

Analysis of AI-Enhanced Next-Generation Wireless Communication Systems

Xinyue Fu

*College of Electronics and Information, (Micro-Nano Technology College), Qingdao University, Qingdao, 266071, China
3341724281@qq.com*

Abstract:

The transition from 5G to 6G in mobile communication technology presents challenges, including achieving ultra-high spectral efficiency, minimizing latency, and supporting large-scale connectivity. In dynamic and complex wireless settings, traditional model-driven methods are approaching their performance limits. Based on existing literature, this paper analyzes the core technical pathways through which artificial intelligence and machine learning can enhance 6G communications. The results indicate that, at the physical layer, data-driven deep learning models achieve accuracy levels comparable to or exceeding those of traditional methods in channel estimation and signal detection tasks, while simultaneously reducing pilot overhead. Meanwhile, reinforcement learning enables real-time intelligent optimization of massive multiple-input multiple-output (MIMO) beam management, increasing spectral efficiency by 15-30%. At the network layer, federated learning serves as a distributed intelligence framework that safeguards data privacy. Its collaborative optimization approach improves model convergence, accelerating the process to twice the rate of conventional approaches. In addition, the semantic communication paradigm focuses on transmitting the meaning of information instead of raw bits, which reduces data volume by more than 80% in bandwidth-limited scenarios, thus boosting communication efficiency. The study further suggests that future networks will be AI-native, with AI deeply integrated into network architectures and protocols. This shift will transform the network's role from a passive tool to an active AI-enabled service. poor model interpretability and high computational complexity, AI-driven enhancement remains a key pathway to advancing 6G performance.

Keywords: Deep Learning; Reinforcement Learning; Federated Learning; Semantic Communication; 6G.

1. Introduction

Fifth-generation mobile communication technology has been widely deployed, laying the infrastructure for societal digital transformation. However, to support future applications such as holographic communication, the metaverse, and smart industries by 2030, networks must meet extremely high-performance requirements. These include peak data rates exceeding 1 Tbps, air interface latency below 0.1 milliseconds, and connection densities surpassing 10 million devices per square kilometer. Traditional communication systems typically rely on mathematical models, such as Shannon's theorem and electromagnetic wave propagation models, to design and optimize processes. Nonetheless, in environments featuring extremely large antenna arrays, ultra-dense networks, and high-frequency communication bands, channels become complex, time-varying, and high-dimensional, rendering accurate modeling extremely challenging. This limits the performance and efficiency of conventional methods. In contrast, artificial intelligence, particularly deep learning and reinforcement learning, has made significant strides in fields such as image recognition and natural language processing. Its robust pattern recognition capabilities provide promising solutions to the performance bottlenecks in wireless communication. The role of AI in communications has shifted from a mere auxiliary tool to a core component of system design, facilitating embedded solutions. It is now advancing toward "AI-native communication," aiming to create intelligent networks capable of self-sensing, self-learning, and self-optimization. Thus, this paper examines the role of artificial intelligence in next-generation wireless communication systems. It reviews basic communication theories and key algorithms, analyzes AI applications and optimization techniques across the physical, access, and network layers, explores the concept of AI-native communication and its associated technical challenges, and seeks to offer insights for the design of future intelligent wireless networks.

2. Wireless Communication Fundamentals and Intelligent Algorithms

2.1 Technical Fundamentals and Performance Challenges

The evolution of next-generation wireless communication systems is built on core technologies including orthogonal frequency division multiplexing (OFDM), multiple-input multiple-output (MIMO), massive MIMO, beamforming techniques, advanced channel coding schemes, and dynamic radio spectrum allocation. However, these tech-

nologies confront serious performance bottlenecks on the path to 6G. In massive MIMO systems, traditional channel estimation methods based on compressed sensing or least squares demand pilot overhead that grows linearly with the number of antennas, which results in a significant increase in signaling load. Research shows that in complex, high-mobility scenarios, the estimation accuracy of such model-driven methods drop noticeably. In contrast, data-driven deep learning models exhibit better robustness and higher accuracy [1]. High-frequency band communication requires precise beam alignment. Traditional beam training methods, however, induce significant initial access delays and inefficient use of time-frequency resources. For emerging technologies such as reconfigurable intelligent surfaces, optimizing phase shifts is a high-dimensional, non-convex problem and traditional optimization algorithms fail to satisfy real-time requirements [2]. In addition, the traditional communication paradigm has fundamental limitations. While Shannon's bit transmission model focuses on error-free bit delivery, it overlooks the meaning of the information, leading to low efficiency in bandwidth-limited environments, for instance, IoT or multimedia transmission. In contrast, semantic communication addresses these shortcomings. Moreover, in dynamic network environments, like massive IoT, traditional resource allocation algorithms struggle to balance objectives such as connection success rate and throughput effectively [3]. Network operating environments are highly dynamic and uncertain. For example, in massive IoT scenarios, traditional model-based or heuristic resource allocation algorithms find it challenging to achieve a dynamic and optimal balance between multiple objectives, including connection success rate and system throughput [4].

2.2 Algorithm Principles and Characteristic Analysis

In the field of artificial intelligence, several core algorithms have proven effective in addressing typical challenges of 6G wireless communication, due to their unique technical mechanisms. The main advantage of deep learning models is their ability to automatically learn complex nonlinear mappings from data. Specifically, Convolutional Neural Networks excel at extracting local spatial features, rendering them ideal for channel estimation and signal modulation recognition, as they capture the two-dimensional spatial correlations of received signals in the time-frequency domain [1]. Besides, Recurrent Neural Networks (RNNs) and their variants are adept at processing time-series data, effectively capturing the temporal correlation of Channel State Information, which makes them well-suited for short-term channel prediction. The

Transformer model, with its self-attention mechanism, can directly capture long-range global dependencies within sequences. Studies show that in wireless channel prediction tasks, its performance often surpasses that of RNNs/LSTMs, modeling the long-term evolution of channels over time with higher accuracy [5]. These models provide a technical foundation for end-to-end system optimization, spanning from received signals to channels or original data. Deep Reinforcement Learning offers another solution path. Its core mechanism involves an agent learning the optimal strategy through continuous interaction with the environment, endowing it with excellent adaptability in highly dynamic environments where precise mathematical models are unavailable. As a result, the algorithm is widely used in scenarios that require real-time decision-making and resource optimization. For instance, it enables massive IoT networks to autonomously learn efficient access and power control strategies, facilitating the management of large-scale device connectivity [4]. Additionally, it can optimize the phase shift matrix of a Reconfigurable Intelligent Surface in real-time, actively shaping the wireless channel to maximize the system sum rate [2]. Matching appropriate intelligent algorithms to problems across different layers of the communication system is key to fully harnessing AI technology and achieving breakthroughs in communication performance. For processing signal data with spatial structure, CNNs are preferred. For modeling the long-term evolution of channels, the Transformer is often a better choice. In dynamic resource allocation problems, reinforcement learning is generally considered the more suitable option.

3. Key Applications and Intelligent Optimization in Wireless Communications

3.1 Channel Processing and Signal Detection

In channel processing, traditional model-driven methods based on compressed sensing or linear estimation encounter unacceptable overhead from pilots and feedback overhead in massive MIMO systems. Data-driven deep learning methods offer an efficient solution by establishing nonlinear mappings from channel data to its essential characteristics. For channel estimation and feedback, studies demonstrate that deep learning models based on CNNs or autoencoders achieve accuracy similar to or higher than traditional compressed sensing methods, while utilizing ten times fewer pilots [1]. The core mechanism is that such models deeply learn the spatiotemporal structural features of the channel from limited pilot data, enabling

highly efficient compression and accurate reconstruction of Channel State Information. In channel prediction, traditional RNN/LSTM-based predictors struggle to capture long-term evolutionary dependencies. In contrast, the Transformer, with its self-attention mechanism, directly model long-range global dependencies within the channel state sequence. Consequently, Transformer-based predictors outperform LSTMs, achieving lower prediction errors, a longer look-ahead window, and more reliable support for precoding and resource reservation. In signal detection, the computational complexity of the optimal maximum likelihood detector in massive MIMO systems grows exponentially with the number of antennas, hampering practical deployment difficult. Modeling the signal detection challenge as a classification task and solving it with deep learning models has proven to be an effective alternative. For instance, by taking the received signal vector as input and directly mapping it to the transmitted symbol vector, such data-driven detectors achieve performance close to maximum likelihood detection. Meanwhile, they reduce computational complexity from exponential to polynomial growth, striking a favorable balance between performance and complexity [1]. More importantly, this end-to-end learning paradigm equips detectors with strong environmental adaptability. Unlike traditional algorithms that heavily rely on precise channel models and ideal assumptions, deep learning detectors learn directly from data containing various impairments, automatically compensating for non-ideal factors. This yields excellent robustness under real-world channel conditions [1].

3.2 Beam Management and End-to-End Optimization

In communication systems assisted by Reconfigurable Intelligent Surfaces, the joint optimization of active beamforming at the base station and passive beamforming at the RIS is a high-dimensional non-convex optimization problem. Traditional iterative algorithms are difficult to apply in real-time due to their high computational complexity. Deep Reinforcement Learning addresses this challenge by constructing a joint action space that includes both beamforming vectors and RIS phase shifts. This allows the agent to independently learn the optimal configuration strategy through interaction with the environment, without requiring an explicit channel model. Research indicates that DRL-based IRS optimization can converge within milliseconds, improving system spectral efficiency by 15-30% compared to traditional methods [2]. Additionally, deep learning-based beam prediction reduces training overhead by directly mapping user location, sub-band

Channel State Information, or historical measurements to the optimal beam index. It cuts initial access delay by 60–80% and lowers signaling overhead. End-to-end communication systems represent a fundamental paradigm shift enabled by AI in communications. This architecture employs a joint deep learning framework consisting of an encoder, channel, and decoder. It integrates traditionally separate modules (e.g., coding, modulation, equalization) into a differentiable whole, optimized globally via gradient descent. Within this framework, the system not only automatically learns noise-resistant coding schemes for specific channel statistics but also discover irregular symbol mapping patterns that significantly surpass traditional modulation constellations. This achieves a signal-to-noise ratio gain of approximately 2-4 dB under the same bit error rate requirement [1]. Going further, semantic communication extends this concept to transmitting the meaning of information. By jointly training semantic encoder and decoder networks, it enables the extraction and reconstruction of task-critical features. In scenarios with low signal-to-noise ratio or limited bandwidth, it reduces the number of bits to be transmitted to less than one-fifth of that required by traditional methods, while maintaining comparable task performance [3].

3.3 Resource Management and Network Scheduling

In complex scenarios such as dense heterogeneous networks and large-scale Internet of Things, wireless resource allocation becomes highly challenging due to its high dimensionality and strong dynamics. Deep reinforcement learning, particularly multi-agent deep reinforcement learning, provides an effective approach to address such problems. This method typically abstracts each entity in the network as an independent agent, enabling it to learn and execute power control and subchannel allocation strategies based solely on local observations within a distributed decision-making framework. Through interactions and iterative learning among agents, the system can naturally attain efficient collaborative resource allocation. In large-scale IoT scenarios, such MADRL-based resource allocation schemes have been demonstrated to improve the connection success rate by approximately 20% while enhancing system throughput by over 15%, significantly outperforming traditional game theory or heuristic-based methods [4]. Network slicing management requires dynamic resource allocation to meet diverse service quality requirements of different applications. Graph Neural Networks naturally represent network topology by modeling physical nodes and logical slices as nodes in a graph, and their connections and interference relationships as edges.

Employing message-passing mechanisms to aggregate neighborhood information, GNNs accurately learn complex inter-slice resource competition and interference relationships, facilitating intelligent and dynamic slice resource allocation. Research indicates that GNN-based slice resource allocation strategies can improve overall resource utilization by about 25% while satisfying diverse Service Level Agreements [6]. For mobility management, leveraging Recurrent Neural Networks or Long Short-Term Memory networks to model user historical trajectories and real-time channel quality sequence data can accurately predict short-term movement paths and channel states. This capability, combined with DRL, enables proactive handover decision-making. Such data-driven intelligent mobility management solutions can reduce unnecessary handovers by approximately 40% while maintaining handover failure rates below 1% in high-speed mobile scenarios, facilitating intelligent and dynamic slice resource allocation [7].

4. The Frontier Paradigms and Challenges of AI-native Communication

4.1 Native Air Interface and Frame Structure Advantages

“AI-native” design denotes that AI is not an afterthought added to an existing framework, but is integrated as a core component from the outset of the communication system design, deeply embedded across all layers of the air interface waveform, frame structure, and protocol stack [8]. Specifically, the future air interface will not strictly depend on fixed waveforms. Instead, using agents based on deep reinforcement learning or generative AI, it will dynamically generate and configure optimal waveform patterns and frame structure parameters according to real-time service demands and channel conditions. For the bursty short-packet traffic common in the Internet of Things, AI can autonomously generate frame structures with shorter training sequences and narrower guard intervals. This can reduce control overhead from over 25% in traditional OFDM systems to below 10%. The key advantage of this end-to-end, AI-driven air interface optimization lies in its ability to achieve high efficiency in time, frequency, and spatial resource utilization. In contrast to the traditional “one-size-fits-all” OFDM framework, this dynamic and adaptive AI-native air interface is expected to improve system-level spectral efficiency by 20-30% and reduce transmission latency by an order of magnitude, providing essential support for meeting the extreme performance goals of 6G [8,9].

4.2 Semantic Communication and Task-Oriented Models

Unlike the traditional reliable bit transmission model in Shannon's information theory, semantic communication focuses on conveying the meaning of information. It focuses on directly conveying the "meaning" of information, which is a key pathway to achieving a leap in 6G communication efficiency [3]. At the sender side, the Transformer model, leveraging its self-attention mechanism, identifies key semantic features in the input data and filters out unnecessary information, enhancing both task relevance and data efficiency. For instance, in video surveillance scenarios, it can directly recognize and extract event semantics such as "person intruding" instead of transmitting all the raw pixels. At the receiver side, the system employs a pre-established task knowledge base, containing domain-specific common sense, to conduct contextual reasoning and reconstruct information from received semantic features. It can also infer and fill in missing information, making the "extraction-transmission-reconstruction" approach particularly well-suited for task-oriented communication. For example, in collaborative perception for autonomous vehicles, employing a semantic encoder to transmit structured semantic information can reduce the required data volume to 1/100th or even less of traditional lossless transmission methods. Meanwhile, it guarantees the successful execution of critical safety tasks. This signifies a fundamental shift in communication design's core philosophy—shifting from "bit error-free" to "task efficiency" [3,9].

4.3 Integration of Communication and Computing with System Constraints

Federated learning relies on robust wireless networks for periodic aggregation, and optimizing distributed AI training requires effective management of communication and computing resources. This involves integrating asynchronous aggregation, model compression, quantization, and user scheduling. Prioritizing users with high channel and data quality enables the system to not only accelerates global model convergence but also ensures efficient use of bandwidth and energy [2]. This trend of integration has given rise to the concept of the "computing-power network," where network node computing power is regarded as a core resource, alongside communication, and can be scheduled on demand [9]. Despite its potential, AI-wireless integration confronts key challenges. In particular, AI performance hinges on data quality and quantity. Techniques like transfer learning and meta-learning can help models adapt to new scenarios. Second, the computational cost of deep learning models conflicts with 6G's

low-latency demands. Lightweight models, developed via knowledge distillation, pruning, and quantization, are required to address this issue. Such models reduce the computational load and energy consumption of the original model with minimal performance degradation. Furthermore, the "black-box" nature of models raises concerns about interpretability and trust. Future research should develop explainable AI methods or construct hybrid expert systems that integrate data-driven learning with established communication knowledge. From an engineering perspective, integrating dynamic AI models into strictly standardized protocols presents a significant challenge for standardization and integration. This requires defining open AI function interfaces in future standards. Additionally, AI models are susceptible to emerging security threats such as adversarial attacks, requiring the development of targeted defense mechanisms. These include adopting adversarial training to enhance model robustness and establishing an AI security awareness system [6].

5. Conclusion

This paper systematically analyzes the application of artificial intelligence in next-generation wireless communication systems and their emerging paradigms. The results indicate that deep learning and reinforcement learning technologies can effectively address complex challenges such as high-dimensional, nonlinear, and dynamic optimization across the physical layer, access layer, and network layer. These AI methods exhibit performance significantly superior to traditional approaches in areas like channel processing, beam management, and resource allocation. However, current research still has limitations, including the dependence of AI models on large-scale training data, their high computational complexity, limited generalization capability in unseen scenarios, and insufficient interpretability. These factors restrict the practical deployment of such models in terminal devices and for mission-critical tasks. Future research should focus on developing lightweight, efficient, and interpretable models to meet computational and energy constraints. It should also explore cross-layer, end-to-end global optimization methods, improve AI-oriented communication protocols and standardization, and establish secure and reliable distributed learning mechanisms. These efforts will advance the deep integration of AI and wireless communications, laying the technical foundation for developing intelligent, task-oriented, and highly efficient 6G networks.

References

- [1] O'shea, T., & Hoydis, J. (2017). An introduction to deep

- learning for the physical layer. *IEEE Transactions on Cognitive Communications and Networking*, 3(4), 563-575.
- [2] Naeem, F., Kaddoum, G., Khan, S., & Khan, K. S. (2022). Intelligent reflective surface deployment in 6G: A comprehensive survey. *arXiv preprint arXiv:2204.01152*.
- [3] Wang, Y., Han, H., Feng, Y., Zheng, J., & Zhang, B. (2025). Semantic communication empowered 6g networks: Techniques, applications, and challenges. *IEEE Access*.
- [4] Luong, N.C., Hoang, D.T., Gong, S., et al. (2019). Applications of deep reinforcement learning in communications and networking: A survey. *IEEE communications surveys & tutorials*, 21(4), 3133-3174.
- [5] Zheng, F. (2022). 6G potential key technologies [M]. Electronics Industry Press.
- [6] Wang, J., Jiang, C., Zhang, H., Ren, Y., Chen, K. C., & Hanzo, L. (2020). Thirty years of machine learning: The road to Pareto-optimal wireless networks. *IEEE Communications Surveys & Tutorials*, 22(3), 1472-1514.
- [7] Saad, W., Bennis, M., & Chen, M. (2019). A vision of 6G wireless systems: Applications, trends, technologies, and open research problems. *IEEE network*, 34(3), 134-142.
- [8] Letaief, K. B., Chen, W., Shi, Y., Zhang, J., & Zhang, Y. J. A. (2019). The roadmap to 6G: AI empowered wireless networks. *IEEE communications magazine*, 57(8), 84-90.
- [9] Jiang, W., Han, B., Habibi, M. A., & Schotten, H. D. (2021). The road towards 6G: A comprehensive survey. *IEEE Open Journal of the Communications Society*, 2, 334-366.
- [10] Lim, W.Y.B., Luong, N.C., Hoang, D.T., et al. (2020). Federated learning in mobile edge networks: A comprehensive survey. *IEEE communications surveys & tutorials*, 22(3), 2031-2063.