TSINGHUA SCIENCE AND TECHNOLOGY
ISSN 1007-0214 01/37 pp1–15
DOI: 10.26599/TST.2025.9010127
Volume 31, Number 1, February 2026

# EIoT: Embodied Intelligence of Things

Yunhao Liu*, Xu Wang*, Yunhuai Liu, Kebin Liu, Shuai Tong, Jinliang Yuan, and Li Liu

**Abstract:** The integration of embodied intelligence into physical environments marks a new frontier in the evolution of intelligent systems. While the Internet of Things (IoT) connects devices and Artificial Intelligence of Things (AIoT) embeds intelligence into them, we argue that a further conceptual leap is required—one that enables the composition of intelligence itself through real-world embodiment, interaction, and evolution. We introduce the paradigm of Embodied Intelligence of Things (EIoT) as a foundational framework for distributed, physically grounded intelligent systems. EIoT systems are structured across three essential dimensions: *Enacted*, where devices are transformed into embodied agents; *Engaged*, where agents interact opportunistically based on physical and contextual constraints; and *Evolutionary*, where intelligence adapts and self-organizes through continuous experience. We further outline a developmental trajectory for EIoT based on the openness of perception and decision spaces, providing a conceptual map from tightly constrained agents to fully autonomous and adaptive systems. This work aims to establish EIoT as a core architectural and theoretical direction for future embodied intelligent systems.

**Key words:** embodied intelligence of things; compositional intelligence; multi-agent coordination; edge intelligence; physical–digital interaction

## 1 Introduction

The advent of deep learning and Large Language Models (LLMs) has propelled Artificial Intelligence (AI) technology into a new phase, demonstrating the potential to help human beings with scientific understanding and cognitive abilities. Mario Krenn et al.[1] mapped out three dimensions in which AI assists humans in scientific understanding: the Computational Microscope, the Resource of Inspiration and the Agent of Understanding. Numerous examples of the first two dimensions have already been observed, including the prediction and design of protein structures[2], the accurate forecasting of weather changes[3], and the control of nuclear fusion reactions using AI technology[4]. The third dimension, however, would require AI agents to be able to automatically acquire scientific understanding and transfer insights to humans. In order to achieve this, it is necessary for the AI agents to possess a physical form and the capacity to interact directly with the physical world. For example, an LLM that displays a text message saying "introduce reagents into a container" does not really complete the experiment. The same is true of an AI agent trying to analyze and draw conclusions without doing field observations. It is thus hypothesized that the capacity for direct interaction with the physical world is a pivotal element in the subsequent evolution of AI, which is also considered to be a prerequisite for the actualization of Artificial General Intelligence (AGI).

Indeed, as far back as 1950, Turing[5] had envisioned such a route when he proposed the equipping of AI agents with sense organs and the training of AIs as if

• Yunhao Liu, Xu Wang, Kebin Liu, Shuai Tong, Jinliang Yuan, and Li Liu are with Tsinghua University, Beijing 100084, China. E-mail: yunhao@tsinghua.edu.cn; xu_wang@tsinghua.edu.cn; kebinliu2021@tsinghua.edu.cn; tongshuai.ts@gmail.com; yuanjinliang@tsinghua.edu.cn; liuli95@tsinghua.edu.cn.
• Yunhuai Liu is with Peking University, Beijing 100871, China. E-mail: yunhuai.liu@pku.edu.cn.
* To whom correspondence should be addressed.
Manuscript received: 2025-07-15; accepted: 2025-08-11

they were human children. Many different types of embodied agents have been proposed, such as humanoid robots[6] and quadrupedal robots[7]. There have also been studies on Multi-agents systems, such as UAV swarms[8]. Much progress has been made in the above studies, but only some of the embodied intelligence scenarios have been covered. As we all know, most of current AI models are trained with data from the Internet, and with the development of IoT technology[9], the Internet has been largely extended from connecting human users to connecting ordinary objects in the physical world. As a result, we observe that a large number of ordinary objects in the physical world have been able to sense, establish connections, and even have certain computational capabilities. These objects are usually interconnected and constitute systems, even a system of systems. An example is an intelligent factory in which there are a large number of sensors, actuators, computing devices, etc. In this case, each object can be an embodied agent, while they are constantly exchanging information with other objects through different connections, interacting with the external environment and humans as a whole. We refer to such systems as Embodied Intelligence of Things (EIoT), which we believe will be an important research direction for Embodied Intelligence.

We define EIoT as a new class of intelligent systems that emerge from the convergence of AI, IoT, and embodiment. These systems consist of physically grounded agents that are capable of perception, decision-making, and interaction—both with the environment and with each other. The defining characteristics of EIoT can be understood through three foundational domains, namely Enacted, Engaged, and Evolutionary. First, devices are **enacted**, meaning they become autonomous agents through the integration of sensing, actuation, and real-world interaction. Intelligence is realized not through preprogrammed rules, but through embodied perception–action loops. Second, agents are **engaged**—they form peer-to-peer, opportunistic interactions based on spatial and temporal constraints. EIoT interactions are typically decentralized, event-driven, and context-dependent, contrasting sharply with the persistent connectivity model of traditional IoT. Third, EIoT systems are fundamentally **evolutionary**: they learn and adapt continuously through feedback, accumulating experience to refine behaviors, develop modular capabilities, and support scalable compositional intelligence.

Together, these three domains represent a shift from connecting things (IoT), to making them intelligent (AIoT), to enabling the composition of embodied intelligence (EIoT). While progress has been made across all three layers[9–12], EIoT remains an emerging and largely underexplored field, with numerous open challenges related to architecture, coordination, scalability, and evaluation.

The rest of this paper is structured as follows: Sections 2–4 trace the conceptual evolution from IoT to EIoT and elaborate the key enabling technologies across the three domains of Enacted, Engaged, and Evolutionary. Section 5 discusses challenges and frontier issues. We present our forward-looking perspective on the developmental trajectory of EIoT and outlines priorities for future research in Section 6.

## 2 Enacted Agent

Unlike conventional AI systems that operate within closed digital or simulated environments, EIoT introduces the concept of *agentization*—the transformation of physical devices or subsystems into intelligent, embodied agents. These agents are not passive data collectors or remote-controlled endpoints; rather, they actively sense, think, and act within their local environments, serving as the foundational units of distributed, embodied intelligence.

Inspired by Brooks's subsumption architecture[13], Enacted agents realize intelligence through embodied, reactive perception–action couplings, where the environment itself serves as the agent's reference frame. In this context, each agent functions as a digital–physical proxy for a physical entity. It manages sensor data, executes control logic, coordinates with other agents, and autonomously adapts to changes in its environment. EIoT agent architectures are similarly hierarchical, as shown in Fig. 1: low-level agents interface directly with individual devices, while mid- and high-level agents manage coordination across clusters or entire systems. This layered structure enables scalability, modularity, and context-aware behavior—mirroring principles of biological organization such as cellular autonomy[14], tissue-level coordination[15], and organism-wide coherence[16].

By replacing the notion of merely "instrumented" devices with that of "enacted" agents, EIoT pushes intelligence closer to the physical substrate. Each agent encapsulates localized intelligence, enabling responsive

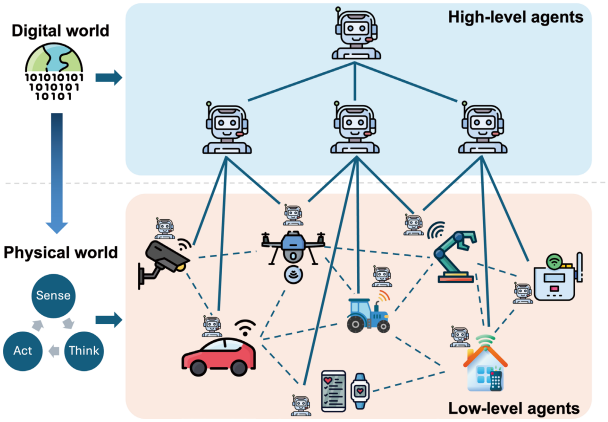



**Fig. 1 EIoT agentizes the physical world into hierarchical agent systems.**

behavior grounded in physical embodiment. Collectively, these agents form a dynamic, extensible, and situated system capable of interacting with the real world in real time.

However, building such hierarchical, agentized systems that faithfully represent and govern physical environments remains a core challenge in EIoT. Unlike traditional AI models that process abstract data, EIoT agents must operate under real-world constraints—dealing with uncertainty, noise, latency, and environmental variability. Three fundamental challenges define this problem space: *(i) Accurate Physical-to-Digital Mapping*: Agents must maintain a reliable and granular understanding of their physical counterparts through sensor fusion and environment modeling. *(ii) Understanding the Physical World*: Agents must interpret physical signals and contextual cues using models that can reason about dynamics, constraints, and interaction affordances. *(iii) Real-Time Closure of the Physical–Digital Loop*: Agents must sustain a tight perception–cognition–action loop, responding to physical changes with minimal latency and maximal fidelity.

The remainder of this section discusses the enabling technologies that support enacted agents in EIoT systems.

## 2.1 Physical-to-digital mapping

Accurate and semantically expressive mapping between the physical world and its digital counterpart forms the foundation for effective operation of EIoT agents. Achieving this requires embodied IoT devices to perceive their environments through rich, multimodal sensing, enabling them to act as autonomous, context-aware agents. Recent research has explored a broad spectrum of sensing modalities, Radio Frequency (RF)[17], acoustic[18], visual[19], tactile[20], and even smart material-based[21]. To push the limits of spatial perception, millimeter-wave radars and reconfigurable metasurfaces[22] have been leveraged to achieve near-LiDAR resolution RF imaging and non-line-of-sight object detection. For instance, a rotating mmWave array can produce high-fidelity 3D depth maps comparable to a 64-beam LiDAR[23], while metasurface beamformers can steer signals around corners to see hidden targets[24]. Such RF-based methods are inherently robust to poor lighting, dust, or smoke, giving IoT agents "X-ray" vision where optical sensors fail[23]. Complementing RF, acoustic sensing is being reinvented for IoT applications: ultra-low-power acoustic systems like SPiDR[25] use a single transducer to emit and capture spatially coded ultrasound, allowing micro-robots to navigate by sensing 3D obstacles at minimal energy cost. Innovative signal processing[26] can exploit hardware non-linearities to boost acoustic vibration sensing to sub-millimeter granularity, enabling detection of fine mechanical movements. These advances in better sensing are fundamentally elevating IoT devices from passive observers to context-aware, interactive participants in their environment.

## 2.2 Physical world understanding

Understanding the physical world remains a central challenge for EIoT agents tasked with interpreting complex, noisy, and dynamic sensory input. Recent researches have demonstrated that Large Language Models can be adapted to interpret IoT sensor data, forming a new class of applications[27]. AutoIoT[28] proposed a systematic framework that leverages LLMs to synthesize interpretable programs from natural language input, enabling agents to reason about ECG signals, IMU data, and mmWave radar streams. Babel[29] addressed multi-modal grounding by pre-training a scalable transformer model across sensor modalities (vision, audio, motion), enabling cross-modal alignment and transfer. LLMSense[30] extended this direction by enabling LLMs to reason over structured temporal data from wearable sensors. ChatIoT[31] tackled physical understanding from a system integration standpoint, proposing a security-aware LLM architecture that maintains contextual memory across edge deployments. IoT-LM[32]

introduced a foundation model pre-trained on 1.15 million sensor-language pairs across 12 modalities. With a multisensory multitask adapter layer, IoT-LM allows agents to fuse IMU, visual, audio, thermal, and spatial data into coherent high-order representations.

Together, these works point to three pillars of physical understanding in EIoT: (i) Grounded perception via sensor-aligned representations, (ii) Semantic abstraction through promptable reasoning and concept extraction, and (iii) Multi-agent cooperation enabled by distributed models with shared context and task-awareness.

### 2.3 Real-time closure of the physical-digital loop

Achieving real-time responsiveness across sensing, inference, and actuation is essential for embodied agents operating in dynamic physical environments. Central to this requirement is Edge Artificial Intelligence (Edge AI)[33], which enables local, low-latency decision-making by shifting computation from cloud servers to the edge of the network. Over the past five years, a large body of work has emerged exploring system, model, and infrastructure-level innovations to reduce end-to-end delay in physical–digital loops.

Techniques like model compression, pruning, and hardware acceleration now enable even deep neural networks to run on IoT-class microcontrollers[34]. For instance, Mistify[35] introduced an automated DNN porting system that takes a cloud-trained model and generates a suite of optimized smaller models for different edge devices, drastically reducing manual effort in deploying AI across heterogeneous hardware. Edge computing testbeds[36] have demonstrated orders-of-magnitude latency improvements for tasks like object detection using such co-location of computation. Moreover, edge AI frameworks increasingly leverage on-device TinyML capabilities: models like MCUNet[37] showed that even microcontrollers can run surprisingly advanced vision models by automating neural architecture search under tight memory constraints. These technologies enable faster, more scalable, and privacy-preserving intelligence in IoT systems, forming a cornerstone of modern EIoT deployments.

## 3 Engaged Interaction

Interaction is essential to the EIoT paradigm. While embodied agents possess the ability to sense and react upon the physical world independently, many real-world scenarios require them to work together. Whether coordinating to accomplish complex tasks or managing encounters in shared environments, agents must be able to communicate, align their actions, and adapt to one another in real time.

These interactions are governed by agents' mobility, resource limitations, and the dynamic nature of their surrounding environment. Consequently, communication is typically *opportunistic* and *intermittent*, occurring only when collaboration is necessary and physical proximity makes it feasible. Such behavior reflects the principles of Opportunistic IoT paradigms[38] and embodies EIoT's emphasis on context-aware, transient networking that minimizes connectivity overhead while maximizing task relevance, as illustrated in Fig. 2.

What sets EIoT apart from traditional IoT systems is not only the agents' enhanced autonomy, but also the shift away from continuous connectivity and centralized coordination. Unlike conventional IoT nodes that rely on persistent network links for data aggregation and control, EIoT agents interact in a more flexible, peer-to-peer manner, enabling local decision-making and collective adaptation.

EIoT's paradigm of intermittent, on-demand interaction drastically reduces the overhead of constant connectivity. Yet, it imposes specific demands on the underlying communication substrate that enables these engaged interactions:

**(1) Lightweight data exchange.** EIoT agents operate under strict energy constraints. Efficient, ultra-low-power methods are essential for exchanging necessary data during brief, opportunistic encounters with nearby agents.

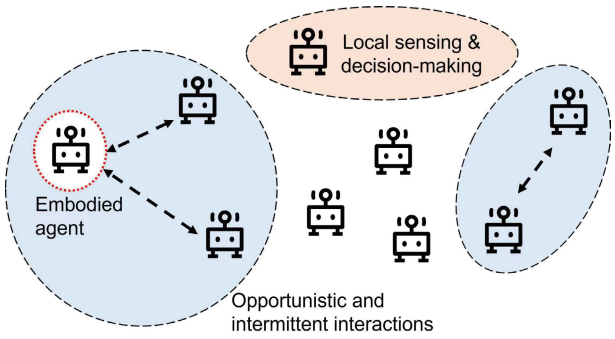



**Fig. 2 Illustration of engaged interactions: Agents interact conditionally with opportunistic and intermittent connections.**

**(2) Heterogeneous interconnection.** Agents for diverse applications may use different communication technologies. Solutions enabling direct information exchange across these heterogeneous protocols are critical for seamless collaboration.

**(3) Concurrent connections.** Multiple agents may interact simultaneously in dense deployments. Thus, robust techniques are needed to manage signal collisions, decode overlapping transmissions, and maintain reliable communication.

The rest of this section details key technologies designed to meet these communication challenges, thereby supporting the intermittent, low-power, and flexible interactions fundamental to realizing collective embodied intelligence in EIoT.

### 3.1 Backscatter communication

EIoT demands lightweight communication to support energy-constrained agents and opportunistic interactions. Backscatter communication meets this need by enabling devices to reflect existing signals instead of generating their own, drastically reducing power consumption. Its passive, short-range nature fits well with on-demand, proximity-based connections among EIoT agents.

Recent backscatter techniques leverage ambient or existing infrastructure signals for low-power and opportunistic communication. Ambient backscatter[39] utilizes pervasive TV and cellular signals as excitation sources for passive communication. FM backscatter[40] reflects FM radio signals to embed data, offering a simple and energy-efficient solution for low-complexity devices. Wi-Fi backscatter[41] reuses existing Wi-Fi infrastructure, by modulating RSSI to different levels for representing different data bits. Advancements like BackFi[42], which manipulates signal phase, and WiTag[43], which selectively corrupts subframes and observes ACKs, further improve encoding efficiency by using ambient signals in the environment as excitation sources.

Ambient backscatter typically supports only short ranges (a few meters), which limits its applicability in complex EIoT environments. To address this, researchers have enhanced both range and scalability by designing new backscatter modulation schemes. Hitchhike[44] increases communication distance by piggybacking data on existing packets through codeword translation. LoRea[45] pushes the range to kilometers using narrowband modulation. PPLoRa[46] concentrates the energy from the whole spectrum for long-range weak signal demodulation. For dense deployments, OFDMA backscatter[47] and Digiscatter[48] introduce OFDMA techniques to support concurrent transmissions from multiple EIoT agents, enhancing scalability of passive networks.

Backscatter is becoming a key EIoT communication technology, meeting low-power needs while ensuring range and scalability.

### 3.2 Cross technology communication

EIoT comprises diverse agents using different wireless protocols. They require flexible and low-overhead communication without relying on centralized infrastructure. Cross Technology Communication (CTC) addresses this need by enabling direct interaction across heterogeneous protocols, eliminating the need for external gateways. Existing CTC methods mainly fall into two categories:

**(1) Signal emulation.** This approach modifies signal modulation, allowing a signal from one technology to mimic the format of another. A representative example is LTE2B[49], which enables ZigBee devices to communicate with LTE via physical-layer signal emulation. Similarly, XFi[50] decodes LoRa packets using Wi-Fi receivers by recovering chirp signals from CSI traces, further expanding the interoperability of EIoT devices.

**(2) Feature-based encoding.** This approach embeds data by creating unique, distinctive features in the transmitted signal, such as signal energy, packet length, or the time between packets. Even if heterogeneous receivers cannot decode the raw data packet, they can interpret the message by recognizing these distinctive signal features. This approach is particularly valuable for resource-constrained EIoT devices which may lack the processing power for full protocol stack emulation. Recent works primarily focused on packet-level information[51–52] to convey data, leveraging characteristics like signal strength[53–54], transmission time[55–56], and packet length[57].

These CTC approaches align well with opportunistic and low-bandwidth EIoT communication, enabling efficient data exchange across diverse agents with minimal overhead.

### 3.3 Collision-resilient communication

With the increasing density of EIoT deployment, concurrent transmissions among agents become

common. Collision-resilient communication enables reliable connectivity by extracting individual packets from overlapped signals at the physical layer, using features in time, frequency, or energy domains. This allows multiple agents to transmit simultaneously without loss of critical data. Current research for collision-resilient communication primarily falls into two categories based on their approach to handling physical layer interference:

**(1) Hardware-assisted communication.** Hardware assisted solutions leverage specific physical layer designs or hardware capabilities to mitigate or resolve collisions. MIMO/MU-MIMO[58–59] techniques utilize multiple antennas on wireless devices to synchronize signal phases from different transmitters, enabling concurrent transmissions without inter-packet interference. Successive Interference Cancellation (SIC)[60–61] aims to decode signals even when collisions occur. SIC iteratively estimates and extracts decoded symbols, effectively subtracting decoded strong signals to reveal weaker ones underneath. Capture effect approaches leverage signal strength differences to decode the strongest signal in a collision[62], treating others as interference. Hardware assisted approaches incur high burden on EIoT agents, including multiple RF chains and complex processing, generally unsuitable for low-power devices.

**(2) Software based communication.** Software solutions offer greater adaptability to low-power EIoT agents through advanced signal processing and protocol design. ZigZag[63] resolves Wi-Fi collisions by leveraging timing offsets to extract non-overlapping segments. mZig[64] applies similar techniques to separate ZigBee packets from a single collision. Netscatter[30] modifies LoRa encoding to support backscatter devices, enabling hundreds of concurrent transmissions via frequency discrimination. DeepSense[65] uses neural networks to identify and decode collided frames, supporting random access and cross-protocol coexistence.

Collision-resilient communication techniques enhance the communication reliability in dense deployments, paving the way for large-scale and low-latency EIoT agents collaboration.

## 4 Evolutionary Intelligence

In EIoT, the challenge lies in endowing embodied systems with sustained, adaptive, and collaborative intelligence—capabilities that go beyond the task-specific inference common in traditional IoT. Intelligence that is pre-trained once and fixed at deployment is inadequate for agents operating in dynamic, uncertain physical environments. These agents must be able to learn continually, adapt to environmental drift, and reorganize their collective behavior in response to new demands.

We frame this as an *evolutionary* shift in EIoT, as depicted in Fig. 3. First, we trace the architectural transition from lightweight Deep Neural Networks (DNNs) toward Transformer-based[66] and foundation-model[67] approaches, enabling embodied agents to integrate multimodal perception, temporal reasoning, and contextual awareness—an essential step noted in embodied intelligence literature[11]. Second, we examine how agents achieve self-evolution through continuous perception–action loops, updating models, representations, and control policies in situ, as emphasized in continual and lifelong learning frameworks[68]. Third, we investigate the emergence of collective intelligence from inter-agent coordination, where decentralized decision-making, shared perception, and adaptive role allocation allow the group to accomplish tasks beyond the capacity of any single agent—aligning with swarm intelligence studies[69].

These three threads—multimodal reasoning, continual adaptation, and emergent collectivity—demonstrate that EIoT intelligence is no longer defined by computational scale alone, but by its capacity to evolve: to refine its capabilities through embodied experience, adapt to shifting contexts, and recompose itself in collaboration with others.

### 4.1 Smarter model

Within the developmental framework of EIoT, IoT devices are undergoing a fundamental paradigm shift in intelligent modeling. The core evolutionary trajectory manifests as a historic transition: from traditional lightweight DNNs in resource-constrained scenarios to Transformer-based LLMs. This transformation is propelled by two synergistic forces:

(1) Escalating Environmental Perception Demands. Embodied IoT requires devices to fuse diverse sensor data (e.g., sound, light) with high precision, shifting from simple event detection to complex environmental understanding for decision-making. M4[70] proposes a unified mobile intelligence framework using a foundation model and lightweight adapters to handle

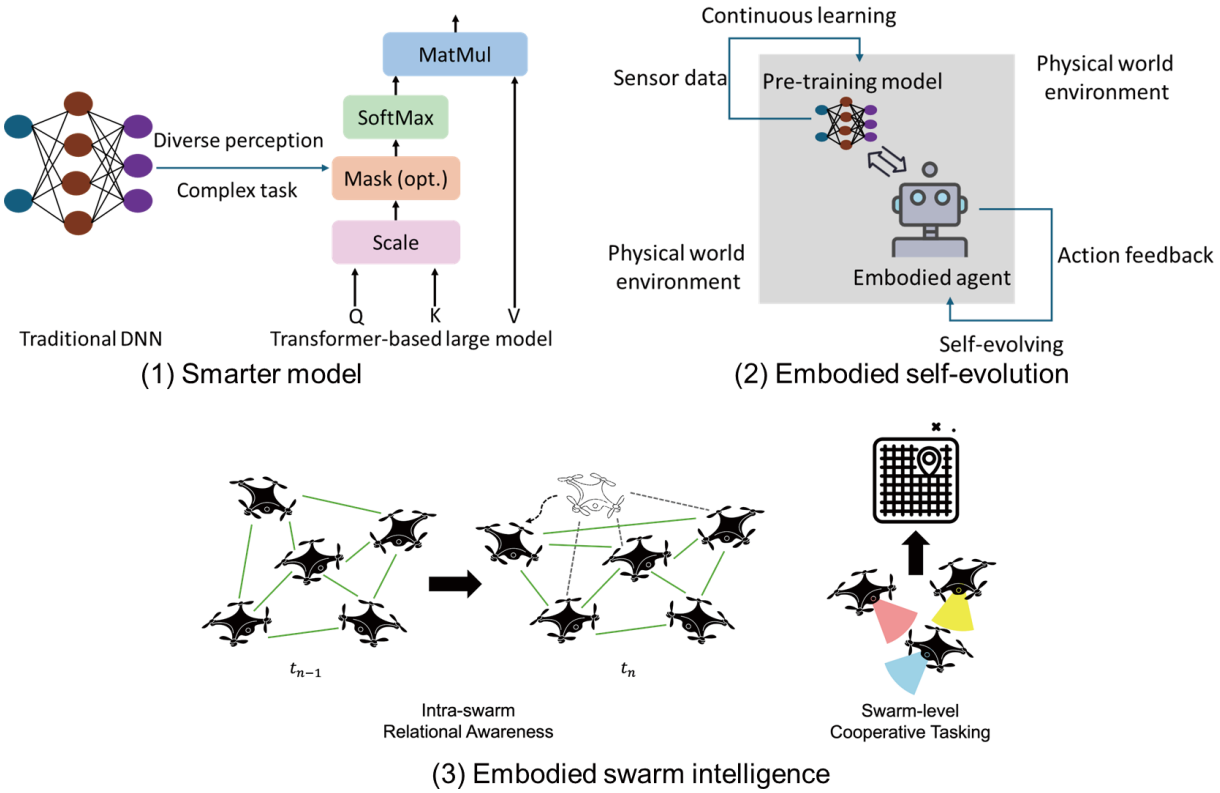



**Fig. 3 Illustration of evolutionary intelligence.**

multimodal inputs, reducing system fragmentation efficiently. Penetrative AI[28] extends LLMs to interpret sensor signals, demonstrating semantic analysis of real-world data for context-aware actions and cyber-physical intelligence.

(2) Increasing Task Interaction Complexity. Establishing a unified language-driven interface across diverse sensory, control, and computational domains by embedding LLMs as semantic planners and communicators. Through structured task decomposition, intermediate reasoning formats (e.g., FSMs, UI graphs), and knowledge augmentation, they enable embodied agents to interpret, coordinate, and generalize complex real-world tasks under dynamic, user-centric scenarios[71–73].

### 4.2 Embodied self-evolution

Although Transformer architectures empower edge devices with cognitive-level intelligence, their capabilities remain constrained by non-stationary properties of the physical world (e.g., environmental parameter drift, long-tail user behavior distributions) and scenario fragmentation (e.g., defect pattern variations in industrial inspection). We posit that the ultimate value of large models in AIoT terminals lies not in static superiority but in establishing self-evolution capability embedded within physical feedback loops. This evolution manifests through three core mechanisms:

(1) Environment-driven continuous learning. Some studies adopt online meta-learning for continuous learning. LifeLearner[74] leverages real-time sensor data streams to dynamically reparameterize computational graphs (e.g., via MAML algorithms), enabling synchronous weight updates during inference (field tests in industrial robotics: 7% average latency increase vs. 83% reduction in accuracy decay). Federated continual learning offers decentralized adaptation mechanism for continuous learning. FedINC[75] and Cross-FCL[76] trigger federated continual learning based on confidence thresholds for out-of-distribution (OOD) samples encountered post-deployment, enabling cross-node knowledge transfer of critical cases.

(2) Agent-environment co-evolution architecture. Transformer-driven embodied agents[77–79] integrate pre-trained knowledge with real-time sensory data to dynamically adapt to changing environments (e.g., lighting variations, moving objects), achieving an intelligent leap from rule-based responses to

autonomous decision-making. These systems continuously adjust decision weights to balance prior knowledge with real-time observations, ultimately enabling more natural and safer environmental interactions. Intelligent agents achieve self-evolution[80–82] by continuously interacting with their environment through language, vision, or physical actions. Using unsupervised or reinforcement learning mechanisms, they autonomously acquire new knowledge and dynamically optimize their behavioral strategies. This iterative learning process enables agents to progressively enhance their environmental adaptability—transitioning from fixed response patterns to context-aware decision-making, and ultimately developing generalized problem-solving capabilities that transcend initial programming constraints.

(3) Evolutionary efficacy optimization. To realize sustained intelligence in EIoT systems, the escalating demands introduced by large-scale models and continual learning must be addressed, particularly under the tight constraints of IoT devices in terms of memory, processing power, energy consumption, and network bandwidth. Miro[83] dynamically profiles energy–accuracy tradeoffs to adapt continual learning strategies on edge device. E-DomainIL[84] handles domain shifts in IoT tasks without task labels or memory overhead by reusing frozen parameters within fixed architectures, maintaining parameter efficiency under limited memory. Q-FedUpdate[85] enables energy-efficient federated learning using low-precision DSP-based training, preserving full-precision model quality via accumulated gradient tracking and quantization–training pipelining, reducing on-device energy by 21×. ELITE[86] integrates edge–cloud collaboration for online continual learning, using a model zoo with on-device selection and latency-aware fine-tuning, achieving a 16.3% accuracy gain and 1.98× latency reduction on dynamic task streams.

### 4.3 Embodied swarm intelligence emergence

In physically grounded environments, embodied agents rarely operate in isolation. From biological systems to human societies, intelligence often emerges not from individual capability alone, but through collective organization and inter-agent interaction. Reflecting this principle, swarm intelligence, emerging from the physical and semantic interactions among multiple agents, marks a natural progression from single-agent autonomy to collaborative intelligence. Beyond the capabilities of isolated agents, these distributed systems exhibit novel behaviors that arise from collaboration, including shared perception, decentralized decision-making, and coordinated actuation within a common physical environment. Its emergence is primarily driven by two interelated dimensions:

(1) Intra-swarm relational awareness. Agents in a swarm must perceive, interpret, and respond to the spatial configurations and interactive behaviors of other agents within the swarm. This dimension of intelligence is critical for enabling local coordination strategies such as formation control, collision avoidance, and dynamic topology adaptation. By maintaining an accurate understanding of relative positions, velocities, intentions, and communication states, agents are able to act coherently as a group without relying on centralized control. EvolveGraph[87] constructs latent interaction graphs to represent evolving spatial and behavioral relationships among heterogeneous agents. HEAT[88] models heterogeneous agents' dynamics, interactions, and map context in connected autonomous driving by an edge-aware attention graph network. Tested on urban traffic, it achieves 0.66 m displacement error, supporting safe planning under complex scenarios. MTR/MTR++[89] proposes a transformer-based framework that performs global intent localization and local trajectory refinement to enable accurate and multimodal future prediction. MTR++ extends this to simultaneous multi-agent forecasting, leveraging mutual intention guidance and symmetric context modeling. ORI[90] adopts a trainable adjacency matrix with a novel AdaRelation optimizer to dynamically learn interaction graphs for relational inference in evolving multi-agent systems. It enables real-time adaptation to changing inter-agent dynamics and generalizes across trajectory patterns, outperforming offline methods in dynamic settings. TransformLoc[91] enables accurate and low-cost localization in heterogeneous drone swarms by turning advanced MAVs (AMAVs) into mobile localization infrastructures for lightweight, resource-constrained MAVs (BMAVs), allowing real-time collaborative positioning. Deployed on industrial drones, it improves localization accuracy by up to 68% and boosts navigation success rates by 60% compared to prior methods.

(2) Swarm-level cooperative tasking. While intra-swarm awareness enables agents to understand each

other's states, effective operation in real-world environments also demands coordinated action toward shared objectives. Agents in a swarm must collaborate across space and time to accomplish tasks such as area exploration, target tracking, cooperative transport, and distributed decision-making under partial observability. SwarmMap[92] enables efficient multi-agent visual SLAM for tasks like drone coordination and autonomous exploration by using log-based delta synchronization, where only incremental updates are shared between agents to reduce redundant transmission. Kouzeghar et al.[93] propose to achieve multi-agent UAV pursuit-evasion in unknown, dynamic environments with evasive targets. It extends MADDPG with a role-based reinforcement learning framework guided by a Voronoi-based reward, where UAVs assume heterogeneous roles for balanced exploration and tracking. CoELA[94] is a modular framework that integrates LLMs (e.g., GPT-4) into multi-agent systems for decentralized cooperation in embodied tasks with raw perception and costly communication, so that enables agents to coordinate long-horizon goals more effectively. MAEN[95] uses multi-agent reinforcement learning to coordinate UAVs as mobile access points for emergency communication in disaster scenarios. It introduces grouping and information sharing to scale collaboration to dozens of UAVs, and employs reward decomposition for effective task allocation. REMAC[96] proposes a self-reflective and self-evolving framework for multi-agent collaboration in long-horizon robotic manipulation. It combines offline skill learning with online reflective adaptation, allowing agents to improve coordination over time without explicit supervision.

As demonstrated across recent works, the collective capabilities of a swarm are not merely an aggregation of individual functions, but an emergent property that reflects contextual awareness, relational reasoning, and task-level cooperation. It is a critical direction for realizing embodied, scalable, and self-organizing intelligence across physical environments, thus is becoming increasingly important to the evolution of EIoT systems.

# 5 Challenges and Frontier Issue

While EIoT presents a transformative vision for distributed, agent-based intelligence rooted in physical environments, its realization is accompanied by a range of technical and theoretical challenges. These challenges span multiple layers of system design—from perception and communication to coordination, learning, and system-level evolution. In this section, we identify several critical bottlenecks and frontier issues that must be addressed to advance EIoT from a conceptual framework to a robust, scalable paradigm.

## 5.1 Compositionality of agents and intelligence

A central challenge in EIoT lies in the compositionality of agents—how multiple embodied agents, each with partial capabilities and local views, can dynamically coordinate to accomplish complex tasks. Unlike monolithic AI systems, EIoT emphasizes distributed autonomy and modular interaction. However, existing approaches to agent design often assume fixed roles, centralized orchestration, or predefined protocols. Achieving compositional intelligence requires agents to flexibly discover, negotiate, and bind their functions in real time, forming transient coalitions and collective behaviors. This composability must be supported across both behavioral and architectural dimensions, including shared semantics, role abstraction, task delegation, and multi-agent planning. Moreover, the composition process itself must be situated, meaning that it must respect embodiment constraints—such as physical proximity, actuation range, and sensing limitations. Developing lightweight, decentralized mechanisms for emergent composition—particularly under uncertainty and intermittent connectivity—remains a key open question.

## 5.2 Context-aware connectivity

Embodied agents do not communicate continuously, but instead interact based on physical context, task relevance, and environmental conditions. For example, an agent may prioritize connecting with a nearby peer carrying complementary sensor data, or defer communication when engaged in critical motor control tasks. This creates a need for connectivity decisions that are not only driven by link quality, but also informed by situational awareness—such as location, object proximity, task urgency, or shared mission goals. Designing such context-aware connectivity mechanisms introduces significant complexity: agents must integrate multi-modal sensory data into communication logic, assess the value of each potential connection, and adapt their interaction frequency accordingly. Unlike traditional IoT where communication is periodic and infrastructure-driven,

EIoT demands adaptive, perception-driven communication strategies that can filter, defer, or accelerate interactions in response to changing physical and task-specific contexts.

### 5.3 Intermittent links

Embodied agents often move through dynamic physical environments, resulting in inherently intermittent and unpredictable short-lived links. At the same time, these agents typically operate on strict energy budgets, which constrain their ability to maintain active radios or engage in continuous neighbor discovery. This dual challenge of transient connectivity and tight energy constraints makes it difficult to ensure timely and reliable information exchange. Conventional networking techniques—such as persistent beacons, routing table maintenance, or active scanning—are too resource-intensive for such scenarios. Therefore, EIoT requires new lightweight and predictive coordination strategies that can opportunistically initiate communication during brief contacts, while minimizing idle listening and unnecessary transmissions. Solutions must also account for variable encounter durations, uncertain mobility patterns, and the need for fast data exchange before disconnection occurs.

### 5.4 Guaranteeing efficient online learning under physical constraints

A fundamental challenge in EIoT is enabling embodied agents to perform online learning during continuous interaction with dynamic physical environments, while strictly adhering to inherent constraints. Unlike controlled simulations, real-world interactions involve irreversible consequences (e.g., collision, system damage) and operate under severe resource limitations (compute, power, latency). Agents must learn and adapt in situ from sparse, noisy data streams generated during task execution. This demands sample-efficient algorithms (e.g., meta-learning, Bayesian optimization) capable of rapid adaptation to unforeseen conditions with minimal trials. Crucially, the learning process itself must incorporate intrinsic safety mechanisms. Techniques include constrained reinforcement learning, formal verification of learned policies before deployment, uncertainty-aware exploration to avoid high-risk states, and real-time monitoring to override unsafe actions. The key is dynamically balancing exploration (seeking new knowledge for better long-term performance) and exploitation (relying on known safe policies), ensuring task completion is never compromised by the learning process. Achieving constrained, real-time interaction-based learning is imperative for trustworthy and resilient evolution of embodied intelligence in the physical world.

### 5.5 Security, safety, and embodied ethics

As EIoT systems begin to act autonomously in physical environments—especially in safety-critical domains such as healthcare, manufacturing, or urban infrastructure—they raise urgent concerns related to security, safety, and ethical behavior. Threat vectors in EIoT are more complex than in traditional networks, as adversaries may exploit both cyber and physical vulnerabilities. Furthermore, embodied agents can unintentionally cause harm through faulty perception, poor generalization, or emergent behavior that violates human norms. Ensuring safe operation requires new paradigms for verifiability, explainability, and fail-safe mechanisms tailored to embodied agents. Additionally, ethical frameworks must account for distributed agency and collective accountability, especially when decision-making is decentralized and emergent.

## 6 Toward Embodied Intelligence of Things

The emerging paradigm of EIoT represents a fundamental shift in how we conceive and construct intelligent systems—moving beyond centralized inference and disembodied computation toward a future where intelligence is physically grounded, dynamically composed, and environmentally situated. Building upon the foundations of IoT and AIoT, EIoT introduces a new ambition: not just to connect or augment things with intelligence, but to enable the composition of intelligence itself—through embodied agents that sense, act, collaborate, and evolve within the physical world.

In this work, we proposed a triadic framework—Enacted, Engaged, and Evolutionary—to characterize the core dimensions of EIoT systems. Enacted agents transform passive devices into locally autonomous entities capable of real-time perception–action loops. Engaged agents interact through opportunistic, peer-to-peer coordination shaped by physical context and task demands. Evolutionary agents adapt and self-organize over time, allowing distributed intelligence to emerge from experience, feedback, and composition.

Looking ahead, we envision the development of

EIoT systems as a progression across the openness of perception and decision spaces. Initial deployments may involve tightly scoped environments with fixed sensing and actuation (e.g., data agents in industrial monitoring). Subsequent stages introduce greater autonomy—either in decision-making or perception—such as smart home managers or autonomous vehicles. The long-term goal is open–open systems that continuously discover, learn, and act within complex, unstructured, and evolving environments. These agents will not only perform tasks but also define them—collaborating with humans and with each other to form new, compositional forms of intelligence.

Achieving this vision will require significant advances in agent composition, embodied learning, decentralized coordination, and system-level evaluation. It also calls for a deeper integration of insights from AI, robotics, IoT, systems engineering, and cognitive science. But the promise of EIoT is clear: a new class of intelligent systems that are not merely embedded in the world—but arise from it.

## References


[1] M. Krenn, R. Pollice, S. Y. Guo, M. Aldeghi, A. Cervera-Lierta, P. Friederich, G. dos Passos Gomes, F. Häse, A. Jinich, A. Nigam, et al., On scientific understanding with artificial intelligence, *Nature Reviews Physics*, vol. 4, no. 12, pp. 761–769, 2022.

[2] J. Abramson, J. Adler, J. Dunger, R. Evans, T. Green, A. Pritzel, O. Ronneberger, L. Willmore, A. J. Ballard, J. Bambrick, et al., Accurate structure prediction of biomolecular interactions with alphafold 3, *Nature*, vol. 630, no. 8016, pp. 493–500, 2024.

[3] R. Lam, A. Sanchez-Gonzalez, M. Willson, P. Wirnsberger, M. Fortunato, F. Alet, S. Ravuri, T. Ewalds, Z. Eaton-Rosen, W. Hu, et al., Learning skillful medium-range global weather forecasting, *Science*, vol. 382, no. 6677, pp. 1416–1421, 2023.

[4] J. Seo, S. Kim, A. Jalalvand, R. Conlin, A. Rothstein, J. Abbate, K. Erickson, J. Wai, R. Shousha, and E. Kolemen, Avoiding fusion plasma tearing instability with deep reinforcement learning, *Nature*, vol. 626, no. 8000, pp. 746–751, 2024.

[5] A. Turing, Computing machinery and intelligence, *Mind*, vol. 59, no. 236, pp. 433–460, 1950.

[6] S. Saeedvand, M. Jafari, H. S. Aghdasi, and J. Baltes, A comprehensive survey on humanoid robot development, *The Knowledge Engineering Review*, vol. 34, p. e20, 2019.

[7] P. Biswal and P. K. Mohanty, Development of quadruped walking robots: A review, *Ain Shams Engineering Journal*, vol. 12, no. 2, pp. 2017–2031, 2021.

[8] S. Javed, A. Hassan, R. Ahmad, W. Ahmed, R. Ahmed, A. Saadat, and M. Guizani, Stateof-the-art and future research challenges in uav swarms, *IEEE Internet of Things Journal*, vol. 11, no. 11, pp. 19023–19045, 2024.

[9] S. H. Shah and I. Yaqoob, A survey: Internet of things (iot) technologies, applications and challenges, *2016 IEEE Smart Energy Grid Engineering* (*SEGE*), pp. 381–385, 2016.

[10] S. I. Siam, H. Ahn, L. Liu, S. Alam, H. Shen, Z. Cao, N. Shroff, B. Krishnamachari, M. Srivastava, and M. Zhang, Artificial intelligence of things: A survey, *ACM Transactions on Sensor Networks*, vol. 21, no. 1, pp. 1–75, 2025.

[11] H. Liu, D. Guo, and A. Cangelosi, Embodied intelligence: A synergy of morphology, action, perception and learning, *ACM Computing Surveys*, vol. 57, no. 7, pp. 1–36, 2025.

[12] Y. Liu, L. Liu, Y. Zheng, Y. Liu, F. Dang, N. Li, and K. Ma, Embodied navigation, *Science China Information Sciences*, vol. 68, no. 4, pp. 1–39, 2025.

[13] R. Brooks, A robust layered control system for a mobile robot, *IEEE journal on robotics and automation*, vol. 2, no. 1, pp. 14–23, 2003.

[14] J. Russo, Basis of cellular autonomy in the subsceptibility to carcinogenesis, *Toxicologic Pathology*, vol. 11, no. 2, pp. 149–166, 1983.

[15] L. E. O'Brien and D. Bilder, Beyond the niche: tissue-level coordination of stem cell dynamics, *Annual review of cell and developmental biology*, vol. 29, no. 1, pp. 107–136, 2013.

[16] A. Przysinda, W. Feng, and G. Li, Diversity of organism-wide and organ-specific endothelial cells, *Current cardiology reports*, vol. 22, no. 4, p. 19, 2020.

[17] X. Wang, X. Wang, and S. Mao, Rf sensing in the internet of things: A general deep learning framework, *IEEE Communications Magazine*, vol. 56, no. 9, pp. 62–67, 2018.

[18] G. Sessler, Acoustic sensors, *Sensors and Actuators A: Physical*, vol. 26, no. 1-3, pp. 323–330, 1991.

[19] D. H. Ballard and C. M. Brown, *Computer vision*. Prentice Hall Professional Technical Reference, 1982.

[20] B. D. Argall and A. G. Billard, A survey of tactile human-robot interactions, *Robotics and autonomous systems*, vol. 58, no. 10, pp. 1159–1176, 2010.

[21] B. Sunil, M. A. Hasan, A. Jain, N. Chahuan, et al., Smart materials for sensing and actuation: state-of-the-art and prospects, in *E3S Web of Conferences*, vol. 505, p. 01034, EDP Sciences, 2024.

[22] Y. Liu, X. Liu, X. Mu, T. Hou, J. Xu, M. Di Renzo, and N. Al-Dhahir, Reconfigurable intelligent surfaces: Principles and opportunities, *IEEE Communications Surveys & Tutorials*, vol. 23, no. 3, pp. 1546–1577, 2021.

[23] H. Lai, G. Luo, Y. Liu, and M. Zhao, Enabling visual recognition at radio frequency, in *Proceedings of the 30th Annual International Conference on Mobile Computing and Networking*, pp. 388–403, 2024.

[24] C. Feng, X. Li, Y. Zhang, X. Wang, L. Chang, F. Wang, X. Zhang, and X. Chen, Rflens: metasurface-enabled

beamforming for iot communication and sensing, in *Proceedings of the 27th Annual International Conference on Mobile Computing and Networking*, pp. 587–600, 2021.

[25] Y. Bai, N. Garg, and N. Roy, Spidr: Ultra-lowpower acoustic spatial sensing for micro-robot navigation, in *Proceedings of the 20th Annual International Conference on Mobile Systems, Applications and Services*, pp. 99–113, 2022.

[26] X. Chen, D. Li, Y. Chen, and J. Xiong, Boosting the sensing granularity of acoustic signals by exploiting hardware non-linearity, in *Proceedings of the 21st ACM Workshop on Hot Topics in Networks*, pp. 53–59, 2022.

[27] H. Xu, L. Han, Q. Yang, M. Li, and M. Srivastava, Penetrative AI: making llms comprehend the physical world, in *Findings of the Association for Computational Linguistics, ACL 2024*, Bangkok, Thailand and virtual meeting, 2024, pp. 7324–7341, 2024.

[28] L. Shen, Q. Yang, Y. Zheng, and M. Li, Autoiot: Llm-driven automated natural language programming for aiot applications, arXiv preprint arXiv: 2503.05346, 2025.

[29] S. Dai, S. Jiang, Y. Yang, T. Cao, M. Li, S. Banerjee, and L. Qiu, Babel: A scalable pre-trained model for multi-modal sensing via expandable modality alignment, in *Proceedings of the 23rd ACM Conference on Embedded Networked Sensor Systems*, pp. 240–253, 2025.

[30] X. Ouyang and M. Srivastava, Llmsense: Harnessing llms for high-level reasoning over spatiotemporal sensor traces, arXiv preprint arXiv: 2403.19857, 2024.

[31] Y. Gao, K. Xiao, F. Li, W. Xu, J. Huang, and W. Dong, Chatiot: Zero-code generation of trigger-action based iot programs, *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 8, no. 3, pp. 1–29, 2024.

[32] S. Mo, R. Salakhutdinov, L.-P. Morency, and P. P. Liang, Iot-lm: Large multisensory language models for the internet of things, arXiv preprint arXiv: 2407.09801, 2024.

[33] R. Singh and S. S. Gill, Edge ai: A survey, *Internet of Things and Cyber-Physical Systems*, vol. 3, pp. 71–92, 2023.

[34] L. Deng, G. Li, S. Han, L. Shi, and Y. Xie, Model compression and hardware acceleration for neural networks: A comprehensive survey, *Proceedings of the IEEE*, vol. 108, no. 4, pp. 485–532, 2020.

[35] P. Guo, B. Hu, and W. Hu, Mistify: Automating {DNN} model porting for {On-Device} inference at the edge, in *18th USENIX Symposium on Networked Systems Design and Implementation* (*NSDI 21*), pp. 705–719, 2021.

[36] N. Abbas, Y. Zhang, A. Taherkordi, and T. Skeie, Mobile edge computing: A survey, *IEEE Internet of Things Journal*, vol. 5, no. 1, pp. 450–465, 2017.

[37] J. Lin, W.-M. Chen, Y. Lin, C. Gan, S. Han, et al, Mcunet: Tiny deep learning on iot devices, *Advances in Neural Information Processing Systems*, vol. 33, pp. 11711–11722, 2020.

[38] B. Guo, D. Zhang, Z. Wang, Z. Yu, and X. Zhou, Opportunistic iot: Exploring the harmonious interaction between human and the internet of things, *Journal of Network and Computer Applications*, vol. 36, no. 6, pp. 1531–1539, 2013.

[39] V. Liu, A. Parks, V. Talla, S. Gollakota, D. Wetherall, and J. R. Smith, Ambient backscatter: Wireless communication out of thin air, in *Proceedings of the ACM SIGCOMM Conference* (*SIGCOMM*), Hong Kong, China, pp. 39–50, 2013.

[40] B. J. Parks, K. R. Nigam, Y. Zhao, M. Reynolds, and S. Gollakota, Fm backscatter: Enabling battery-free mobile sensors, in *Proceedings of the 13th International Conference on Mobile Systems, Applications and Services* (*MobiSys*), Florence, Italy, pp. 243–256, 2015.

[41] B. J. Parks, S. Gollakota, J. R. Smith, S. Chandrasekaran, N. Jain, and E. Wu, Wi-fi backscatter: Passive communication with commodity 802.11 devices, in *Proceedings of the 12th USENIX Symposium on Networked Systems Design and Implementation* (*NSDI*), Santa Clara, CA, USA, pp. 75–88, USENIX Association, 2016.

[42] J. R. Smith, B. Kellogg, A. Parks, S. Gollakota, D. Wetherall, and V. Talla, Backfi: A backscatterbased communication system using ambient wi-fi signals, in *Proceedings of the 12th USENIX Symposium on Networked Systems Design and Implementation* (*NSDI*), Santa Clara, CA, USA, USENIX Association, 2015.

[43] V. Liu, A. Parks, V. Talla, S. Gollakota, and J. R. Smith, Witag: A tag for your wi-fi devices, in *Proceedings of the 23rd Annual International Conference on Mobile Computing and Networking* (*MobiCom*), Snowbird, UT, USA, pp. 386–398, 2017.

[44] D. J. Yang, V. Talla, S. Gollakota, J. R. Smith, and D. Wetherall, Hitchhike: Piggybacking data on 802.11b, in *Proceedings of the 20th Annual International Conference on Mobile Computing and Networking* (*MobiCom*), Chicago, IL, USA, pp. 423–434, 2014.

[45] A. Varshney, O. Harms, C. Perez-Penichet, x C. Rohner, F. Hermans, and T. Voigt, Lorea: A backscatter architecture that achieves a long communication range, in *Proceedings of the ACM Conference on Embedded Networked Sensor Systems* (*SenSys*), Delft, Netherlands, pp. 145–158, 2017.

[46] J. Jiang, Z. Xu, F. Dang, and J. Wang, Long-range ambient lora backscatter with parallel decoding, in *Proceedings of the 27th Annual International Conference on Mobile Computing and Networking* (*MobiCom*), New Orleans, LA, USA, pp. 1–13, 2021.

[47] E. T. Kang, M. Pattankar, B. Kellogg, and S. Gollakota, Ofdma backscatter: Enabling scalable passive communication in wi-fi networks, in *Proceedings of the 16th USENIX Symposium on Networked Systems Design and Implementation* (*NSDI*), Boston, MA, USA, pp. 379–394, USENIX Association, 2019.

[48] F. Zhu, Y. Feng, Q. Li, X. Tian, and X. Wang, Digiscatter:

Efficiently prototyping large-scale ofdma backscatter networks, in *Proceedings of the 18th International Conference on Mobile Systems, Applications, and Services* (*MobiSys*), Toronto, Canada, pp. 42–53, 2020.

[49] R. Liu, Z. Yin, W. Jiang, and T. He, Lte2b: time-domain cross-technology emulation under lte constraints, SenSys' 19, New York, NY, USA, pp. 179–191, Association for Computing Machinery, 2019.

[50] R. Liu, Z. Yin, W. Jiang, et al., XFi: Crosstechnology iot data collection via commodity wifi, in *Proceedings of the 28th International Conference on Network Protocols* (*ICNP*), Madrid, Spain, pp. 1–11, 2020.

[51] Z. Li, Z. Yin, L. Liu, et al., Demo: WEBee: Physical-layer cross-technology communication via emulation, in *Proceedings of the 23rd Annual International Conference on Mobile Computing and Networking* (*MobiCom*), New York, NY, USA, pp. 2–14, 2017.

[52] D. Xia, X. Zheng, F. Yu, et al., WiRa: Enabling cross-technology communication from wifi to lora with ieee 802.11ax, in *IEEE INFOCOM 2022 - IEEE Conference on Computer Communications*, Piscataway, NJ, USA, pp. 430–439, 2022.

[53] X. Guo, Y. He, X. Zheng, et al., Zigfi: Harnessing channel state information for cross-technology communication, *IEEE/ACM Transactions on Networking*, vol. 28, no. 1, pp. 301–311, 2020.

[54] S. Tong, Y. He, Y. Liu, et al., De-spreading over the air: Long-range ctc for diverse receivers with lora, in *Proceedings of the 28th Annual International Conference on Mobile Computing and Networking* (*MobiCom*), New York, NY, USA, pp. 42–54, 2022.

[55] S. Kim and T. He, Freebee: Cross-technology communication via free side-channel, in *Proceedings of the 21st Annual International Conference on Mobile Computing and Networking* (*MobiCom*), New York, NY, USA, pp. 317–330, 2015.

[56] Z. Yin, W. Jiang, S. Kim, et al., C-morse: Crosstechnology communication with transparent morse coding, in *IEEE INFOCOM 2017 - IEEE Conference on Computer Communications*, Piscataway, NJ, USA, pp. 1–9, 2017.

[57] X. Zheng, Y. He, and X. Guo, Stripcomm: Interference-resilient cross-technology communication in coexisting environments, in *IEEE INFOCOM 2018 - IEEE Conference on Computer Communications*, Piscataway, NJ, USA, pp. 171–179, 2018.

[58] X. Xia, Q. Chen, N. Hou, et al., Hylink: Towards high throughput lpwans with lora-compatible communication, in *Proceedings of the 20th ACM Conference on Embedded Networked Sensor Systems* (*SenSys*), New York, NY, USA, pp. 578–591, 2022.

[59] X. Xia, N. Hou, Y. Zheng, et al., Pcube: Scaling lora concurrent transmissions with reception diversities, *ACM Transactions on Sensor Networks* (*TOSN*), vol. 18, no. 4, pp. 1–25, 2022.

[60] M. O. Shahid, M. Srivastava, K. Whitehouse, et al., Concurrent interference cancellation: Decoding multi-packet collisions in lora, in *Proceedings of the ACM SIGCOMM 2021 Conference*, New York, NY, USA, pp. 503–515, 2021.

[61] B. Hu, Z. Yin, S. Wang, et al., Sclora: Leveraging multi-dimensionality in decoding collided lora transmissions, in *2020 IEEE 28th International Conference on Network Protocols* (*ICNP*), Piscataway, NJ, USA, pp. 1–11, 2020.

[62] R. Eletreby, D. Bharadia, S. Gopi, et al., Empowering low-power wide area networks in urban settings, in *Proceedings of the ACM SIGCOMM 2017 Conference*, New York, NY, USA, pp. 309–321, 2017.

[63] S. Gollakota and D. Katabi, Zigzag decoding: Combating hidden terminals in wireless networks, in *Proceedings of the ACM SIGCOMM 2008 Conference on Data Communication*, New York, NY, USA, pp. 159–170, 2008.

[64] L. Kong and X. Liu, mzig: Enabling multipacket reception in zigbee, in *Proceedings of the 21st Annual International Conference on Mobile Computing and Networking* (*MobiCom*), New York, NY, USA, pp. 552–565, 2015.

[65] J. Chan, A. R. Wang, A. Krishnamurthy, et al., Deepsense: Enabling carrier sense in low-power wide area networks using deep learning, 2019, arXiv preprint arXiv: 1904.10607.

[66] A. Gillioz, J. Casas, E. Mugellini, and O. Abou Khaled, Overview of the transformerbased models for nlp tasks, in *2020 15th Conference on computer science and information systems* (*FedCSIS*), pp. 179–183, 2020.

[67] H. Chen, H. Chen, Z. Zhao, K. Han, G. Zhu, Y. Zhao, Y. Du, W. Xu, and Q. Shi, An overview of domain-specific foundation model: key technologies, applications and challenges, arXiv preprint arXiv: 2409.04267, 2024.

[68] G. I. Parisi, R. Kemker, J. L. Part, C. Kanan, and S. Wermter, Continual lifelong learning with neural networks: A review, *Neural networks*, vol. 113, pp. 54–71, 2019.

[69] H. Iba, *AI and SWARM*: *Evolutionary approach to emergent intelligence*. CRC Press, 2019.

[70] J. Yuan, C. Yang, D. Cai, S. Wang, X. Yuan, Z. Zhang, X. Li, D. Zhang, H. Mei, X. Jia, S. Wang, and M. Xu, Mobile foundation model as firmware, in *Proceedings of the 30th Annual International Conference on Mobile Computing and Networking, ACM MobiCom 2024*, Washington D.C., DC, USA, 2024, pp. 279–295, 2024.

[71] E. King, H. Yu, S. Lee, and C. Julien, Sasha: Creative goal-oriented reasoning in smart homes with large language models, *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, vol. 8, no. 1, pp. 12:1–12:38, 2024.

[72] H. Cui, Y. Du, Q. Yang, Y. Shao, and S. C. Liew, Llmind: Orchestrating AI and iot with LLM for complex task execution, *IEEE Commun. Mag.*, vol. 63, no. 4, pp. 214–220, 2025.

[73] H. Wen, Y. Li, G. Liu, S. Zhao, T. Yu, T. J. Li, S. Jiang, Y. Liu, Y. Zhang, and Y. Liu, Autodroid: Llm-powered task automation in android, in *Proceedings of the 30th Annual International Conference on Mobile Computing*

*and Networking, ACM MobiCom 2024*, Washington D.C., DC, USA, 2024, pp. 543–557, 2024.

[74] Y. D. Kwon, J. Chauhan, H. Jia, S. I. Venieris, and C. Mascolo, Lifelearner: Hardware-aware meta continual learning system for embedded computing platforms, in *Proceedings of the 21st ACM Conference on Embedded Networked Sensor Systems*, pp. 138–151, 2023.

[75] Y. Deng, S. Yue, T. Wang, G. Wang, J. Ren, and Y. Zhang, Fedinc: An exemplar-free continual federated learning framework with small labeled data, in *Proceedings of the 21st ACM Conference on Embedded Networked Sensor Systems*, pp. 56–69, 2023.

[76] Z. Zhang, B. Guo, W. Sun, Y. Liu, and Z. Yu, Cross-fcl: Toward a cross-edge federated continual learning framework in mobile edge computing systems, *IEEE Transactions on Mobile Computing*, vol. 23, no. 1, pp. 313–326, 2022.

[77] Z. Li, K. Yu, S. Cheng, and D. Xu, League++: Empowering continual robot learning through guided skill acquisition with large language models, in *ICLR 2024 Workshop on Large Language Model* (*LLM*) *Agents*, 2024.

[78] B. Yang, L. He, N. Ling, Z. Yan, G. Xing, X. Shuai, X. Ren, and X. Jiang, Edgefm: Leveraging foundation model for open-set learning on the edge, in *Proceedings of the 21st ACM Conference on Embedded Networked Sensor Systems*, pp. 111–124, 2023.

[79] Z. Xi, Y. Ding, W. Chen, B. Hong, H. Guo, J. Wang, D. Yang, C. Liao, X. Guo, W. He, et al., Agentgym: Evolving large language modelbased agents across diverse environments, arXiv preprint arXiv: 2406.04151, 2024.

[80] N. Potteiger and X. Koutsoukos, Safeguarding autonomous uav navigation: Agent design using evolving behavior trees, in *2024 IEEE International Systems Conference* (*SysCon*), pp. 1–8, 2024.

[81] S. Mohan, W. Piotrowski, R. Stern, S. Grover, S. Kim, J. Le, Y. Sher, and J. de Kleer, A domain-independent agent architecture for adaptive operation in evolving open worlds, *Artificial Intelligence*, vol. 334, p. 104161, 2024.

[82] J. Xu, H. Zhang, X. Li, H. Liu, X. Lan, and T. Kong, Sinvig: A self-evolving interactive visual agent for human-robot interaction, arXiv preprint arXiv: 2402.11792, 2024.

[83] X. Ma, S. Jeong, M. Zhang, D. Wang, J. Choi, and M. Jeon, Cost-effective ondevice continual learning over memory hierarchy with miro, in *Proceedings of the 29th Annual International Conference on Mobile Computing and Networking*, pp. 1–15, 2023.

[84] Y. Zhao, D. Saxena, and J. Cao, Memoryefficient domain incremental learning for internet of things, in *Proceedings of the 20th ACM Conference on Embedded Networked Sensor Systems*, pp. 1175–1181, 2022.

[85] J. Yuan, S. Wang, H. Li, D. Xu, Y. Li, M. Xu, and X. Liu, Towards energy-efficient federated learning via int8-based training on mobile dsps, in *Proceedings of the ACM Web Conference 2024*, pp. 2786–2794, 2024.

[86] H. Liu, C. Gong, Z. Zheng, S. Liu, and F. Wu, Enabling real-time inference in online continual learning via device-cloud collaboration, in *Proceedings of the ACM on Web Conference 2025*, pp. 2043–2052, 2025.

[87] J. Li, F. Yang, M. Tomizuka, and C. Choi, Evolvegraph: Multi-agent trajectory prediction with dynamic relational reasoning, *Advances in Neural Information Processing Systems*, vol. 33, pp. 19783–19794, 2020.

[88] X. Mo, Z. Huang, Y. Xing, and C. Lv, Multiagent trajectory prediction with heterogeneous edge-enhanced graph attention network, *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 7, pp. 9554–9567, 2022.

[89] S. Shi, L. Jiang, D. Dai, and B. Schiele, Mtr++: Multi-agent motion prediction with symmetric scene modeling and guided intention querying, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 46, no. 5, pp. 3955–3971, 2024.

[90] B. Kang, P. Saha, S. Sharma, B. Chakraborty, and S. Mukhopadhyay, Online relational inference for evolving multi-agent interacting systems, arXiv preprint arXiv: 2411.01442, 2024.

[91] H. Wang, J. Xu, C. Zhao, Z. Lu, Y. Cheng, X. Chen, X.-P. Zhang, Y. Liu, and X. Chen, Transformloc: Transforming mavs into mobile localization infrastructures in heterogeneous swarms, in *IEEE INFOCOM 2024-IEEE Conference on Computer Communications*, pp. 1101–1110, 2024.

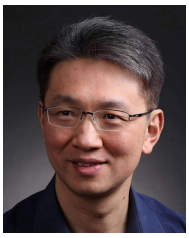
[92] J. Xu, H. Cao, Z. Yang, L. Shangguan, J. Zhang, X. He, and Y. Liu, {SwarmMap}: Scaling up real-time collaborative visual {SLAM} at the edge, in *19th USENIX Symposium on Networked Systems Design and Implementation* (*NSDI 22*), pp. 977–993, 2022.

[93] M. Kouzeghar, Y. Song, M. Meghjani, and R. Bouffanais, Multi-target pursuit by a decentralized heterogeneous uav swarm using deep multi-agent reinforcement learning, in *2023 IEEE International Conference on Robotics and Automation* (*ICRA*), pp. 3289–3295, 2023.

[94] H. Zhang, W. Du, J. Shan, Q. Zhou, Y. Du, J. B. Tenenbaum, T. Shu, and C. Gan, Building cooperative embodied agents modularly with large language models, arXiv preprint arXiv: 2307.02485, 2023.

[95] Y. Xu, J. Zha, J. Ren, X. Jiang, H. Zhang, and X. Chen, Scalable multi-agent reinforcement learning for effective uav scheduling in multihop emergency networks, in *Proceedings of the 30th Annual International Conference on Mobile Computing and Networking*, pp. 2028–2033, 2024.

[96] P. Yuan, A. Ma, Y. Yao, H. Yao, M. Tomizuka, and M. Ding, Remac: Self-reflective and self-evolving multi-agent collaboration for longhorizon robot manipulation, arXiv preprint arXiv: 2503.22122, 2025.



**Yunhao Liu** received the BEng degree from the Department of Automation, Tsinghua University, Beijing, China, in 1995, and the MA degree from Beijing Foreign Studies University, Beijing, in 1997, and the MS and PhD degrees in computer science and engineering from Michigan State University, East Lansing, MI, USA, in 2003 and 2004, respectively. He is currently a professor with the Department of Automation and the Dean of the Global Innovation Exchange, Tsinghua University, Beijing. His research interests include Internet of Things, wireless sensor networks, indoor localization, the Industrial Internet, and cloud computing. He is a Fellow of CCF, IEEE, and ACM.

**Xu Wang** received the BEng and PhD degrees in software engineering from Tsinghua University, Beijing, in 2015 and 2020, respectively. He is a research assistant professor at the Global Innovation Exchange, Tsinghua University, Beijing. He is a member of CCF, ACM, and IEEE. His research interests include the industrial Internet, edge computing, and Internet of Things.

**Yunhuai Liu** is now a full professor with the School of Computer Science at the Peking University, China. He received the PhD degree from Hong Kong University of Science and Technology, and is the recipient of the National Distinguished Young Scholar of NSFC (2019), and National Talented Young Scholar program (2015), and Boya Professorship (2021) of Peking University. He is now serving as the Vice chair of ACM China Council, Director of Beijing Institute of Big Data Research. He received the Outstanding Paper Award at the 28th IEEE ICDCS 2008, the 25th SANER 2018, and the 63th ACL. He has published over 150 peer-reviewed technical papers with over 6000 citations (google scholar).

**Kebin Liu** received the BEng degree from Tongji University, China and PhD degree from Shanghai Jiao Tong University, China. He is a research associate professor in Global Innovation Exchange, Tsinghua University, Beijing, China. His research interests include machine learning and Internet of Things.

**Shuai Tong** received the BE degree from Nankai University in 2019, and the PhD degree from Tsinghua University in 2024. He is currently a postdoctoral researcher at Tsinghua University. His research interests include low-power wide-area networks and Internet of Things.

**Jinliang Yuan** is currently a research assistant with the Department of Automation, Tsinghua University, working with Prof. Yunhao Liu. He received the PhD degree from the School of Computer Science, Beijing University of Posts and Telecommunication, Beijing, China. His research interests are in the area of service computing, mobile edge computing, and federated learning.

**Li Liu** received the PhD degree in computer science from Michigan State University and is currently a postdoctoral researcher at the Department of Automation, Tsinghua University. Her research interests include low-power wide-area network communications, mobile computing, and embodied intelligence. She has served as a Program Committee member for IEEE MASS 2023, IEEE ICPADS 2024, and IEEE MSN 2024. She has also reviewed papers for prestigious conferences and journals such as IEEE INFOCOM, NeurIPS D&B, EAI MobiQuitous, and ACM Transactions on Sensor Networks (TOSN).