AIoT and Edge Computing: The Next Frontier of Intelligent, Real-Time Operations

Ali Imam Sunny[1]*, Mehadi Hasan Bijoy[2], Md Musfiqur Rahman[3] and Tanvir Ibn Touhid[4]

[1]Ernst & Young, 1 More London Pl, London SE1 2AF

[2]Department of Electrical and Electronic Engineering, Chittagong University of Engineering and Technology, Bangladesh

[3]Department of Computer Science and Engineering, Chittagong University of Engineering and Technology, Bangladesh

[4]Software Engineer, Indetechs Software Limited, House: 31, Road: 20, Block: K, Dhaka, 1212, BangladeshTechnology, Bangladesh

*Corresponding author: Ali Imam Sunny, Ernst & Young, 1 More London Pl, London SE1 2AF

Introduction

Over the past decade, Artificial Intelligence (AI) and the Internet of Things (IoT) have fundamentally transformed how industries monitor assets, operate processes and make decisions. AI provides intelligence in the form of recognition, prediction and optimization, while IoT supplies connectivity and vast data streams from billions of devices [1]. The convergence of these technologies, termed Artificial Intelligence of Things (AIoT), is revolutionizing operational capabilities. When combined with edge computing, which processes data closer to its source rather than relying on distant cloud servers, AIoT's potential is significantly amplified [2]. This integration reduces latency, enhances performance, improves resilience and enables a new era of real-time intelligent operations. This article argues that AIoT with edge computing is not merely an incremental improvement but a strategic imperative for industries pursuing speed, security and scalability.

The Business Imperative for AIoT with Edge Computing

Traditional IoT architectures typically transmit raw data to centralized cloud systems for analysis, creating two main bottlenecks: latency and bandwidth costs. For latency-sensitive applications such as autonomous vehicles, predictive maintenance in manufacturing, or remote patient monitoring, delays of even milliseconds can be critical [1,2]. Edge computing addresses these challenges by analyzing data at or near the data source. Deploying AI models at the edge enables instant decision-making without the latency associated with cloud inference. This approach not only improves operational responsiveness but also enhances privacy by keeping sensitive data local [3].

The benefits of this architecture include:

A. Reduced downtime: AI-powered anomaly detection at the edge can pre-empt equipment failures.

B. Operational cost savings: Lower network traffic reduces bandwidth and storage expenses.

C.   Regulatory compliance: Local processing supports data sovereignty and privacy regulations.

Industry Impact

Manufacturing and industrial IoT

In advanced manufacturing, unplanned downtime can result in significant financial losses. AIoT integrated with edge computing facilitates real-time equipment health monitoring and enables self-optimizing production lines. Sensors can detect anomalies in vibration, temperature or quality, triggering immediate corrective actions without cloud delays. For example, a predictive maintenance system can process vibration data locally to detect early signs of bearing wear and schedule maintenance before catastrophic failure, directly improving profitability [1] (Figure1).
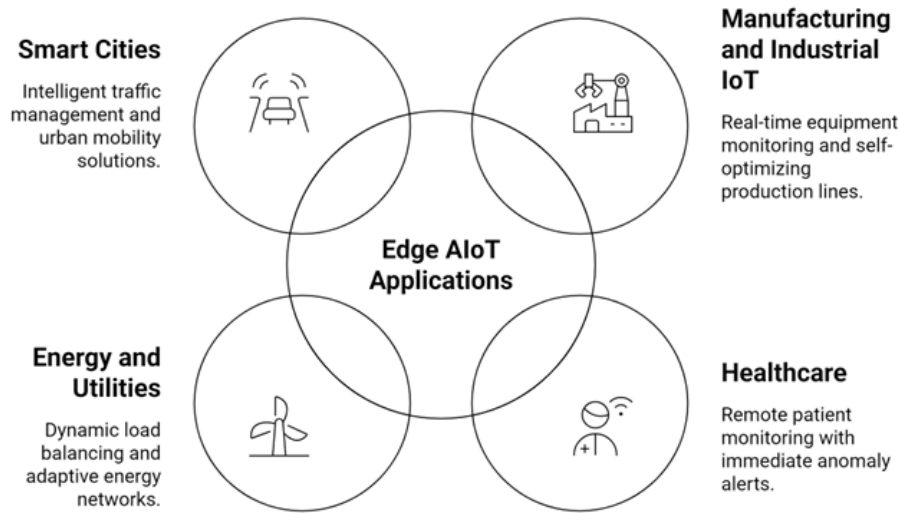
Figure 1: Edge AIoT applications in industry.

Healthcare

Remote patient monitoring is rapidly growing within telemedicine. Edge AIoT devices can continuously analyze vital signs, transmitting alerts only when anomalies occur, thus reducing unnecessary data transmission and improving patient privacy. Consider a wearable ECG patch that locally processes heart rhythms and instantly alerts medical teams upon detecting irregularities such as atrial fibrillation, where latency could mean the difference between life and death [1].

Energy and utilities

Renewable energy grids benefit from edge AI's ability to balance load in real-time by dynamically adjusting storage and distribution. IoT sensors on wind turbines and solar farms optimize energy capture, forecast output and trigger dynamic grid rebalancing, resulting in self-healing and adaptive energy networks resilient to supply fluctuations [1].

Smart cities

AIoT at the edge empowers smart city infrastructure to respond in real time. For instance, an intelligent traffic light system can predict congestion and adjust signal timings locally, without waiting for centralized control instructions, improving urban mobility and reducing emissions [1].

Why Edge Computing is the Catalyst

While AIoT promises intelligence, edge computing makes it practical and scalable. The latency advantage is critical industrial robots, for example, cannot wait hundreds of milliseconds for cloud commands when making split-second decisions. The bandwidth advantage is equally important, as IoT deployments generate vast amounts of data daily; transmitting all raw data to the cloud is inefficient and costly. Edge computing filters and preprocesses data, transmitting only essential information for centralized analysis or archiving. From a security standpoint, edge computing limits the exposure of sensitive raw data by keeping processing local. This is increasingly relevant as IoT devices are frequent targets of cyberattacks, often exploited as entry points into larger networks [1,3].

Challenges and Risks

The integration of AI, IoT, and edge computing presents several challenges:

a)   Model deployment and updates: AI models must be centrally trained but deployed across many distributed edge devices. Updating models without service disruption remains complex [1].

b)   Hardware constraints: Edge devices often have limited processing power, necessitating optimized AI models that

balance low latency with accuracy [1].

c) Interoperability: Heterogeneous IoT ecosystems involve diverse devices and vendors, making seamless communication and data standardization difficult [1].

d) Security: Although edge computing reduces data transfer risks, it increases the number of attack surfaces. Each edge node must be secured against tampering and intrusion [1,3].

e) Skills gap: Developing and maintaining AIoT systems requires expertise across AI engineering, embedded systems and cybersecurity, a combination currently in short supply [1].

Strategic Recommendations

To realize the full potential of AIoT with edge computing, organizations should:

A. Identify high-value use cases: Focus on latency-critical applications or those constrained by bandwidth, such as quality inspection, real-time hazard detection or fraud monitoring.

B. Adopt hybrid architectures: Combine edge inference for real-time operations with cloud-based model training and big data analytics to leverage the strengths of both environments.

C. Optimize AI models: Utilize model compression techniques like pruning, quantization and knowledge distillation to enable efficient edge deployment.

D. Prioritize security-by-design: Implement encryption, secure boot mechanisms and routine patching to protect distributed edge nodes.

E. Plan governance frameworks: Ensure transparency and accountability in AI decision-making, integrating Explainable AI to build trust and comply with regulations.

F. Invest in talent and collaboration: Build cross-disciplinary teams to bridge AI, embedded systems and cybersecurity skills.

Conclusion

The convergence of AI, IoT and edge computing represents a strategic inflection point for industries seeking agility, operational excellence and resilience. By processing intelligence close to the data source, organizations can transition from reactive to proactive operations, reduce reliance on centralized infrastructure and deliver faster, more secure outcomes. However, the path forward demands a balanced approach, leveraging technological innovation alongside robust security and governance frameworks. Mastery of AIoT at the edge will define the competitive advantage and operational capabilities of future-ready organizations.

References

1. Chang Z, Liu S, Xiong X, Cai Z, Tu G (2021) A survey of recent advances in edge-computing-powered artificial intelligence of things. IEEE Internet of Things Journal 8(18): 13849-13875.

2. Bourechak A, Zedadra O, Kouahla MN, Guerrieri A, Seridi H, et al. (2023) At the confluence of artificial intelligence and edge computing in IoT-based applications: A review and new perspectives. Sensors 23(3): 1639.

3. Rupanetti D, Kaabouch N (2024) Combining edge computing-assisted internet of things security with artificial intelligence: Applications, challenges and opportunities. Applied Sciences 14(16): 7104.