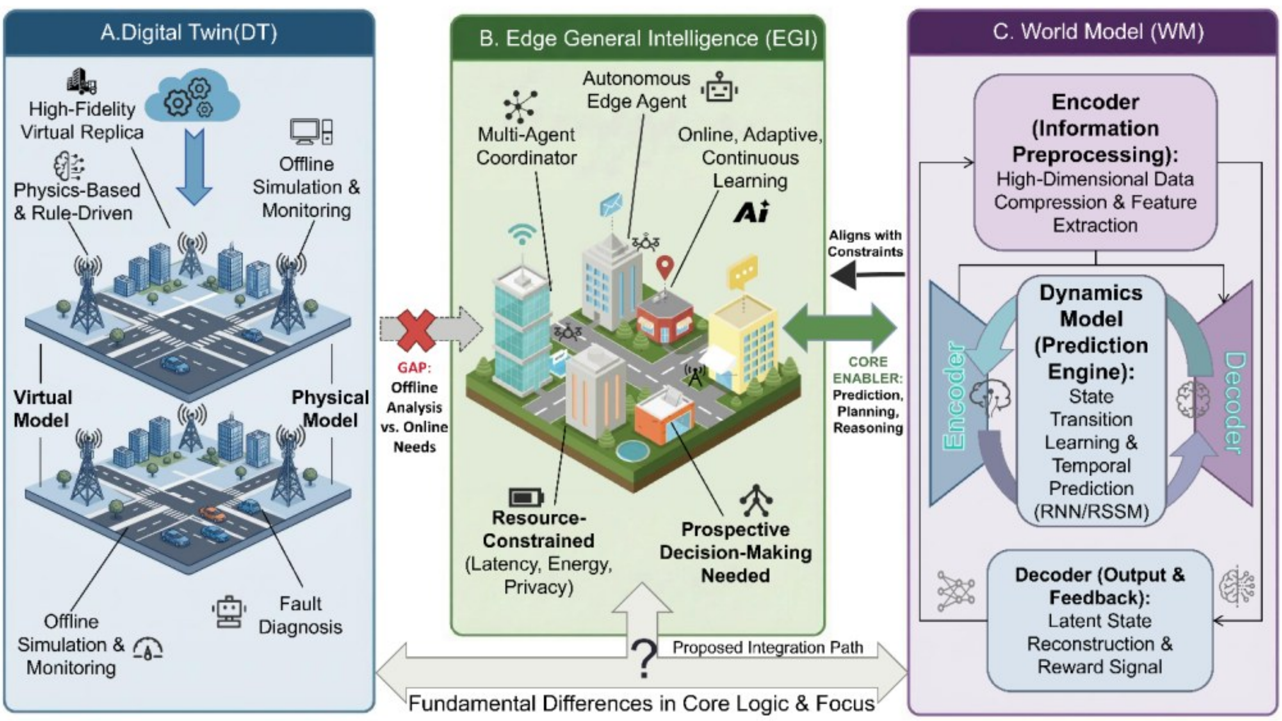

Fig. 2 A conceptual framework illustrating how digital twins and world models enable EGI. (A): Digital twin offline physics replica feeds online monitoring. (B): EGI agents compress data and pick actions via reward. (C): world model hierarchy bridges

models in complex edge environments[25].

However, their strength in offline engineering analysis does not translate directly to the needs of online, autonomous EGI agents that must continuously learn, adapt, and act, motivating a reassessment of world modeling for EGI.

1.2 Overview of World Models

World models are internal simulations of environmental dynamics constructed by intelligent systems[26]. By learning from real-world data and implicit laws, they capture key dynamic properties of the environment, predict future states, and provide agents with the ability to understand, reason about, and plan for the physical world. Figure 3 illustrates the world model framework, where the upper layer represents core capabilities, and the lower layer provides end-to-end support through multi-modal large language models and generative models. This capability of learning and representing physical dynamics enables world models to excel in computer vision tasks such as video generation. For instance, by internalizing spatio-temporal relationships from driving data, world models can generate realistic 4D scenes where objects follow consistent physical trajectories and spatial layouts over time, demonstrating their potential as physical simulators for visual content creation[30].

World models are not high-fidelity reproductions of reality. Instead, they focus on dynamic aspects directly related to agent behavior, building an internal cognitive framework for imaginative reasoning[9]. This abstraction avoids the computational cost of full-scale replication, making world models suitable for resource-constrained scenarios like EGI.

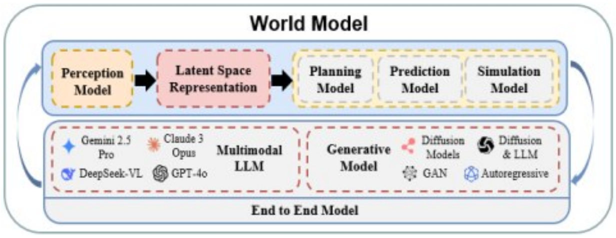

Fig. 3　Overview of the world model framework: core capability modules and end-to-end model support

Data-driven dynamic modeling is central to the adaptability of world models. Unlike traditional models that depend on preset physical formulas, world models automatically extract implicit environmental laws—including physical regularities and temporal correlations—via unsupervised or self-supervised learning from multi-modal interaction data[18]. For example, in UAV flight scenarios, they learn the mapping between actions and states from data reflecting channel variations and meteorological interference[18]. Even under previously unseen low-altitude weather conditions, they can still make reasonable predictions based on learned general laws, overcoming the generalization limits of traditional models and adapting flexibly to edge environment dynamics.

World models aim to perceive the current state and learn latent environment dynamics over time to support simulation, planning, and decision-making. From a mathematical perspective, several variants have been developed based on the basic formulation,

TABLE I: Differences Between Digital Twins and World Models

Name	Ref	Definition	Technical Foundation	Advantages
World models	[1] [9] [26] [27]	• Learns compact latent representations from multi modal sensory inputs • Models temporal state transitions to capture environment physics • Forecasts future latent states conditioned on planned actions	• Representation learning • Dynamics models • Self-supervised learning • Latent-state compression	• Resource-efficient • Self-adaptive • Foresighted • Task-centric
Digital twins	[11] [16] [28] [29]	• Constructs physics-based virtual replicas of physical assets • Maintains bidirectional synchronization between physical and virtual states • Propagates state changes via real-time data streaming	• IoT sensory ingestion • Bidirectional synchronization • Physics-based simulation • Edge-cloud computing	• Cyber-physical fidelity • Life cycle governance • Anticipatory risk mitigation • Holistic visualization

including observation-level generative, latent-space, and object-centric world models. In general, the basic world model predicts the future environment state at time step □ given historical states and optional actions

$$z_{t+1} \sim \pi_\theta\left(z_{t+1} | z_t, a_t, o_{t+1}\right) \tag{1}$$

where $\pi_\theta(\cdot)$ denotes the world model. z_t and z_{t+1} represent the latent environment states at time t and $t+1$, respectively a_t denotes the optional agent action, and o_{t+1} represents the observation at time $t+1$.

Observation-level generative world models decode latent states into future observations for simulation. Based on the general formulation (1), a decoder $\Phi_\theta(\cdot)$ maps the latent state z_{t+1} to the observation o_{t+1}. The objective is given as follows

$$\hat{o}_{t+1} = \Phi_\theta\left(z_{t+1}\right), z_{t+1} \sim \pi_\theta\left(z_{t+1} | z_t, a_t, c_t\right) \tag{2}$$

where $\hat{o}_{t+1}$ denotes the predicted observation at time $t+1$. During simulation, observation-level generative world models generate multiple imagined trajectories $\{\zeta_i\}$, each consisting of a sequence of observations with length T.

Latent-space world models learn dynamics in a high-dimensional latent space for simulation and planning. Based on formulation (1), a predictor $\Omega_\theta(\cdot)$ is introduced to estimate the latent state at time $t+1$,and the objective is to minimize the difference from the ground-truth z_{t+1}

TABLE III: From digital twins to world models: Enabling edge general intelligence

Scope	Ref	Key insight	From digital twins to world models
Edge general intelligence	[33]	• Semi-synchronous edge intelligence using digital twins • Virtual-physical interaction for real-time decisions	Digital twins continuously supply boundary data, laying an online learning foundation for later integration with world models
	[18]	• Lightweight recurrent state-space model with fast reasoning • Real-time prediction for autonomous edge nodes	World models give edge nodes real-time look-ahead reasoning ability, providing a decision brain for subsequent connection to digital twins
	[34]	• Compress LLM knowledge into world models • Digital twins calibrate for trustworthy edge AI	World models compress and offload large model capabilities to the edge, while digital twins correct physical errors in real time, making edge general intelligence both efficient and trustworthy

TABLE II: Quantitative comparison trends between digital twins and world models in EGI scenarios

Quantitative Indicator	Ref	Digital Twin (DT)	World Model (WM)	Typical Improvement
Computational Resource (FLOPs per inference step)	[26] [27]	High (physics simulation dominant, ~ 10^9 + scale)	Low (latent compression, ~ 10^6--10^8 scale)	Significant order-of-magnitude reduction
End-to-End Latency (ms)	[31] [15]	Higher (synchronization updates, second-scale)	Low (online reasoning, 10--200 ms)	Substantial improvement
Energy Consumption (mJ per decision)	[31] [16]	Higher (continuous simulation)	Low (inference-dominant)	Notable reduction
Samples Needed for New Scenario Adaptation	[26] [27]	Fewer (mainly parameter tuning)	More (interactive data fine-tuning)	WM more data-dependent

The values shown are indicative trends summarized from representative studies. Absolute values may vary depending on hardware platforms and deployment scenarios.

$$\hat{z}_{t+1} \sim \Omega_\theta\left(\hat{z}_{t+1} | z_{1:t}, s_{1:t}, a_{1:t}\right) \tag{3}$$

$$min\mathcal{L} = \mathbb{E}\left[// \hat{z}_{t+1} - z_{t+1} //^2 \right] \tag{4}$$

where $z_{1:t}$, $s_{1:t}$, and $a_{1:t}$ denote the sequences of latent representations, end-effector states, and agent actions from time step 1 to t,respectively.

Object-centric world models use slot attention to represent scenes as composable object slots and model object-level dynamics. The objective minimizes the difference between predicted and ground-truth slots at future time steps

$$\hat{h}_{t+1} \sim \lambda_\theta\left(\hat{h}_{t+1} | \mathcal{A}\left(o_t\right), a_t\right) \tag{5}$$

$$min\mathcal{L} = \mathbb{E}\left[\mathcal{D}\left(\hat{h}_{t+1}, \mathcal{A}\left(o_{t+1}\right)\right)\right] \tag{6}$$

where $\mathcal{A}(\cdot)$ denotes the slot attention function that maps an observation to a set of object slots. $\mathcal{A}\left(o_t\right)$ and $\mathcal{A}\left(o_{t+1}\right)$ represent the slot sets at time t and $t + 1$, respectively. $\lambda_\theta(\cdot)$ is a slot predictor that estimates the slot set at time $t + 1$, and $\mathcal{D}(\cdot)$ denotes a commonly used distance metric.

Prospective decision-making and planning are the core objectives of world models[9]. Rather than passive response, they follow a prediction – imagination – decision pipeline that provides forward-looking cognitive capabilities, enabling simulation of multi-step environment evolution, assessment of action sequences, and optimization to avoid short-sighted behavior. This aligns well with the latency, energy and privacy constraints of EGI[1], offering core support for long-term autonomous decision-making at the edge. This adaptability comes from three key design characteristics for edge deployment.

Efficiency comes from abstraction of latent-space, extraction of decision-relevant information without high-fidelity reconstruction, significantly reducing computation, storage, and communication overhead for edge devices[18]. The agent-centric nature focuses only on the dynamics relevant to the agent’s observations and actions, forming a localized task-oriented model that avoids wasting limited resources on redundant information[32]. Generative imagination uses generative AI to infer unseen scenario trajectories, improving sample efficiency for edge scenarios[18]. These characteristics complement the data-driven and prospective nature of world models, strengthening their applicability in edge environments.

Thus, world models learn compact representations of environmental dynamics from data and, through generative imagination and prospective reasoning, provide efficient,

autonomous, and adaptive decision support for resource-constrained edge agents, becoming an indispensable cognitive pillar for EGI.

1.3 Unified Mathematical Formulation of World Models

Consider an agent operating in a Markov Decision Process (MDP), where s_t, a_t, and r_t denote the state, action, and reward at time step t, respectively. Since the true transition function T is unknown and observations o_t are high-dimensional, the world model learns an internal simulator to support policy optimization without excessive real-world interaction.

● *State Representation.* The encoder compresses observation o_t, previous latent state h_{t-1}, and action a_{t-1} into a compact latent state h_t, while the decoder reconstructs $\hat{o}_t$ to preserve information fidelity

$$h_t = f_{\mathrm{enc}}\left(h_{t-1}, o_t, a_{t-1}\right), \hat{o}_t = f_{\mathrm{dec}}\left(h_t\right) \tag{7}$$

● *Action-Conditioned Transitions.* The dynamics model f_{dyn} predicts the next latent state, and the reward predictor $\hat{R}_\theta$ estimates immediate rewards, enabling multi-step imagined rollouts $\hat{\tau}$ within the latent space

$$h_{t+1} = f_{\mathrm{dyn}}\left(h_t, a_t\right), \hat{R}_\theta\left(h_t, a_t\right) \approx r_t \tag{8}$$

$$\hat{\tau} = \left(h_t, a_t, \hat{r}_t; h_{t+1}, a_{t+1}, \hat{r}_{t+1}; \cdots\right) \tag{9}$$

● *Edge Resource Constraints.* The agent maximizes expected cumulative reward over imagination horizon H under edge resource constraints $\mathcal{C}_k(\cdot)$, covering communication rate, energy, and transmit power limits

$$\max_\pi \mathbb{E}_{\hat{T}_\theta}\left[\sum_{k=0}^{H-1} r_{t+k}\right] \# (10)$$

$$\text{s.t.} \mathcal{C}_k\left(h_t, a_t\right) \leqslant 0, \forall k \in \{1, \cdots, K\}, \forall t$$

This unified formulation enables edge agents to make proactive decisions under resource constraints, overcoming the sample inefficiency of conventional model-free approaches.

1.4 Differences Between World Models and Digital Twins

Both world models and digital twins aim to model environmental dynamics and support decision-making, but their core logic and application focus are fundamentally different.

A world model is an internal cognitive framework of an agent that distills environmental dynamics relevant to its actions and decisions, learning implicit laws from multi-modal interaction data without relying on preset physical formulas[27]. To fit resource-constrained EGI settings, it reduces dimensionality via latent-space abstraction, reducing

computation and communication overhead, and enables long-term autonomous decision-making in dynamic edge environments[35].

By contrast, a digital twin builds a high-fidelity virtual

mirror of a physical entity, relying heavily on preset physical formulas, rule bases, and rich sensor data to restore real system states through physics-based modeling[28],[29]. This makes digital twins effective for intelligent asset management, but their high-fidelity requirements lead to massive data trans-

mission and complex simulations, typically relying on cloud or high-performance servers, making them more suitable for Industry 4.0 rather than resource-constrained edge scenarios.

The essential difference is that digital twins emphasize restoring reality, making physical states observable, while world models emphasize predicting the future to enable forward-looking decisions[27]. This core orientation allows world models to better match the latency, energy, and privacy constraints of edge scenarios through law distillation and imaginative reasoning[34]. Quantitative trends confirm this distinction, as summarized in Table II.

1.5 World-Model-Enabled Edge General Intelligence

Through leveraging the four core capabilities, such as imagination, prediction, planning, and reasoning, world models meet the autonomous decision-making needs of EGI under resource constraints and dynamic uncertainty[36].

● Imagination: Imagination enables world models to synthesize novel scenarios, allowing agents to mentally explore hypothetical futures without physical interaction[26]. This supports creative problem-solving and risk-free experimentation in resource-constrained edge environments. For instance, navigation agents can leverage learned visual priors to imagine trajectories from a single image and plan safe paths[30].

● Prediction: Prediction enables world models to infer future environmental states by learning implicit laws from multi-modal interaction data[18]. In EGI scenarios, a world model can predict channel-quality fluctuations based on historical data and device trajectories, providing time margins to avoid link outages[20]. In low-altitude networks, it can fuse historical and real-time weather data to predict wind changes and provide early warnings for UAV flight safety[37].

● Planning: Planning uses predicted outcomes to simulate multi-step scenario evolution and optimize decisions[38]. For UAV delivery or inspection, the world model combines weather and channel predictions to plan trajectories balancing safety, energy, and throughput. It also dynamically allocates computing, storage, and communication resources across edge nodes based on task load, preventing overload[38]. World models enhance EGI’ s decision-making through latent-space simulation, as demonstrated by a generative AI framework for satellite communication that improves planning via multi-agent policy optimization[39].

● Reasoning: Reasoning enables world models to handle unknowns using learned general laws[40]. In vehicular communications, edge nodes on board infer communication routes and trajectory adjustments by combining real-time road conditions with link states, maintaining data delivery and safety[41]. In industrial maintenance, a world model analyzes abnormal sensor readings, distinguishes faults from interference or false alarms, and supports accurate maintenance decisions to reduce production loss[40].

EGI faces challenges such as time-varying channels and dynamic node mobility. The prediction, planning and reasoning capabilities of world models help ensure link stability, resource optimization, and rapid emer-gency response, meeting the low-latency and high-reliability requirements of wireless EGI networks[42]. In low-altitude cooperative UAV scenarios, these capabilities jointly support flight safety, mission effi-ciency, and adaptation to complex edge environments[20].

2 FROM DIGITAL TWIN TO WORLD MODEL

This section first introduces the core components of world models and then discusses the key technical shifts that drive the evolution from digital twins to world models.

2.1 Core Components of World Models

The perception - prediction - decision pipeline of world models relies on three tightly coupled modules: the encoder, the dynamics model, and the decoder, which together form an efficient closed-loop workflow[18].

● *Encoder:* The encoder serves as the front-end processing module, compressing high-dimensional multimodal inputs—such as images and radar signals—into compact low-dimensional latent states by removing redundant details and retaining decision-relevant features[43]. This compression significantly reduces the computational overhead of downstream modules, which is critical for resource-constrained edge devices.

● *Dynamics model*: The dynamics model is the core prediction engine of the world model, learning the environmental state transition function to predict the next latent state given the current state and the agent's action[44]. It supports both deterministic architectures (e.g., RNN-based models) and stochastic architectures (e.g., RSSM), and generates multi-step la-

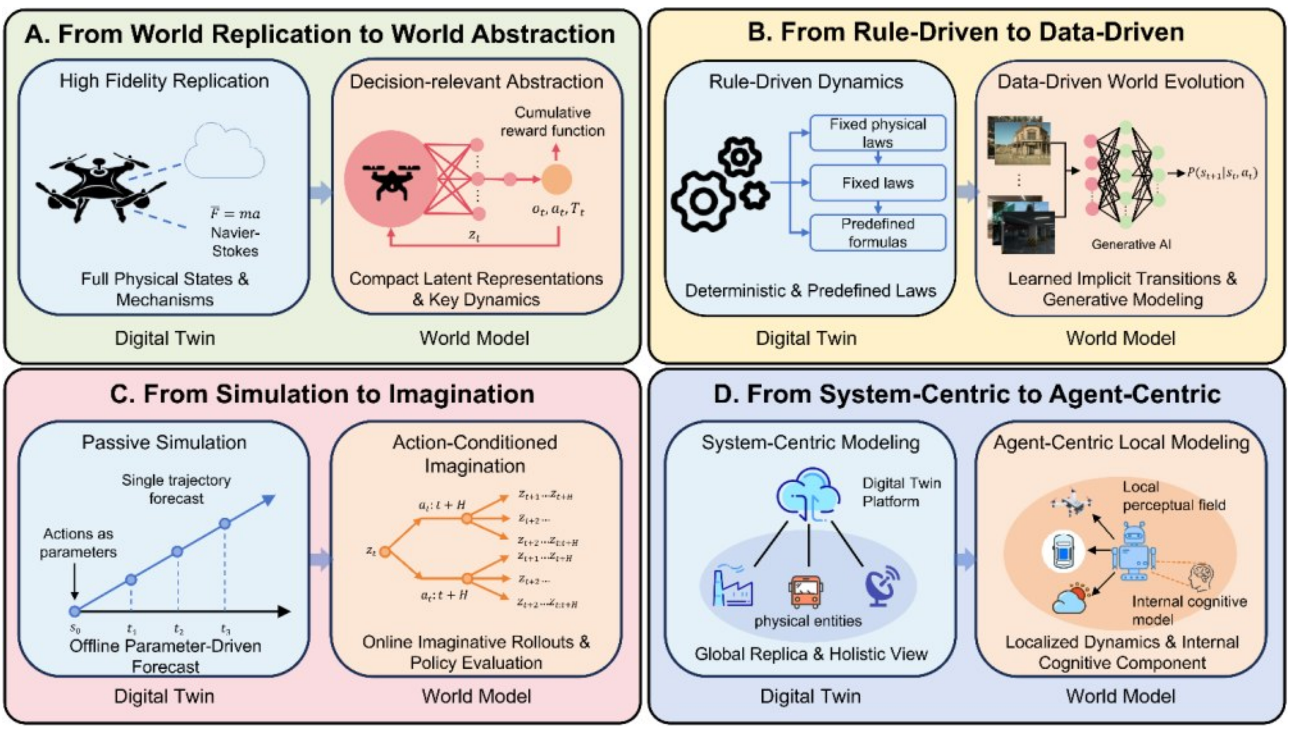

Fig. 4 Evolution from Digital Twin to World Model for EGI. (A) From world replication to world abstraction, (B) From rule-driven to data-driven, (C) From passive simulation to active imagination, (D) From system-centric to agent-centric.

tent trajectory predictions to inform agent decision-making[24].

● *Decoder:* The decoder maps predicted latent states back to observable quantities, reconstructing future environment frames and producing reward signals[9]. These outputs support policy evaluation and also serve as feedback to refine the encoder and dynamics model, thereby closing the perception – prediction – decision loop and ensuring pipeline integrity[19].

Together, these three modules drive the transition from digital twins—characterized by rule-driven, system-centric replication—toward world models that enable data-driven, agent-centric abstraction with active imagination, as illustrated in Figure 4 and Table IV.

2.2 From World Replication to World Abstraction

The core principle of digital twins is world replication constructing virtual systems that are highly consistent with the physical world in structure, state, and mechanisms to support system-level simulation and monitoring[16]. In contrast, world models pursue world abstraction, retaining only those environmental dynamics that affect an agent's future cumulative rewards. For instance, the framework in [26] uses a variational autoencoder (VAE) to learn low-dimensional latent representations sufficient for complex control tasks, while the Dreamer family[27] demonstrates that imaginative rollouts in learned world models significantly improve policy optimization efficiency.

When the modeling objective shifts toward online decision-making for EGI, state representation criteria change fundamentally. EGI aims to deliver general, adaptive cognitive capabilities on resource-constrained edge devices, emphasizing long-term autonomous decision-making under tight computation, storage, and communication budgets[48]. Maintaining a globally synchronized physical replica in such settings introduces heavy overhead and decision-irrelevant information[18]. Moreover, traditional digital twins rely on prior physical models that are difficult to maintain in dynamic edge environments and often lack generalization[49].

Abstraction-based world models offer clear advantages for EGI. By discarding decision-irrelevant details, they reduce computation and communication costs[50], and by learning dynamics directly from interaction data, they reduce dependence on precise physical priors while enabling strong generalization in un-

TABLE IV: Comparison between digital twin and world model

Feature	Characteristics	Replication to Abstraction	Rule-based to Data-driven	Passive to Active	System-centric to Agent-centric
Digital Twin	World Replication [16] Rule-driven [45] Passive Simulation [46] System-centric [47]	Achieve high-fidelity mapping that reflects precise physical reality Maintain structural isomorphism to ensure geometric consistency with objects Reproduce intricate physical details and geometry for total synchronization	Operate based on deterministic rules to ensure predictable outcomes Models are strictly governed by explicit physical laws and mechanisms Utilize complex mechanism equations to simulate exact physical behaviors	Run simulations based on fixed parameters to verify designs Simulations rely on specific initial states to predict results Treat agent actions as external inputs within a static system	Provide comprehensive global modeling for the entire system landscape Create a complete virtual replica encompassing all system components Ensure full coverage of complex system behaviors and interactions
World Model	World Abstraction [26] Data-driven [27] Active Imagination [18] Agent-internal cognitive component [9]	Abstract complex environments into compact, task-oriented representations Retain only the critical information necessary for reward prediction Extract low-dimensional latent features to simplify environmental complexity	Learn versatile environment representations through advanced generative AI Evolve through continuous learning from massive environmental interaction data Capture implicit state transition dynamics without explicit physical formulas	Predict future environment states conditioned on specific agent actions Imagine virtual trajectories within latent space to evaluate plans Enable autonomous exploration through internal "what-if" active simulations	Focus on ego-centric local modeling relevant to the agent Dynamically adapt modeling to match the agent's real-time perception Function as the internal cognitive brain for autonomous decision-making

known environments[51]. The evolution from digital twins to world models thus reflects a fundamental shift from high-fidelity physical replication to decision-oriented abstraction that supports efficient prediction, planning, and autonomous learning at the edge.

2.3 From Rule-Driven Dynamics to Data-Driven World Evolution

Traditional digital twins are grounded in physics-based mechanistic models, where environmental evolution is governed by predefined physical laws and system-level equations, providing deterministic descriptions through high-fidelity virtual replicas[16]. In contrast, world models adopt a data-driven paradigm, compressing environmental dynamics into compact latent spaces via generative AI techniques and autonomously learning implicit state transition laws from agent--environment interaction data[26].

A key requirement of EGI is long-horizon autonomous decision-making in dynamic and uncertain environments[52]. In practical scenarios such as UAV networks and intelligent transportation systems[53], fixed rule-based models frequently fail to capture complex long-term system evolution, limiting generalization. Furthermore, the heavy numerical solvers required for high-fidelity physical modeling impose substantial computational costs that conflict with the limited resources of edge devices.

Data-driven world evolution modeling directly addresses these limitations. EGI agents can continuously refine their internal world models using local interaction data, adapting to environmental changes without redesigning physical equations[18]. By performing inference in low-dimensional latent spaces, world models avoid pixel-level simulation and significantly reduce computational overhead. For example, aerial network world models trained on large-scale trajectories and semantic labels can predict plausible long-range scenarios under unseen meteorological conditions, supporting navigation path generation that balances obstacle avoidance and semantic objectives[37]. This shift from rule-driven to data-driven modeling thus represents a fundamental evolution of EGI, enabling proactive prediction rather than passive reproduction in open-world environments.

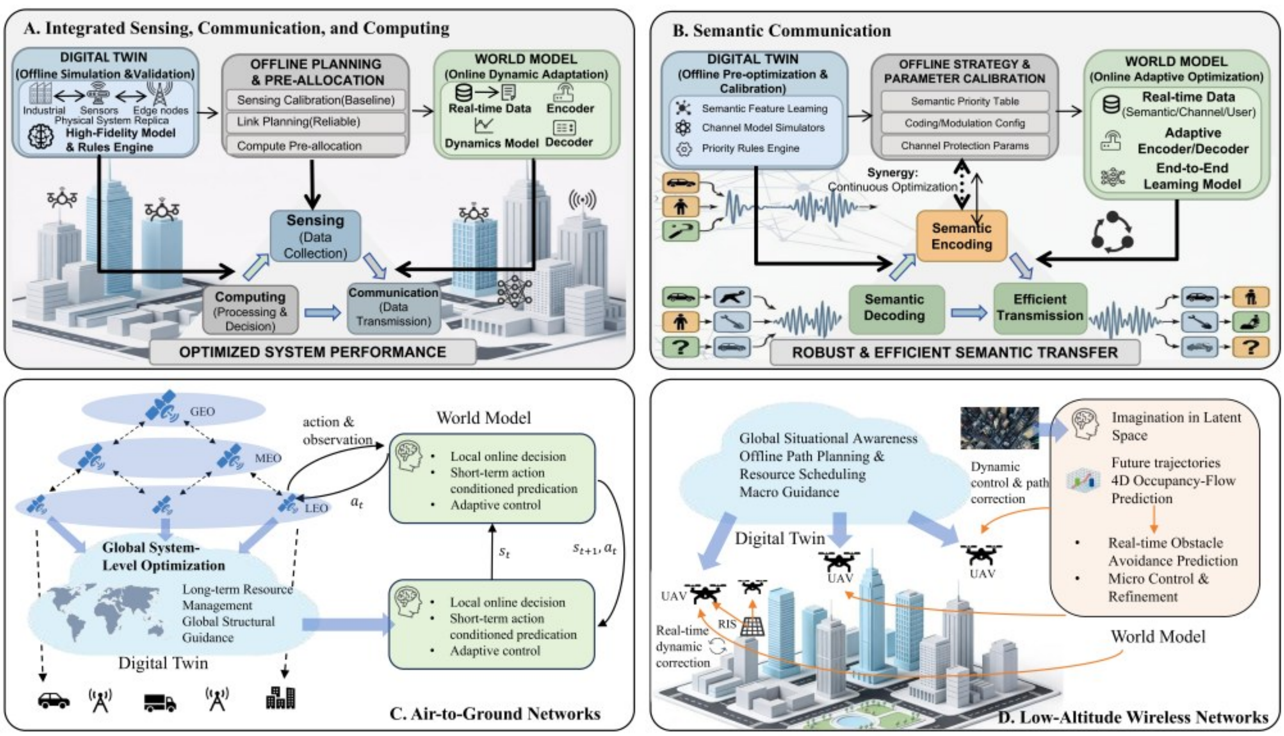

Fig. 5 Evolution from digital twin to world model for EGI. (A) From world replication to world abstraction, (B) From rule-driven to data-driven, (C) From passive simulation to active imagination, (D) From system-centric to agent-centric.

2.4 Action-Conditioned Prediction and Imagination

As EGI evolves toward complex autonomous decision-making, the modeling objective shifts from describing current states to reasoning about the consequences of actions[54]. Digital twins and world models represent two distinct predictive paradigms in this regard. Digital twins rely on passive, parameter-driven simulation, treating agent actions as external inputs rather than intrinsic drivers of system evolution[55]. World models, by contrast, explicitly incorporate agent actions as key variables in state transitions, enabling imaginative trajectory rollouts in compact latent spaces without direct interaction with the real environment[9].

This distinction is especially significant for EGI, where edge devices face strict constraints on computing, storage, and communication, and tasks often involve long time horizons[56]. High-fidelity simulation is too costly for evaluating multiple candidate action sequences under such constraints. World models address this by performing action-conditioned prediction in low-dimensional latent spaces, avoiding pixel-level simulation while preserving the essential dynamics needed for decision-making. For example, FSDrive[57] employs a world model to generate spatiotemporally coherent future scenes in latent space, enabling trajectory planning and policy evaluation without explicit physical simulation. This shift from system-level reproduction to action-driven imagination provides a more efficient and scalable decision-support paradigm for edge intelligence.

2.5 Agent-Centric and Localized World Modeling

Digital twin technology, originating in Industry 4.0, follows a system-centric philosophy by creating global virtual replicas for holistic management[46]. In contrast, world models adopt an agent-centric and localized paradigm, serving as internal cognitive components that capture only environmental dynamics relevant to an agent’ s specific sensing, actions, and tasks[58]. This localization effectively decouples model complexity from the overall environmental scale, enabling resource-constrained edge devices to perform complex modeling[59].

In EGI scenarios, agents prioritize rapid local understanding and action-prediction under partial observability over full-scale environment reconstruction[60]. The agent-centric nature of world models aligns with these requirements by reducing maintenance overhead and supporting continuous online updates[61]. Furthermore, world models integrate directly with decision-making frameworks such as reinforcement learning and model predictive control, offering better scalability and robustness in dynamic, multi-agent environments[62]. Consequently, shifting from system-centric to agent-centric modeling minimizes resource consumption while providing essential internal support for autonomous reasoning at the edge.

2.6 Summary of the World Model Paradigm

This section explores the paradigm shift from digital twins to world models, moving from high-fidelity replication to decision-oriented abstraction. This transition establishes a cognitive framework better suited for resource-constrained EGI. Accordingly, we summarize the defining characteristics of world models across four key dimensions.

- *Decision-Oriented Abstract Modeling:* Unlike high-fidelity digital twins, world models employ decision-oriented abstraction via compact latent representations. By preserving only dynamics relevant to future rewards, they reduce dimensionality and computational overhead for resource-constrained edge devices[26].

- *Data-Driven Evolutionary Learning:* Shifting from rule-based physical equations, world models use generative modeling to learn dynamics from agent – environment interactions. This captures complex, non-analytical patterns and supports continual learning to adapt to environmental changes[18].

- *Action-Conditioned Imaginative Reasoning:* World models enable action-conditioned imagination by iterating latent-space predictions. This allows agents to evaluate long-term trajectories and consequences without real-world interaction, significantly improving planning and sample efficiency[51].

Agent-Centric Local Modeling: As internal cognitive components, world models focus on local dynamics relevant to the agent's perception and goals. This decouples model complexity from the global environment and facilitates heterogeneous deployment in distributed edge systems[9].

3 APPLICATIONS OF FROM DIGITAL TWINS TO WORLD MODELS

This section discusses the application of world models to integrated sensing, communication, and computing

(ISCC), semantic communication, air-to-ground networks, and low-altitude wireless networks. Table V compares digital twins and world models based on their main functions and related studies in these scenarios, while Fig 5 illustrates the conceptual evolution from digital twins to world models that underpin these applications.

3.1 Integrated Sensing, Communication, and Computing

Integrated sensing, communication, and computation (ISCC) orchestrates end-to-end resource optimization under edge constraints[63]. Digital twins and world models jointly support ISCC: the digital twin enables offline verification for sensing, communication, and computation, while the world model dynamically adjusts sensing frequency, communication parameters, and offloading policies online, forming a complementary co-design paradigm (Fig. 6).

The digital twin performs offline system-level modeling and validation of sensing accuracy, link quality, and resource allocation[64]. Using high-fidelity virtual replicas, it reproduces sensor coverage, signal attenuation, and compute-node capacity for ex-ante policy evaluation[65]. The world model focuses on online co-optimization through prospective planning and real-time adaptation[66], learning latent dependencies among sensing, communication, and computation from multi-modal data via an encoder – dynamics – decoder pipeline[67]. In UAV swarms, it forecasts traffic and resource variations[68]; in industrial edge, it tunes sampling rates, protocols, and offloading fractions[69]. For physical layer security, APEG uses generative AI for adaptive authentication[70].

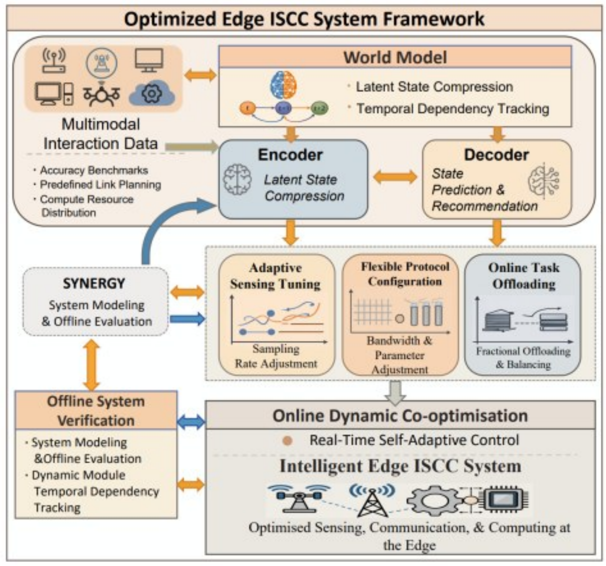

Fig. 6 A conceptual framework illustrating how the optimized edge ISCC system enables intelligent edge computing and sensing through multi-modal data processing, world model-driven perception, and dynamic edge co-optimization

The synergy between digital twins and world models enables a shift from offline static to online dynamic optimization[18]. Digital twins provide calibration benchmarks and resource pre-allocation, while world models continuously refine operational parameters[71]. This integration achieves 33% end-to-end latency reduction and 30% tail latency reduction at 2 – 7% energy cost via digital twin frameworks[64], while world models deliver 15 – 20% lower execution latency[69].

Case Study of ISCC: In high-speed smart transportation, conventional sensing suffers from blind spots and high latency, inadequate for L4+ cooperative obstacle avoidance. ISCC constructs a persistent electromagnetic world model that replaces snapshot imaging, integrated via semantic compression for multi-vehicle interaction. A dual-radio-frequency chain synthesizes 64 virtual aperture elements with full-polarization observation. Simulations show the architecture requires only ×2 frame overhead for model updates, achieving higher update rates than time-division multiplexing MIMO with lower hardware cost and calibration complexity[72].

3.2 Semantic Communication

Semantic communication transmits meaning rather than raw bits, overcoming throughput limits of conventional schemes under limited bandwidth or hostile channels[87], [88]. The digital twin handles offline pre-optimization and parameter calibration of semantic policies, while the world model performs online adaptation by tracking end-to-end dynamics and refining encoding, transmission, and decoding, forming a complementary offline – online evolution loop (Fig. 7).

The world model captures real-time dynamics along the semantic link, autonomously learning latent relationships among semantic content, channel state, and user demand from live data for end-to-end optimization[76],[77]. It compresses semantic data via an encoder, adapts modulation and power during transmission, and corrects distortion at the decoder. Its lightweight design enables edge deployment for efficient cloud-edge-terminal architectures[78].

In the offline phase, the digital twin generates semantic priority tables, code modulation settings, and channel protection parameters[75]. During online operation, the world model dynamically updates these parameters to adapt to real-time noise, traffic, and resource conditions, ensuring near-optimal performance[76]. This yields a scalable, reliable architecture for next-generation intelligent networks.

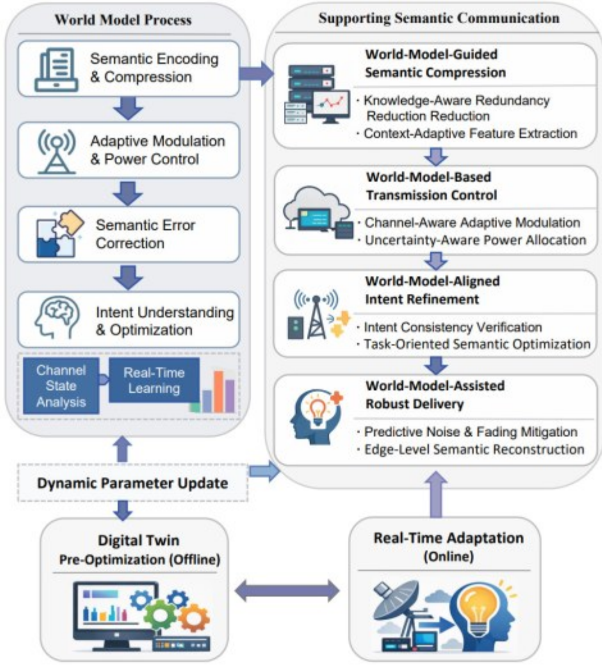

Fig. 7 A conceptual framework depicting the integration of world model processing, digital twin pre-optimization, and real-time adaptation for enabling robust and intelligent semantic communication.

Case Study of Semantic Communication: In se-

TABLE V: Functional comparison of digital twin and world model in edge intelligence

Scenario	Paradigm	References	Key Functions and Contributions
ISCC	Digital Twin	[18], [64], [65]	Supports high-fidelity simulation for offline sensing calibration, link configuration, and resource pre-allocation.
	World Model	[66] – [69]	Enables online coordination via latent dynamics learning and adaptive sampling and offloading.
Semantic Communication	Digital Twin	[73] – [75]	Provides semantic policy calibration through priority setting and channel loss simulation.
	World Model	[76] – [78]	Supports dynamic semantic adaptation through implicit channel learning and encoding adjustment.
A2G Networks	Digital Twin	[79] – [81]	Facilitates global multi-layer scheduling and cross-layer system optimization.
	World Model	[37], [82]	Enables low-latency local control via action – link modeling and latent roll out.
LAWNs	Digital Twin	[83], [84]	Supports 3D environment modeling and trajectory optimization for coverage planning.
	World Model	[35], [85], [86]	Enables onboard real-time control with obstacle avoidance and channel prediction.

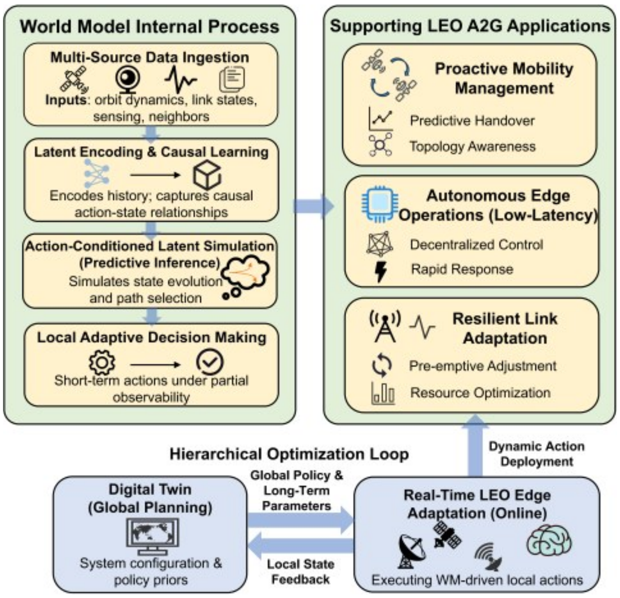

Fig. 8　Overview of world model - enabled framework for LEO - A2G networks: Latent simulation and causal learning with digital twin - edge hierarchical optimization supporting mobility, low-latency operations, and link adaptation.

mantic video transmission, traditional bit-level coding suffers from high bandwidth and rapid degradation under dynamic channels. Semantic communication with world foundation models transmits only a small set of semantic features, using a receiver-side world model as a shared prior within a prediction-feedback-adaptation framework: after the initial frame, subsequent frames are predicted, with retransmission triggered when feedback error exceeds a threshold. Simulation results under 5 dB SNR show the framework requires only 6 transmissions (5 KBytes) to maintain LPIPS at 0.2330, while limiting 20-frame bandwidth to under 10 KBytes[76].

3.3　Air-to-Ground Networks

Air-to-Ground (A2G) networks integrate satellites, high-altitude platforms, and UAVs into a space - air - ground (SAG) architecture, supporting wide-area connectivity, low-latency services, and edge intelligence. However, dynamic aerial topologies, heterogeneous resources, and strict latency requirements make resource allocation and task offloading challenging[89]. Digital twins provide global modeling of SAG structures and resource constraints, enabling system-level optimization. The integrated Digital Twin - World Model framework is shown in Fig.8.

Digital twins support resource management and computation offloading by building high-fidelity virtual replicas of SAG networks. Hevesli et al.[79] proposed a digital twin framework for air - ground cooperation, enabling real-time state prediction. Gong et al.[80] introduced a SAG digital twin with blockchain to support global resource scheduling and optimal task allocation.

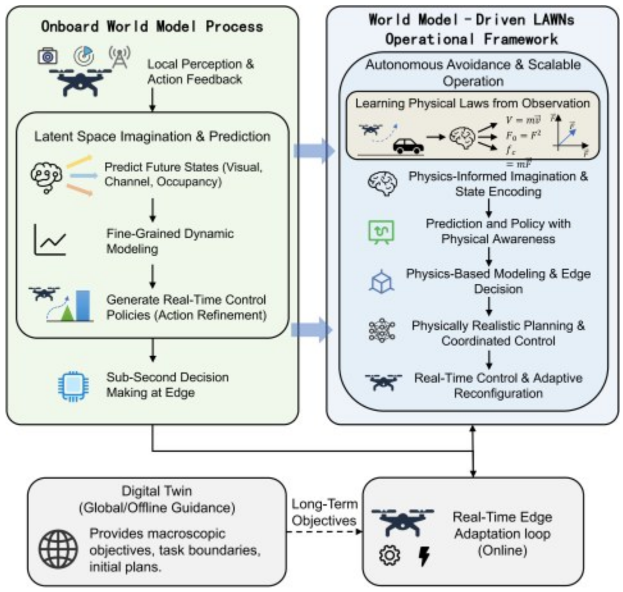

Fig. 9　Overview of the world model - driven framework for LAWNs: Onboard latent prediction, physics-aware control, and digital twin - guided edge adaptation.

World models act as local simulators for A2G edge nodes, focusing on the relation between agent actions and local link states under partial observability. Zhang et al.[37] developed a world model that encodes link states, trajectories, and actions into latent representations for path selection. Lu et al.[82] enabled action-conditioned prediction of spatial states to reduce reliance on centralized control. Zhang et al.[90] proposed an air - ground edge - cloud framework that improves multi-modal inference under limited bandwidth.

In A2G networks, digital twins and world models have distinct roles. Digital twins serve as global planning tools for system configuration[81], while world models enable local decision-making based on current conditions. Together, they form a hierarchical framework from system-level planning to link-level control.

Case Study of A2G: In SWIPT-enabled satellite-terrestrial HetNets, time-varying channels and multi-tier interference make resource allocation challenging for conventional model-free methods. To address this, a world model is deployed for each agent to encode channel and interference data into compact latent representations. By generating multi-step imaginary trajectories, the model minimizes real-world interactions while enabling joint optimization of beamforming and power allocation under QoS and energy constraints. Simulations show this approach converges nearly five times faster than baselines and reduces constraint violations from 4.11% to 0.73%, validating its effectiveness in enhancing sample efficiency and proactive planning[91].

3.4 Low-Altitude Wireless Networks

Low-Altitude Wireless Networks (LAWNs), comprising UAVs and eVTOLs, face highly unstructured environments and strict resource constraints that require both global planning and rapid local responses[92] . To address this, a hierarchical closed-loop architecture combining offline digital twins with online world models is utilized, as shown in Fig.9.

Digital twins provide global situational awareness by maintaining high-fidelity virtual replicas of urban environments and UAV states. This macroscopic view supports system-level optimization for trajectory planning, coordinated task management[83], and joint resource allocation[84].

Conversely, world models function as onboard embodied cognition for real-time local control[93]. By imagining state transitions in latent space, world models predict channel dynamics and environmental disturbances. This enables UAVs to perform online path planning in unfamiliar areas[35] and execute sub-second obstacle avoidance without frequent synchronization with centralized digital twins[85]. Additionally, world models can optimize reconfigurable intelligent surface (RIS) policies end-to-end by learning implicit channel mappings without requiring complex analytical models[86].

Ultimately, this collaborative framework leverages digital twins for global, long-term network configuration and world models for real-time, link-level adaptation[94]. This integration ensures efficient system-level resource utilization while granting aerial edge nodes the autonomy to dynamically navigate transient environmental changes.

Case Study of LAWNs: In low-altitude wireless networks, UAV swarms must plan trajectories under dynamic obstacles and changing missions, while traditional methods lack generalization and real-time decision support. The multi-UAV task is formulated as a multi-traveling salesman problem, and a hierarchical world model encodes swarm behavior into three probabilistic levels: mission, route, and motion. Each UAV adapts by minimizing the KL divergence between its belief and the world model reference, enabling real-time updates when new targets appear. Simulation results in a 40-target scenario show that the proposed framework outperforms GA-RF and modified Q-Learning in mission completion time, flight distance, convergence speed, and stability, confirming that the world model improves UAV swarm coordination in complex low-altitude environments[95].

3.5 Industrialization, Standardization, and Prototype Implementation

To enable the evolution from digital twins to world models for EGI, it is essential to move from laboratory concepts to real-world deployment. This transition depends on three interrelated components.

- Industrialization: Industry is integrating world models with edge computing to enhance deployability, combining them with digital twins via lightweight architectures and edge–cloud deployment. Progress exists in drones and robotics[8], though challenges remain in data scarcity and generalization.

- Standardization: Standardization efforts remain early, focusing on interface design and interoperability. Unified APIs and latent-state exchange protocols (e.g., RESTful, MQTT) should be established for world models and digital twins, compatible with existing standards like 3GPP and IEEE, alongside dedicated EGI benchmarks and interpretable interfaces for safety[96].

● Prototype Implementation: Several prototypes validate the approach: DreamerV3 for sub-second edge planning[27], aerial world models for UAV control[37], and industrial prototypes using physics-informed maintenance[69]. Key bottlenecks remain in real-time accuracy trade-offs and cross-vendor compatibility.

Industrialization, standardization, and prototype implementation bridge theory and practice. Current approaches center on latent-space abstraction and action-conditioned inference, with digital twins providing physical grounding. Future efforts should promote cross-industry collaboration to establish unified benchmarks and interface standards for large-scale deployment.

3.6 Limitations and Trade-offs of World Models

Current world models face several critical challenges in EGI. Limited by the computational and storage resources of edge devices, high-capacity models are difficult to deploy directly. The distribution gap between training data and real-world environments is harder to mitigate in open and dynamic scenarios[97]. In addition, error accumulation in multi-step prediction and insufficient long-term dynamics modeling further reduce reliability in complex tasks. Moreover, the limited ability to capture physical laws and causal relationships, together with the difficulty of multimodal information fusion, constrains both generalization and interpretability, making it difficult to meet the practical requirements of general intelligence[98].

At the same time, the design of world models involves multiple inherent trade-offs. There is a direct conflict between model representation capability and computational cost on edge devices. Achieving high-accuracy long-term prediction is difficult to reconcile with strict latency requirements for real-time response. Furthermore, balancing generality and task specificity makes it challenging to achieve both strong cross-scenario generalization and high performance in specific tasks. These factors jointly limit the reliable deployment of world models in EGI scenarios.

4 OPEN RESOURCES PROJECT

This section provides related open-source projects of digital twin, world model, and EGI across various fields.

4.1 Digital Twin Framework

Digital twins have advanced in multiple dimensions, including generation, behavior modeling, robust tracking, platform support, and object reconstruction. The following work highlights representative frameworks and methods that illustrate progress in these areas.

Deep learning enables digital twin generation. *FaceChain*[99] creates identity-preserving portraits using decoupled training and face perception models, with Classifier-Free Guidance and models like DamoFD[113] and M2FP[114]. LLMs enhance personalization: *PsyDT*[100] captures counselors’ linguistic patterns and therapy styles via dynamic one-shot learning and multi-turn dialogue synthesis. Robust tracking is provided by *DTTDNet*[101], a Transformer-based 6DoF network with geometric filtering that handles sensor noise and outperforms prior methods. Composable platforms like *DTaaS*[102] manage models, data, functions, and tools, supporting digital twins as services. Digital twins also model articulated objects: *Ditto*[103] reconstructs part-level geometry and articulation from visual observations using implicit neural representations for real-world simulation.

These advances improve digital twins in generation, personalization, tracking, platform management, and complex-object modeling, supporting EGI applications in perception, reasoning, and autonomous decision-making.

4.2 World Model Framework

World models represent and predict complex environments by integrating perception, planning, and simulation. They enable agents to learn and act efficiently across long tasks, high-dimensional spaces, and multiple domains, supporting AI and robotics ap-

TABLE VI: A summary of frameworks for digital twin, world model, edge general intelligence

Field	Method	Characteristic	Related Resource Link
Digital Twin	FaceChain[99]	Training-free and compatible	https://github.com/modelscope/facechain
	PsyDT[100]	LLMs and psychological counseling	https://github.com/scutcyr/SoulChat2.0
	DTTD2[101]	Robust and object tracking	https://github.com/augcog/DTTD2
	DTaaS[102]	Management and services	https://github.com/INTO-CPS-Association/DTaaS
	Ditto[103]	PointNet++ and articulated object	https://github.com/UT-Austin-RPL/Ditto
World Model	LWM[104]	Context understanding and training	https://github.com/LargeWorldModel/LWM
	DreamerV3[27]	RL and imagination-based planning	https://github.com/danijar/dreamerv3
	LingBot-World[105]	High-fidelity and long-horizon	https://github.com/Robbyant/lingbot-world
	IRIS[106]	Data-efficient and Sequence-modeling	https://github.com/eloialonso/iris
	GigaBrain-0[107]	Policy robustness and spatial reasoning	https://github.com/open-gigaai/giga-brain-0
Edge General Intelligence	LotteryFL[108]	Personalized and Low-Comm FL	https://github.com/charleslipku/LotteryFL
	Neurosurgeon[109]	Fine-grained and Layer-wise	https://github.com/Tjyy-1223/Neurosurgeon
	FedCache[110]	Device-Fit and Personalized	https://github.com/wuzhiyuan2000/FedCache
	ORRIC[111]	Adaptive Inference and Retraining	https://github.com/caihuaiguang/ORRIC
	pFedSD[112]	Faster personalization and robustness	https://github.com/CGCL-codes/pFedSD

plications.

Multimodal models with extended context have advanced. *LWM*[104] handles text, images, and videos across long- and short-context inputs with an open-source pipeline. World model-driven RL like *DreamerV3*[27] adapts to 150+ tasks without tuning, learning efficiently via reconstruction and actor – critic methods, and outperforming PPO and MuZero. Cross-domain simulations benefit from *LingBot-World*[105], which enables low-latency multi-domain simulations, agent training, and 3D reconstruction. Visual and long-horizon learning is enhanced by *IRIS*[106], using a discrete autoencoder and autoregressive Transformer to predict pixels, rewards, and terminations, surpassing humans in Atari 100k games. Efficient robot learning leverages *GigaBrain-0*[107], which combines RGBD data and chain-of-thought supervision to reason about geometry and long-horizon tasks for robust dexterous and mobile performance.

These studies show that world models support EGI by enabling reasoning, planning, simulation, and robot learning, improving efficiency and generalization for robust perception, decision-making, and autonomous actions.

4.3 Edge Artificial Intelligence Framework

EGI faces challenges in efficient, personalized learning across distributed devices. Using world models, federated learning, resource-aware scheduling, and knowledge distillation enhances efficiency, generalization, and supports diverse intelligent applications.

Reducing communication and enabling personalization are key in edge federated learning. *LotteryFL*[108] trains client-specific subnetworks using the Lottery Ticket hypothesis, cutting communication while improving accuracy on non-IID data with real-time edge deployment. Resource-aware frameworks like *Neurosurgeon*[109] partition networks across edge and cloud, adapting to hardware, architecture, and network conditions for better latency and efficiency. For heterogeneous clients, *FedCache*[110] uses a server knowledge cache and ensemble distillation to enhance communication while maintaining performance. Balancing computation and accuracy, *ORRIC*[111] models resource competition between retraining and inference to optimize long-term accuracy and latency. Self-distilled methods such as *pFedSD*[112] guide local training with client knowledge, improving personalization, convergence, and privacy with low overhead.

These studies highlight the importance of communication-efficient, resource-aware, and personalized strategies in edge environments. By reducing overhead, managing model drift, optimizing inference, and supporting diverse devices, they provide a solid basis for efficient and reliable EGI systems.

5 FUTURE DIRECTIONS

Advancing world models for EGI requires realism, adaptability, and efficiency in dynamic and constrained environments. Edge systems need models that combine physical knowledge with data-driven learning, support continual adaptation, and operate at different spatial and temporal scales. The following directions outline key areas for developing robust, explainable, and cooperative world models.

5.1 Hybrid Physics-Driven and Data-Driven World Models

Future work should investigate hybrid architectures that combine explicit physical knowledge (e. g., radio propagation models, mobility laws, power constraints) with learned latent dynamics. Purely data-driven world models often struggle with extrapolation under sparse, non-stationary, or shifted data, and their generalization and scalability in complex real-world robotic scenarios remain in question[115]. In contrast, purely physics-based digital twins can be brittle and computationally expensive. Promising directions include physics-informed latent spaces, generative models regularized by conservation laws[116], and differentiable simulators coupled with neural world models. For EGI, such hybrid designs can improve robustness and interpretability while remaining lightweight enough for deployment at the edge.

5.2 Federated World Modeling at the Edge

EGI requires world models that evolve as environments, traffic patterns, and hardware conditions change. Future research should address continual and lifelong learning of world models under strict resource and privacy constraints, so that intelligent systems can retain existing knowledge while continuously acquiring and integrating new information[117]. This includes mechanisms for online adaptation without catastrophic forgetting, efficient model versioning across heterogeneous edge nodes, and federated learning protocols tailored to world-model training (e.g., regularizing latent dynamics to improve agent behavior[118]). Handling non-IID data, enabling communication-efficient aggregation, and supporting privacy-preserving updates will be central challenges.

5.3 Multi-Agent and Multi-Scale World Models

Edge scenarios such as 6G, air-to-ground networks, and low-altitude operations inherently involve many interacting agents (devices, base stations, UAVs, vehicles) evolving across multiple temporal and spatial scales. Future research should explore world models that capture multi-agent interactions (e. g., via graph-structured agent-level interaction modules[119]) and multi-scale processes (fast wireless-channel fluctuations vs. slower mobility and traffic patterns). Agent-centric models with interactive perception capabilities can support decentralized collaboration among distributed edge nodes, enable predicting emergent behaviors, and facilitate cooperative planning. Decentralized agents may form collusion through covert communication[120], highlighting the need to move beyond single-agent, local-view world models and to develop mechanisms for effective decentralized planning.

5.4 Explainable and Trustworthy World Models

As EGI systems are deployed in high-stakes environments, it is crucial that their decisions are understandable, reliable, and trustworthy. Current world models, often based on deep neural networks, tend to act as black boxes, making failures difficult to interpret and diagnose[121]. A key research direction is the development of explainable artificial intelligence (XAI) techniques tailored to world models. Potential approaches include methods to visualize the model's imagined futures, to identify which environmental features are most influential for its predictions, and to quantify uncertainty in its forecasts. Building trust also requires mechanisms to detect out-of-distribution conditions and to decide when the world model should not be used for planning, as such detection is a key component of trusted machine learning systems[122] and enables graceful degradation, conservative fallback strategies, or safe human or rule-based intervention.

6 CONCLUSION

This survey has outlined a unified perspective on the transition from digital twins to world models for EGI. Digital twins remain indispensable for high-fidelity engineering analysis, lifecycle management, and system-level optimization. However, their reliance on explicit modeling, centralized computation, and continuous synchronization limits their suitability for autonomous and real-time operations at the edge. World models address these limitations by learning compact, action-conditioned representations of the environment and by enabling imagination-based planning and self-supervised adaptation. This survey has reviewed core architectures and algorithms for world models, their coupling with communication, sensing, and control, and their emerging role in future wireless networks and cyber – physical systems. We have also highlighted open issues, including hybrid physics – data integration, federated and continual world modeling under non-IID edge data, multi-agent and multi-scale modeling, as well as safety, explainability, and standardization. These challenges define a rich research agenda for the coming years. Thus, the synergistic use of digital twins and world models is expected to provide a key technological pillar for robust, efficient, and intelligent edge systems in 6G and beyond.

References:

[1] H. Chen, W. Deng, S. Yang, J. Xu, Z. Jiang, E. C. H. Ngai, J. Liu, and X. Liu, "Towards edge general intelligence via large language models: Opportunities and challenges," IEEE Network, vol. 39, no. 5, pp. 263 – 271, 2025.

[2] N. Syed, A. Anwar, Z. Baig, and S. Zeadally, "Artificial intelligence as a service (aiaas) for cloud, fog and the edge: State-of-the-art practices," vol. 57, no. 8, Mar. 2025.

[3] D. Katare, D. Perino, J. Nurmi, M. Warnier, M. Janssen, and A. Y. Ding, "A survey on approximate edge ai for energy efficient autonomous driving services," IEEE Communications Surveys & Tutorials, vol. 25, no. 4, pp. 2714 – 2754, 2023.

[4] X. Wang, Z. Tang, J. Guo, T. Meng, C. Wang, T. Wang, and W. Jia, "Empowering edge intelligence: A comprehensive survey on on-device ai models," ACM Comput. Surv., vol. 57, no. 9, Apr. 2025.

[5] G. K. Pandey, D. S. Gurjar, S. Yadav, Y. Jiang, and C. Yuen, "Uav assisted communications with rf energy harvesting: A comprehensive survey," IEEE Communications Surveys & Tutorials, vol. 27, no. 2, pp. 782 – 838, 2025.

[6] B. Liu, H. Shi, D. Jia, E. Wang, W. Han, K. Zhong, L. Wu, S. Chen, C. Qiao, and J. Wang, "Collaborative sensing and communication for intelligent connected vehicles: A comprehensive survey," IEEE Communications Surveys & Tutorials, vol. 28, pp. 3125 – 3164, 2026.

[7] S. V. Balkus, H. Wang, B. D. Cornet, C. Mahabal, H. Ngo, and H. Fang, "A survey of collaborative machine learning using 5g vehicular communications," IEEE Communications Surveys & Tutorials, vol. 24, no. 2, pp. 1280 – 1303, 2022.

[8] X. Long, Q. Zhao, K. Zhang, Z. Zhang, D. Wang, Y. Liu, Z. Shu, Y. Lu, S. Wang, X. Wei et al., "A survey: Learning embodied intelligence from physical simulators and world models," arXiv preprint arXiv: 2507.00917, 2025.

[9] J. Ding, Y. Zhang, Y. Shang, Y. Zhang, Z. Zong, J. Feng, Y. Yuan, H. Su, N. Li, N. Sukiennik et al., "Understanding world or predicting future? a comprehensive survey of world models," ACM Computing Surveys, vol. 58, no. 3, pp. 1 – 38, 2025.

[10] M. Goff, G. Hogan, G. Hotz, A. du Parc Locmaria, K. Raczy, H. Sch¨afer, A. Shihadeh, W. Zhang, and Y. Yousfi, "Learning to drive from a world model," in Proceedings of the Computer Vision and Pattern Recognition Conference, 2025, pp. 1964 – 1973.

[11] P. Empl, D. Koch, M. Dietz, and G. Pernul, "Digital twins in security operations: State of the art and future perspectives," ACM Comput. Surv., vol. 58, no. 1, Sep. 2025.

[12] D. Li, D. Han, N. Crespi, R. Minerva, S. M. Raza, R. Farahbakhsh, W. Liang, and Z. Zheng, "Blockchain in the digital twin context: A comprehensive survey," ACM Comput. Surv., vol. 58, no. 6, Dec. 2025.

[13] Y. Mu, T. Chen, Z. Chen, S. Peng, Z. Lan, Z. Gao, Z. Liang, Q. Yu, Y. Zou, M. Xu, L. Lin, Z. Xie, M. Ding, and P. Luo, "Robotwin: Dualarm robot benchmark with generative digital twins," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 2025, pp. 27649 – 27660.

[14] J. Chen, W. Wang, B. Fang, Y. Liu, K. Yu, V. C. M. Leung, and X. Hu, "Digital twin empowered wireless healthcare monitoring for smart home," IEEE Journal on Selected Areas in Communications, vol. 41, no. 11, pp. 3662 – 3676, 2023.

[15] Z. Yang, M. Chen, Y. Liu, and Z. Zhang, "A joint communication and computation framework for digital twin over wireless networks," IEEE Journal of Selected Topics in Signal Processing, vol. 18, no. 1, pp. 6 – 17, 2024.

[16] S. Mihai, M. Yaqoob, D. V. Hung, W. Davis, P. Towakel, M. Raza, M. Karamanoglu, B. S. Barn, D. Shetve, R. V. Prasad, H. Venkataraman, R. Trestian, and H. X. Nguyen, "Digital twins: A survey on enabling technologies, challenges, trends and future prospects," IEEE Communications Surveys & Tutorials, vol. 24, pp. 2255 – 2291, 2022.

[17] G. Dagnaw, R. Capuano, and H. Muccini, "Digital twins for cultural heritage: A systematic analysis of the state of the art," ACM Comput. Surv., 2026, just Accepted.

[18] C. Zhao, R. Zhang, J. Wang et al., “World models for cognitive agents: Transforming edge intelligence in future networks,” arXiv preprint arXiv:2506.00417, 2025.

[19] G. Zhao, X. Wang, Z. Zhu, X. Chen, G. Huang, X. Bao, and X. Wang, “Drivedreamer-2: Llm-enhanced world models for diverse driving video generation,” Proceedings of the AAAI Conference on Artificial Intelligence, vol. 39, no. 10, pp. 10412 – 10420, Apr. 2025.

[20] Q. He, W. Liang, C. Hao, G. Sun, and J. Tian, “Glam: Global-local variation awareness in mamba-based world model,” Proceedings of the AAAI Conference on Artificial Intelligence, vol. 39, no. 16, pp. 17105 17113, Apr. 2025.

[21] Y. Yang, J. Mei, Y. Ma, S. Du, W. Chen, Y. Qian, Y. Feng, and Y. Liu, “Driving in the occupancy world: Vision-centric 4d occupancy forecasting and planning via world models for autonomous driving,” Proceedings of the AAAI Conference on Artificial Intelligence, vol. 39, no. 9, pp. 9327 – 9335, Apr. 2025.

[22] S. Xiong, A. Payani, Y. Yang, and F. Fekri, “Deliberate reasoning in language models as structure-aware planning with an accurate world model,” in Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). Vienna, Austria: Association for Computational Linguistics, Jul. 2025, pp. 31900 – 31931.

[23] Q. Jia, J. Zheng, L. Gao, J. Niu, R. Cao, and J. Ren, “Satellite aided low-altitude uav service migration with semantic extraction and generated graphs,” IEEE Transactions on Cognitive Communications and Networking, vol. 12, pp. 5136 – 5147, 2026.

[24] C. Hao, W. Lu, Y. Xu, and Y. Chen, “Neural motion simulator pushing the limit of world models in reinforcement learning,” in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 2025, pp. 27608 – 27617.

[25] X. Huang, H. Yang, C. Zhou, M. He, X. Shen, and W. Zhuang, “When digital twin meets generative AI: Intelligent closed-loop network management,” IEEE Network, 2025, to appear.

[26] D. Ha and J. Schmidhuber, “Recurrent world models facilitate policy evolution,” in Advances in Neural Information Processing Systems (NeurIPS), vol. 31, Montr´eal, Canada, 2018.

[27] D. Hafner, J. Pasukonis, J. Ba, and T. Lillicrap, “Mastering diverse control tasks through world models,” Nature, vol. 640, no. 8059, pp. 647 – 653, 2025.

[28] F. Tao, J. Qi, L. Zhang et al., “Digital twin modeling: A systematic literature review and meta-analysis,” ACM Computing Surveys, vol. 57, no. 3, pp. 1 – 34, 2025.

[29] L. Tao, Y. Zheng, J. Cao, and F. Tao, “Digital twin-driven smart manufacturing: A review and future directions,” IEEE Transactions on Industrial Informatics, vol. 21, pp. 1 – 11, 2025.

[30] Y. Zhang et al., “DriveDreamer4D: World models are effective data machines for 4D driving scene representation,” in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), 2025, p. to appear.

[31] Y. Shi, K. Yang, T. Jiang, J. Zhang, and K. B. Letaief, “Communication efficient edge ai: Algorithms and systems,” IEEE Communications Surveys & Tutorials, vol. 22, no. 4, pp. 2167 – 2191, 2020.

[32] Y. Guan, Liao et al., “World models for autonomous driving: An initial survey,” IEEE Transactions on Intelligent Vehicles, pp. 1 – 17, 2024.

[33] M. Tao, L. Liao, R. Xie, Y. Zhang, G. Min, and Y. Zhang, “SAEIDT: Semi-asynchronous edge intelligence for industrial digital twin networks in 6g,” IEEE Network, vol. 39, no. 5, pp. 138 – 144, Sep. 2025.

[34] R. Zhou, D. Chen, Z. Jia et al., “Digital twin AI: Opportunities and challenges from large language models to world models,” arXiv preprint arXiv:2601.01321, 2026.

[35] A. Bar, G. Zhou, D. Tran, T. Darrell, and Y. LeCun, “Navigation world models,” in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2025, pp. 15791 – 15801.

[36] L. He, L. Fan, X. Lei, P. Fan, A. Nallanathan, and G. K. Karagiannidis, “The road toward general edge intelligence: Standing on the shoulders of foundation models,” IEEE Communications Magazine, vol. 63, no. 9, pp. 1 – 9, 2025.

[37] W. Zhang, P. Tang, X. Zeng, F. Man, S. Yu, Z. Dai, B. Zhao, H. Chen, Y. Shang, W. Wu, C. Gao, X. Chen, X. Wang, Y. Li, and W. Zhu, “Aerial world model for long-horizon visual generation and navigation in 3d space,” arXiv preprint arXiv:2512.21887, 2025.

[38] Y. Wu, L. Ma, R. Zhang et al., “Towards edge general intelligence: Knowledge distillation for mobile agentic AI,” arXiv preprint arXiv: 2511.19947, 2025.

[39] R. Zhang, H. Du, Y. Liu, D. Niyato, J. Kang, Z. Xiong, A. Jamalipour, and D. In Kim, “Generative ai agents with large language model for satellite networks via a mixture of experts transmission,” IEEE Journal on Selected Areas in Communications, vol. 42, no. 12, pp. 3581 – 3596, 2024.

[40] T. Susnjak, T. R. McIntosh, A. L. C. Barczak et al., “Over the edge of chaos? excess complexity as a roadblock to artificial general intelligence,” IEEE Transactions on Cybernetics, vol. 56, no. 1, pp. 1 – 12, 2025.

[41] Y. Tang, J. Yu, K. Gai et al., “Missing target-relevant information prediction with world model for accurate zero-shot composed image retrieval,” in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2025, pp. 24785 – 24795.

[42] Z. Li, J. Lou, Z. Tang, J. Guo, T. Wang, W. Jia, and W. Zhao, “Online layer-aware joint request scheduling, container placement, and resource provision in edge computing,” IEEE Transactions on Services Computing, vol. 18, no. 1, pp. 328 – 341, 2025.

[43] Y. Shen, H. Liu, K. Pei et al., “MetaWorld: Skill transfer and composition in a hierarchical world model for grounding high-level instructions,” arXiv preprint arXiv:2601.17507, 2026.

[44] R. Worden, “AI and world models,” arXiv preprint arXiv:2601.17796, 2026.

[45] W. Liu, Y. Fu, Z. Shi, and H. Wang, “When digital twin meets 6g: Concepts, obstacles, and research prospects,” IEEE Communications Magazine, vol. 63, pp. 16 – 22, 2024.

[46] F. Tao, H. Zhang, and C. Zhang, “Advancements and challenges of digital twins in industry,” Nature Computational Science, vol. 4, pp. 169 – 177, 2024.

[47] J. Li, S. Guo, W. Liang, J. Wang, Q. Chen, Y. Zeng, B. Ye, and X. Jia, “Digital twin-enabled service provisioning in edge computing via continual learning,” IEEE Transactions on Mobile Computing, vol. 23, pp. 7335 – 7350, 2024.

[48] R. Zhang, G. Liu, Y. Liu, C. Zhao, J. Wang, Y. Xu, D. Niyato, J. Kang, Y. Li, S. Mao, S. Sun, X. Shen, and D. I. Kim, “Toward edge general intelligence with agentic ai and agentification: Concepts, technologies, and future directions,” IEEE Communications Surveys & Tutorials, vol. 28, pp. 4285 – 4318, 2025.

[49] T. Meuser, L. Lov´en, M. H. Bhuyan, S. G. Patil, S. Dustdar, A. Aral, S. Bayhan, C. Becker, E. de Lara, A. Y. Ding, J. Edinger, J. Gross, N. Mohan, A. D. Pimentel, E. Rivi`ere, H. Schulzrinne, P. Simoens, G. Solmaz, M. Welzl, and S. Dustdar, "Revisiting edge ai: Opportu nities and challenges," IEEE Internet Computing, vol. 28, pp. 49–59, 2024.

[50] Y. Zheng, P. Yang, Z. Xing, Q. Zhang, Y. Zheng, Y. Gao, P. Li, T. Zhang, Z. Xia, P. Jia, and D. Zhao, "World4drive: End-to-end autonomous driving via intention-aware physical latent world model," 2025.

[51] Q. Fang, W. Du, H. Wang, and J. Zhang, "Towards unraveling and improving generalization in world models," ArXiv, vol. abs/2501.00195, 2024.

[52] S. Deng, H. Zhao, J. Yin, S. Dustdar, and A. Y. Zomaya, "Edge intelligence: The confluence of edge computing and artificial intelligence," IEEE Internet of Things Journal, vol. 7, pp. 7457–7469, 2019.

[53] E. Baccour, N. Mhaisen, A. A. Abdellatif, A. Erbad, A. Mohamed, M. Hamdi, and M. Guizani, "Pervasive ai for iot applications: A survey on resource-efficient distributed artificial intelligence," IEEE Communications Surveys & Tutorials, vol. 24, no. 4, pp. 2366–2418, 2022.

[54] M. A. Ali and F. Dornaika, "Edge artificial intelligence: A systematic review of evolution, taxonomic frameworks, and future horizons," ArXiv, vol. abs/2510.01439, 2025.

[55] Q. He, J. Lin, H. Fang, X. Wang, M. Huang, X. shuang Yi, and K. Yu, "Integrating iot and 6g: Applications of edge intelligence, challenges, and future directions," IEEE Transactions on Services Computing, vol. 18, pp. 2471–2488, 2025.

[56] Z. Liu, X. Chen, H. Wu, Z. Wang, X. Chen, D. T. Niyato, and K. Huang, "Integrated sensing and edge ai: Realizing intelligent perception in 6g," IEEE Communications Surveys & Tutorials, vol. 28, pp. 2725–2770, 2025.

[57] S. Zeng, X. Chang, M. Xie, X. Liu, Y. Bai, Z. Pan, M. Xu, and X. Wei, "Futuresightdrive: Thinking visually with spatio-temporal cot for autonomous driving," in The Thirty-ninth Annual Conference on Neural Information Processing Systems, 2025.

[58] L. Baraldi, Z. Zeng, C. Zhang, A. Nayak, H. Zhu, F. Liu, Q. Zhang, P. Wang, S. Liu, Z. Hu, A. Cangelosi, and L. Baraldi, "The safety challenge of world models for embodied ai agents: A review," ArXiv, vol. abs/2510.05865, 2025.

[59] X. Li, X. He, L. Zhang, and Y. Liu, "A comprehensive survey on world models for embodied ai," ArXiv, vol. abs/2510.16732, 2025.

[60] H. Luo, Y. Liu, R. Zhang, J. Wang, G. Sun, D. Niyato, H. Yu, Z. Xiong, X. Wang, and X. Shen, "Toward edge general intelligence with multiple-large language model (multi-llm): Architecture, trust, and orchestration," IEEE Transactions on Cognitive Communications and Networking, vol. 11, pp. 3563–3585, 2025.

[61] X. Wang and W. Jia, "Optimizing edge ai: A comprehensive survey on data, model, and system strategies," ArXiv, vol. abs/2501.03265, 2025.

[62] P. Yang, B. Lu, Z. Xia, C. Han, Y. Gao, T. Zhang, K. Zhan, X. Lang, Y. Zheng, and Q. Zhang, "Worldrft: Latent world model planning with reinforcement fine-tuning for autonomous driving," ArXiv, vol. abs/2512.19133, 2025.

[63] D. Wen, Y. Zhou, X. Li, Y. Shi, K. Huang, and K. B. Letaief, "A survey on integrated sensing, communication, and computation," IEEE Communications Surveys & Tutorials, vol. 27, no. 5, 2025.

[64] B. Li, W. Liu, W. Xie, N. Zhang, and Y. Zhang, "Adaptive digital twin for UAV-assisted integrated sensing, communication, and computation networks," IEEE Transactions on Green Communications and Networking, vol. 7, no. 4, pp. 1996–2009, Aug. 2023.

[65] Y. Li, W. Liang, Z. Xu, W. Xu, and X. Jia, "Budget-constrained digital twin synchronization and its application on fidelity-aware queries in edge computing," IEEE Transactions on Mobile Computing, vol. 24, no. 1, pp. 165–182, Jan. 2025.

[66] J. Del Ser et al., "World models in artificial intelligence: Sensing, learning, and reasoning like a child," Mar. 2025.

[67] X. Chen, K. Huang et al., "Distributed integrated sensing and edge AI exploiting prior information," 2025.

[68] Y. Ma, B. Ai, J. Li et al., "Integrated sensing, communication, computing, and control meets UAV swarms in 6g," 2025.

[69] C. Deng et al., "Integrated sensing, communication, and computation with adaptive DNN splitting in multi-UAV networks," IEEE Transactions on Wireless Communications, 2024, early access.

[70] X. Cheng, R. Meng, X. Xu, H. Gao, P. Zhang, and D. Niyato, "Apeg: Adaptive physical layer authentication with channel extrapolation and generative ai," IEEE Transactions on Information Forensics and Security, vol. 21, pp. 1257–1272, 2026.

[71] D. Wen et al., "Integrated sensing, communication, and computation for over-the-air federated edge learning," IEEE Transactions on Wireless Communications, vol. 25, pp. 2748–2762, 2026.

[72] P.-H. Ho, H. Mei, L. Peng, Y. Miao, K. Liang, and Y. Jiao, "From snapshot sensing to persistent EM world modeling: A generative-space perspective for ISAC," arXiv preprint arXiv:2602.13554, 2026.

[73] B. Li, H. Cai, L. Liu, and Z. Fei, "Delay-aware digital twin synchronization in mobile edge networks with semantic communications," IEEE Transactions on Vehicular Technology, vol. 74, no. 7, pp. 10974 10983, Jul. 2025.

[74] S. D. Okegbile, H. Gao, and J. Cai, "A novel secure split federated semantic learning framework and its optimization for digital twin network evolution," IEEE Transactions on Mobile Computing, vol. 25, no. 1, pp. 1302–1319, Jan. 2026.

[75] F. Tang, L. Luo, Z. Guo, M. Zhao, and N. Kato, "Semantic twin network: Bridging real-world and virtual networks with semantics," 17 IEEE Wireless Communications, vol. 32, no. 6, pp. 224–232, Dec. 2025.

[76] P. Jiang, J. Guo, C.-K. Wen, S. Jin, and J. Zhang, "Semantic communications with world models," arXiv preprint arXiv:2510.24785, 2025.

[77] S. Tan et al., "SceneDiffuser++: City-scale traffic simulation via a generative world model," in Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2025.

[78] Y. Yang et al., "Integrated sensing, computing, and semantic communication with fluid antenna for metaverse," arXiv preprint arXiv: 2504.07656, Apr. 2025.

[79] M. Hevesli, A. M. Seid, A. M. Erbad, and M. M. Abdallah, "Task offloading optimization in digital twin assisted mec-enabled air–ground iiot 6g networks," IEEE Transactions on Vehicular Technology, vol. 73, pp. 17527–17542, 2024.

[80] Y. Gong, H. Yao, Z. Xiong, C. L. P. Chen, and D. Niyato, "Blockchainaided digital twin offloading mechanism in space-air-ground networks," IEEE Transactions on Mobile Computing, vol. 24, pp. 183–197, 2025.

[81] Z. Lin, Z. Feng, K. Guo, A. Nauman, D. T. Niyato, and J. Wang, "Ai-driven seamless and massive access in space-air-ground integrated net-

works," IEEE Wireless Communications, vol. 32, pp. 72–79, 2025.

[82] Y. Lu, B. Wu, Z. Li, K. Li, C. Huang, H. Wang, Q. Lan, R. Chen, L. Chen, and B. Liang, "Remote sensing-oriented world model," ArXiv, vol. abs/2509.17808, 2025. [Online]. Available: https://api.semanticscholar.org/CorpusID:281420476

[83] W. Xie, F. Qi, L. Liu, and Q. Liu, "Radar imaging based uav digital twin for wireless channel modeling in mobile networks," IEEE Journal on Selected Areas in Communications, vol. 41, pp. 3702–3710, 2023.

[84] C. Wang, Y. Han, L. Zhang, Z. Jia, H. Zhang, C. S. Hong, and Z. Han, "Computing power in the sky: Digital twin-assisted collaborative computing with multi-uav networks," IEEE Transactions on Vehicular Technology, vol. 74, pp. 14 466–14 482, 2025.

[85] C. Diehl, Q. Sykora, B. Agro, T. Gilles, S. Casas, and R. Urtasun, "Dio: Decomposable implicit 4d occupancy-flow world model," 2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 27 456–27 466, 2025.

[86] S. Chen, J. Gu, W. Duan, M. Wen, G. Zhang, and P.-H. Ho, "Hybrid near- and far-field communications for ris-uav system: Novel beamfocusing design," IEEE Transactions on Intelligent Transportation Systems, vol. 26, pp. 17866–17878, 2025.

[87] C. Chaccour, W. Saad, M. Debbah, Z. Han, and H. V. Poor, "Less data, more knowledge: Building next-generation semantic communication networks," IEEE Communications Surveys & Tutorials, vol. 27, no. 1, pp. 37–76, Feb. 2025.

[88] J. Zheng, B. Du, H. Du, J. Kang, D. Niyato, and H. Zhang, "Energyefficient resource allocation in generative ai-aided secure semantic mobile networks," IEEE Transactions on Mobile Computing, vol. 23, no. 12, pp. 11 422–11 435, 2024.

[89] C. Hu, R. Zhang, B. Li, X. Jiang, N. Zhao, M. D. Renzo, D. Niyato, A. Nallanathan, and G. K. Karagiannidis, "Generative ai-empowered secure communications in space–air–ground integrated networks: A survey and tutorial," IEEE Communications Surveys & Tutorials, vol. 28, pp. 4156–4194, 2025.

[90] S. Zhang, Q. Liu, K. Chen, B. Di, H. Zhang, W. Yang, D. Niyato, Z. Han, and H. V. Poor, "Large models for aerial edges: An edgecloud model evolution and communication paradigm," IEEE Journal on Selected Areas in Communications, vol. 43, no. 1, pp. 21–35, 2025.

[91] G. Liu, Y. Liu, R. Zhang, D. Niyato, J. Kang, S. Sun, A. Jamalipour, and P. Zhang, "Dwm-ro: Decentralized world models with reasoning offloading for swipt-enabled satellite-terrestrial hetnets," ArXiv, vol. abs/2511.05972, 2025.

[92] L. Cai, J. Wang, R. Zhang, Y. Zhang, T. Jiang, D. Niyato, X. Wang, A. Jamalipour, and X. Shen, "Secure physical layer communications for low-altitude economy networking: A survey," IEEE Communications Surveys & Tutorials, vol. 28, pp. 2497–2530, 2025.

[93] B. Zhao, R. Tang, M.-M. Jia, Z. Wang, F. Man, X. Zhang, Y. Shang, W. Zhang, W. Wu, C. Gao, X. Chen, and Y. Li, "Airscape: An aerial generative world model with motion controllability," Proceedings of the 33rd ACM International Conference on Multimedia, 2025.

[94] J. Wu, Y. Yang, W. Yuan, W. Liu, J. Wang, T. Mao, L. Zhou, Y. Cui, F. Liu, G. Sun, Y. Ma, N. Wu, D. Zheng, J. Xu, N. Ma, Z. Feng, W. Xu, D. Niyato, C. Yuen, X. Jing, Z. Shi, Y. Liang, B. Ai, S. Jin, D. I. Kim, J. Wang, P. Zhang, H. Yin, and J. Zhang, "Low-altitude wireless networks: A comprehensive survey," 2025.

[95] K. Arshid, A. Krayani, L. Marcenaro, D. M. G´omez, and C. S. Regazzoni, "Active inference-driven world modeling for adaptive uav swarm trajectory design," ArXiv, vol. abs/2601.12939, 2026.

[96] IEEE Standard for Orchestration of Digital Synchronization Between Cyber and Physical Worlds, IEEE Std. IEEE Std 2888.3-2024, Nov. 2024.

[97] Z. Zhu, X. Wang, W. Zhao, C. Min, N. Deng, M. Dou, Y. Wang, B. Shi, K. Wang, C. Zhang, Y. You, Z. Zhang, D. Zhao, L. Xiao, J. Zhao, J. Lu, and G. Huang, "Is sora a world simulator? a comprehensive survey on general world models and beyond," ArXiv, vol. abs/2405.03520, 2024.

[98] S. Tu, X. Zhou, D. Liang, X. Jiang, Y. Zhang, X. Li, and X. Bai, "The role of world models in shaping autonomous driving: A comprehensive survey," ArXiv, vol. abs/2502.10498, 2025.

[99] Y. Liu, C. Yu, L. Shang, Z. Wu, X. Wang, Y. Zhao, L. Zhu, C. Cheng, W. Chen, C. Xu, H. Xie, Y. Yao, W. Zhou, C. Yingda, X. Xie, and B. Sun, "Facechain: A playground for identity-preserving portrait generation," arXiv preprint arXiv:2308.14256, 2023.

[100] H. Xie, Y. Chen, X. Xing, J. Lin, and X. Xu, "PsyDT: Using LLMs to construct the digital twin of psychological counselor with personalized counseling style for psychological counseling," in Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). Association for Computational Linguistics, Jul. 2025, pp. 1081–1115.

[101] Z. Huang, K. Yao, Z. Zhao, C. Pan, and A. Yang, "Dttdnet: Robust 6dof pose estimation against depth noise and a comprehensive evaluation on a mobile dataset," in Proceedings of the Computer Vision and Pattern Recognition Conference (CVPR) Workshops, June 2025, pp. 1848–1857.

[102] P. Talasila, C. Gomes, L. B. Vosteen, H. Iven, M. Leucker, S. Gil, P. H. Mikkelsen, E. Kamburjan, and P. G. Larsen, "Composable digital twins on digital twin as a service platform," SIMULATION, p. 00375497241298653, 2024.

[103] Z. Jiang, C.-C. Hsu, and Y. Zhu, "Ditto: Building digital twins of articulated objects from interaction," in arXiv preprint arXiv: 2202.08227, 2022.

[104] H. Liu, W. Yan, M. Zaharia, and P. Abbeel, "World model on millionlength video and language with ringattention," arXiv preprint, 2024.

[105] R. Team, Z. Gao, Q. Wang, Y. Zeng, J. Zhu, K. L. Cheng, Y. Li, H. Wang, Y. Xu, S. Ma, Y. Chen, J. Liu, Y. Cheng, Y. Yao, J. Zhu, Y. Meng, K. Zheng, Q. Bai, J. Chen, Z. Shen, Y. Yu, X. Zhu, Y. Shen, and H. Ouyang, "Advancing open-source world models," arXiv preprint arXiv:2601.20540, 2026.

[106] V. Micheli, E. Alonso, and F. Fleuret, "Transformers are sampleefficient world models," in The Eleventh International Conference on Learning Representations, 2023.

[107] GigaAI, "Gigabrain-0: A world model-powered vision-language-action model," arXiv, 2025.

[108] A. Li, J. Sun, B. Wang, L. Duan, S. Li, Y. Chen, and H. Li, "Lotteryfl: Empower edge intelligence with personalized and communicationefficient federated learning," in 2021 IEEE/ACM Symposium on Edge Computing (SEC), 2021, pp. 68–79.

[109] Y. Kang, J. Hauswald, C. Gao, A. Rovinski, T. Mudge, J. Mars, and L. Tang, "Neurosurgeon: Collaborative intelligence between the cloud and mobile edge," ACM SIGARCH Computer Architecture News, vol. 45, no. 1, pp. 615–629, 2017.

[110] Z. Wu, S. Sun, Y. Wang, M. Liu, K. Xu, W. Wang, X. Jiang, B. Gao, and J. Lu, "Fedcache: A knowledge cache-driven federated learning architecture for personalized edge intelligence," IEEE Transactions on Mobile Computing, pp. 1–15, 2024.

[111] H. Cai, Z. Zhou, and Q. Huang, "Online resource allocation for edge intelligence with colocated model retraining and inference," in IEEE INFOCOM 2024 - IEEE Conference on Computer Communications, 2024, pp. 1900–1909.

[112] H. Jin, D. Bai, D. Yao, Y. Dai, L. Gu, C. Yu, and L. Sun, "Personalized edge intelligence via federated self-knowledge distillation," IEEE Transactions on Parallel and Distributed Systems, vol. 34, no. 2, pp. 567–580, 2023.

[113] Y. Liu, J. Deng, F. Wang, L. Shang, X. Xie, and B. Sun, "DamoFD: Digging into backbone design on face detection," in The Eleventh International Conference on Learning Representations, 2023.

[114] B. Cheng, I. Misra, A. G. Schwing, A. Kirillov, and R. Girdhar, "Masked-attention mask transformer for universal image segmentation," in 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2022, pp. 1280–1289.

[115] Z. Zhang, R. Chen, J. Ye, Y. Sun, H. Ren, X. Du, P. Wang, J.-C. Pang, K. Li, T.-S. Liu, H. Lin, Y. Yu, and Z.-H. Zhou, "WHALE: Towards generalizable and scalable world models for embodied decisionmaking," in NeurIPS 2025 Workshop on Embodied World Models for Decision Making, 2025.

[116] U. Utkarsh, P. Cai, A. Edelman, R. Gomez-Bombarelli, and C. V. Rackauckas, "Physics-constrained flow matching: Sampling generative models with hard constraints," in The Thirty-ninth Annual Conference on Neural Information Processing Systems, 2025.

[117] J. Zheng, C. Shi, X. Cai, Q. Li, D. Zhang, C. Li, D. Yu, and Q. Ma, "Lifelong learning of large language model based agents: A roadmap," IEEE Transactions on Pattern Analysis and Machine Intelligence, pp. 1–20, 2026.

[118] T. Saanum, P. Dayan, and E. Schulz, "Simplifying latent dynamics with softly state-invariant world models," in Advances in Neural Information Processing Systems, GlobersonA., MackeyL., BelgraveD., FanA., PaquetU., TomczakJ., and ZhangC., Eds., vol. 37. Curran Associates, Inc., 2024, pp. 38 355–38 382.

[119] J. Jeong, D. Park, and K.-J. Yoon, "Multi-agent long-term 3d human pose forecasting via interaction-aware trajectory conditioning," in Proceedings ofthe IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 2024, pp. 1617–1628.

[120] S. Motwani, M. Baranchuk, M. Strohmeier, V. Bolina, P. Torr, L. Hammond, and C. Schroeder de Witt, "Secret collusion among ai agents: Multi-agent deception via steganography," Advances in Neural Information Processing Systems, vol. 37, pp. 73 439–73 486, 2024.

[121] Z. Aghababaeyan, M. Abdellatif, L. Briand, R. S, and M. Bagherzadeh, "Black-box testing of deep neural networks through test case diversity," IEEE Transactions on Software Engineering, vol. 49, no. 5, pp. 3182– 3204, 2023.

[122] M. Sun, J. Zheng, H. Du, H. Zhang, D. Niyato, J. Kang, J. Wang, J. Ren, L. Gao, and Z. Wang, "Trust online over-the-air computation for wireless federated learning," IEEE Transactions on Mobile Computing, vol. 24, no. 8, pp. 7152–7170, 2025.