AI-Powered 6G Network Energy Efficiency optimization: A Survey

Fengze Xie*

Nanjing University of Posts and Telecommunications, Nanjing, Jiangsu, China

* Corresponding author email: 2825698964@qq.com

Abstract

The emergence of sixth generation (6G) networks aims not only to satisfy the massive connectivity, ultra-low latency, and high throughput requirements of next-generation applications but also to align with global sustainability goals. Artificial intelligence (AI) plays a crucial role in facilitating this dual objective by enabling intelligent, adaptive, and energy-efficient network management. This review synthesizes current research in AI-empowered resource optimization, lightweight infrastructure deployment, and green network architectures, with a special focus on energy efficiency for sustainable 6G networks. Drawing insights from three pivotal studies, this article outlines key innovations, challenges, and future opportunities in building eco-friendly, intelligent 6G systems.

Keywords

6G Networks, Artificial Intelligence, Energy Efficiency, Resource Optimization.

1. Introduction

As wireless communication technologies evolve toward 6G, the demand for higher data rates, massive connectivity, and seamless integration with intelligent systems such as the Internet of Things , autonomous vehicles, and smart infrastructure grows rapidly. This evolution poses a serious challenge: how to meet performance demands while minimizing the environmental and energy impact of increasingly dense network deployments.

Energy efficiency is not just a technical requirement; it is a societal imperative. With 6G expected to integrate deeply into cities, industries, and homes, its carbon footprint could be significant if not carefully managed [1]. Artificial Intelligence offers a promising solution by enabling smart, adaptive systems that can learn and optimize energy usage across networks [2]. The introduction of AI transforms network management from a traditional passive or pre-programmed mode to an active, self-optimizing mode. Traditional networks rely on preset rules or manual intervention, while the literature emphasizes that AI can dynamically allocate, intelligently manage and optimize. This means that the network will transform from a static, rule-based system to an adaptive, learning-based system that can predict and respond to changing network conditions, thereby achieving continuous energy optimization without constant human supervision. This autonomy is critical to the scale and complexity of 6G networks.

This review systematically reviews recent progress in AI-driven energy efficiency solutions for 6G networks, focusing on three core pillars. The first is intelligent resource optimization, in which AI algorithms dynamically allocate spectrum, power, and compute resources to achieve a balance between high performance and minimal energy waste. The second is lightweight deployment strategies, leveraging model compression, pruning, quantization, and distributed inference to enable AI execution on constrained edge and end devices without sacrificing accuracy or responsiveness. The third is green architecture design, extending energy efficiency principles to the entire system level, integrating AI into physical infrastructure, network topology planning, and cross-domain ecosystem management. In addition to synthesizing representative advances in these areas, this review also explores typical application scenarios, including autonomous vehicles, integrated space-ground networks, holographic communications, and smart cities, where AI plays a decisive role in achieving sustainable 6G operations. The discussion further identifies key challenges, such as the high energy consumption of AI models, insufficient model generalization, the lack of a unified green evaluation framework, and data privacy concerns. Finally, this article outlines future research directions aimed at developing green AI algorithms, enabling cross-layer collaboration, establishing standardized evaluation systems, and integrating privacy protection mechanisms. By providing comprehensive analysis at the technical, architectural, and application levels, this work aims to lay a solid foundation for building smart, energy-efficient, and environmentally sustainable 6G networks.

Frame of the article is shown in Figure 1.

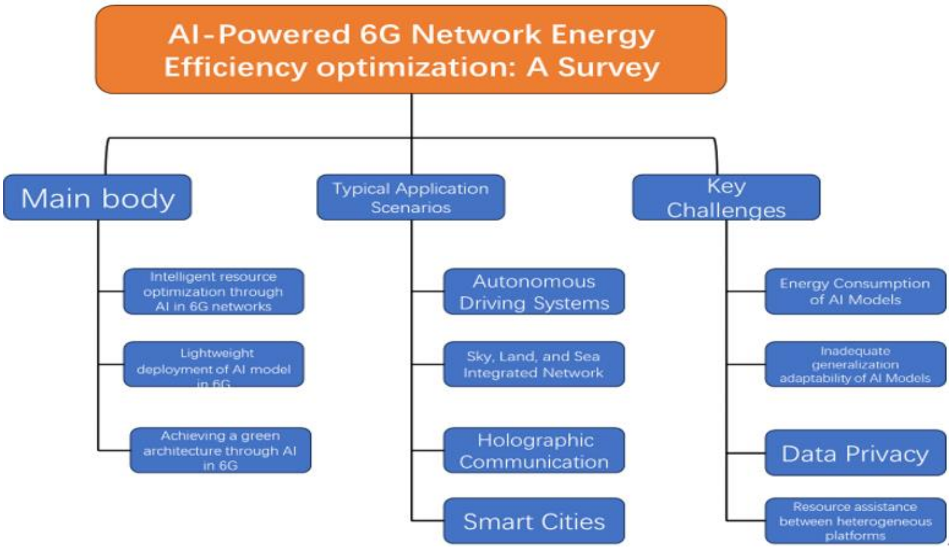

Figure 1. Frame of the article

2. Main body

This section will explore how AI technology can be used to optimize various network resources, thereby significantly saving energy.

2.1. Intelligent resource optimization through AI in 6G networks

Intelligent resource optimization involves the use of artificial intelligence (AI) algorithms to perform real-time scheduling of multi-dimensional resources, such as spectrum allocation, power control, and task offloading, in the complex and dynamic environments of 6G networks, to achieve a balance between minimizing energy consumption and maximizing performance. In this area, Sodhro et al. proposed the HABPA and DSA algorithm framework for smart city IoT scenarios, significantly improving battery life and reducing end-to-end latency[3]. Liu et al. constructed a reinforcement learning model for joint spectrum and power optimization in cognitive IoT, effectively alleviating communication congestion and improving energy efficiency per bit[4]. Furthermore, Yao et al. implemented energy-efficient scheduling strategies in CIoT scenarios under uncertain network conditions by introducing a load prediction mechanism, thereby improving task completion efficiency[5]. Raval's research, targeting large-scale IoT systems, combined AI algorithms to implement an energy-aware scheduling scheme based on task priority, providing a paradigm for collaborative processing among distributed devices[6]. Negi et al. further introduced a multi-agent AI architecture to optimize the joint power control strategy for multiple base stations and terminals in 6G networks, significantly reducing energy consumption at the macro network level [7].

The above body of research collectively demonstrates that AI-based approaches can achieve remarkable efficiency in dynamic resource allocation within complex and heterogeneous 6G network environments. By leveraging predictive analytics, reinforcement learning, and multi-agent coordination, these methods are able to adapt resource scheduling, spectrum assignment, and power control to fluctuating network demands, thereby reducing energy waste while maintaining high quality of service. Nevertheless, several notable limitations remain. Many of the current solutions exhibit insufficient versatility, as they are often tailored to specific network scenarios, traffic patterns, or hardware configurations, which limits their applicability in diverse deployment environments. Furthermore, the energy consumption of the AI models themselves, particularly during model training and high-frequency inference, is often overlooked in optimization objectives, potentially offsetting the energy savings achieved at the network level. Addressing these issues requires not only the development of lightweight AI models with reduced computational complexity but also the integration of cross-domain optimization strategies that jointly consider physical-layer constraints, network-layer policies, and application-layer requirements. In the future, such holistic approaches, supported by adaptive algorithms and hardware–software co-design, may break through current bottlenecks and enable AI systems that are both energy-efficient and universally adaptable across different 6G application domains.

2.2. Lightweight deployment of AI model in 6G

Deploying AI in 6G networks faces challenges such as limited computing resources and stringent latency constraints. Lightweight AI deployment has become a key technical direction. In their research, Letaief et al. proposed an edge AI system architecture for 6G, advocating the use of model pruning and local inference mechanisms to reduce AI deployment energy consumption and improve system responsiveness [8]. Shi et al. emphasized efficient communication in edge AI design, proposing an inference framework that combines model compression and federated learning to reduce inference communication overhead [9]. The work 7 focused on intelligent resource collaborative management in fog computing environments, demonstrating how virtualization technology and an AI scheduler can effectively control power consumption and network latency between edge nodes [10]. Furthermore, the work 8’s earlier review detailed the system concept for AI-enabled 6G, specifically noting that low-power AI modules will serve as "green hubs" between network layers [11]. Guo's research, taking a systemic perspective, summarized lightweight AI inference engine design strategies for 6G large-scale IoT scenarios, emphasizing both model portability and energy-efficient deployment [12].

The deployment of edge AI has become a key path to achieving the dual goals of ultra-low latency and low energy consumption in 6G networks. It enables intelligent processing closer to the data source, reducing reliance on centralized cloud infrastructure. By supporting localized inference and decision-making, edge AI significantly reduces response times while minimizing the communication overhead and associated power consumption of transmitting large amounts of data across the network. Despite its promising prospects, several key challenges must be addressed to fully realize its potential. The lack of standardized models and interfaces for edge AI hinders interoperability between heterogeneous devices and vendors, making large-scale deployment more complex and costly. Furthermore, algorithm stability remains a pressing issue, as lightweight models deployed at the edge often operate in highly variable environments, such as fluctuating network quality, diverse hardware capabilities, and dynamic workloads—which can lead to inconsistent performance or decreased accuracy. Equally important is the need for cross-layer collaboration—that is, seamlessly integrating AI optimization strategies into the physical, network, and application layers to achieve end-to-end energy and latency advantages. Therefore, future research should focus on developing universally compatible model standards, robust adaptive algorithms that can adapt to changing environments, and holistic optimization frameworks that can leverage cross-layer synergies. These advances will help transform edge AI from an experimental technology into a foundational component of sustainable, high-performance 6G systems.

2.3. Achieving a green architecture through AI in 6G

Green architecture not only focuses on energy efficiency control during communications but also extends to the ecological design of physical infrastructure, building systems, and even network protocols. Abu-shaikha explored how to optimize energy consumption in smart buildings through machine learning algorithms, including the dynamic control of lighting, HVAC systems, and network nodes [13]. Their digital twin model effectively supported real-time feedback on building energy conservation. Feng et al. constructed a green platform architecture that integrates AI and 6G [14]. Through system-level intelligent scheduling and inter-module linkage control, they achieved coordinated energy management among different network components. In another review, Huang et al. proposed a green 6G network concept, emphasizing the use of AI for topology planning, cooling mechanism adjustment, and protocol energy-saving control [15]. Goyal's research further proposed that AI should not only empower the communication layer but also extend to the facility operation and maintenance layer, forming an end-to-end green design closed loop centered on energy efficiency [16]. Nguyen's proposed 6G IoT architecture integrates an AI task scheduling module, systematically implementing green control strategies from the terminal to the core network, demonstrating the potential for high integration of AI in system-level energy efficiency control [17].

The green architecture track in 6G research prioritizes system-level energy conservation, focusing on the holistic optimization of physical infrastructure, network topology, and operational processes. This approach aims to integrate sustainability principles into every phase of network planning, deployment, and management, thereby reducing operational costs and environmental impact. However, despite significant progress, a unified green assessment standard remains lacking to serve as a common benchmark for evaluating the environmental performance of different network designs and technologies. Without such standardized metrics, it is difficult to objectively compare solutions, track progress toward sustainability goals, or guide investment and policy decisions. Furthermore, most architectural optimization frameworks do not systematically quantify the energy consumption of AI itself—covering model training, parameter updates, and ongoing inference. This oversight can result in energy savings achieved through AI-driven network-level optimization being partially offset by the computational costs of running complex AI models. To fully realize the potential of AI-enabled green architectures, future work must prioritize the development of widely accepted sustainability assessment frameworks and incorporate detailed accounting of AI-related energy expenditures into the design and assessment process. These measures will ensure that green 6G architectures are not only conceptually sustainable but also significantly effective in reducing the overall carbon footprint of next-generation communication systems.

3. Typical Application Scenarios

With the gradual advancement of 6G technology, its application in multiple key scenarios has placed higher demands on energy efficiency optimization. Artificial intelligence, as a core supporting technology for energy-efficient scheduling and adaptive resource control, demonstrates tremendous potential in areas such as autonomous driving, integrated space-ground networks, immersive media, and smart cities. This chapter systematically analyzes the specific role and application value of AI in the 6G energy efficiency revolution, focusing on these four typical application scenarios [18].

3.1. Autonomous Driving Systems

Autonomous driving places extremely high demands on communication latency and reliability, especially in high-speed scenarios, where vehicles need to obtain real-time environmental information and make collaborative decisions with surrounding vehicles and infrastructure. The ultra-low latency and highly reliable links provided by 6G networks enable collaboration between vehicles and edge nodes. In this context, AI can be used for data preprocessing locally in vehicles or at edge nodes, reducing the communication energy consumption associated with uploading raw data. Furthermore, through reinforcement learning, autonomous driving systems can dynamically select the optimal offloading strategy, achieving intelligent allocation of computing tasks. Furthermore, predictive models can be used to proactively analyze traffic conditions and road conditions, optimizing vehicle communication intervals and behavioral control, further conserving communication resources.

3.2. Sky, Land, and Sea Integrated Network

As communication coverage expands from the ground to the airspace and sea, the 6G network builds a space-ground integrated communication architecture consisting of satellites, drones, high-altitude balloons, ground base stations, and offshore platforms. In this complex system, nodes are highly mobile and link switching is frequent, making energy management a problem. The energy efficiency optimization of AI in such networks is reflected in many aspects. Through the spatiotemporal trajectory prediction model, AI can make accurate judgments on the position changes of satellites or high-altitude platforms, thereby realizing resource pre-configuration. At the same time, AI algorithms can assist in formulating link selection strategies, giving priority to low-power, high-stability transmission paths. In addition, AI can also classify and schedule link loads in real time to ensure that critical data transmission is completed first, thereby improving overall system efficiency.

3.3. Holographic Communication

Holographic communication, as one of the most representative immersive applications in the 6G era, places extreme demands on bandwidth, computing power and storage. A typical 3D holographic transmission consumes more than 10 times more resources than traditional high-definition video. If there is a lack of energy efficiency optimization mechanism, its large-scale commercial deployment will be seriously restricted. The energy efficiency advantage of AI in holographic communication is mainly reflected in three aspects: streaming media compression, user behavior prediction and edge collaborative rendering. First, the image coding model based on deep learning can effectively identify key frames and redundant frames, and realize on-demand coding and dynamic resolution adjustment. Second, through multimodal behavior modeling, the system can predict the user's line of sight and interactive behavior, reducing data generation and transmission in non-focus areas. Finally, the edge node can deploy a lightweight AI model to determine whether to trigger a remote call, thereby avoiding unnecessary data scheduling.

3.4. Smart Cities

Smart cities cover a variety of urban infrastructure systems such as transportation, electricity, water, construction, and security. The number of terminal devices is huge, and energy consumption is widely distributed and fragmented. The ultra-large connectivity and wide coverage provided by 6G provide a physical basis for comprehensive urban perception and scheduling.AI technology plays an important role in energy efficiency control in smart cities.

By modeling the urban environment and resource use through machine learning algorithms, intelligent switching control of energy-consuming equipment such as lighting, air conditioning, and drainage systems can be achieved. At the same time, reinforcement learning is used to predict regional energy consumption peaks, adjust load distribution strategies in advance, and reduce power grid pressure. In addition, the AI scheduling model of edge nodes can dynamically offload perception and processing tasks according to real-time load conditions, reduce the processing pressure of central servers, and improve energy efficiency response capabilities.

4. Key Challenges

Although AI holds great potential for improving 6G network energy efficiency, its practical implementation still faces multiple technical and systemic challenges, hindering its full deployment within the future 6G architecture.

4.1. Energy Consumption of AI Models

Currently, mainstream deep learning models consume high energy during both training and inference. Especially during the pre-training phase, the carbon emissions caused by large-scale GPU parallel computing are significant. One study indicates that the carbon emissions from training a single model are equivalent to the carbon emissions of a car over its entire lifecycle. The energy consumed by training the model itself is even higher than the energy it saves. Therefore, developing more energy-efficient AI model structures, training frameworks, and hardware platforms is a prerequisite for achieving a green, intelligent 6G network.

4.2. Inadequate generalization adaptability of AI Models

Current AI models typically rely on specific scenarios and static datasets for training, resulting in poor adaptability to dynamically changing and heterogeneous network environments. For example, a resource scheduling model trained in urban scenarios may fail completely in rural, marine, or aerial environments. This lack of model transferability limits its feasibility for large-scale deployment. To improve the versatility of AI systems across multiple scenarios, new methods such as transfer learning, meta-learning, and domain adaptation are needed to enhance the model's generalization robustness.

4.3. Data Privacy

The optimization process of AI models often relies on sensitive data such as user location, usage behavior, and device status. This is particularly prominent in scenarios such as smart cities, connected vehicles, and telemedicine. The collection and processing of large amounts of data not only poses privacy risks but also increases the vulnerability of the system to attacks. Therefore, ensuring data security while maintaining energy efficiency optimization capabilities is a key bottleneck in AI deployment. Addressing this issue requires the development of technologies such as federated learning, differential privacy, and trusted execution environments to achieve "data-native" AI model training and inference.

4.4. Resource assistance between heterogeneous platforms

6G networks are built on a large-scale distributed architecture, encompassing multiple layers of heterogeneous computing units, including terminals, edge computing, fog computing, and a core cloud. However, the resource requirements of AI models are often concentrated on GPUs and high-performance computing platforms, making their deployment on low-power edge devices challenging. Furthermore, deploying AI at the edge can result in some computation being offloaded to local servers, leading to load imbalance in the cloud. Achieving cross-layer and cross-domain resource coordination to avoid wasted computing power and bottleneck congestion has become a key issue in building green 6G networks.

5. Future Development Directions

To address the aforementioned challenges and fully unleash the potential of AI in improving 6G energy efficiency, future research should advance from multiple perspectives, including algorithm design, system architecture, evaluation systems, and privacy protection. The following four areas deserve particular attention.

5.1. Building Green AI Algorithms and Low-Power Hardware Systems

To achieve green communications goals, AI models themselves must also be "green." Future research should focus on developing AI algorithms with more compact structures and lower computational complexity, such as Tiny ML, pruned networks, and quantized neural networks, while incorporating energy-constrained objective functions into the training process. Furthermore, hardware-specific low-power AI chips (such as Google TPU, Cambrian, and Ascend) can be deployed to further reduce unit computing energy consumption. Co-optimization of hardware and software will become the foundation of green AI deployment.

5.2. Cross-layer Collaborative Intelligent Resource Management Mechanisms

AI models should enable full-link information fusion and optimized decision-making, from the physical layer to the link layer to the application layer. For example, channel state information at the physical layer can be used by the MAC layer for dynamic spectrum allocation, which is then passed to the application layer for task prioritization and edge offloading strategy formulation. This cross-layer intelligent control minimizes redundant overhead, improves energy efficiency, and enhances network robustness.

5.3. Build a Standardized Green Assessment System and Testing Platform

There is an urgent need to establish a unified green AI assessment platform, including standard datasets and simulation toolchains covering multiple typical 6G scenarios (smart cities, autonomous driving, space communications, etc.). Furthermore, comprehensive energy efficiency evaluation metrics should be defined, such as unit transmission energy consumption, unit task completion energy consumption, and carbon neutrality time, to enhance the comparability and practicality of research and promote the transition of green AI from research to implementation.

5.4. Deeply Integrate Privacy Protection and Security Mechanisms

To balance energy efficiency and data security, future AI models should adopt federated learning for distributed training, ensuring that user data is always stored locally. Differential privacy mechanisms and encrypted inference technologies (such as homomorphic encryption and TEE) can also be introduced to enhance model confidentiality during inference. Combined with the local awareness capabilities of 6G networks, this "privacy-aware AI" will become the core support for building trusted, green, and intelligent systems.

6. Conclusion

This article introduces the key advancements of AI-driven 6G network energy efficiency revolution: AI optimizes spectrum, power, and computing resource allocation, and low-power and lightweight models empower edge devices to achieve multi-layer energy efficiency improvement through local inference and green architecture design. This technology is of significant value to scenarios such as autonomous driving and air-space-ground networks, but it still faces challenges such as high energy consumption of models and lack of standards. In the future, it is necessary to break through green AI design, cross-layer collaboration, unified evaluation, and privacy protection to support the construction of intelligent, green and efficient 6G networks.

References

[1] L. Khaloopour et al., "Resilience-by-Design in 6G Networks: Literature Review and Novel Enabling Concepts," in IEEE Access, vol. 12, pp. 155666-155695, 2024.

[2] P. Zhang, Y. Xiao, Y. Li, X. Ge, G. Shi and Y. Yang, "Toward Net-Zero Carbon Emissions in Network AI for 6G and Beyond," in IEEE Communications Magazine, vol. 62, no. 4, pp. 58-64, April 2024.

[3] Sodhro A H, Pirbhulal S, Luo Z, et al. Towards an optimal resource management for IoT based Green and sustainable smart cities[J]. Journal of Cleaner Production, 2019, 220: 1167-1179.

[4] Liu X, Li Y, Zhang X, et al. Energy-efficient resource optimization in green cognitive internet of things[J]. Mobile Networks and Applications, 2020, 25(6): 2527-2535.

[5] Yao, W., Khan, F., Jan, M.A. et al. Artificial intelligence-based load optimization in cognitive Internet of Things. Neural Comput & Applic 32, 16179–16189 (2020).

[6] Maulin Raval, Shubhendu Bhardwaj, Aparna Aravelli, Jaya Dofe, Hardik Gohel,Smart energy optimization for massive IoT using artificial intelligence, Internet of Things, Volume 13,2021,100354,ISSN 2542-6605.

[7] D. Negi, F. Prakash and A. Gupta, "Influence of AI and MAS in Enhancing Energy Efficiency of 6G Communications," 2024 International Conference on Communication, Computer Sciences and Engineering (IC3SE), Gautam Buddha Nagar, India, 2024, pp. 234-238.

[8] K. B. Letaief, Y. Shi, J. Lu and J. Lu, "Edge Artificial Intelligence for 6G: Vision, Enabling Technologies, and Applications," in IEEE Journal on Selected Areas in Communications, vol. 40, no. 1, pp. 5-36, Jan. 2022.

[9] Y. Shi, K. Yang, T. Jiang, J. Zhang and K. B. Letaief, "Communication-Efficient Edge AI: Algorithms and Systems," in IEEE Communications Surveys & Tutorials, vol. 22, no. 4, pp. 2167-2191, Fourthquarter 2020.

[10] Monali Gulhane, Dr. Ankita Tiwari, Saurabh Bhattacharya, Shraddha Satish Kashid, Dharmesh Dhabliya, Yatin Gandhi, "Developing Energy-Efficient IoT Architecture with Edge and Fog Computing for Smart Cities", 2025 International Conference on Emerging Smart Computing and Informatics (ESCI), pp.1-6, 2025.

[11] K. B. Letaief, W. Chen, Y. Shi, J. Zhang and Y. -J. A. Zhang, "The Roadmap to 6G: AI Empowered Wireless Networks," in IEEE Communications Magazine, vol. 57, no. 8, pp. 84-90, August 2019.

[12] F. Guo, F. R. Yu, H. Zhang, X. Li, H. Ji and V. C. M. Leung, "Enabling Massive IoT Toward 6G: A Comprehensive Survey," in IEEE Internet of Things Journal, vol. 8, no. 15, pp. 11891-11915, 1 Aug.1, 2021.

[13] Abu-shaikha M. Smart sustainable architecture: leveraging machine learning for adaptive digital design and resource optimization[J]. Asian Journal of Civil Engineering, 2025, 26(1): 147-158.

[14] H. Feng, Z. Cui, C. Han, J. Ning and T. Yang, "Bidirectional Green Promotion of 6G and AI: Architecture, Solutions, and Platform," in IEEE Network, vol. 35, no. 6, pp. 57-63, November/December 2021.

[15] T. Huang, W. Yang, J. Wu, J. Ma, X. Zhang and D. Zhang, "A Survey on Green 6G Network: Architecture and Technologies," in IEEE Access, vol. 7, pp. 175758-175768, 2019,

[16] Goyal, S., Sharma, N., Kaushik, I., Bhushan, B., Kumar, N. (2021). A Green 6G Network Era: Architecture and Propitious Technologies. In: Khanna, A., Gupta, D., Pólkowski, Z., Bhattacharyya, S., Castillo, O. (eds) Data Analytics and Management. Lecture Notes on Data Engineering and Communications Technologies, vol 54. Springer, Singapore.

[17] D. C. Nguyen et al., "6G Internet of Things: A Comprehensive Survey," in IEEE Internet of Things Journal, vol. 9, no. 1, pp. 359-383, 1 Jan.1, 2022.

[18] Y. Zhao, G. Yu, and H. Xu, "6G mobile communication network: Vision, challenges and key technologies," 2019.