

Special Topic: Integration of Large AI Model and 6G

# Towards wireless native big AI model: the mission and approach differ from large language model<sup>†</sup>

Zirui CHEN<sup>1,2</sup>, Zhaoyang ZHANG<sup>1,2,3\*</sup>, Chenyu LIU<sup>1,2</sup> & Ziqing XING<sup>1,2,3</sup>

<sup>1</sup>College of Information Science and Electronic Engineering, Zhejiang University, Hangzhou 310027, China

<sup>2</sup>Zhejiang Provincial Key Laboratory of Multi-Modal Communication Networks and Intelligent Information Processing, Hangzhou 310027, China

<sup>3</sup>Institute of Fundamental and Transdisciplinary Research, Zhejiang University, Hangzhou 310058, China

Received 14 November 2024/Revised 21 February 2025/Accepted 3 June 2025/Published online 19 June 2025

**Abstract** Research on leveraging big artificial intelligence model (BAIM) technology to drive the intelligent evolution of wireless networks is emerging. However, breakthroughs in generalization brought about by BAIM techniques mainly occur in natural language processing. There is a lack of a clear technical direction on how to efficiently apply BAIM techniques to wireless systems, which typically have many additional peculiarities. To this end, this paper reviews recent research on BAIM for wireless systems and assesses the current state of the field. It then analyzes and compares the differences between language intelligence and wireless intelligence on multiple levels, including scientific foundations, core usages, and technical details. It highlights the necessity and scientific significance of developing wireless native BAIM technologies, as well as specific issues that need to be considered for technical implementation. Finally, by synthesizing the evolutionary laws of language models with the particularities of wireless systems, this paper provides several instructive methodologies for developing wireless native BAIM.

**Keywords** wireless artificial intelligence, foundation model, wireless big AI model

**Citation** Chen Z R, Zhang Z Y, Liu C Y, et al. Towards wireless native big AI model: the mission and approach differ from large language model. *Sci China Inf Sci*, 2025, 68(7): 170303, <https://doi.org/10.1007/s11432-024-4464-8>

## 1 Introduction

Given the outstanding potential of artificial intelligence (AI) technologies in complex feature extraction, high-dimensional data representation, and adaptive decision-making, the application of AI technology to resolving problems in wireless systems has been a critical topic in wireless technology research [1–6]. However, simple migrating or reusing successful approaches from computer vision (CV) or natural language processing (NLP) to wireless systems tends to create serious bottlenecks in terms of performance and generalizability [7]. This “incompatibility” is because AI for wireless is not directly equivalent to AI for NLP or CV. Indeed, it is necessary to incorporate electromagnetic physics and engineering methods, which makes modeling wireless AI substantially different from modeling human cognition and interaction capabilities in NLP and CV. High-quality wireless AI technologies must adapt to the characteristics of wireless data and the unique requirements of wireless systems, a concept known as becoming “wireless native.” In recent years, the incorporation of wireless native methods like physics-inspired structures [7–12] and cross-module combination [13, 14] has significantly improved the performance and generalization of wireless AI. Meanwhile, the introduction of prior information and advanced learning architectures can reduce the demand for training data and computational resources. These fundamental advantages are of great significance for advancing the usability of wireless AI, whether in centralized or distributed usage scenarios [2].

Although much progress has been made in wireless AI research, advancements have still only produced intelligence limited to specific tasks and scenarios, which creates obvious weaknesses in generalizability and stability. Meanwhile, in the NLP field, recent large language models (LLMs) have demonstrated the outstanding potential of big AI models (BAIMs) in broad generalization and high reliability. With the

\* Corresponding author (email: ning\_ming@zju.edu.cn)

<sup>†</sup> Special Topic: The 27th Annual Meeting of the China Association for Science and Technology

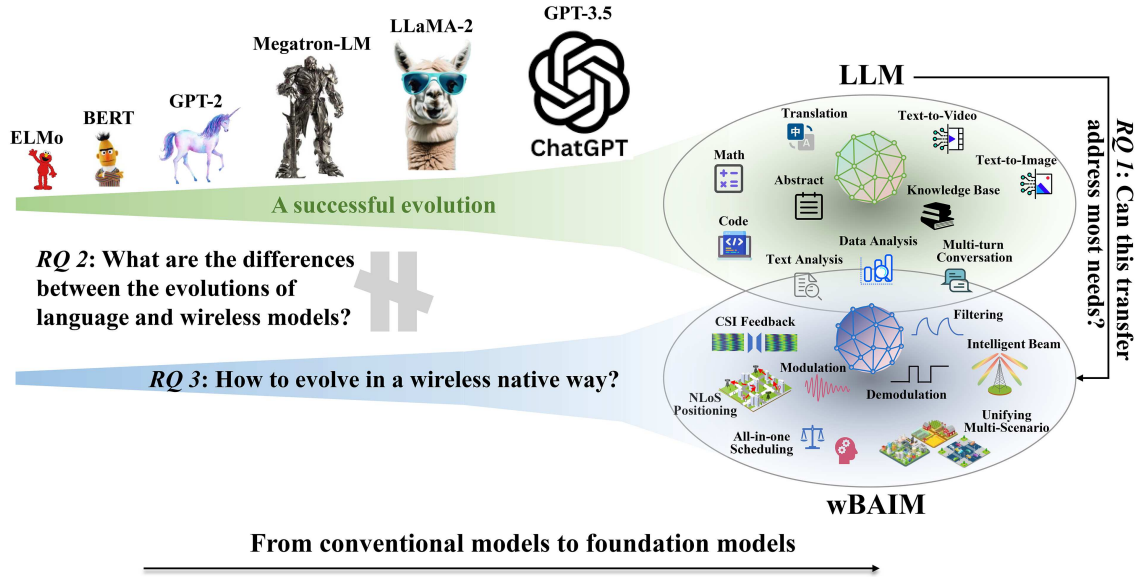


Figure 1 Research backgrounds and critical questions of ‘BAIM for wireless’.

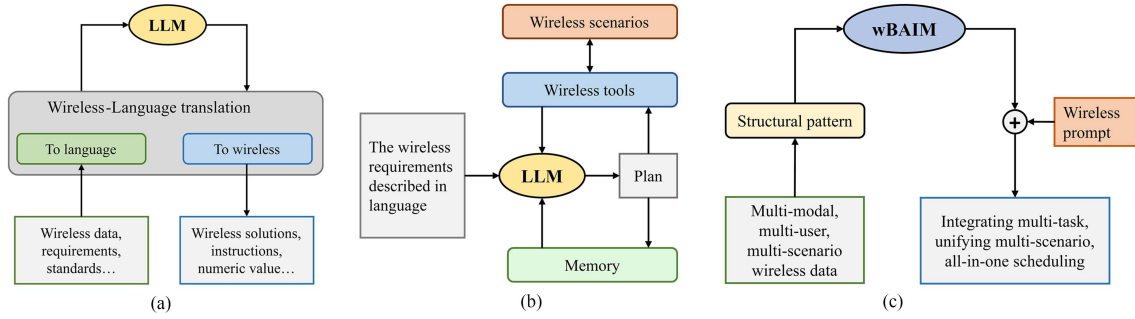
help of a pre-trained LLM, numerous language tasks like translation, question answering (Q&A), and summarization can be accomplished with state-of-the-art performance [15]. Moreover, the emergence of advanced LLMs like DeepSeek has significantly reduced the training and deployment costs of such large-scale AI models [16–18], further expanding the applicability of this cross-task general intelligence. Wireless AI desperately needs similar advancements in generalization. Researchers in [19] first defined wireless BAIM (wBAIM) as having the typical features of integrating multitasks, unifying multiscenarios, and all-in-one scheduling. This definition emphasizes the development of a BAIM that focuses on the physical nature of wireless data and the fundamental requirements of wireless systems. Realizing these promising visions would represent a quantum leap in functionality, universality, and reliability for wireless AI, but it requires the support of comprehensive, in-depth technical exploration. Although Ref. [19] has summarized the technical challenges and potential solutions of wBAIM, several issues almost immediately emerged with wBAIM that still require clarification and supplementation. These issues informed the research questions (RQs) for this study (see Figure 1).

- RQ 1: What are the most necessary and significant aspects when investigating wBAIM in the context of powerful LLMs and their transferability?
- RQ 2: What are the technical peculiarities of wBAIM, especially in comparison to conventional wireless models and existing LLMs?
- RQ 3: What methodologies for developing wBAIM can be derived by synthesizing the evolutionary laws of language models and the peculiarities revealed by RQ 2?

To answer these RQs, this paper is organized as follows. Section 2 provides an overview of the current research landscape and recent advancements in the field. Next, in Section 3, the fundamental differences between language and wireless are analyzed, highlighting the necessity, challenges, and value of studying wBAIM. Section 4 then examines, from a system-level perspective, the unique technical considerations essential for wBAIM research. Section 5 presents several forward-looking insights into how to develop wBAIM. Finally, the conclusion is drawn in Section 6.

## 2 Literature review

The success of LLMs has motivated numerous studies on BAIM for wireless to accelerate wireless systems towards ubiquitous intelligence. Existing studies can be generally categorized into the following three research paradigms: wireless LLM, LLM-based wireless agent, and wBAIM. The overall architectures of these three paradigms are shown in Figure 2.



**Figure 2** Common research paradigms on ‘BAIM for wireless’. (a) Wireless LLM; (b) LLM-based wireless agent; (c) wBAIM.

## 2.1 Wireless LLM

Wireless LLM refers to applying existing LLMs to wireless problems through reuse, fine-tuning, retrieval augmented generation (RAG) techniques, and so on. These methods translate wireless system requirements into language-based descriptions, positioning them within (or approximately within) LLMs’ generalization capabilities. Leveraging LLMs’ advanced information comprehension and vast knowledge base, initial successes have been achieved in wireless systems with protocol understanding, command interaction, and resource allocation [20, 21]. Moreover, some researchers have applied LLMs to NLP-related tasks in wireless systems, such as text categorization [22], text summarization, and Q&A [23, 24], with specific applications including the creation of AI platforms for customer interaction, virtual assistants for technicians, network diagnostic management, and telecom-standard Q&A bots [25–27]. While these studies have demonstrated the value of LLMs in wireless systems, they primarily remain within the scope of traditional NLP. Furthermore, LLMs have been applied to practical requirements in wireless systems such as power control, intrusion detection, resource allocation, and symbol demodulation [28–31]. By constructing appropriate prompts, LLMs based on in-context learning can understand wireless problems to a certain extent, thus enabling initial wireless-specific functions without the need for parameter adjustments.

Meanwhile, to further enhance the applicability of LLMs in wireless communications, some researchers have begun exploring the integration of wireless domain knowledge into LLMs while focusing on the role of multimodal information [32]. These researchers have first pre-trained generative models on specific telecom knowledge datasets, followed by supervised fine-tuning (SFT) of the models with telecom instruction sets [33], and aligning the model’s outputs with human preferences through reinforcement learning from human feedback (RLHF) [34–36]. In addition to direct retraining, the RAG mechanism and the LangChain framework have also been identified as important supplements for a telecom knowledge base [37–41]. Additionally, some researchers have constructed foundational telecom text datasets to serve as material for expanding LLMs’ wireless knowledge. These include SPEC5G, a text categorization dataset derived from 3rd Generation Partnership Project (3GPP) technical reports [42]; TeleQnA, a Q&A dataset extracted from technical documents and research materials [43]; TSPEC-LLM, a dataset based on 3GPP specifications that preserves table formulas [44]; Tele-Data, a LaTeX-formatted dataset obtained from sources like arXiv, 3GPP standards, and Wikipedia; and Tele-Eval, a Q&A evaluation dataset [45].

## 2.2 LLM-based wireless agent

From the perspective of model architecture, LLMs are primarily designed to meet the requirements of language tasks. As a result, they have inherent limitations in data representation, inference patterns, and other aspects when applied to wireless systems. To address these shortcomings, some researchers have begun integrating AI agent technology with wireless LLMs, exploring the concept of LLM-based wireless agents [46]. Unlike traditional wireless LLM, an LLM-based wireless agent uses LLM as a core component, combined with various functional modules to form a complete wireless agent. In [47], the authors proposed a 6G LLM agent, which utilizes an LLM pre-trained on a wireless text dataset as its “brain” for environment perception and action planning, and retrieves available tool libraries to accomplish wireless tasks. Ref. [48] further introduced the experience accumulation module to enable the continuous evolution of the AI agent. Additionally, in [49, 50], researchers further analyze the significance of multi-agent collaboration in wireless systems. It is important to note that, by invoking external tools,

LLM-based agents can potentially outperform wireless LLMs in terms of functionality. However, the upper limits of their understanding and decision-making capabilities are still constrained by the LLM itself. As such, they still struggle to serve complete wireless systems or solve wireless problems that cannot be solved by conventional algorithms. Moreover, the intricate logical architecture of the agent further increases the reasoning complexity beyond that of the LLM alone.

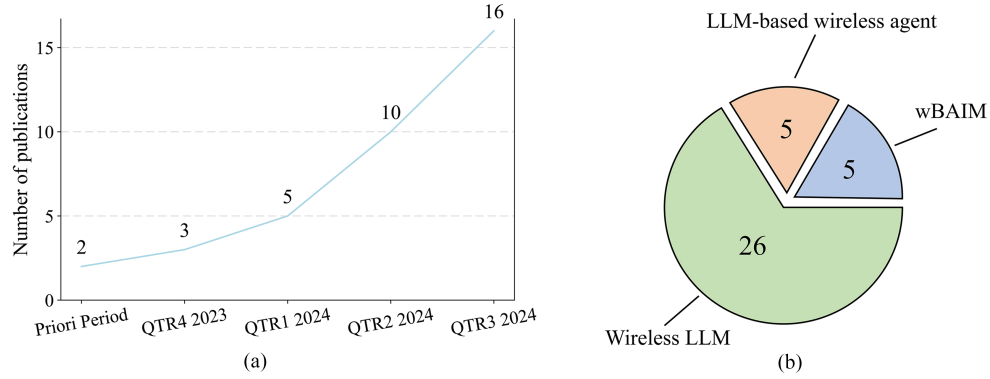
### 2.3 wBAIM

Although LLM-based methods have been somewhat explored, the current technical progress primarily focuses on instruction design, resource scheduling, and a few simple wireless problems. These approaches still lack effective solutions for numerous tough wireless challenges, such as scenario sensing and high-rate transmission. This limitation arises from two primary factors: firstly, linguistic expressions are frequently insufficient in precisely characterizing multimodal signals and their intricate correlations; secondly, the language intelligence derived from textual training lacks adequate knowledge reserves to comprehend the data structures and underlying mechanisms inherent in wireless data. Therefore, instead of embracing LLMs, some researchers have explored BAIM in a wireless native way. Wireless native AI refers to an intelligent framework based on regular patterns of AI technologies, such as deep neural networks, while incorporating the peculiarities of wireless systems. It tailors model design to the nature of wireless data and system requirements, and adapts through re-learning on wireless data. Conventional wireless AI has also followed this approach in many studies. For example, in model design, researchers initially leveraged the numerical characteristics of wireless data, such as exploiting channel sparsity in the angle-delay domain and transforming it into the angle domain to make it more image-like [51, 52]. Later, researchers delved deeper into the physical mechanisms underlying the numerical characteristics and employed more fundamental, physics-inspired techniques to guide model optimization [7–10, 12]. In terms of architecture, some cross-module [13, 14] and cross-layer [53] joint techniques have also been developed to facilitate applications in wireless systems.

However, given that wBAIM has a broader functional scope, relying solely on existing wireless AI technologies is insufficient. Therefore, in [19], the authors first outlined three key generalizability indicators of wBAIM: multi-task integration, multi-scenario unification, and all-in-one scheduling. Meanwhile, Ref. [19] also provides a demonstrative case study where a CMixer [11] model is trained across multiple scenarios to simultaneously accomplish high-dimensional channel estimation and channel compression feedback tasks, illustrating the feasibility and significance of wBAIM. Building on this basis, Ref. [54] introduced SWTCAN, a network based on the advanced swin transformer, which integrates downlink pilot patterns, channel estimation, and uplink channel feedback into a joint design. Ref. [55] proposed Csi-LLM, a downlink channel prediction model based on GPT2, which supports variable-length inputs and enables continuous prediction. Similarly, in [56], a pre-training and fine-tuning-based channel prediction method was presented, enhancing performance in few-shot learning. Meanwhile, [57] evaluated the similarity and diversity across multi-scenario datasets, offering valuable insights into developing multi-scenario generalization techniques from a data-based perspective.

### 2.4 Summary of the literature

Figure 3 presents statistical information on the 36 publications related to BAIM in wireless systems included in this paper. It shows that research interest in BAIM for wireless is growing, and there is a growing consensus on the importance of BAIM techniques for the evolution of 6G. However, studies based on LLMs dominate among the existing paradigms, while research on wBAIM remains limited, and technical advances in wBAIM are still in the early stages. Furthermore, current LLM-based solutions primarily focus on a narrow set of wireless tasks, due to substantial differences in core principles and application contexts between language and wireless domains. The integration of BAIM into wireless systems has not yet reached the depth and breadth it has achieved in language and other fields. Therefore, the following sections highlight the necessity and significance of wBAIM, offering critical insights into its unique challenges and methodologies to foster a deeper integration of BAIM and wireless technologies.



**Figure 3** Statistics of reviewed publications related to ‘BAIM for wireless’. (a) Recent quantitative trends; (b) categorization of studies.

### 3 Comparisons with LLM: the ambitions of wBAIM

#### 3.1 Differences in intelligence orientation

Leveraging existing LLMs, the research value of integrating BAIM techniques and wireless networks can be initially evaluated with low resources and uncomplicated technology adaptation, which is an important motivation for why existing research primarily investigates wireless LLM and LLM-based wireless agents. However, there are many intuitionistic differences between language and wireless. What is the scientific nature of such differences? Can one expect to build complete universal wireless intelligence based on LLM’s transfer in the presence of such differences? This is one of the important questions to be answered in this paper.

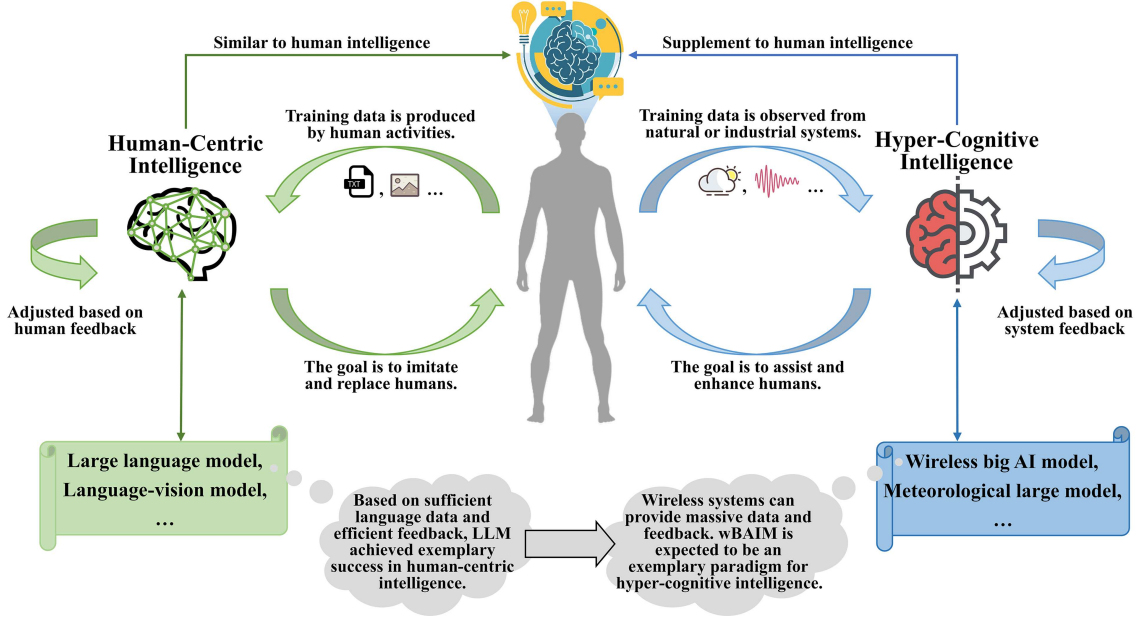
In fact, the intelligence orientation and applications of AI technologies can be divided into two categories based on human cognition, which we term “human-centric” and “hyper-cognitive”, as illustrated in Figure 4. The human-centric model primarily learns from human-generated data, such as images and language, and is continually refined through human feedback, such as RLHF, ultimately aiming to replicate humanoid intelligence. In this sense, the “teacher” of the human-centric model is humanity itself, and its ultimate goal is to replace and liberate humans. However, due to the inherent limitations of biological intelligent agents in complex perception and intensive computation, humans lack the deep understanding and control over many natural and industrial systems. As a result, hyper-cognitive intelligence becomes essential. The hyper-cognitive model primarily learns from phenomena in natural and industrial systems, such as weather patterns and electromagnetic signals, and is optimized through system feedback, ultimately producing intelligence that surpasses the limits of human cognition. In this process, humans typically act as assistants, observing and interacting with information, while the model draws knowledge from systematic phenomena to help humans understand and manipulate complex scientific laws beyond their existing cognitive frameworks.

The distinction between these two forms of intelligence can be intuitively grasped by referencing human intelligence. Rooted in human-activity-and-feedback-based learning, this type of intelligence is inherently constrained by the limits of human cognition, making it challenging to fully address problems that exceed human understanding. Intuitively, language is a human-centric domain, whereas electromagnetic sciences and systems, which are not directly perceivable or recognizable by humans, represent hyper-cognitive challenges. Consequently, while LLMs excel at language-related tasks, they struggle to fully engage with electromagnetic intelligence—an area even beyond human capability. The identified differences address Question 1: LLMs and their transfer can simplify human interaction with wireless systems and manage certain simple wireless use cases. However, for complex wireless problems that are difficult to solve even with the mathematical and physical methods available to humans, LLMs cannot provide a complete solution.

#### 3.2 Scientific significance of wBAIM

Given these substantial differences, an in-depth understanding of electromagnetic laws is essential to more freely and accurately process, represent, and utilize electromagnetic signals. This necessitates the development of a BAIM that is native to electromagnetic sciences and systems—referred to as wBAIM.





**Figure 4** Graphical illustration of human-centric and hyper-cognitive intelligence, which reflects substantial differences in intelligence orientation between LLM and wBAIM.

However, the black-box nature of AI models, coupled with the limitations of human cognition, makes it impossible to directly assess whether a given model has truly acquired the required capabilities. As a result, implicit assessment metrics become essential. This is precisely one of the reasons why Ref. [19] defined multi-task integration, multi-scenario unification, and all-in-one scheduling as the key features of wBAIM. If a single model can be used to generalize across multiple wireless tasks and scenarios, it must inherently capture the commonalities among them—that is, the general electromagnetic laws. In other words, this is the “compression is intelligence” in the wireless context. Likely, this intelligence not only serves wireless communication systems but also provides technical support for other electromagnetics-related fields such as remote sensing [58, 59] and radar systems [60].

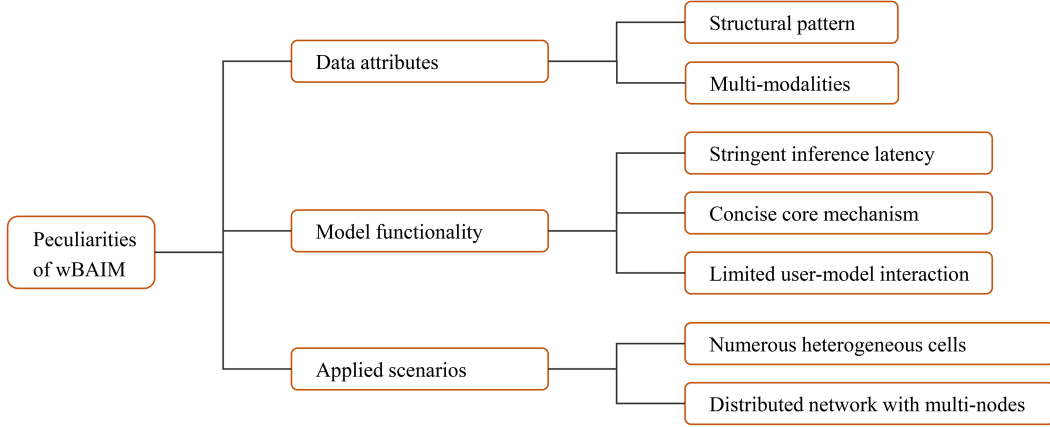
Compared to human-centric intelligence, hyper-cognitive intelligence is also crucial for the advancement of human society, though its development is still in relatively early stages. Given that the wireless network is a widely-used, scientifically well-defined, and highly digitized field within common natural and industrial systems, their advantages in data, computational power, and system feedback are particularly pronounced. As such, wBAIM serves as an ideal candidate for studying hyper-cognitive intelligence, with significant potential to lead technological breakthroughs and provide a valuable reference framework for AI-driven science and engineering.

## 4 Peculiarities of wBAIM

As discussed in Section 3, the intelligence inherent in natural language does not entirely satisfy the functional requirements of wireless systems. To fully leverage the potential of BAIM in wireless systems, it is necessary to develop wBAIM in a wireless native manner. Before creating practical implementation methods, we must evaluate critical aspects of AI technology development, including data, models, and applications, as summarized in Figure 5. This evaluation will help us understand how wBAIM differs from conventional wireless models and LLMs.

### 4.1 Data attribute

In conventional wireless models, the focus is often limited to specific tasks, meaning these models typically process only unimodal data in input or output. However, one key feature of wBAIM is its integration of multiple wireless tasks, which requires representing and processing multimodal data, such as channel state information (CSI), modulation symbols, and positional coordinates. In contrast, language data are inherently unimodal, and multimodal data processing is not a crucial requirement in LLMs. Other BAIMs,



**Figure 5** Technical peculiarities of wBAIM in terms of data, model, and application.

such as language-vision models (LVMs), primarily rely on the alignment of semantic information to handle multimodal data, yet there remains a need to better exploit the correlations and complementarities among these modalities. As a result, addressing the multimodality of wireless data requires further technical research. Moreover, wireless data are typically raw data obtained through sampling and observation, which impose strict logical order, correspondence, and structural properties. In comparison, language data, which have been pre-processed by humans, do not require the same strict order and logic, as they can convey semantic information accurately even when somewhat rearranged. This fundamental difference places higher demands on feature extraction and representation learning structures in wBAIM.

## 4.2 Model functionality

Unlike language models, which encompass a wide variety of knowledge, the fundamental laws of electromagnetic waves are relatively simple. The complexity of wireless signals often comes from the deep coupling of these simple laws with scattering environments and transmission mechanisms. This characteristic should be reflected in the development of wBAIM, i.e., the key to the functionality of the model is its reasoning ability rather than its memorization. Specifically, this involves how to guide the model to understand the concise mechanism behind complex data and how to couple the learned mechanism with the features of the target scenario.

Additionally, wireless systems require real-time transmission, making it essential to strictly control inference time to avoid issues like outdated state information and additional transmission latency. This means that wBAIM urgently needs lightweight reasoning mechanisms under a large information capacity. Furthermore, since obtaining real-time state information in wireless systems often incurs additional signaling overhead, the interactions between wBAIM, the system, and users are inherently limited. As a result, a mechanism capable of deep information extraction and high-density data representation is more needed for wBAIM than for processing extremely long information sequences.

## 4.3 Applied scenarios

The unique networking architecture of wireless systems also brings peculiarities to wBAIM. On the one hand, since the major objective of wBAIM is to unify various scenarios and provide all-in-one scheduling, it requires wBAIM to be generalizable across diverse and heterogeneous cellular scenarios. The transmission configurations and scatterer environments within different scenarios are usually different, making it a key technical challenge to enable wBAIM to quickly adapt to various scenarios. On the other hand, as wireless networks are evolving towards decentralization, the wireless data and computational resources are inherently distributed across scenarios. This decentralized nature makes it impractical for wBAIM to perform training and inference in a fully centralized manner, which differs from the mainstream LLM paradigm. Therefore, developing an efficient, multi-node architecture for collaborative computation is another critical technical requirement for advancing wBAIM.

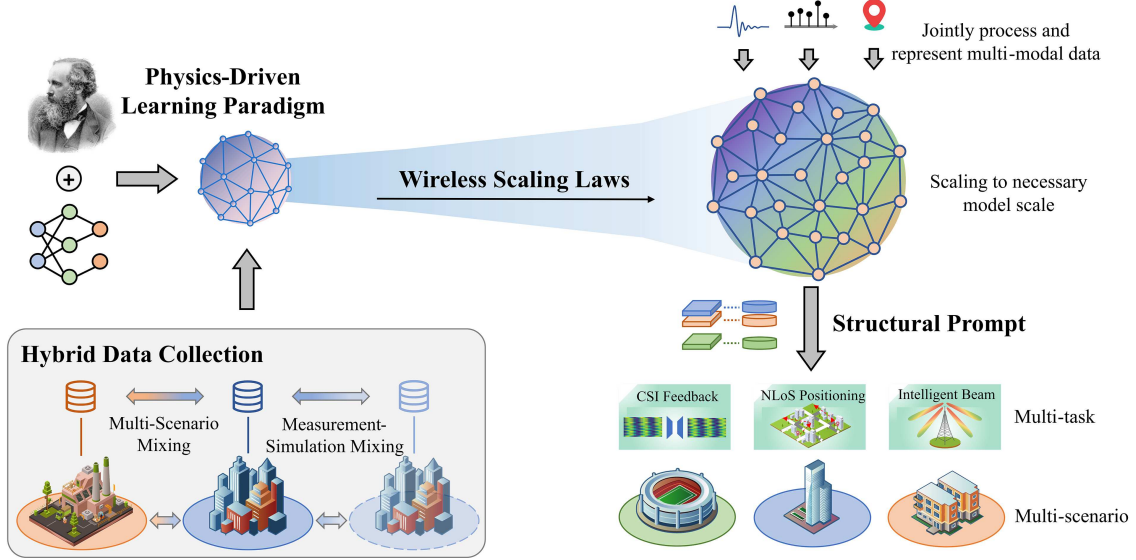


Figure 6 Systematic methodologies in developing wBAIM.

## 5 Methodologies to build wBAIM

Despite the essential differences between wireless and language, scientific development often has strong commonalities across different domains. The successful evolution of LLMs from conventional language models is essentially driven by four key elements: the extensive data as a source of intelligence, the learning paradigm oriented towards general intelligence, the emergence effect brought about by the expansion of model structure and scale, and the efficient method for use case adaptation. In this paper, we map these four elements to the wireless case, merge the evolutionary laws of LLMs with wireless peculiarities, and provide a methodology to guide the development of wBAIM, as shown in Figure 6.

### 5.1 Hybrid data collection

A rich knowledge base of training data is the foundation for BAIM to achieve wide-ranging intelligence. However, since wBAIM aims to serve massive heterogeneous cellular scenarios, the wireless datasets used for pre-training must adequately reflect the scenario diversity, which requires mixing the data from multiple scenarios. Meanwhile, to maximize the diversity of the dataset, an initial assessment of the similarity between scenarios, such as the data distribution sampling [57], should be performed before the data collection, and the scenarios with more significant differences should be prioritized. In addition, given privacy and security considerations, federated learning [61], which employs model interaction rather than direct data mixing, represents a potential technical approach.

In addition, considering the cost of the real data collection, efficient wireless data augmentation techniques, such as data conversion based on a priori physical properties of electromagnetic signals and data generation based on generative models, are worth exploring. Meanwhile, the significance of simulation data should be especially emphasized. Advanced simulation techniques, like ray-tracing, can generate high-fidelity wireless datasets at a low cost. Moreover, they can obtain additional information difficult to obtain in real measurements, such as the multi-path component of the channel and the unbiased user location. Therefore, the efficient mixing of simulated and measured data can significantly reduce the cost of the pre-training for wBAIM.

### 5.2 Physics-driven learning paradigm

To build broad-domain generalization intelligence for BAIM, it is essential to establish a learning paradigm that enables the model to learn various regular phenomena. In wireless networks, the regulations of different tasks and scenarios are derived from the physical laws of electromagnetic waves. Therefore, the ideal learning paradigm for wBAIM should be physics-driven. On the one hand, incorporating known physical laws into the model's a priori design can significantly improve learning efficiency and generalization ability, reducing parameter redundancy and effectively controlling inference complexity.



For instance, one could design the neural network's representation structure according to the physical structure of wireless data, define the optimization objective based on the distribution of wireless features, or select the activation function based on the mapping relationship of wireless information [8].

On the other hand, the architecture design of wireless AI should shift from local module-oriented to global function-oriented, from task-driven to physics-driven. This transformation would facilitate better utilization of the physical correlations across multiple domains and dimensions of wireless data through necessary module reorganization. For example, in [13], researchers proposed a novel channel acquisition method, channel deduction, by combining channel estimation and prediction. This approach overcomes the limitations of traditional channel acquisition models, leveraging the physical correlation of wireless data across multiple dimensions of space, time, and frequency, leading to a low-overhead, continuous autoregressive channel acquisition technique.

### 5.3 Wireless scaling laws

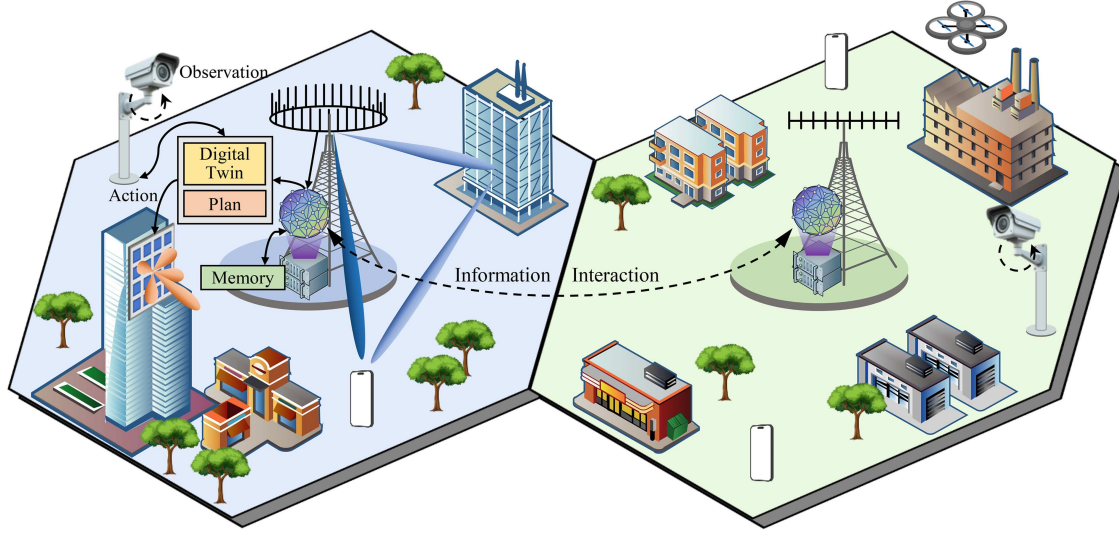
Another typical feature of BAIM is that it continuously improves performance through the structure or scale expansion of the model, even producing the “emergence” effect. In language models, researchers have summarized the statistical relationships between parameter count, dataset size, computational cost, and loss value, broadly predicting the performance gains that can result from expanding the model scale. This has also become an important theoretical foundation for the evolution of LLMs. This paper argues that increasing model size is also essential for the performance of wireless models, especially in terms of generalization, because sufficient neurons are necessary for the model to store diverse knowledge and perform complex reasoning. This principle is preliminarily confirmed in [19].

Meanwhile, it is necessary to point out that the capability to jointly represent and process multi-modal, multi-user, and multi-scenario data is also crucial for wireless models to break through the current performance bottlenecks and generate the emergence effect of ‘more is different’. In wireless models, only limited real-time state information is available due to naturally constrained signaling costs, which differs greatly from language models, where sufficient information is usually accessible. This scarcity of known information can significantly restrict the model's performance, even if the model has extreme depth in information processing. Therefore, for wBAIM, the ability to simultaneously process multi-view information across a wide range of dimensions will significantly enhance their capacity to represent complex wireless environments, which will be a fundamental composition of the wireless scaling laws.

### 5.4 Structural prompt

After learning a wealth of knowledge and intelligence, the model's efficient adaption to specific use cases is crucial for the deployment and application of BAIM. Common adaptation methods include fine-tuning, feature extraction, and prompting. Both fine-tuning and feature extraction require back-propagation and gradient descent in batches, which are demanding on hardware devices and computational resources. Meanwhile, this retraining process often demands considerable time, during which the model's intelligence expression may fail to meet current scenario requirements—a particularly critical issue for wireless communication systems with frequent environmental and configuration changes. Therefore, prompting, which does not require retraining, is the preferred adaptation method for wireless systems with highly decentralized computational resources.

Unlike LLMs, where the prompt can linguistically be described in various forms, the ideal prompt for wBAIM should be strictly structured. The model must design specific data reading and processing patterns to extract essential features from the wireless prompt, facilitating an understanding of and adaptation to the target scenario. For instance, in wireless prompting, the data with different modalities should be processed through corresponding representation interfaces, and the orders and relationships between sub-elements must be strictly maintained during processing. This additional design is necessary due to the peculiarities of wireless data. As described in Subsection 4.1, the meaning of wireless data depends on its strict logic and structure. In contrast, language has robust overall semantics and can convey core information even without strict logic and structure, which significantly simplifies the structure requirements for the prompt to LLMs. While in the development of wBAIM, researchers need to pay extra attention to developing and utilizing structural prompts.



**Figure 7** Overall example of wBAIM's collaboration with AI-agents.

### 5.5 Collaboration with AI-agents

After constructing wBAIM, the next step is to consider the integration with AI-agents, an important booster for expanding the practical applications of current BAIM technology. Building upon wBAIM's deep understanding of wireless regulations, adding modules such as planning, observation, memorization, and small model tool libraries will create wireless intelligent agents, significantly enhancing the usability and flexibility of wireless intelligence. For instance, a planning module can improve wBAIM's logical reasoning capabilities and optimize its global performance; an observation module can enable autonomous acquisition of necessary state information, reducing signaling overhead; a memory module can help the agent to evolve continuously and reduce repetitive thinking; and small-tool libraries can lower the energy consumption required for intelligent computations and optimize the system configurations. An overall example is shown in Figure 7. Furthermore, the technical crux of coordinating such multifunctional modules lies in controlling the overall computational latency and achieving effective fusion and alignment of multi-modal information.

Meanwhile, wireless intelligence is not an isolated intelligence. It can collaborate deeply with linguistic and visual intelligence to further facilitate the management of wireless systems. For example, by aligning wireless commands with natural language through techniques such as the projection layer, wireless agents can be effectively linked with other commonly used agents, improving the interactivity of wireless systems and reducing the technical barriers to operation. In summary, the core objective of integrating wBAIM with AI agents is to avoid the passive and mechanical use of wireless intelligence, instead developing it into an intelligent wireless brain that can actively observe, control, optimize, and interact with the system.

## 6 Conclusion

Ubiquitous intelligence will become an indispensable, fundamental enabling technology as the next-generation wireless networks evolve and improve their functionality, performance, and flexibility. To address the current challenges that wireless AI technologies face in terms of unit cost and reliability, BAIM, which has recently made significant breakthroughs in versatility and generalizability, has become a promising new approach. Therefore, the question of building a BAIM suitable for wireless networks has become a topic of great interest in wireless AI research.

In this study, we focus on creating a research roadmap for developing native wBAIM. We provide a detailed description of its scientific objectives, dissect the peculiarities, and indicate instructive methodologies for the evolution of wireless intelligence. This study can provide academic researchers and industry experts with insightful ideas on the evolution of wBAIM. With further efforts in these research directions,

wBAIM is anticipated to create a revolution in the intelligence of entire wireless systems.

**Acknowledgements** This work was supported in part by National Natural Science Foundation of China (Grant Nos. 62394292, 624B2129), Ministry of Industry and Information Technology (Grant No. TC220H07E), Zhejiang Provincial Key R&D Program (Grant No. 2023C01021), and Fundamental Research Funds for the Central Universities (Grant No. 226-2024-00069).

## References

- 1 You X H, Zhang C, Tan X S, et al. AI for 5G: research directions and paradigms. *Sci China Inf Sci*, 2019, 62: 021301
- 2 Xu W, Yang Z, Ng D W K, et al. Edge learning for B5G networks with distributed signal processing: semantic communication, edge computing, and wireless sensing. *IEEE J Sel Top Signal Process*, 2023, 17: 9–39
- 3 Yang Z, Chen M, Zhang Z, et al. Energy efficient semantic communication over wireless networks with rate splitting. *IEEE J Sel Areas Commun*, 2023, 41: 1484–1495
- 4 Letaief K B, Chen W, Shi Y, et al. The roadmap to 6G: AI empowered wireless networks. *IEEE Commun Mag*, 2019, 57: 84–90
- 5 You X H, Huang Y M, Zhang C, et al. When AI meets sustainable 6G. *Sci China Inf Sci*, 2025, 68: 110301
- 6 Zhu G X, Lyu Z H, Jiao X, et al. Pushing AI to wireless network edge: an overview on integrated sensing, communication, and computation towards 6G. *Sci China Inf Sci*, 2023, 66: 130301
- 7 Chen Z, Zhang Z, Xiao Z, et al. Viewing channel as sequence rather than image: a 2-D Seq2Seq approach for efficient MIMO-OFDM CSI feedback. *IEEE Trans Wireless Commun*, 2023, 22: 7393–7407
- 8 Xiao Z, Zhang Z, Huang C, et al. C-GRBFnet: a physics-inspired generative deep neural network for channel representation and prediction. *IEEE J Sel Areas Commun*, 2022, 40: 2282–2299
- 9 Xiao Z, Zhang Z, Chen Z, et al. From data-driven learning to physics-inspired inferring: a novel mobile MIMO channel prediction scheme based on neural ODE. *IEEE Trans Wireless Commun*, 2024, 23: 7186–7199
- 10 Chen Z, Zhang Z, Xiao Z, et al. Deep learning-based multi-user positioning in wireless FDMA cellular networks. *IEEE J Sel Areas Commun*, 2023, 41: 3848–3862
- 11 Chen Z, Zhang Z, Yang Z, et al. Channel mapping based on interleaved learning with complex-domain MLP-mixer. *IEEE Wireless Commun Lett*, 2024, 13: 1369–1373
- 12 Fan S, Xu W, Xie R, et al. Deep CSI compression for dual-polarized massive MIMO channels with disentangled representation learning. *IEEE Trans Commun*, 2024, 72: 5564–5580
- 13 Chen Z R, Zhang Z Y, Yang Z H, et al. Channel deduction: a new learning framework to acquire channel from outdated samples and coarse estimate. 2024. [ArXiv:2403.19409](#)
- 14 Guo J, Wen C K, Jin S. Deep learning-based CSI feedback for beamforming in single- and multi-cell massive MIMO systems. *IEEE J Sel Areas Commun*, 2021, 39: 1872–1884
- 15 Bommasani R, Hudson D A, Adeli E, et al. On the opportunities and risks of foundation models. 2021. [ArXiv:2108.07258](#)
- 16 Liu A X, Feng B, Wang B, et al. Deepseek-v2: a strong, economical, and efficient mixture-of-experts language model. 2024. [ArXiv:2405.04434](#)
- 17 Liu A X, Feng B, Xue B, et al. Deepseek-v3 technical report. 2024. [ArXiv:2412.19437](#)
- 18 Yang A N, Yang B S, Zhang B C, et al. Qwen2.5 Technical Report. 2024. [ArXiv:2412.15115](#)
- 19 Chen Z, Zhang Z, Yang Z. Big AI models for 6G wireless networks: opportunities, challenges, and research directions. *IEEE Wireless Commun*, 2024, 31: 164–172
- 20 Soman S, Ranjani H G. Observations on LLMs for telecom domain: capabilities and limitations. In: *Proceedings of the 3rd International Conference on AI-ML Systems*, 2023. 1–5
- 21 Bariah L, Zhao Q, Zou H, et al. Large generative AI models for telecom: the next big thing? *IEEE Commun Mag*, 2024, 62: 84–90
- 22 Bariah L, Zou H, Zhao Q Y, et al. Understanding telecom language through large language models. In: *Proceedings of IEEE Global Communications Conference (GLOBECOM)*, 2023. 6542–6547
- 23 Ahmed T, Piovesan N, de Domenico A, et al. Linguistic intelligence in large language models for telecommunications. 2024. [ArXiv:2402.15818](#)
- 24 Cai H R, Wu S J. TKG: telecom knowledge governance framework for LLM application. *Research Square*, 2023. doi: 10.21203/rs.3.rs-3252192/v1
- 25 Lin X Q, Kundu L, Dick C, et al. A primer on generative AI for telecom: from theory to practice. 2024. [ArXiv:2408.09031](#)
- 26 Maatouk A, Piovesan N, Ayed F, et al. Large language models for telecom: forthcoming impact on the industry. *IEEE Commun Mag*, 2024, 63: 62–68
- 27 Karapantelakis A, Thakur M, Nikou A, et al. Using large language models to understand telecom standards. In: *Proceedings of IEEE International Conference on Machine Learning for Communication and Networking (ICMLCN)*, 2024. 440–446
- 28 Zhou H, Hu C M, Yuan D, et al. Large language model (LLM)-enabled in-context learning for wireless network optimization: a case study of power control. 2024. [ArXiv:2408.00214](#)
- 29 Zhang H, Sediq A B, Afana A, et al. Large language models in wireless application design: in-context learning-enhanced automatic network intrusion detection. 2024. [ArXiv:2405.11002](#)
- 30 Lee W, Park J. LLM-empowered resource allocation in wireless communications systems. 2024. [ArXiv:2408.02944](#)
- 31 Abbas M, Kar K, Chen T Y. Leveraging large language models for wireless symbol detection via in-context learning. 2024. [ArXiv:2409.00124](#)
- 32 Shao J W, Tong J W, Wu Q, et al. WirelessLLM: empowering large language models towards wireless intelligence. 2024. [ArXiv:2405.17053](#)
- 33 Kan K B, Mun H, Cao G, et al. Mobile-LLaMA: instruction fine-tuning open-source LLM for network analysis in 5G networks. *IEEE Netw*, 2024, 38: 76–83
- 34 Zou H, Zhao Q Y, Tian Y, et al. TelecomGPT: a framework to build telecom-specific large language models. 2024. [ArXiv:2407.09424](#)
- 35 Zhou H, Hu C M, Yuan Y, et al. Large language model (LLM) for telecommunications: a comprehensive survey on principles, key techniques, and opportunities. 2024. [ArXiv:2405.10825](#)
- 36 Sevim N, Ibrahim M, Ekin S. Large language models (LLMs) assisted wireless network deployment in urban settings. 2024. [ArXiv:2405.13356](#)
- 37 Bornea A L, Ayed F, de Domenico A, et al. Telco-RAG: navigating the challenges of retrieval-augmented language models for telecommunications. 2024. [ArXiv:2404.15939](#)
- 38 Yilma G M, Ayala-Romero J A, Garcia-Saavedra A, et al. TelecomRAG: taming telecom standards with retrieval augmented generation and LLMs. 2024. [ArXiv:2406.07053](#)
- 39 Piovesan N, de Domenico A, Ayed F. Telecom language models: must they be large? 2024. [ArXiv:2403.04666](#)
- 40 Roychowdhury S, Jain N, Soman S. Unlocking telecom domain knowledge using LLMs. In: *Proceedings of IEEE International Conference on Communication Systems and Networks (COMSNETS)*, 2024. 267–269
- 41 Ma X L, Zhao R Q, Liu Y, et al. Design of a large language model for improving customer service in telecom operators. *Electron Lett*, 2024, 60: 13218

- 42 Karim I, Mubasshir K S, Rahman M M, et al. SPEC5G: a dataset for 5G cellular network protocol analysis. 2023. ArXiv:2301.09201
- 43 Maatouk A, Ayed F, Piovesan N, et al. TeleQnA: a benchmark dataset to assess large language models telecommunications knowledge. 2023. ArXiv:2310.15051
- 44 Nikbakht R, Benzaghta M, Geraci G. TSpec-LLM: an open-source dataset for LLM understanding of 3GPP specifications. 2024. ArXiv:2406.01768
- 45 Maatouk A, Ampudia K C, Ying R, et al. Tele-LLMs: a series of specialized large language models for telecommunications. 2024. ArXiv:2409.05314
- 46 Tong J W, Shao J W, Wu Q, et al. WirelessAgent: large language model agents for intelligent wireless networks. 2024. ArXiv:2409.07964
- 47 Xiao Z R, Ye C H, Hu Y B, et al. LLM agents as 6G orchestrator: a paradigm for task-oriented physical-layer automation. 2024. ArXiv:2410.03688
- 48 Cui H, Du Y, Yang Q, et al. LLMind: orchestrating AI and IoT with LLM for complex task execution. *IEEE Commun Mag*, 2025, 63: 214–220
- 49 Xu M, Niyato D, Kang J, et al. When large language model agents meet 6G networks: perception, grounding, and alignment. *IEEE Wireless Commun*, 2024, 31: 63–71
- 50 Jiang F, Peng Y, Dong L, et al. Large language model enhanced multi-agent systems for 6G communications. *IEEE Wireless Commun*, 2024, 31: 48–55
- 51 Wen C K, Shih W T, Jin S. Deep learning for massive MIMO CSI feedback. *IEEE Wireless Commun Lett*, 2018, 7: 748–751
- 52 Sun X, Wu C, Gao X, et al. Fingerprint-based localization for massive MIMO-OFDM system with deep convolutional neural networks. *IEEE Trans Veh Technol*, 2019, 68: 10846–10857
- 53 Tian Y Q, Che J Z, Zhang Z Y, et al. Federated learning with unsourced random access. In: *Proceedings of IEEE 97th Vehicular Technology Conference (VTC2023-Spring)*, 2023. 1–5
- 54 Zeng Y, Qiao L, Gao Z, et al. CSI-GPT: integrating generative pre-trained transformer with federated-tuning to acquire downlink massive MIMO channels. 2024. ArXiv:2406.03438
- 55 Fan S L, Liu Z Y, Gu X Y, et al. Csi-LLM: a novel downlink channel prediction method aligned with LLM pre-training. 2024. ArXiv:2409.00005
- 56 Liu B X, Liu X Y, Gao S J, et al. LLM4CP: adapting large language models for channel prediction. 2024. ArXiv:2406.14440
- 57 Tang H N, Yang L S, Zhou R, et al. Assessing air-interface dataset similarity and diversity for AI-enabled wireless communications. In: *Proceedings of IEEE International Conference on Communications Workshops (ICC Workshops)*, 2024. 1623–1628
- 58 Xing Z, Zhang Z, Chen Z, et al. VBIM-Net: variational born iterative network for inverse scattering problems. *IEEE Trans Geosci Remote Sens*, 2025, 63: 1–16
- 59 Liu X, Hu J, Zhang H, et al. Meta-material sensor-based Internet of Things for environmental monitoring by deep learning: design, deployment, and implementation. *IEEE Trans Wireless Commun*, 2023, 22: 2462–2476
- 60 Zhang C Z, Zhang Z Y, Chen Z R, et al. Stepsize-adaptive SAMP algorithm for fast mmWave radar imaging. In: *Proceedings of IEEE Wireless Communications and Networking Conference (WCNC)*, 2024. 1–6
- 61 Yao J, Xu W, Yang Z, et al. Wireless federated learning over resource-constrained networks: digital versus analog transmissions. *IEEE Trans Wireless Commun*, 2024, 23: 14020–14036